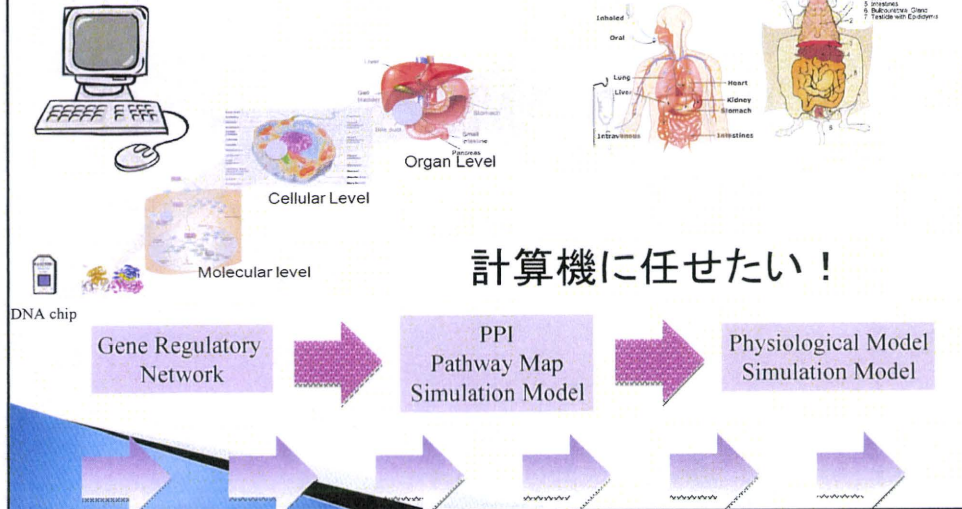


И желают всем счастья!

DATA-DRIVEN APPROACH FOR MULTI-SCALING MODELING



データの現状

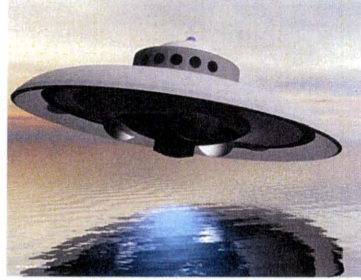
- ▶ 全発現プロファイルが4万プローブを含む
 - 初期反応遺伝子 (responsive transcription)
 - Circadian 遺伝子 (basal transcription)
 - 無反応遺伝子 (not circadian)
 - 実験誤差

Needle in the haystack problem



SYSTEM we NEED

- ▶ Fast
- ▶ Work in uncertainty
- ▶ Macro/micro flexible
- ▶ Rich and interactive visualization
- ▶ Approximate/generate parameters
- ▶ Rapid knowledge annotation

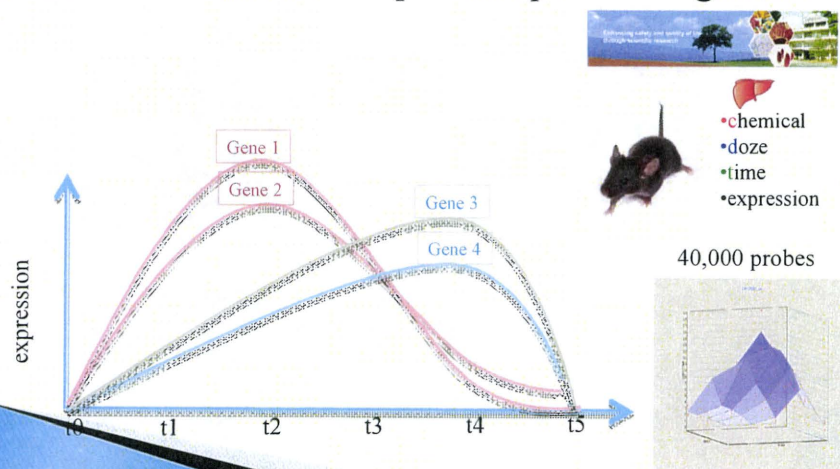


5

AGCT A Geometric Clustering Tool

Atlas of Cell Life by AGCT

- ▶ AGCT reconstructs gene network basing on the similarities of the expression profiles of genes



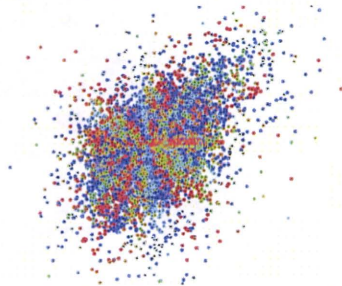
AGCT – A Gene Clustering Tool

- Obtain a global processing chain:
 - 1 **Coding**: from signal to slopes and wavelets
 - 2 **Filtering**: select variables and subsets of genes
 - 3 **Geometric learning**: approximate representations into a low-dimensional manifold
 - 4 **Clustering**: post-process to find clusters of genes with close expressions
- Obtain a user-friendly chain:
 - 1 **Visualization**: all steps must be graphical (3D)
 - 2 **Record-replay**: incorporate a scenario-macro mechanism

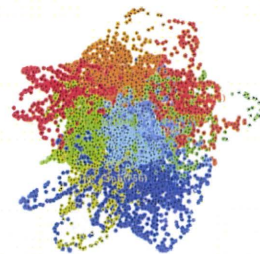
G-protein coupled receptor signaling architecture of mammalian immune cells.
Polouliakh N, Nock R, Helsen F, Kitano H. PLoS One. 2009;4(1):e4189. Epub 2009 Jan 14.

PCA/Manifold [Unfolded origami]

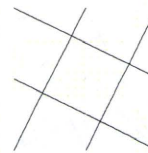
M*N matrix



M*M matrix



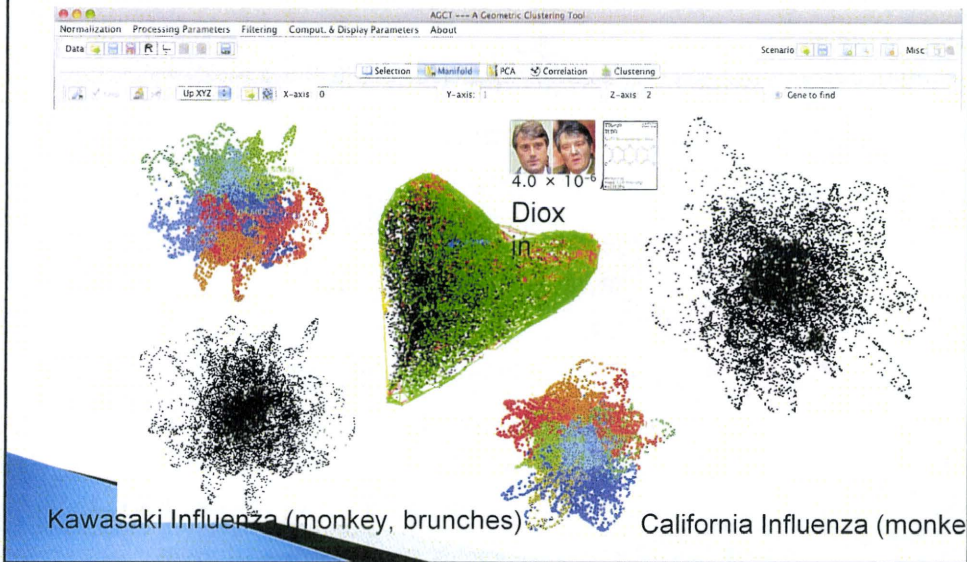
- speedy processing large data
- reduction dimension without information loss
- unsupervised cluster analysis
- simple visualization



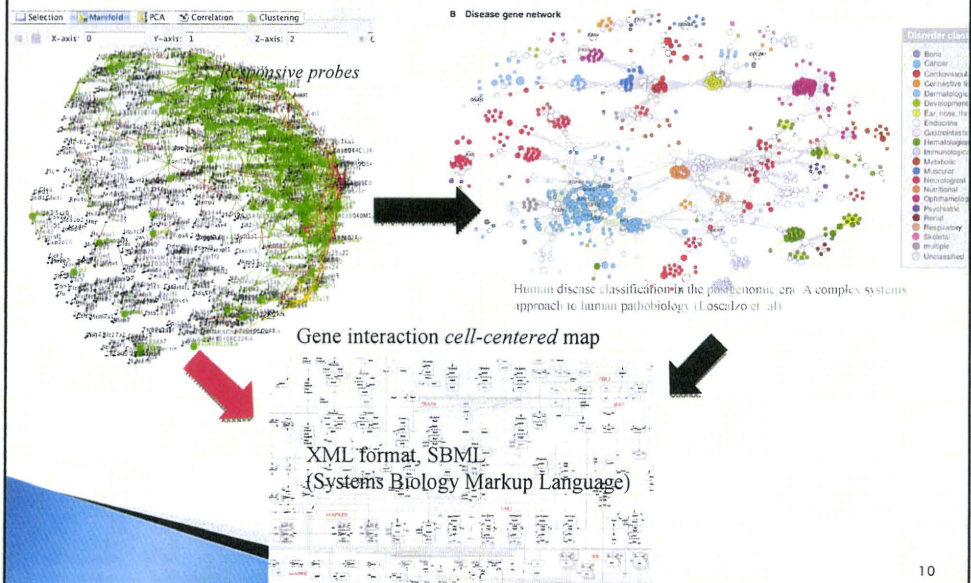
8

AGCT / Geometric Clustering Tool

La vie en rose sur la Manifold



Manifold -> Gene Network -> MAP



Implementations on AGCT in 2010

Pre-processing and Filtering

- ▶ Geometric mean/arithmetic mean
- ▶ Circadian effect subtraction

```
%-profile normalization
subtract_circadian /* 20110127 デフォルトにしないことにした。*/
合のみ使える。
し引き、その後、全時間にわたる中央値を足したも
して、x_i = time_i でのactivation としたとき、
an(xs)
```

やるなら、フル自動で
やりましょう！



- Filtering by Magnitude
- Filtering by Peaks
- Filtering by Responsive probes (a priori knowledge)

Improvements on algorithms

12

Matrix sparsification

(1) Sparse similarity matrix W

$d_{ij} \dots d_{ij2} \dots d_{ij10}$

$O(n^2+10)$ instead of $O(n^2)$

(2) Laplacian matrix L

$L = D^{-1/2} W D^{-1/2}$

The degree of a vertex $v_i \in V$ is defined as $d_i = \sum_{j \in V} w_{ij}$

Degree matrix D is defined as the diagonal matrix with the degrees d_1, \dots, d_n on the diagonal.

m eigenVectors $O(n^2)$ genes $d=10$ $m=100$ where d density for sparsification

(3) Lanczos algorithm
(Transforms the original matrix into a tridiagonal matrix which is real and symmetric.)

Note that (x,y) represents the dot product of vectors x and y here.
After the iteration, we get the α_i and β_i which constructs a tridiagonal matrix

$$T_{m \times m} = \begin{pmatrix} \alpha_1 & \beta_1 & & & 0 \\ \beta_2 & \alpha_2 & \beta_2 & & \\ & \beta_3 & \alpha_3 & \beta_3 & \\ & & \beta_{m-1} & \alpha_{m-1} & \beta_{m-1} \\ 0 & & & \beta_m & \alpha_m \end{pmatrix}$$

The vectors v_i (Lanczos vectors) generated on the fly constructs the transformation $Q = [v_1 \dots v_m]$ which is orthogonal. In practice, it constructs the first vector v_1 and the rest are generated on the fly.

constant factor speedup by memorization

$Q^T = Q^{-1} Q = I$
 $Q^T L Q = Q^T D^{-1/2} W D^{-1/2} Q = LQ$

$Lq_1 = \alpha_1 q_1 + \beta_1 q_2$
 $Lq_2 = \beta_2 q_1 + \alpha_2 q_2 + \beta_3 q_3$
 $Lq_m = \beta_m q_m + \alpha_m q_{m+1}$

(4) e_1, e_2, e_3, \dots

~ 100

12 hours per cell!

160万倍計算UP!

$O(n^3) \rightarrow O(d^* m^* n)$

$d=10$ $m=100$
 n -probes

CellDesigner Plug-in

AGCT

Data loading and coding slopes/wavelets

Data loading and coding slopes/wavelets

Clustering: Bergman & means, Affinity propagation, EM, CP, etc.

Optimization: cluster number

Cluster centroids

PCA view

Positive/negative correlations

CellDesigner

Sony Computer Science Laboratories, Inc.

Cell Designer -> Payao Plug-in

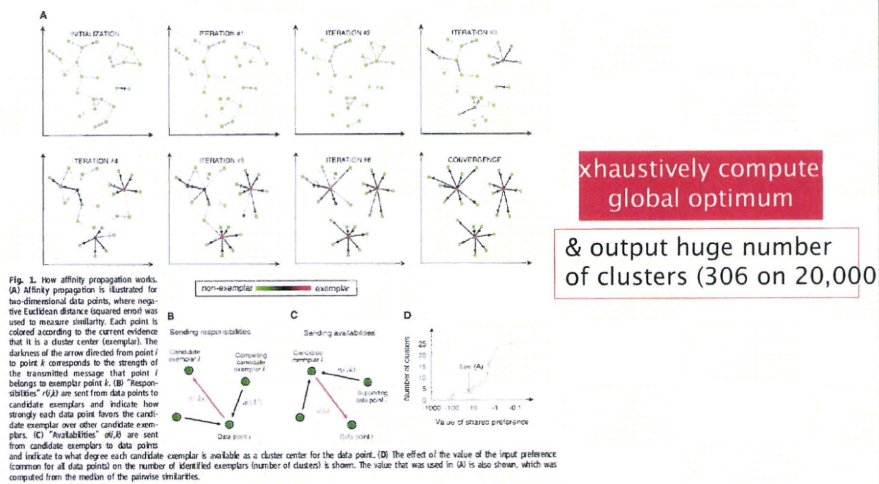
Payao : A Community Platform for SBML Pathway Model Curation
 Yukiko Matsuoka, Samik Ghosh, Norihiro Kikuchi and Hiroaki Kitano

Haystack problem

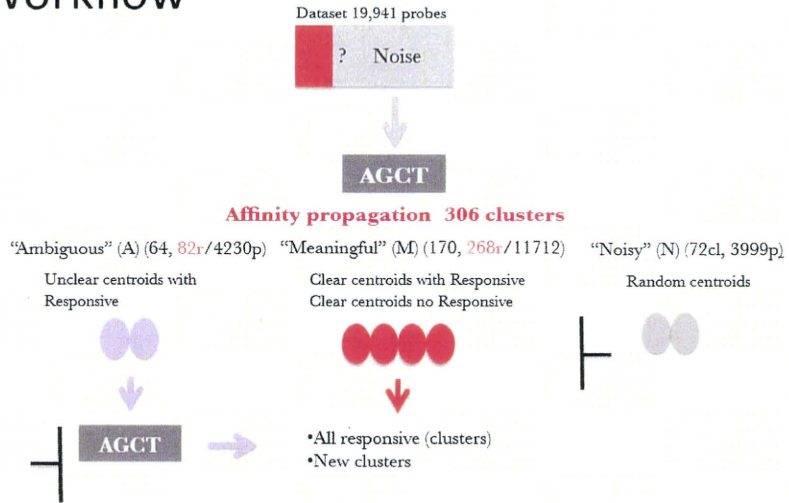


Clustering algorithms

- ▶ Affinity propagation, Bregman K-means, EM, HC, Non-negative matrix factorization, Complete positive factorization

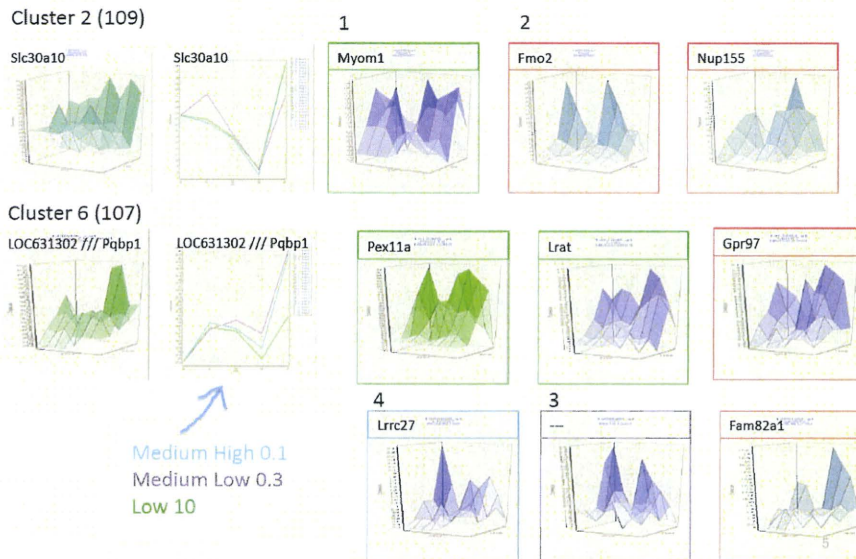


Workflow

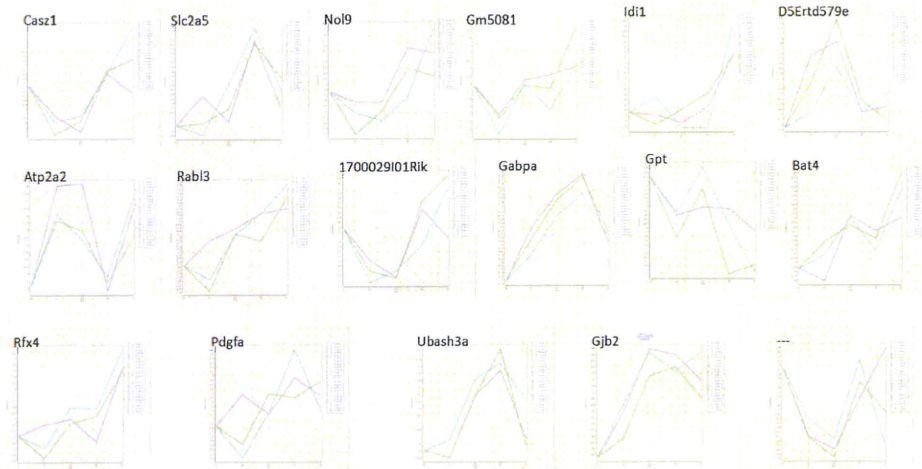


秘 CONFIDENTIAL

“Meaningful” (M) (170)

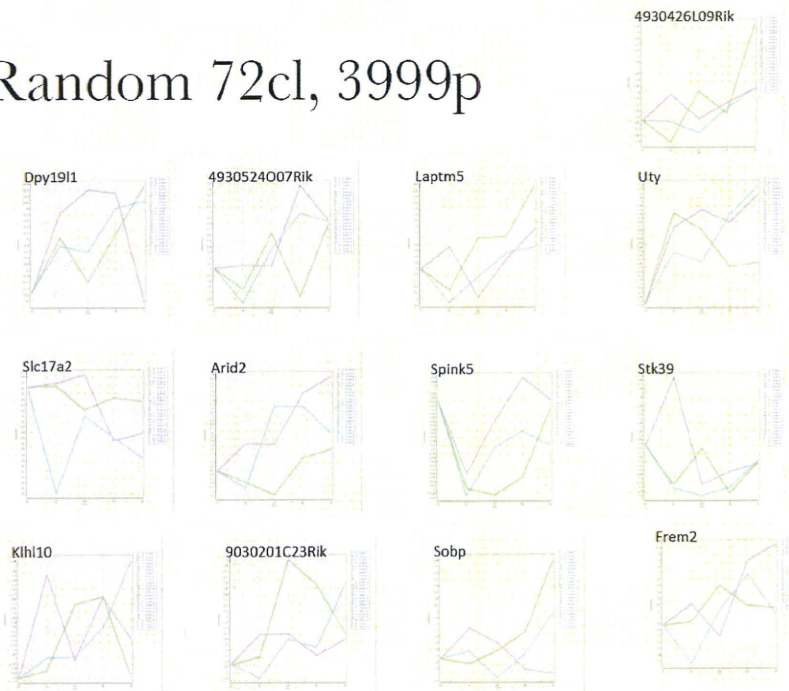


“Meaningful” (M) (170) New



8

Random 72cl, 3999p

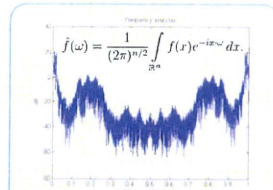


9

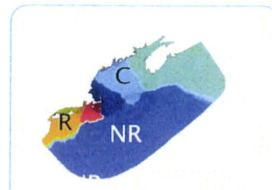
Validation



Cluster density and diversity



Fourier transform for periodicity



Partitioning and causal relations



Acknowledgements

Sony CSL

Hiroaki Kitano
Frank Nielsen
Keigo Oka

University of Martinique

Richard Nock

SBI

Yukiko Matsuoka
Samik Ghosh

厚生労働科学研究費補助金（化学物質リスク研究事業）
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
－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にインフォマティクス技術開発－
(H21-化学-一般-001)

分担研究報告書

胎児、ES細胞、概日変動等の自律的な遺伝子ネットワークの描出研究

研究分担者 北嶋 聡

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 第五室長

研究要旨

本分担研究では、自律的な遺伝子発現が認められる発生過程や概日リズムの局所シグナルネットワークの描出を目的としている。このために先行研究において取得・構築済みの野生型マウス胚、マウスES（胚性幹）・EB（胚葉体）分化系及び、成熟期マウス肝および視交叉上核（SCN）の遺伝子発現の経時データベースを活用する。平成21年度は、マウス野生型胚の経時データベースを用い、T遺伝子に着目し、この発現の経時変化を基としたピアソン相関係数の利用およびISHによる発現部位の可視化を組み合わせる解析を通して、当該遺伝子ネットワークに関連する既知の遺伝子の抽出だけでなく、機能未知の遺伝子の探索を効率よくおこなえることを報告した。

今年度（平成22年度）は、概日変動リズム1周期（=24時間）の成熟期マウスの肝および視交叉上核（SCN）サンプルの網羅的トランスクリプトームデータを用い[Time point: 7点(4時間毎)]、概日変動リズム関連遺伝子のひとつDbp遺伝子について局所ネットワークの描出を検討した。ピアソン相関係数を利用し（ α 値0.9以上）Dbp遺伝子の発現と同様な遺伝子を抽出、さらに目視による検討により生物学的な変化が示唆されたものとして、肝では81プローブセット(ps)（64遺伝子）、SCNでは44ps（39遺伝子）の遺伝子リストが得られた。このうち、概日リズムとの関連が既知の遺伝子は肝では6遺伝子、SCNでは4遺伝子であり、ほとんどが概日リズムとの関連が未報告の遺伝子であった。また、Nr1d2遺伝子（Rev-ErbA beta）のみ両遺伝子リストに共通していた。両リストにつき*in silico*でのプロモーター解析を検討したところ、肝ではKLFの、SCNではEGR family、ETS1及びKLFの結合配列が見いだされ、Dbp遺伝子の関与する局所ネットワーク

の発現制御は肝とSCNでは異なるが、KLFが関与する制御は共通である事が示唆された。加えて、当毒性部が所有する100化合物を超えるPercellomeデータベースを利用し、肝においてDbp遺伝子の発現を抑制する化学物質の探索を経て、その際の経時データベースを活用する事で、抽出した遺伝子リストの妥当性を示す事ができた。肝においてDbp遺伝子の発現を特に強く抑制する化学物質は、化審法での所謂白物質である2-クロロ-4,6-ジメチルアニリン[0、3、10、30mg/kg]、有機リン系殺虫剤であるアセフェート[0、7、20、70mg/kg]及び、食品添加物の保存料であるデヒドロ酢酸ナトリウム[0、30、100、300mg/kg]の3化合物であった。本解析手法は、当該局所ネットワークに関わる既知の遺伝子の抽出だけでなく機能未知の遺伝子の探索を効率よくおこなえ、発現制御を含む多臓器間の連携についても新知見を得ることが出来るものと考えられる。

A. 研究目的

本研究は、先行実施された化学物質リスク評価の基盤整備としてのトキシコゲノミクス研究の成果を受け継ぎ拡充しつつ、毒性分子メカニズムに依拠した毒性予測評価システムの実用化の最終段階としてのインフォマティクス技術を開発することを目的とする。これには、更なる高度な知識と技術を要するが、今までのトキシコゲノミクスが「関連した遺伝子のリスト」や「バイオマーカー」等の静止面的な情報しか提供しない状況を打破し「どのネットワークが反応すると、どのような毒性と直結するか」という時間的要素を含む動的な因果関係を導き出すことが可能となり、遺伝子発現情報からの毒性予測の精度の格段の向上が見込まれる。

本分担研究では、1) 特に自律的な遺伝子発現が認められる発生過程や概日リズムのシグナルネットワークを描出し、また2) マウス胚およびES（胚性幹）・EB（胚葉体）分化系に化学物質を暴露し、遺伝子発現データを得て、発生過程のシグナルネットワー

クに対する影響を検証する。

平成21年度は、先行研究において取得・構築済みのマウス野生型胚およびマウスES・EB分化系の遺伝子発現の経時データベースを利用し、脊索を含む中胚葉の形成と原腸陥入に関わるT(Brachyury)遺伝子に着目し、この局所シグナルネットワークの描出を目的とし検討した。T遺伝子に関しては、その発現時期から、野生型胚およびES・EB分化系、双方の経時データベースを活用できるため、両者共に活用し比較・検討することとした。なお、ES・EB分化系における経時データベースは、ES・EB分化系が、マウス胚の採取が困難な着床前後の胎生4-6日の代替となりうることから作成した経緯がある。T遺伝子の発現の経時変化を基としたピアソン相関係数の利用およびISHによる発現部位の可視化を組み合わせる解析を通して、当該遺伝子ネットワークに関連する既知の遺伝子の抽出だけでなく、機能未知の遺伝子の探索を効率よくおこなえることを報告した。

今年度（平成22年度）は、発生期ではな

く成熟期の、概日リズムに関わるシグナルネットワークに着目することとした。生物は、太陽が作り出す光環境の変化をあらかじめ予想して、体内の代謝や行動を制御することが出来る概日リズム(サーカディアンリズム)というシステムを作り出している。このシステムは視床下部にある視交叉上核で約24時間周期のリズムを発生し、末梢臓器に同期のシグナルを送る。末梢臓器を構成する細胞も約24時間周期のリズムを発生している。この内因性のリズムは、時計遺伝子の転写・翻訳・翻訳後修飾などからなるフィードバック機構によって、周期的な振動が生み出されていることが明らかになってきている。ヒトでは24.8時間周期(24.18時間周期という報告もある)、マウスでは約23時間周期と報告され、これは光や温度など外界の刺激によって修正を受けることが知られている。関与する遺伝子として、昼に発現が高くなる、いわゆる昼時計遺伝子Per1, Per2, Cry1, Cry2等と、夜に高くなる、いわゆる夜時計遺伝子Bmal1 (Arntl)等が報告されている。

当方では、概日変動リズム1周期(=24時間)の成熟期マウス肝および視交叉上核(SCN)の遺伝子発現変動の経時データベース[Time point: 7点(4時間毎)]を有している。そこで今年度は、このデータベースを利用し、概日変動リズム関連遺伝子のひとつDbp遺伝子[bZIP型転写因子Dbp (D-site of albumin promoter (albumin D-box) binding protein)]に着目し、この遺伝子が関わる局所ネットワークの描出を検討した。Dbp遺伝子に着目した理由として、この遺伝子は、視交叉上核をはじめ末梢臓器に発現

するため、肝及びSCNでの発現がシャープな概日振動が報告されているため選択した。加えて、肝およびSCNの組織間連関についても検討した。

B. 研究方法

B-1: 概日変動リズム1周期(=24時間)の成熟期マウス肝および視交叉上核(SCN)の遺伝子発現変動の経時データベースの構築:

12週齢の雄性 C57BL/6NcrSlc マウス(日本エスエルシー)を実験に用いた。個別飼いで3週間馴化したものを実験に供した。経時的(12時間明暗サイクルの下、点灯0、4、8、12、16、20及び24時間後)にサンプリングした肝及びSCN RNA サンプルを用い(n=6)、マイクロアレイ [Affymetrix GeneChip Mouse Genome 430 2.0]を用いて、網羅的遺伝子発現変動解析を検討することにより、遺伝子発現経時データベース [Time point: 7点(4時間毎)]を作製した。

B-2: Total RNA の分離精製

RNA 抽出にあたっては、マウス組織を採取後すみやかに RNA later (Ambion 社)に4°Cで一晩浸漬し、RNase を不活化した。肝は5mm 径の生検トレパンにより3ヶ所を各々別チューブに採取した。肺は気管からRNA later を注入し、RNase の不活化を促した後、採取した。その後、RNA 抽出操作までは-80°Cにて保存した。抽出に当たっては、RNA later を除いた後、RNeasy キット (キアゲン社)に添付される RLT buffer を添加し、ジルコニアビーズを用いて破砕液を調製した。得られた破砕液の10 µL を取り、DNA 定量蛍光試薬 Picogreen を用いて DNA 含量

を測定した。DNA 含量に応じ、臓器毎にあらかじめ設定した割合で Spike cocktail (Bacillus 由来 RNA 5 種類の濃度を変えて混合した溶液) を添加し、TRIZOL により水層を得、RNeasy キットを用いて全 RNA を抽出した。100ng を電気泳動し RNA の純度及び分解の有無を検討した。

SCN サンプルについては、2 段階増幅により cDNA を得た。すなわち、全 RNA の 100ng にとり、T7 プロモーターの付加したオリゴ dT プライマーを用い逆転写し cDNA を得、2 本鎖とした後、T7 RNA ポリメラーゼ (Ambion 社) を用いて cRNA を合成 (この段階ではビオチン化塩基は用いない) した (増幅 1 回目)。その cRNA を鋳型に random primer を用いて逆転写して cDNA を得、2 本鎖にし、T7 RNA ポリメラーゼ (ENZO 社) を用い、ビオチン化 CTP, UTP 共存下 cRNA を合成し、断片化した後、GeneChip へのハイブリダイゼーションに供した。

B-3: GeneChip 解析

全 RNA 5 µg を取り、アフィメトリクス社のプロトコールに従い、T7 プロモーターが付加したオリゴ dT プライマーを用いて逆転写し cDNA を合成し、得た cDNA をもとに第二鎖を合成し、二本鎖 DNA とした。次に T7 RNA ポリメラーゼ (ENZO 社キット) を用い、ビオチン化 UTP, CTP を共存させつつ cRNA を合成した。cRNA は Affymetrix 社キットにて精製後、300-500bp となるよう断片化し、GeneChip ターゲット液とした。GeneChip には Mouse Genome 430 2.0 (マウス) を用いた。ハイブリダイゼーションは 45°C にて 18 時間行い、バッファーによる洗

浄後、phycoerythrin (PE) ラベルストレプトアビジンにて染色し、専用スキャナーでスキャンしてデータを得た。データは当方で開発したソフトウェアならびに、シグナルネットワークの探索のための市販の Ingenuity Pathways Analysis (IPA) (Ingenuity Systems Inc.) を併用して解析した。

(倫理面への配慮)

動物実験の計画及び実施に際しては、科学的及び動物愛護的配慮を十分行い、下記、所属の研究機関が定める動物実験に関する規定、指針を遵守した。「国立医薬品食品衛生研究所・動物実験の適正な実施に関する規程 (平成 19 年 4 月版)」。

C. 研究結果及び考察

C-1: 概日変動リズム 1 周期 (=24 時間) の成熟期マウス肝および視交叉上核 (SCN) の遺伝子発現変動の経時データベースを用いた Dbp 遺伝子に関わる局所シグナルネットワークの描出:

経時的に得た概日変動リズム 1 周期 (=24 時間) の成熟期マウス肝および視交叉上核 (SCN) の RNA サンプルから得た遺伝子発現経時データベース (TIME POINT: 7 点、12 時間明暗サイクルの下、点灯 0、4、8、12、16、20 及び 24 時間後) を用いて、Dbp 遺伝子の発現変動を検討した。Dbp 遺伝子は、SCN では点灯 20 時間後から、肝では点灯 4 時間後から急速に発現が増加しはじめる一過性の発現パターンを示した。この結果を図 1 に示す。Dbp 遺伝子の発現変動は SCN と肝とでは経時的に一致していなかった。

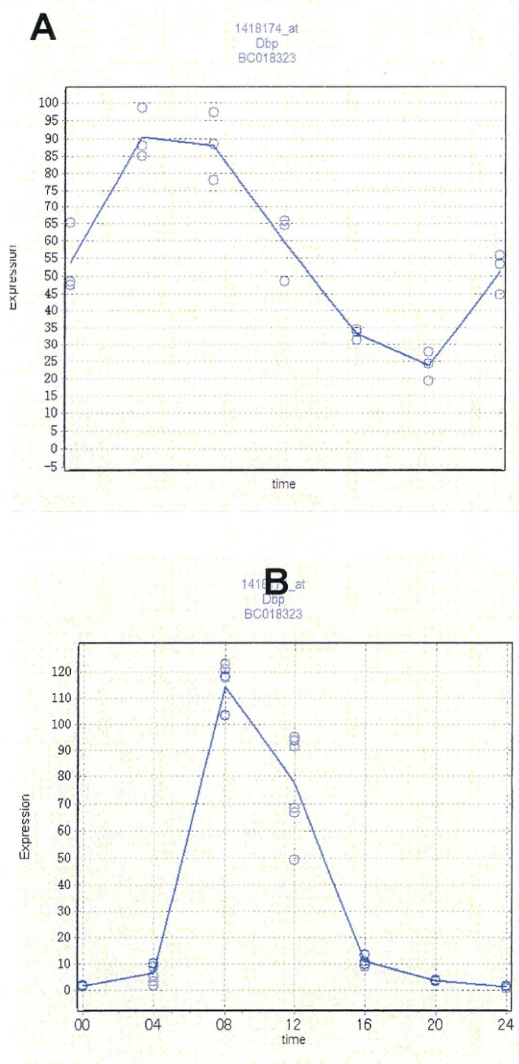


図1 概日変動リズム1周期(=24時間)の成熟期マウス SCN (A)および肝(B)サンプルから得た遺伝子発現経時データベースにおける Dbp 遺伝子の発現変動:

グラフの線は各平均値を結び、点は各データの素点を表す(n=6)。縦軸は、発現量(細胞一個あたりのコピー)、横軸は点灯後の時間(各0、4、8、12、16、20及び24時間後)を表す。

次いで、Dbp 遺伝子に関わる局所シグナルネットワークを描出する目的で、当該遺伝子の発現が、図1で示したような Dbp 遺

伝子と似た経時変化を示す遺伝子の抽出をおこなった。標準化(normalize)した Dbp 遺伝子の各時点の平均値の経時変化を基に、各遺伝子の経時変化についてピアソンの相関係数を求め α 値が高い順に並び替えた。この際、発現コピー数に制限を設けなかった。その結果肝では、 α 値が 1)0.85 以上の場合は 1,002 プローブセット(ps)、2)0.90 以上の場合は 438ps、3)0.95 以上の場合は 103 ps と求められた。

他方 SCN では、 α 値が 1)0.85 以上の場合は、2,054 プローブセット(ps)、2)0.90 以上の場合は 842ps、3)0.95 以上の場合は、159 ps、と求められた。

肝において Dbp 遺伝子を最上段に据え、ピアソンの相関係数順に並べた図(ヒートカラーマップ)を図2に示す。

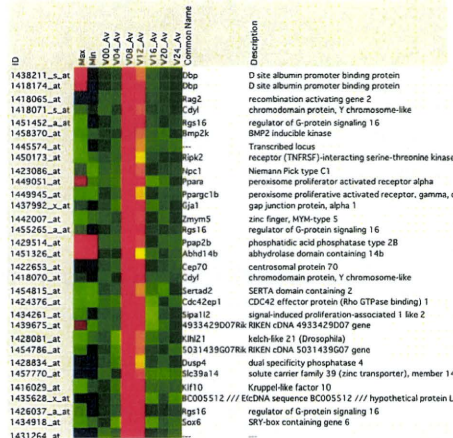


図2 概日変動リズム1周期(=24時間)の成熟期マウス肝サンプルにおける Dbp 遺伝子発現の経時変化を基に、ピアソンの相関係数順に並べたもの。発現の高低を色で示した(ヒートカラーマップ)。発現が高いほど赤く、低いほど青い。

C-1A: 概日変動リズム1周期(=24時間)の成熟期マウス肝における Dbp 遺伝子に関わる局所シグナルネットワークの描出:

肝について、 α 値が 0.90 以上を示す遺伝子リスト (439 ps) について、主として変動幅と標準偏差の大きさをもとに目視による経時変化の再検討をおこない、より生物学的な経時変化を示す遺伝子リストを作成した。その結果、81 ps の遺伝子リストが得られた (Dbp 遺伝子を含まない)。遺伝子数では、64 遺伝子であった。この遺伝子リストについて、市販の IPA により検索したところ、特定のシグナルネットワークは抽出されてこなかった。

そこで、この 81 ps (64 遺伝子) について PubMed による遺伝子機能の検索をおこない、概日リズムとの関連が既知の遺伝子と関連が見いだせなかった遺伝子に分類した。下記にその遺伝子数と既知のものについては遺伝子名を記載する。

・概日リズムとの関連が既知の遺伝子;
6 遺伝子: Rgs16, Ppara, Ppargc1b, Klf10, Gnat1, Nr1d2

・概日リズムとの関連が見いだせなかった遺伝子; 58 遺伝子

以上のことから、得られたリスト中の遺伝子のほとんど (約 90%) が、概日リズムとの関連が未報告の遺伝子であることが示唆された。関連が既知の遺伝子の中で Dbp 遺伝子との関連が報告されている遺伝子は Ppara, Ppargc1b, Nr1d2 遺伝子の 3 つであった。したがって本抽出法により、概日リズムとの関連が既知の遺伝子が実際に 6 遺伝子抽出でき、この中には Dbp 遺伝子との関連が報告されている 3 遺伝子が含まれ、この結果は、本抽出法の技術的な妥当性を

示すものとする。

C-1B: 概日変動リズム 1 周期 (=24 時間) の成熟期マウス SCN における Dbp 遺伝子に関わる局所シグナルネットワークの抽出:

肝と同様に SCN について α 値が 0.90 以上を示す遺伝子リスト (843 ps) について、主として変動幅と標準偏差の大きさをもとに目視による経時変化の再検討をおこない、より生物学的な経時変化を示す遺伝子リストを作成した。その結果、44 ps の遺伝子リストが得られた (Dbp 遺伝子を含まない)。遺伝子数では、39 遺伝子であった。この遺伝子リストについて、市販の IPA により検索したところ、Circadian Rhythm Signaling シグナルネットワークが抽出されてきた。

そこで、この 44 ps (39 遺伝子) について PubMed による遺伝子機能の検索をおこない、概日リズムとの関連が既知の遺伝子と関連が見いだせなかった遺伝子に分類した。下記にその遺伝子数と既知のものについては遺伝子名を記載する。

・概日リズムとの関連が既知の遺伝子;
4 遺伝子: Nr1d2, Tef, Hlf, Nr1d1, Perl

・概日リズムとの関連が見いだせなかった遺伝子; 35 遺伝子

以上のことから、得られたリスト中の遺伝子のほとんど (約 90%) が、概日リズムとの関連が未報告の遺伝子であることが示唆された。関連が既知の遺伝子の中で、Dbp 遺伝子との関連が報告されている遺伝子は 4 遺伝子全てであった。したがって本抽出法により、概日リズムとの関連が既知の遺

伝子が実際に 4 遺伝子抽出でき、これら全ての遺伝子が Dbp 遺伝子との関連が報告され、この結果は、本抽出法の技術的な妥当性を示すものとする。

次に、C-1A で得た肝での経時データベースから得た 81 ps (Dbp を入れて 83ps) の遺伝子リストと、C-1B で得た SCN での 44 ps (Dbp を入れて 46ps) の遺伝子リストとを比較・検討した。結果を図 3 にベン図として示す。遺伝子リストは両者でかなり異なり、両者に共通して含まれていた遺伝子は、Dbp 以外では、Nr1d2 (nuclear receptor subfamily 1, group D, member 2; 別名: Rev-ErbA beta) の 1 遺伝子のみであった。したがって、Dbp 遺伝子と似た発現パターンを示す遺伝子は、SCN と肝とでは、かなり異なる事が明らかとなった。

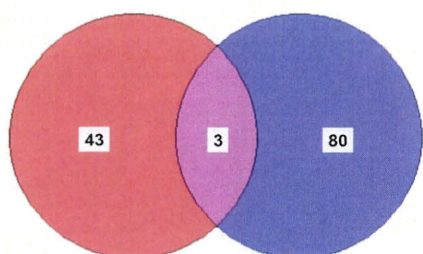


図 3 Dbp 遺伝子発現と似た経時変化を示す遺伝子数について、SCN における遺伝子発現の経時データベースを用いた場合 (赤) と肝における遺伝子発現の経時データベースを用いた場合 (青) の比較

C-1C: SCN 及び肝について得られた Dbp 遺伝子と似た発現パターンを有する遺伝子リストについての、*in silico* でのプロモーター解析:

この Dbp 遺伝子と似た経時変化を示す遺

伝子が同じシグナルネットワークを共有しているか否か、を検討する為に、各遺伝子の転写開始点上流での共通した結合配列の有無を *in silico* でのプロモーター解析により検討した (市販のソフト: Genomatix を使用)。

肝では、83ps、すなわち 65 遺伝子全てのプロモーターに共通する転写因子結合配列は見いだせなかったが、64 遺伝子のプロモーターに共通する転写因子結合配列は見いだされ (V\$KLFS)、結合する転写因子は KLF (Krueppel like transcription factors) のみであった。

SCN では、全 44 ps、すなわち 40 遺伝子のプロモーターに共通する転写因子結合配列は見いだせなかったが、39 遺伝子のプロモーターに共通する転写因子結合配列は見いだされ (V\$EGRF、V\$ETSF 及び V\$KLFS) であった。結合する転写因子は EGR family (EGR/nerve growth factor induced protein C & related factors), ETS1 factors 及び、KLF (Krueppel like transcription factors) であった。

したがって肝、SCN 双方共において、ほぼ全ての遺伝子リストに共通する結合配列が見いだされたことから、これらの遺伝子が Dbp 遺伝子の関わる局所のネットワークと同一のネットワークを利用している事が示唆された。また SCN と肝に共通して結合する配列として KIF 結合配列が見いだされた。ともに、SCN では肝とは異なる結合配列を有しており、このことが肝と SCN とで、遺伝子リストが異なる理由と考えられる。

C-2: Percellome データベースを利用した、Dbp 遺伝子と似た発現パターンを有する遺

伝子リストの妥当性の検証:

肝において得られた Dbp 遺伝子と似た経時変化を示す 81 ps (64 遺伝子) が、Dbp 遺伝子の局所ネットワーク上に存在するかどうかを検証するために、上記の *in silico* でのプロモーター解析とは別の手法により検討した。すなわち、当毒性部が所有する 100 化合物を超える Percellome データベースを利用し、肝において Dbp 遺伝子の発現を抑制する化学物質の探索を経て、その際の経時データベースを活用する手法により検討した。これらの遺伝子が同一の局所ネットワーク上に存在する場合には、得られた 64 遺伝子の経時変化は Dbp 遺伝子と同様に発現が抑制されるはずである。

肝において Dbp 遺伝子の発現を強く抑制する化学物質を検索したところ、化審法での所謂白物質である 2-クロロ-4,6-ジメチルアニリン[0、3、10、30 mg/kg]、有機リン系殺虫剤であるアセフェート[0、7、20、70 mg/kg] 及び、食品添加物の保存料であるデヒドロ酢酸ナトリウム[0、30、100、300 mg/kg] の 3 化合物であった (いずれも単回強制経口投与で、溶媒は 0.5%メチルセルロース)。この 3 化合物について、Dbp 遺伝子の発現変化を図 4 に示す。図は下記のように、各遺伝子につき濃度依存性、経時変化、遺伝子発現量についての 3 次元グラフとして示した。具体的には、縦軸(Z 軸)に絶対値化した (細胞 1 個あたりのコピー数) mRNA の発現量を取り、X、Y 軸にはそれぞれ、投与用量と投与後経過時間を取り、各条件の n=3 の平均値曲面で表示する。加えてこの



平均曲面の上下に標準偏差(SD)平面 (薄い色) を示す。

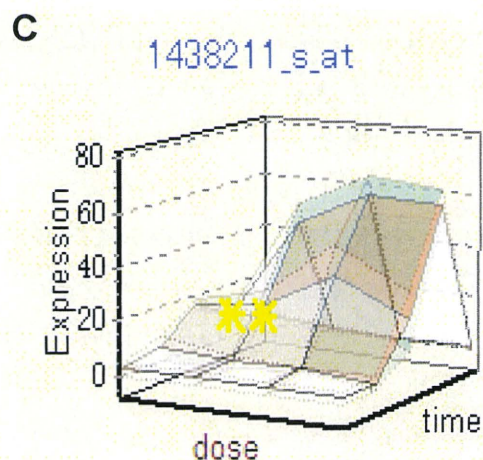
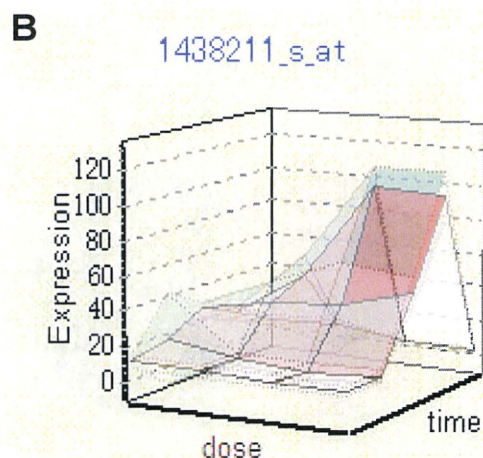
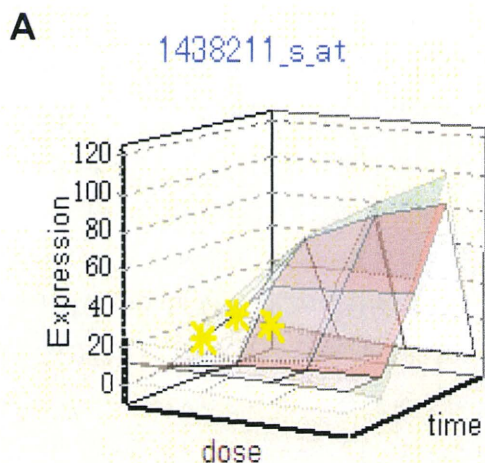
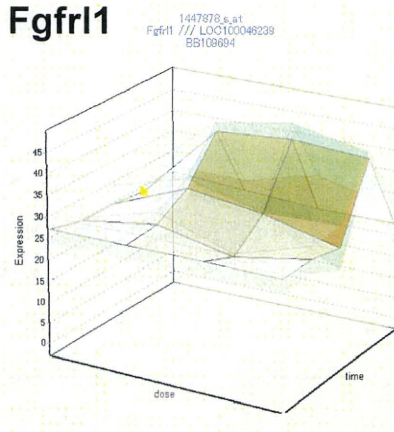


図4 2-クロロ-4,6-ジメチルアニリン(A)、アセフェート(B)及びデヒドロ酢酸ナトリウム(C)を単回強制経口投与した際のマウス肝における Dbp 遺伝子の発現変動。溶媒群と投与群の間の有意差の検定を Student の t 検定によりおこないP 値が0.05 未満の場合を有意と判定し、図中に★を付した。

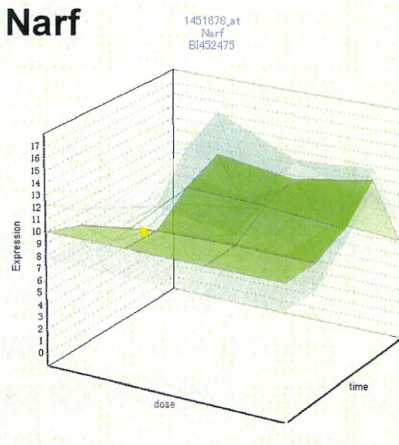
1000化合物を超える Percellome データベースの中で、この3化合物は日内変動を示す Dbp 遺伝子の発現を用量依存的に顕著に抑制した。

3化合物の内、2-クロロ-4,6-ジメチルアニリン投与した場合について検討したところ、肝において得られた Dbp 遺伝子と似た経時変化を示す 81 ps のうち、63 ps、すなわち約8割の ps の発現が抑制された。したがって、少なくともこの63 ps の遺伝子は、Dbp と同一の局所ネットワーク上に存在することが示唆された。例として、Fgfr11、Narf 及び Gne 遺伝子の発現の発現変化を図5に示す。

Fgfr11



Narf



Gne

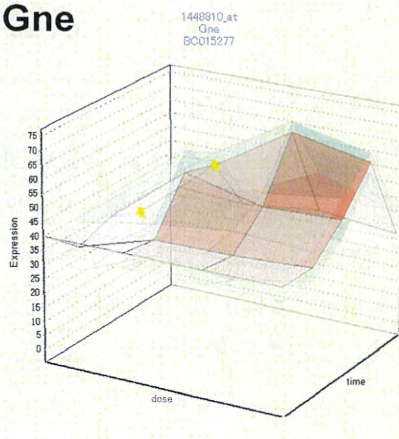


図5 2-クロロ-4,6-ジメチルアニリンを単回経口投与した際のマウス肝における Fgfr11、Narf 及び Gne 遺伝子の発現変動。溶媒群と投与群間の有意差検定を Student の t 検定によりおこないP 値が0.05 未満の場合を有意と判定し、図中に★を付した。

D. 結論

本研究の目的は、毒性分子メカニズムに依拠した毒性予測評価システムの実用化の最終段階としてのインフォマティクス技術を開発することである。今年度の本分担研究では、取得・構築済みの、概日変動リズム1周期(=24時間)の成熟期マウスの肝および視交叉上核(SCN)サンプルの網羅的トランスクリプトームデータ[Time point: 7点(4時間毎)]を利用し、概日変動リズム関連遺伝子のひとつ Dbp 遺伝子に着目し、この局所シグナルネットワークの抽出を目的とし検討した。

概日変動リズム1周期(=24時間)の成熟期マウスの肝およびSCNの遺伝子発現経時データベースを用い、Dbp 遺伝子の発現変動をみたところ、肝では点灯4時間後から、SCNでは点灯20時間後から、急速に発現増加する一過性の発現パターンが観察された。このことから両組織において Dbp 遺伝子の発現変動は経時的に一致しないことが明らかとなった。Dbp 遺伝子の発現の経時変化を基としたピアソン相関係数を利用し(α 値0.9以上)の発現と同様な遺伝子を抽出、さらに目視による検討により生物学的な変化が示唆されたものとして、肝では81プロンプセット(ps)(64遺伝子)、SCNでは44ps(39遺伝子)の遺伝子リストが得られた。このうち、概日リズムとの関連が既知の遺伝子は肝では6遺伝子、SCNでは4遺伝子であった。この結果は、本解析の技術的な妥当性を示すと共に、当該遺伝子ネットワークに関連する既知の遺伝子の抽出だけでなく、機能未知の遺伝子の探索を効率よくおこなえるものと考えられた。また両リストを比較したところ、Nr1d2 遺伝子

(Rev-ErbA beta)のみ両遺伝子リストに共通していたことから、肝とSCNでDbp局所ネットワークの発現制御がかなり異なる事が示唆された。

得られたリストに含まれる遺伝子が Dbp 遺伝子局所ネットワークに含まれるか否か、また肝とSCNとで発現制御が異なるか否か、を明らかにする目的で、両リストに含まれる遺伝子の転写開始点上流に共通した結合配列が存在するか否かを、*in silico*でのプロモーター解析により検討した。その結果、肝では抽出されてきた65遺伝子中64遺伝子にKLFの結合配列が、SCNでは抽出されてきた40遺伝子中39遺伝子に、EGR family、ETS1及びKLFの結合配列が見いだされた。このことは、得られた両遺伝子リストに含まれる遺伝子は Dbp と同じ局所ネットワークに存在する事、及び Dbp 遺伝子の関与する局所ネットワークの発現制御は肝とSCNでは異なるが、KLFが関与する制御は両組織で共通である事が示唆された。転写因子KLFと概日リズムとの関連は、文献検索の結果、ほとんど報告は見いだせなかったが、Guillaumondら(Mol Cell Biol 30(12):30, 3059-3070, 2010)の1報を見いだす事が出来た。この論文では、KLF10遺伝子を欠失したマウスを作製し、KLF10が肝エネルギー代謝の概日リズムと関連する転写因子である事を示している。我々の解析結果では、SCNと肝の双方共ににおいてKLF結合配列が Dbp 遺伝子に関連するネットワークを制御している事が示唆され、加えて肝ではKLF結合配列だけが見いだせた事から、概日リズムにおけるKLFの関与は大きいものと考えられた。今後、この観点から詳細に検討する必要があると考える。