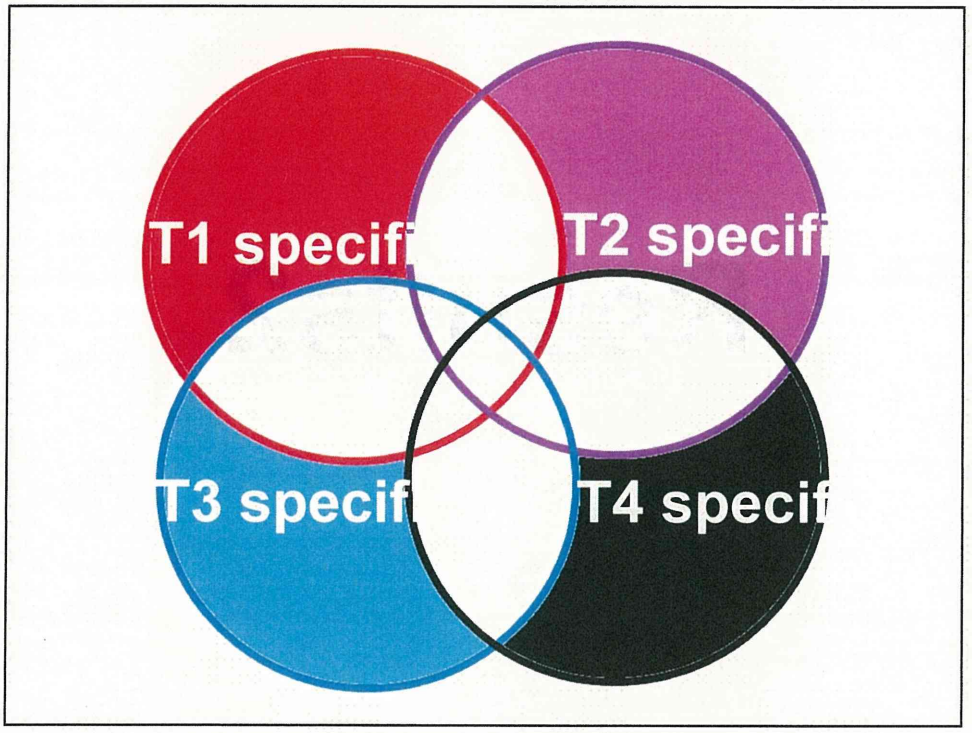
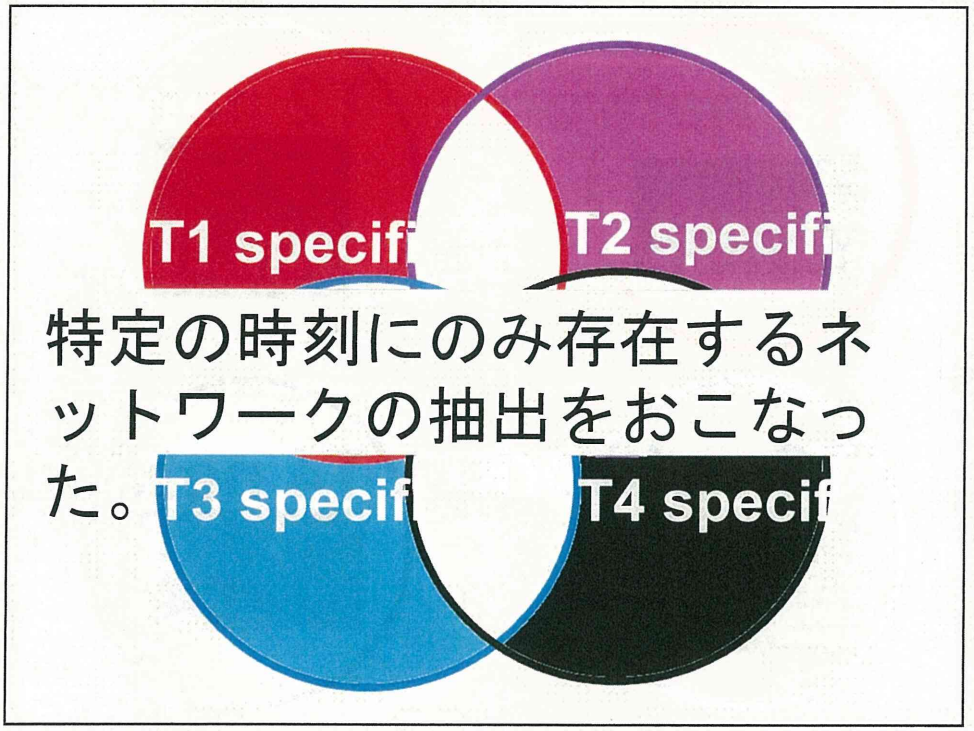
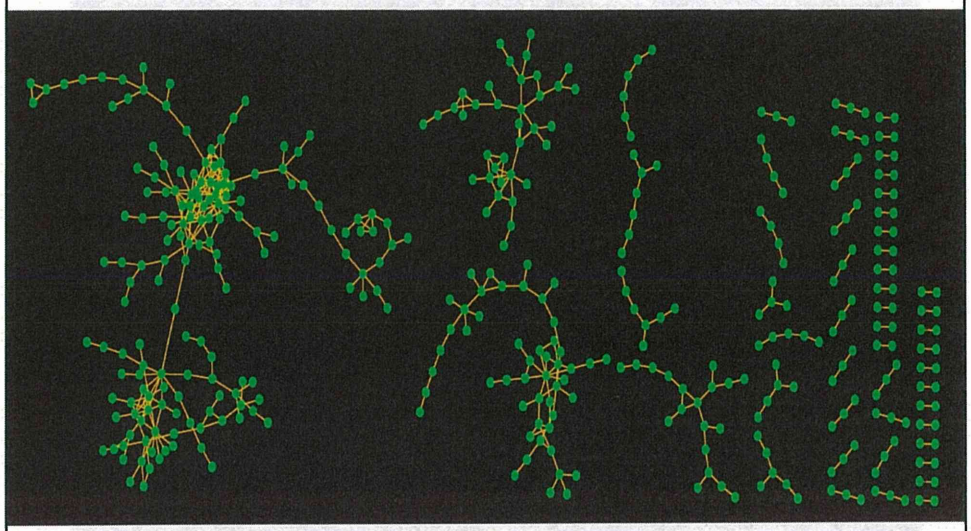
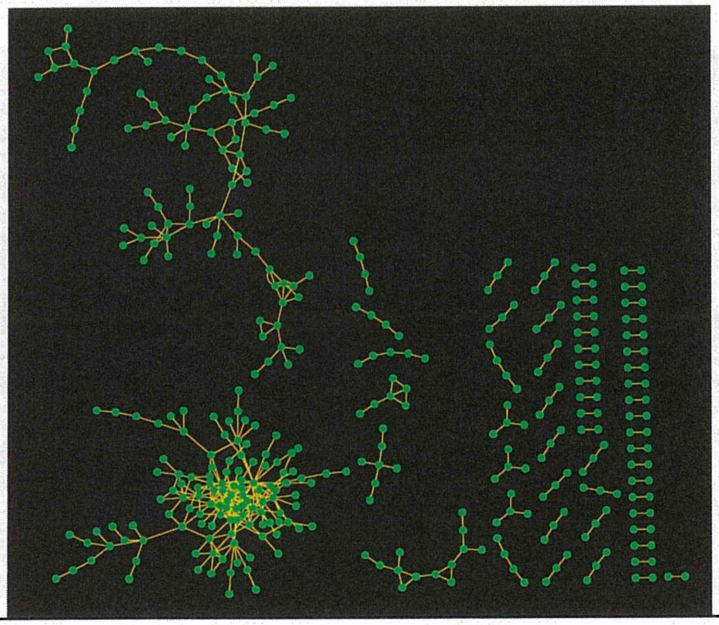

T1 specif
T2 specif
T3 specif
T4 specif


T1 specif
T2 specif
特定の時刻にのみ存在するネットワークの抽出をおこなった。
T3 specif
T4 specif

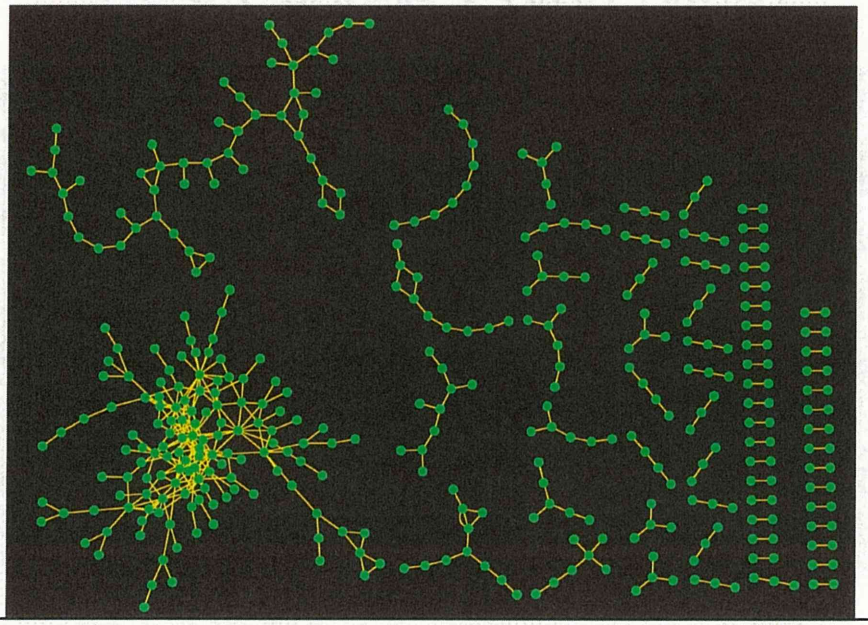
T1 specific gene regulatory network
（時刻T1のみで現れるリンク）



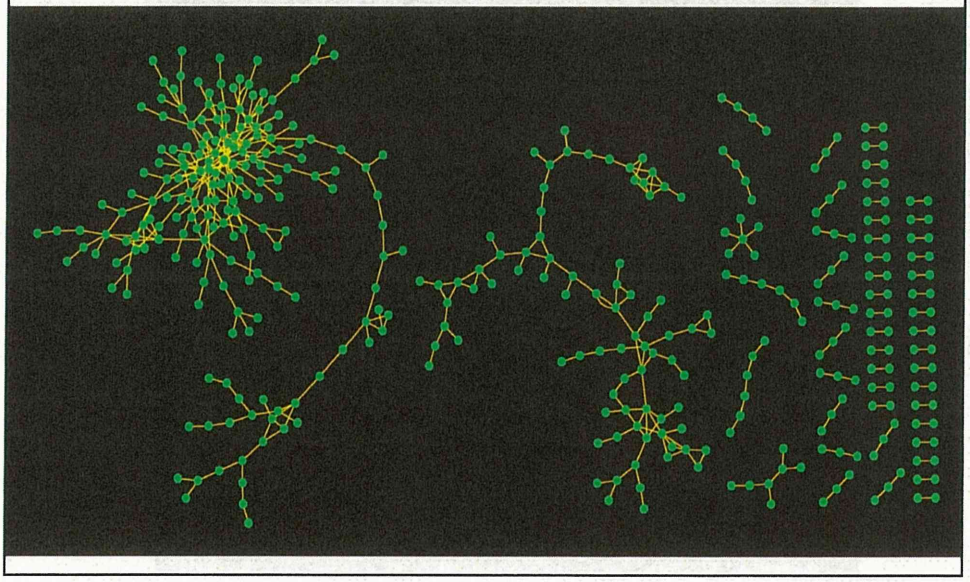
T2 specific gene regulatory network



T3 specific gene regulatory network

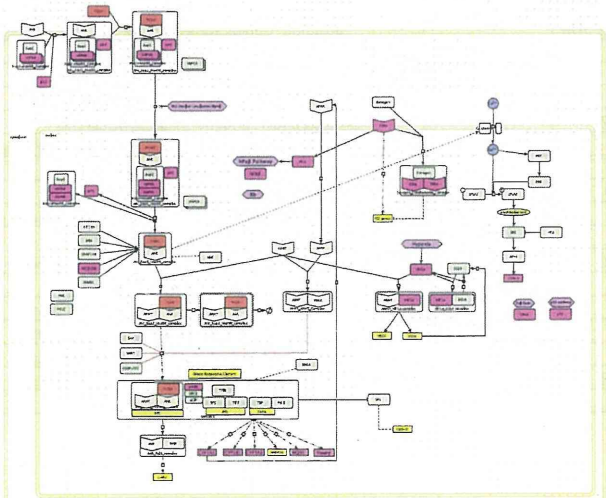


T4 specific gene regulatory network



Results3

パスウェイ情報を用いて重要な遺伝子セットを同定する。



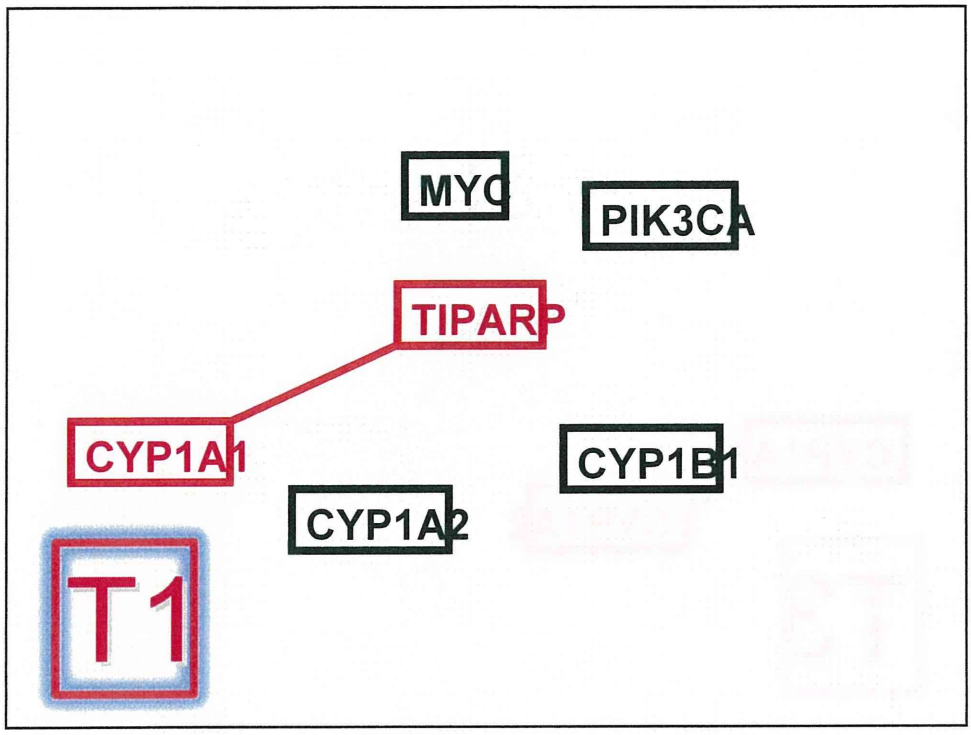
パスウェイ情報を用いて重要な遺伝子セットを同定する。

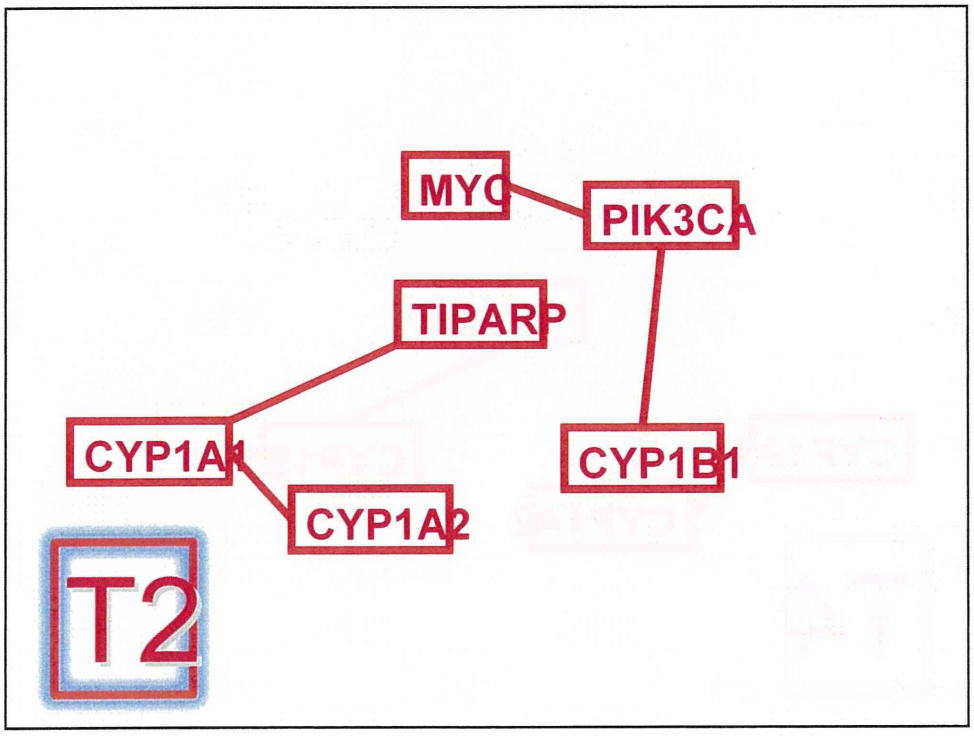
これらの遺伝子間のリンクが、TCDD投与後の時間経過に応じて、どのように変化するのかを探索する。

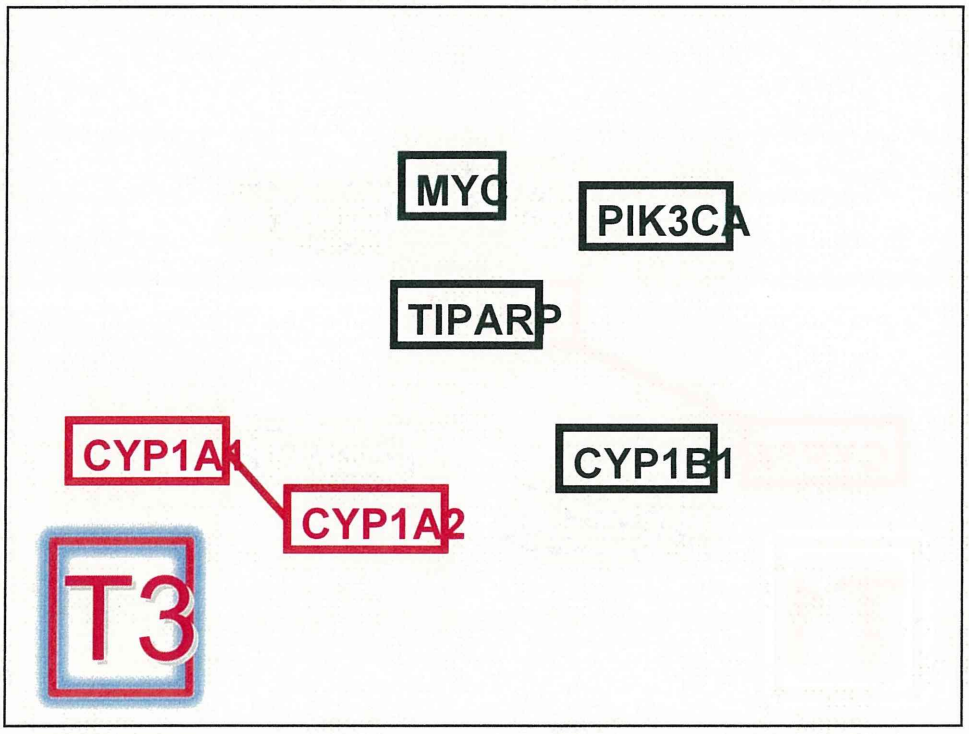
以下、時間変化に応じたネットワークの変化の例を示す。

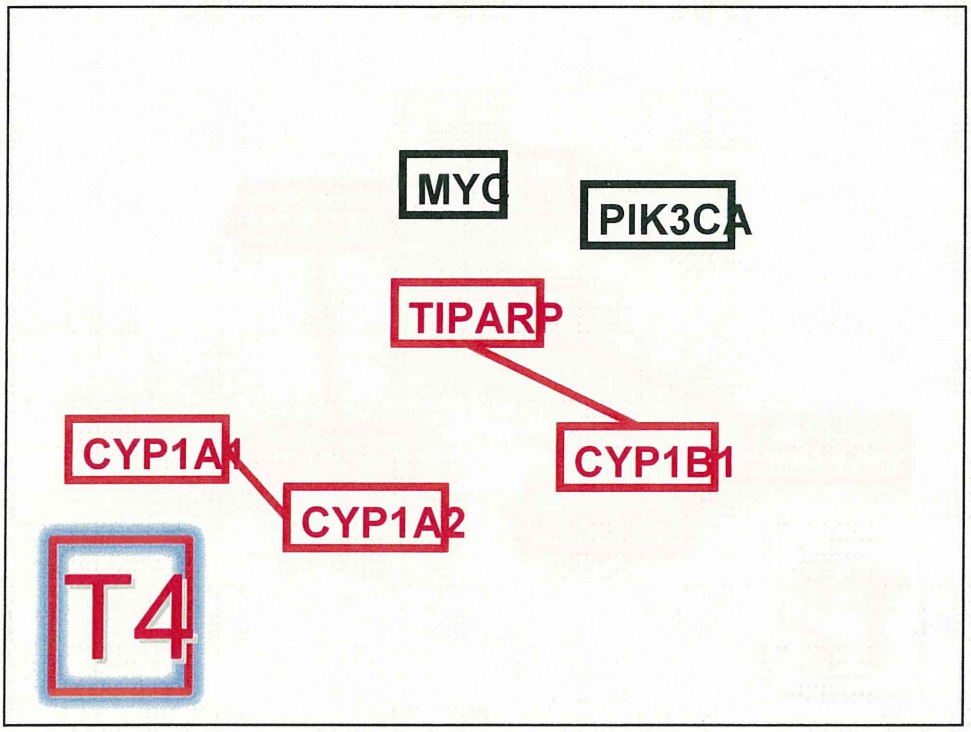
17 important genes

Symbol	Synonym	Description
UBE2J1	UBE2M=NEDD8	ubiquitin-conjugating enzyme E2 J1
TRIAP1	p53	TP53 regulated inhibitor of apoptosis 1
TIPARP		TCDD-inducible poly(ADP-ribose) polymerase
RRAS2	p23	related RAS viral (r-ras) oncogene homolog 2
PTGS2	COX2	prostaglandin-endoperoxide synthase 2
PIK3CA		phosphatidylinositol 3-kinase catalytic alpha polypeptide
nqo1		NAD(P)H dehydrogenasequinone 1
notch1	p300	Notch gene homolog 1 (Drosophila)
NFKBIA		nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor alpha
MYC		myelocytomatosis oncogene
HSP90AA1	HSP90	heat shock protein 90 alpha (cytosolic) class A member 1
esrrg		estrogen-related receptor gamma
Epas1		HIF-1 alpha-like factor endothelial PAS domain protein 1
CYP1B1		cytochrome P450 family 1 subfamily b polypeptide 1
Cyp1a2		cytochrome P450 family 1 subfamily a polypeptide 2
CYP1A1	AHRR	cytochrome P450 family 1 subfamily a polypeptide 1
Cdk2		cyclin-dependent kinase 2


MYC
PIK3CA
TIPARP
CYP1A1
CYP1B1
CYP1A2
T1


MYC
PIK3CA
TIPARP
CYP1A1
CYP1B1
CYP1A2
T2


MYC
PIK3CA
TIPARP
CYP1A1
CYP1B1
CYP1A2
T3


MYC
PIK3CA
TIPARP
CYP1A1
CYP1B1
CYP1A2
T4

Application of network reconstruction technique to Percellome

- 投与量に依存した、遺伝子間制御ネットワークの構造変化の探索。
- 組織特異的な遺伝子間制御ネットワークの構築
- ある投与量と組織において、遺伝子間制御ネットワークが時間とともにどのように変化するのか？

班会議平成２４年２月９日

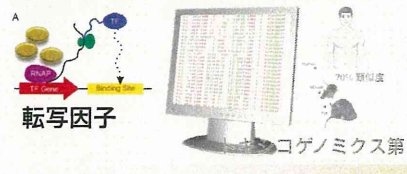
クラスタリングとプロモータ解析に よりネットワーク検出のアプローチ

株式会社ソニーコンピュータサイエンス研究所
ポリュリャーフ・ナターリア

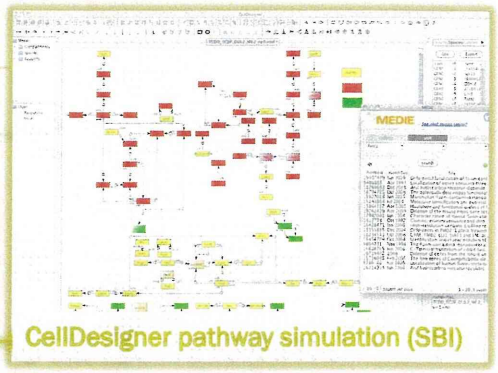
Agenda

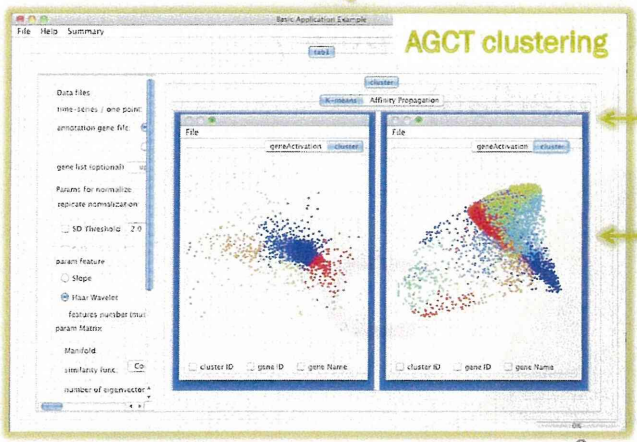
* **AGCT** A Geometric Clustering Tool (from 2008)
  * Clustering Percellome data based on similarity of gene expression profile. Application on TCDD and TCDF (2,3,7,8-Tetrachlorodibenzo-p-dioxin and 2,3,7,8-Tetrafuran) chemicals.
    * Data normalization
    * Unsupervised gene clustering
    * Sorting clusters upon validity
    * Issues remained

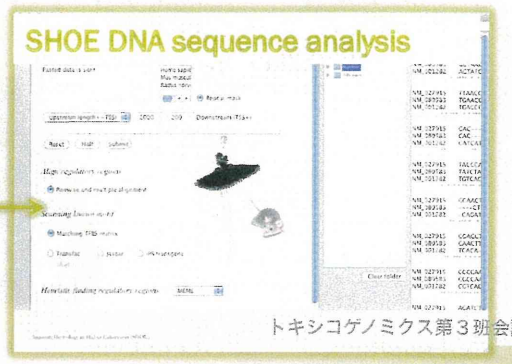
* **SHOE** Sequence Homology in Higher Eukaryote
  * Phylogenetic footprinting for discovery of transcription regulation network (from 2011)



Pecellome analysis pipeline







Plugged-in

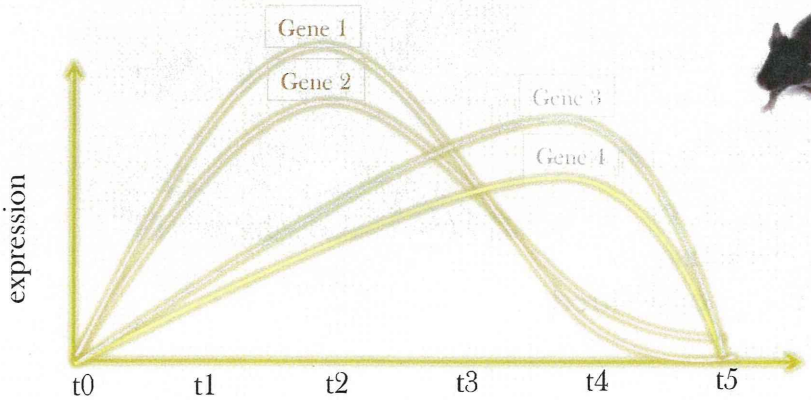
トキシコゲノミクス第３班会議

AGCT A Geometric Clustering Tool

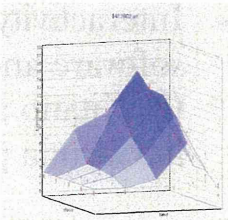
Atlas of Cell Life by AGCT

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

- chemical
- doze
- time
- expression



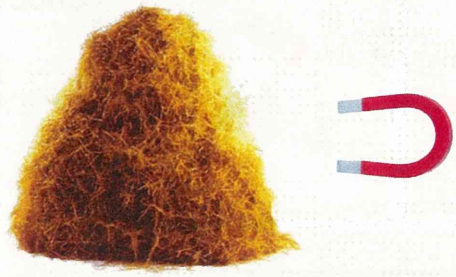
40,000 probes



トキシコゲノミクス第３班会議

データの現状

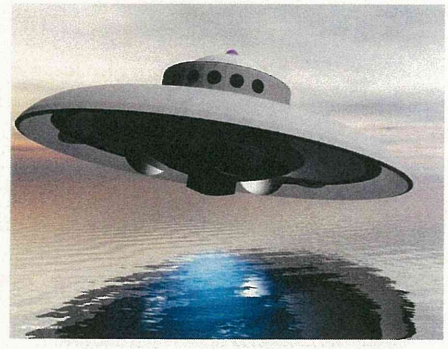
* 全発現プロファールが４万プローブを含む
  * 初期反応遺伝子（responsive transcription)
  * Circadian 遺伝子 (basal transcription)
  * 無反応遺伝子（not circadian)
  * 実験誤差



GARUDA
THE WAY BIOLOGY CONNECTS

Software: robust (independent but interactive)

* Fast
* Work in uncertainty
* Macro/micro flexible
* Rich and interactive visualization
* Approximate/generate parameters
* Rapid knowledge annotation
* Interactivity with other software and database (H.Kitano already explained in Garuda project)



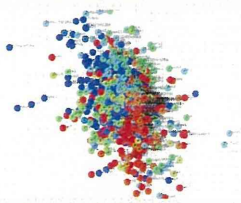
AGCT

A Gene Geometric Clustering Tool

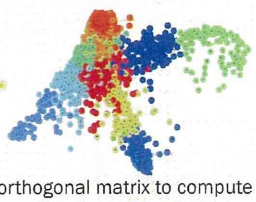
Processing data on AGCT

1. 時系列データ前処理:線形回帰/ウェーブレット変換
2. 遺伝子間の類似度マットリックス
3. 低次元に落とすためにSpectral clusteringを行う。通常の主成分分析も行う。
4. 発見的なClustering法を使って構造上でデータの分割を行う。
5. 結果のinteractive visualizationやscenario 記録を行う。

PCA : M*N matrix



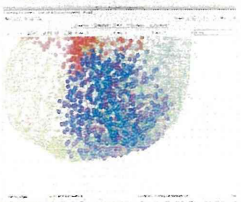
Spectral clustering:
M*M matrix



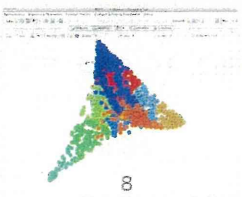
orthogonal matrix to compute
one dimension per cluster/gene

Examples of different network topologies

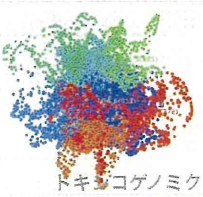
Mouse Stem cell



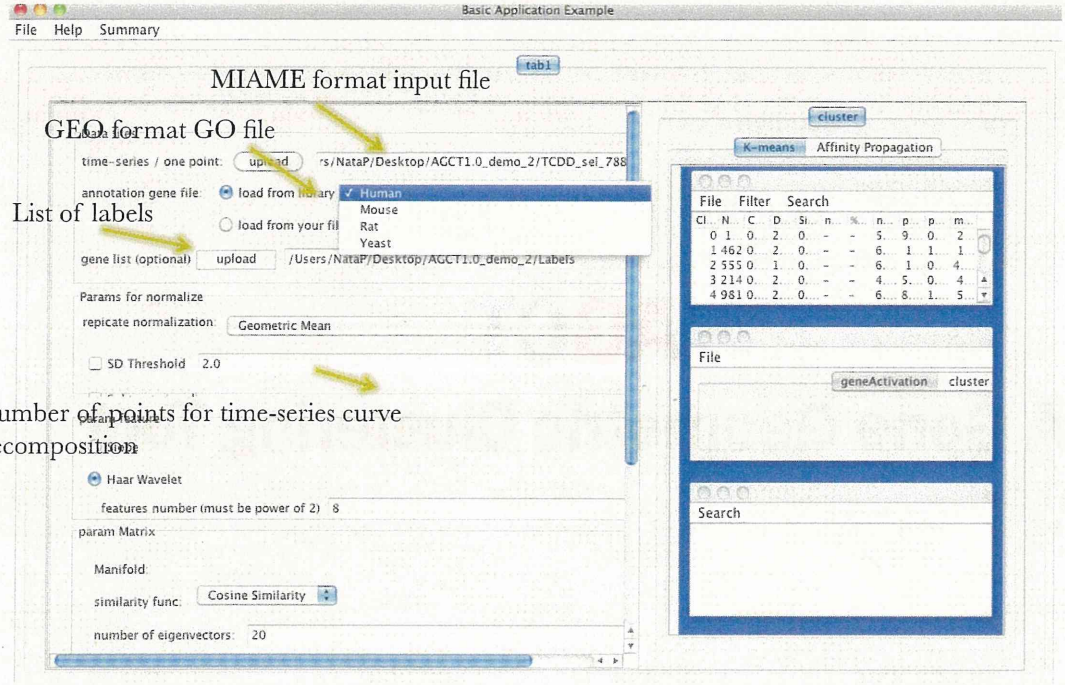
TCDD affected mouse liver cell



Influenza affected mouse bronchi cell



AGCT interface 1



Input file format (tab-delimited .txt)

5*1*4*3=60 columns

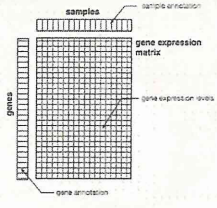
Groups to specify

Species	Mouse430_2.txt								/* Constants	
E (entity)	5	Doze0	Doze0.1	Doze0.3	Doze10	Doze30				
EG (entity group	1	1	1	1	1	1			/* Groups	
T (time)	4	2	4	8	24					
R (replicate)	3									
DATA	7881									
1426408_at	13.68638	13.74065	10.62874	10.29534	12.14933	13.41148	16.33447	16.8473	17.75223	14.50626
1430123_a_at	329.4958	304.0648	317.0229	320.1172	314.5138	331.6112	273.2262	351.2784	337.1702	303.972
1427909_at	79.69938	65.37696	76.33909	74.55177	67.03818	75.51203	79.76383	76.7676	66.0889	65.07177
1449231_at	1.998361	0.6321552	1.147684	0.8781132	1.167874	2.397003	2.210826	0.4704294	2.583703	0.3943034
1419477_at	23.11418	13.63653	16.88693	7.034207	10.33751	11.61575	10.32749	9.681456	14.54885	8.335682
1430115_at	4.780659	4.343996	3.392105	2.823855	3.745805	4.923362	2.997585	3.623773	5.434154	3.351972
1433236_at	0.6044968	2.002586	1.370757	0.2526424	0.247167	0.2072143	0.4506208	1.558007	0.9185823	0.3346251
1448996_at	1.737852	1.323068	0.7732058	1.240441	1.236961	0.4244971	0.6674667	0.6260773	0.9274884	0.4314171
1458025_at	3.499104	2.880943	3.601012	4.745797	4.247523	4.485948	4.44686	3.311373	4.278707	5.641602

© 2001 Nature Publishing Group http://genetics.nature.com

commentary

Minimum information about a microarray experiment (MIAME)—toward standards for microarray data

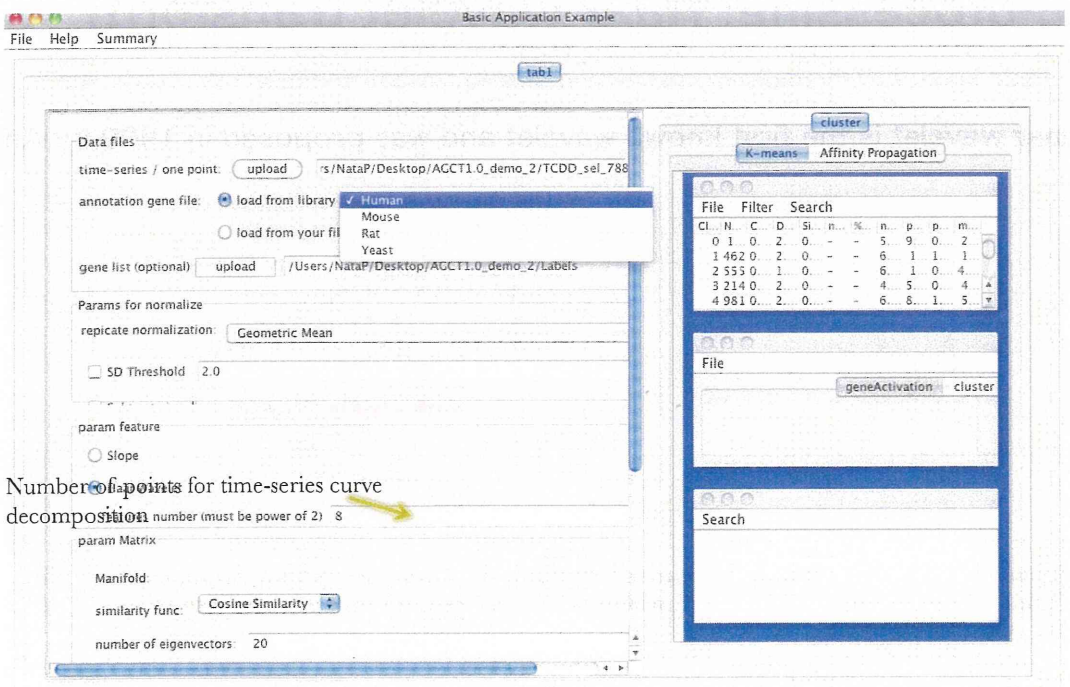


Gene ontology file

1429433_at	BI083827	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:BI083827 /DE=	BI083827	BAT2 domain co	Bat2d	226562	NM_001081290	Not_defined	Not_defined	Not_defined
1429434_at	BE647269	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK018164.1 /DE=	BE647269	phosphatidylinos	Pik3ca	18706	NM_008839	0006006 // glu	0005942 // phos	0004428 // inositol or phosphatid
1429435_x_at	BE647269	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK018164.1 /DE=	BE647269	phosphatidylinos	Pik3ca	18706	NM_008839	0006006 // glu	0005942 // phos	0004428 // inositol or phosphatid
1429436_at	BB626803	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK012731.1 /DE=	BB626803	PRP40 pre-mRN	Prpf40a	56194	NM_018785	0006397 // mRN	0005634 // nuc	0005515 // protein binding // inf
1429437_at	BB626803	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:BB626803 /DE=	BB626803	PRP40 pre-mRN	Prpf40a	56194	NM_018785	0006397 // mRN	0005634 // nuc	0005515 // protein binding // inf
1429438_at	AV318805	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AV318805 /DE=	AV318805	BCL6 interactin	Bcor	71458	NM_029510	0006350 // tran	0005634 // nuc	0003714 // transcription corepre
1429439_at	BF020657	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:BB380720 /DE=	BF020657	excision repair	Ercc6	71991	NM_028042	0000209 // pro	0000109 // nuc	0004842 // ubiquitin-protein liga
1429440_at	BI734299	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:BI734299 /DE=	BI734299	RIKEN cDNA 18	1810041L15Rik	72301	XM_128189 ///	Not_defined	0016020 // mem	Not_defined
1429441_at	AK006369	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:BI686465 /DE=	AK006369	F-box protein 3	Fbxo30	71885	NM_027969	0006511 // ubiq	Not_defined	0008270 // zinc ion binding // m
1429442_at	AA560280	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:BB125490 /DE=	AA560280	RNA pseudouridi	Rpusd2	271842	NM_173450	0001522 // pseu	Not_defined	0003723 // RNA binding // infer
1429443_at	AK014396	Not_defi								74020	NM_028719	Not_defined	Not_defined	Not_defined
1429444_at	AK004971	Not_defi								68885	NM_028864	0007264 // sma	0005622 // intr	0000166 // nucleotide binding //
1429445_at	AK014661	Not_defi								70994	NM_001081046	Not_defined	Not_defined	0005509 // calcium ion binding //
1429446_at	BQ229980	Not_defi								66244	NM_025441	Not_defined	0005634 // nuc	Not_defined
1429447_at	AK003753	Not_defi								68525	NM_145920	Not_defined	0016020 // mem	Not_defined
1429448_s_at	BM240219	Not_defi								52463	XM_125673 ///	Not_defined	Not_defined	Not_defined
1429449_at	BF319665	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:BF319665 /DE=	BF319665	sterile alpha mo	Samd4	74480	NM_001037221	0045727 // posi	0019717 // synt	0030371 // translation repressor
1429450_at	AK006386	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK006386.1 /DE=	AK006386	RIKEN cDNA 17	1700026J04Rik	71864	Not_defined	Not_defined	Not_defined	Not_defined
1429451_at	AK011950	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK011950.1 /DE=	AK011950	RIKEN cDNA 26	2610301B20Rik	67157	NM_026005	Not_defined	Not_defined	Not_defined
1429452_x_at	AK007420	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK009333.1 /DE=	AK007420	RIKEN cDNA 49	4933439C20Rik	68778	NR_003518	0006654 // pho	Not_defined	0004909 // phosphatidylserine de
1429453_a_at	BM209870	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:BM209870 /DE=	BM209870	mitochondrial rib	Mrps5	67212	NM_026035	0006412 // tran	0005739 // mito	0003735 // structural constituen
1429454_at	BM123170	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK014474.1 /DE=	BM123170	GTPase activati	Gapvd1	68691	NM_026709	0006897 // end	0005622 // intr	0005085 // guanyl-nucleotide ex
1429455_at	BM123170	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:BB795549 /DE=	BM123170	GTPase activati	Gapvd1	68691	NM_026709	0006897 // end	0005622 // intr	0005085 // guanyl-nucleotide ex
1429456_a_at	AV251548	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK005034.1 /DE=	AV251548	polymerase (RN	Polr3e	29939	NM_025298	0006350 // tran	0005634 // nuc	0003899 // DNA-directed RNA p
1429457_at	BB001825	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:BB001825 /DE=	BB001825	prosaposin-like	Psapl1	76943	NM_175249	0006629 // lipid	0005576 // extr	Not_defined
1429458_at	AV083806	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AV083806 /DE=	AV083806	RIKEN cDNA 24	2410127L17Rik	67583	5775	NM_026120 ///	Not_defined	Not_defined
1429459_at	BB499147	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AI451684 /DE=	BB499147	sema domain, im	Sema3d	108151	NM_023682	0007275 // mul	0005576 // extr	0004872 // receptor activity //
1429460_at	AK019508	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK019508.1 /DE=	AK019508	G protein-coupl	Gpr115	78249	XM_001476141	0007165 // sign	0005886 // plas	0004871 // signal transducer act
1429461_at	AK013101	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK013101.1 /DE=	AK013101	integrator compl	Ints2	70422	NM_027421	0016180 // snR	0005622 // intr	0005515 // protein binding // inf
1429462_at	AK011759	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:BE374892 /DE=	AK011759	solute carrier fa	Slc25a32	69906	NM_172402	0006810 // tran	0005739 // mito	0005215 // transporter activity /
1429463_at	BB612385	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AI891575 /DE=	BB612385	protein kinase,	Prkaa2	108079	NM_178143	0006468 // prot	0005634 // nuc	0000166 // nucleotide binding //
1429464_at	BB612385	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:BB612385 /DE=	BB612385	protein kinase,	Prkaa2	108079	NM_178143	0006468 // prot	0005634 // nuc	0000166 // nucleotide binding //
1429465_at	AK005847	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK005847.1 /DE=	AK005847	RIKEN cDNA 17	1700010M22Rik	68328	NM_025490	0008152 // met	Not_defined	0005488 // sterol carrier activity
1429466_s_at	AK002310	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK002310.1 /DE=	AK002310	anterior pharynx	Aph1b ///	Aph1208117	/// 883	NM_026674 ///	0001656 // met	0004175 // endopeptidase activi
1429467_s_at	BB118542	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:BB118542 /DE=	BB118542	solute carrier fa	Slc26a3	13497	NM_021353	0006810 // tran	0005887 // int	0005215 // transporter activity /
1429468_at	AK003775	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK003775.1 /DE=	AK003775	RIKEN cDNA 11	1110018F16Rik	68594	Not_defined	Not_defined	Not_defined	Not_defined
1429469_at	BB498937	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:BB498937 /DE=	BB498937	RIKEN cDNA A6	A930013F10Rik	68074	Not_defined	Not_defined	Not_defined	Not_defined
1429470_at	AK014910	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK014910.1 /DE=	AK014910	speedy homolog	Spdya	70891	NM_001142631	0000082 // G1/	0005634 // nuc	0019901 // protein kinase bindin
1429471_at	AK003742	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK003742.1 /DE=	AK003742	RIKEN cDNA 11	1110017D15Rik	73721	NM_001048005	0043623 // cell	0005737 // cyto	0043014 // alpha-tubulin binding
1429472_at	AK014969	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK014969.1 /DE=	AK014969	RIKEN cDNA 49	4921525O09Rik	74050	Not_defined	Not_defined	Not_defined	Not_defined
1429473_at	AV292561	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AV292561 /DE=	AV292561	RIKEN cDNA 12	1200003I07Rik	66969	NM_001039985	0006355 // reg	0005622 // intr	0003676 // nucleic acid binding
1429474_at	BE283373	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:BE283373 /DE=	BE283373	prostaglandin-e	Ptgr2	77219	NM_029680	0008152 // met	0005737 // cyto	0003824 // catalytic activity //
1429475_at	AK013361	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:AK013361.1 /DE=	AK013361	similar to RIKEN	LOC677224 ///	677224	/// 728	NM_176960 ///	Not_defined	0005634 // nuc	Not_defined
1429476_s_at	BG063818	Not_defined	Mus musculus	11-Mar-09	Consensus seqs GenBank	gb:BG063818 /DE=	BG063818	DnaJ (Hsp40) ho	Dnaja2	56445	NM_019794	0006457 // prot	0016020 // mem	0008270 // zinc ion binding // m

Use NCBI GEO annotation file as is,
but substitute blank fields with 'Not_defined'

AGCT interface 1



Number of points for time-series curve decomposition

java -jar -Xmx1G(40G) -Xms1G(40G) AGCT3.jar

Pre-processing: Slopes

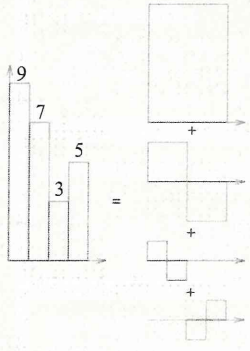
A ***time series*** for gene **i** and ligand **j** is mapped to a slope s_{ij} using a conventional linear regression fit:

$$Sij = \frac{\mathrm{cov}(t, x_{ij})}{\mathrm{var}(t)} = \frac{\Sigma_k (t_k - avg(t))(x_{ijk} - avg(x_{ij}))}{\Sigma_k (t_k - avg(t))^2}.$$

Here, "avg" denotes the average, "var" is the variance and "cov" is the covariance; furthermore, *t* denotes the set of time stamps and x_{ij} denotes the set of measurements for gene *i* and ligand *j*. *k* spans {2, 4, 8, 24}.

Pre-processing: Haar wavelet transform

The Haar wavelet is the first known wavelet and was proposed in 1909 by Alfred Haar.



One-dimentional Haar wavelet transform

Resolution	Averages	Detail coefficients
4	[9 7 3 5]	
2	[8 4]	[1 −1]
1	[6]	[2]

wavelet transform (filter bank)

[(6) 2 1 −1]

basis coefficients

Figure 1: Decomposition of a small histogram (left) into a set of four Haar wavelets components (right). Taking a smooth curve instead of an histogram would necessitate much more coefficients to get an appropriate fitting.

Using the vector space from linear algebra

We can think think of histogram as piecewise-constant functions on the half-open intervals [0, 1].

We'll let V^0 be the vector space of all functions, i.e., block function:

$$\phi(t) = \begin{cases} 1 & \text{if. } t \in [0,1) \\ 0 & otherwise \end{cases}$$

and mother wavelet :

$$\psi(t) = \begin{cases} 1 & \text{if } t \in [0,1/2) \\ -1 & \text{if } t \in [1/2,1) \\ 0 & otherwise \end{cases}$$

The space V^j will include all piecewise constant functions defined on the interval [0, 1] with constant pieces over each of 2^j equal subintervals. Thus, the spaces V^j are nested as:

$$V^0 \subset V^1 \subset V^2 \subset \cdots$$

Any curve can be obtained (eventually up to any finite precision) by the sum of translated/dilated curves that are born from (2), (3):

basis for V^j

$$\phi_i^j = \sqrt{2^j}\phi(2^j t - i) \ , j \in \mathbb{N}, i = 0, 1, ..., 2^j - 1 \quad (4)$$

$$\psi_i^j = \sqrt{2^j}\psi(2^j t - i) \ , j \in \mathbb{N}, i = 0, 1, ..., 2^j - 1 \ . \quad (5)$$

What we want to do is to compute each time serie as a sum of such coefficients, and more precisely, as a sum of the form $\alpha\phi_0^0 + \sum_{i,j} \beta_i^j \psi_i^j, \alpha \in \mathbb{R}, \beta_i^j \in \mathbb{R}$. We have normalized the functions ($\int_0^1 f^2(t)dt = 1$) for orthonormality reasons, but if we had chosen to remove the square-roots above, ϕ_0^0 would just be the global average of the curve. Since the mother wavelet is the curve which gives an indication of the slopes at each scale-dilation factors, this first "average" coefficient is removed from the analysis.

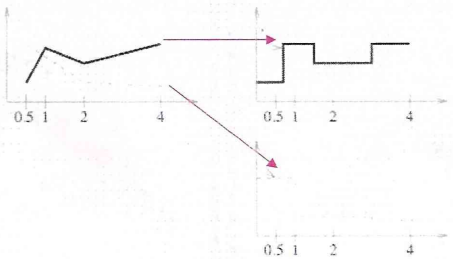
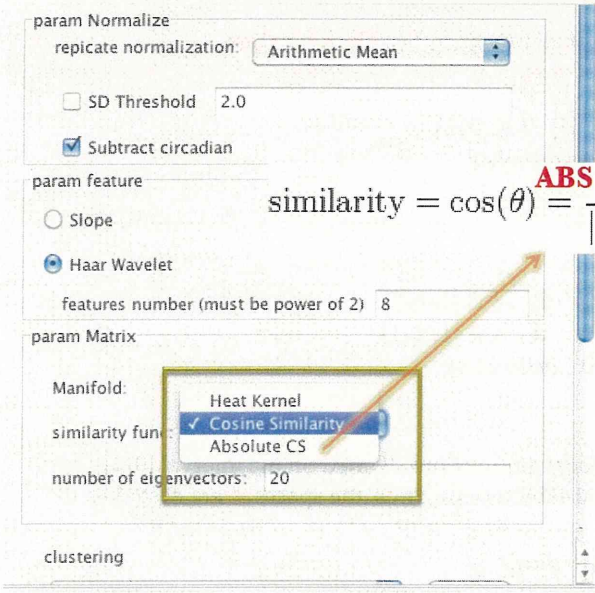


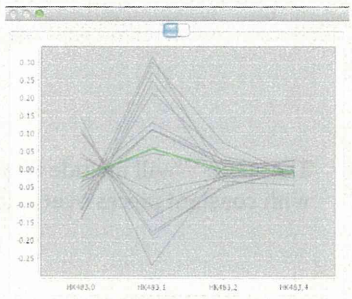
Figure 2: Curves are best fit to histograms on a time step tailored to have exact Haar approximation with few coefficients.

Similarity matrix



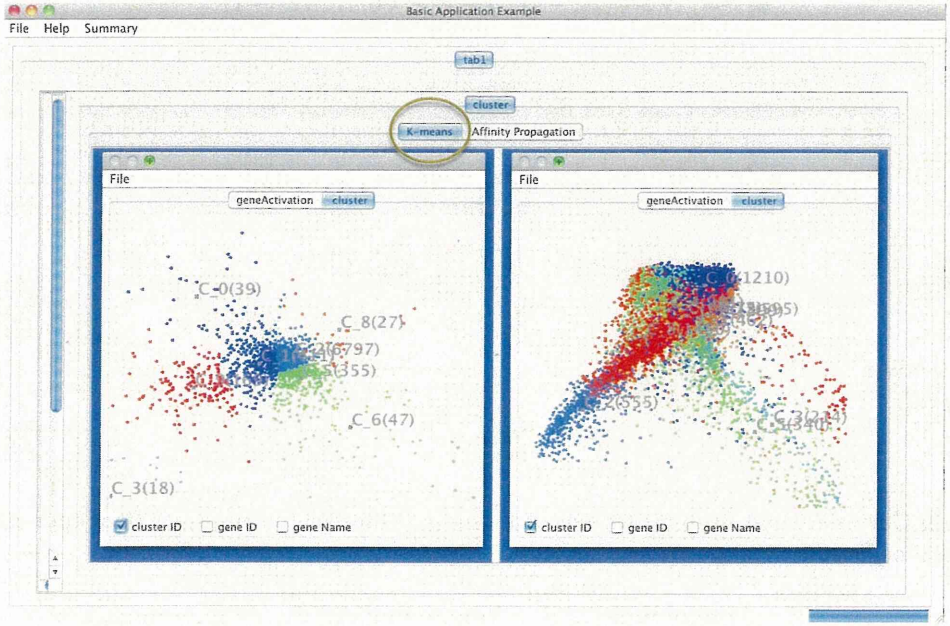
math. artifact serves Biology

$$\text{similarity} = \cos(\theta) \overset{\textbf{ABS}}{=} \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$



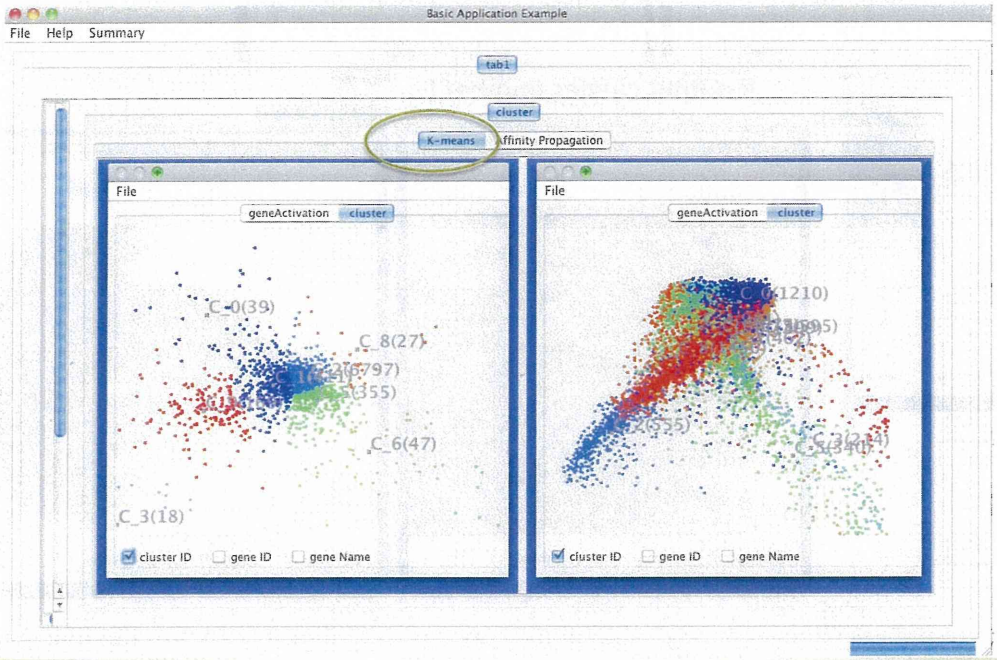
AGCT Results Panel 2

Principal Component structure　　Spectral clustering structure (MANIFOLD)

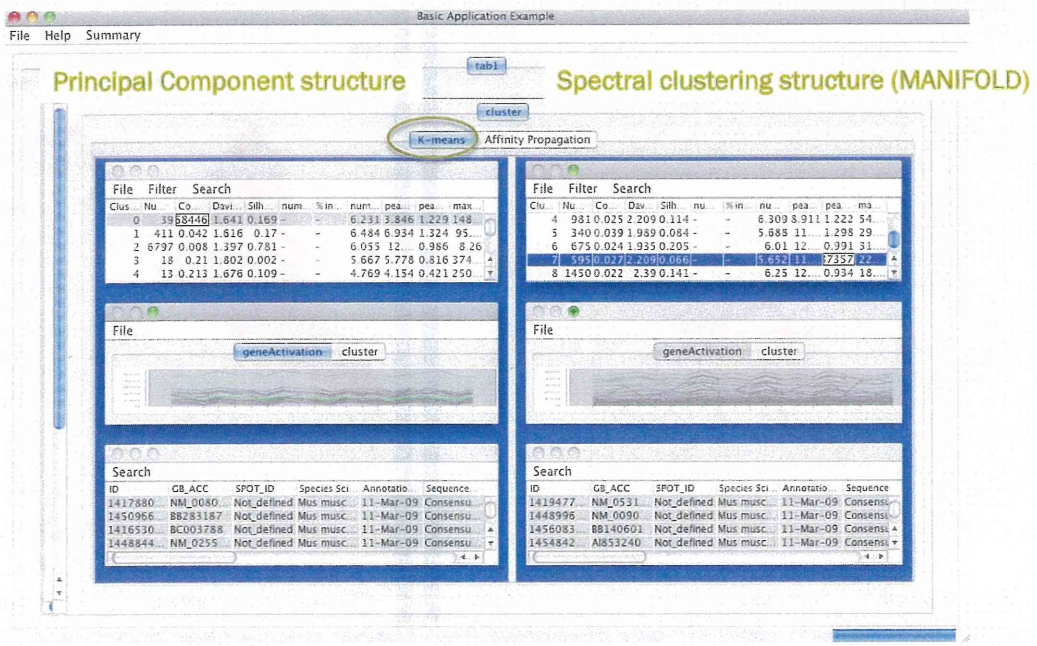


AGCT K-means Results 1

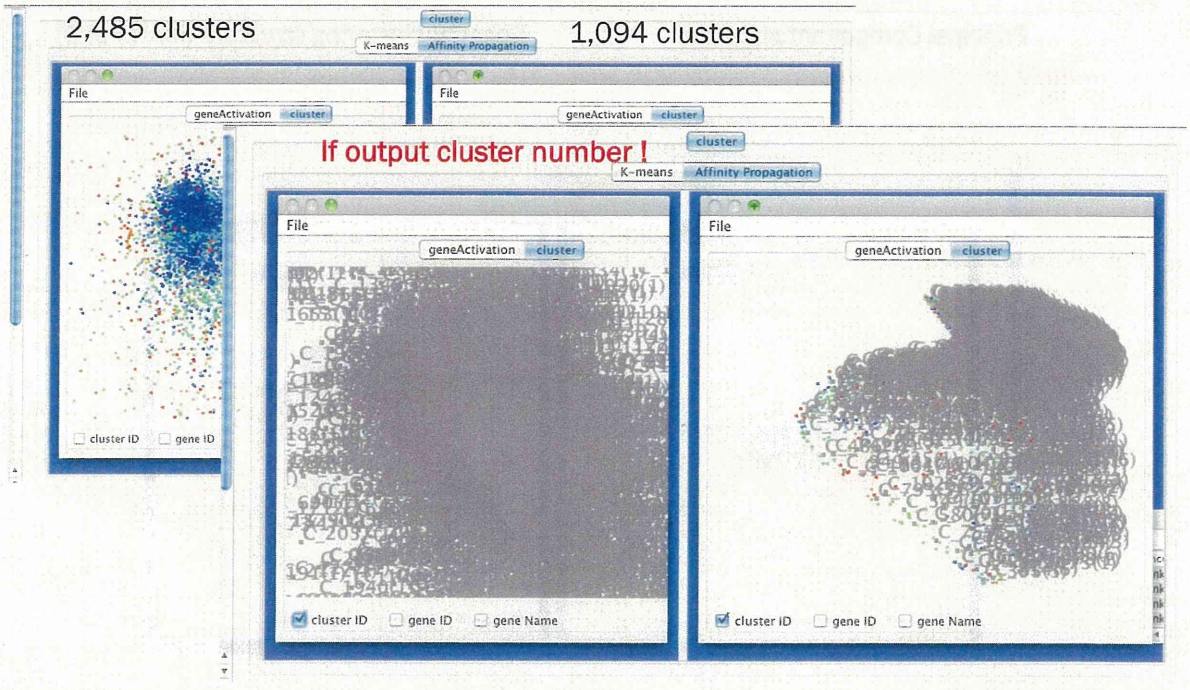
Principal Component structure　　Spectral clustering structure (MANIFOLD)



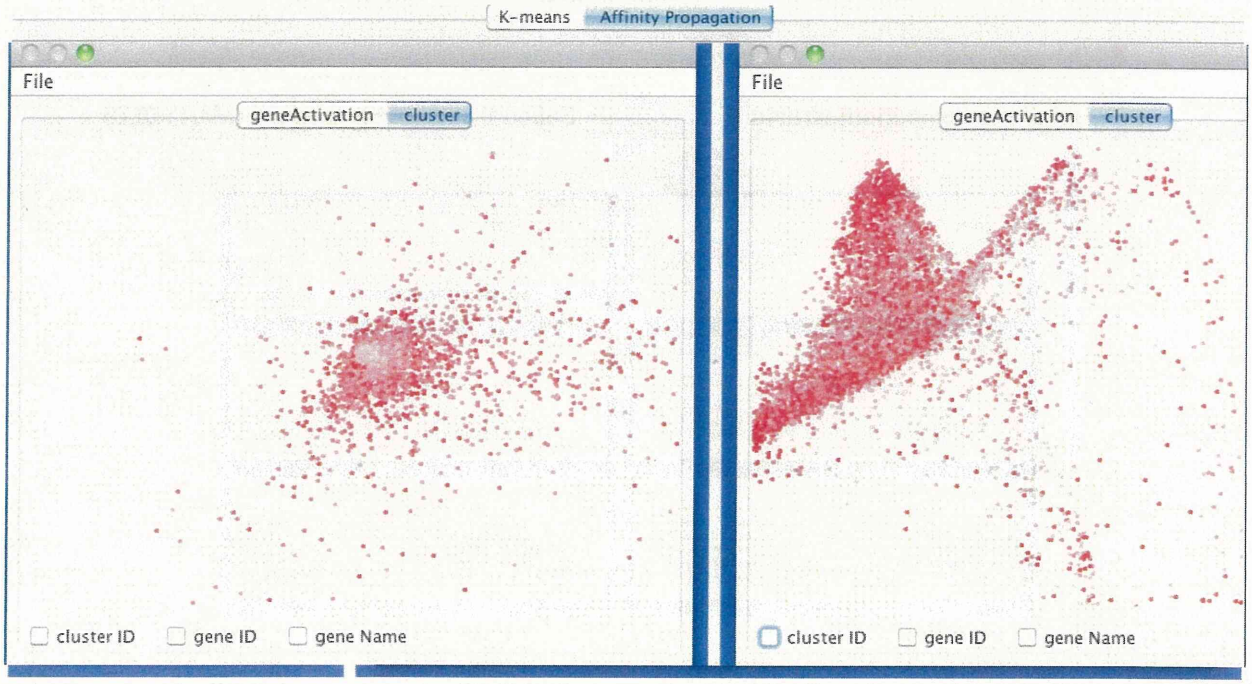
AGCT K-means Results 2



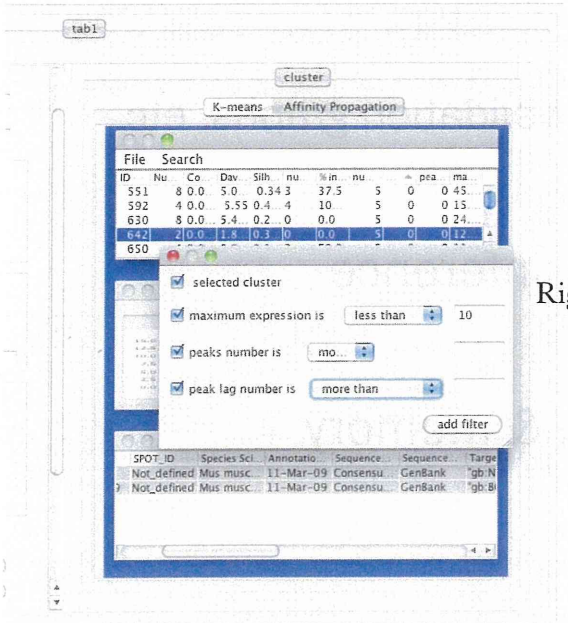
AGCT Affinity Propagation (AP) Results 1



AGCT Cluster Validation Panel



Filtering data



Right click on mouse
and filtering menu shows up !

- Data below the thresholds will disappear.
  You keep the rest of data in clusters.
- Filtering can be done on selected/all clusters.

Scenario