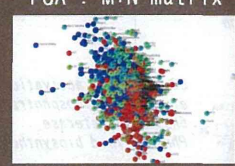
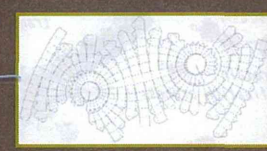
Processing data on AGCT

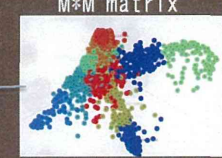
1. 時系列データ前処理:線形回帰/ウェーブレット変換
2. 遺伝子間の類似度マットリックス
3. 低次元に落とすためにSpectral clusteringを行う。通常の主成分分析も行う。
4. 発見的なClustering法を使って構造上でデータの分割を行う。
5. 結果のinteractive visualizationやscenario 記録を行う。

PCA : M*N matrix





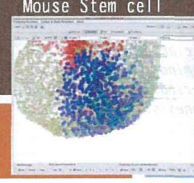
Spectral clustering: M*M matrix



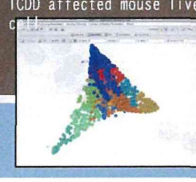
orthogonal matrix to compute one dimension per cluster/gene

Examples of different network topologies

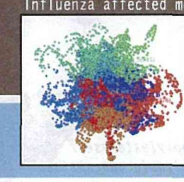
Mouse Stem cell



TCDD affected mouse liver cell

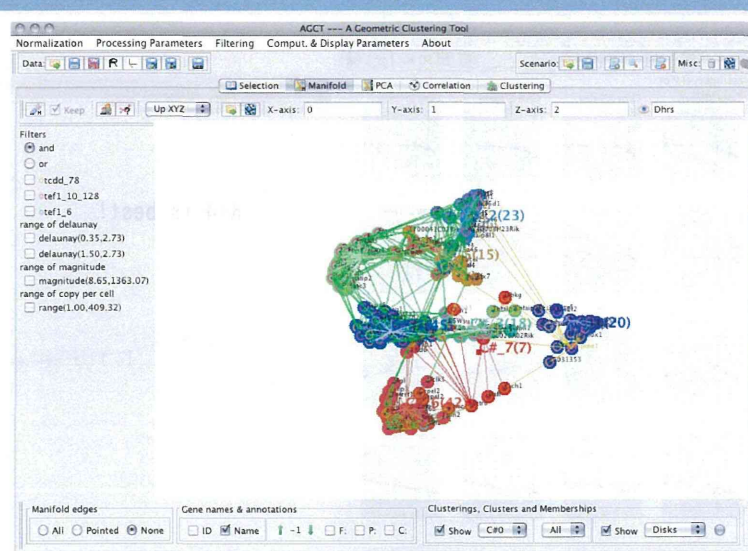


Influenza affected mouse



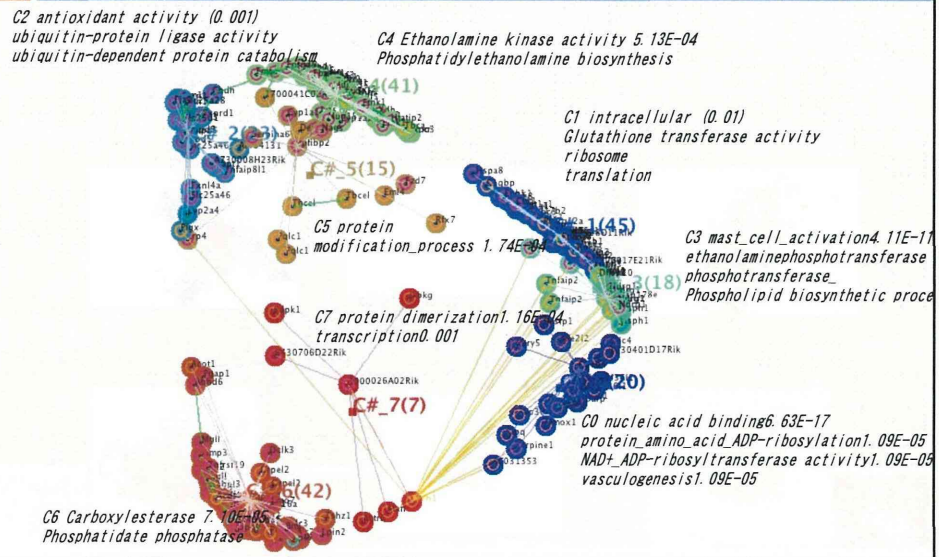
AGCT result: visualization on Manifold

8



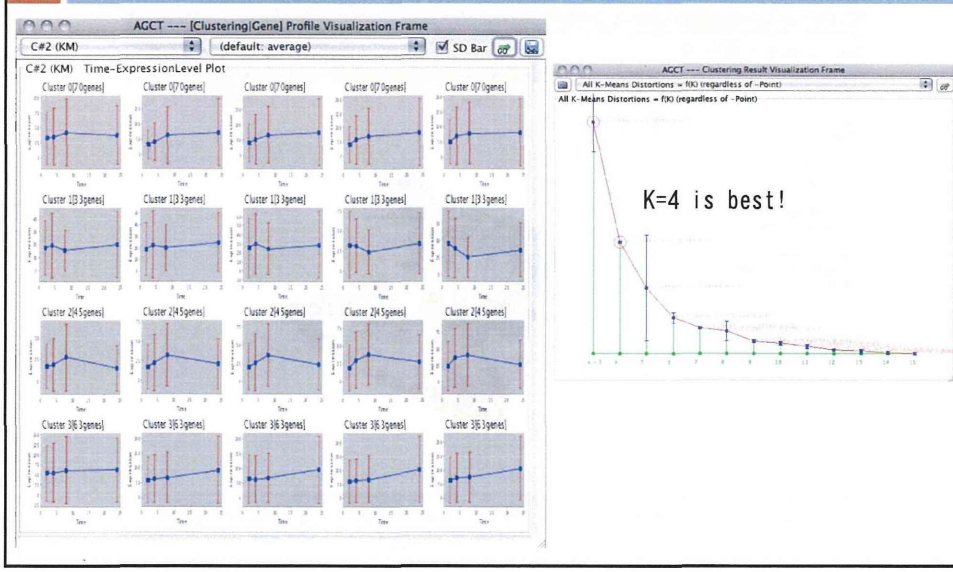
Computing gene ontology for Clusters

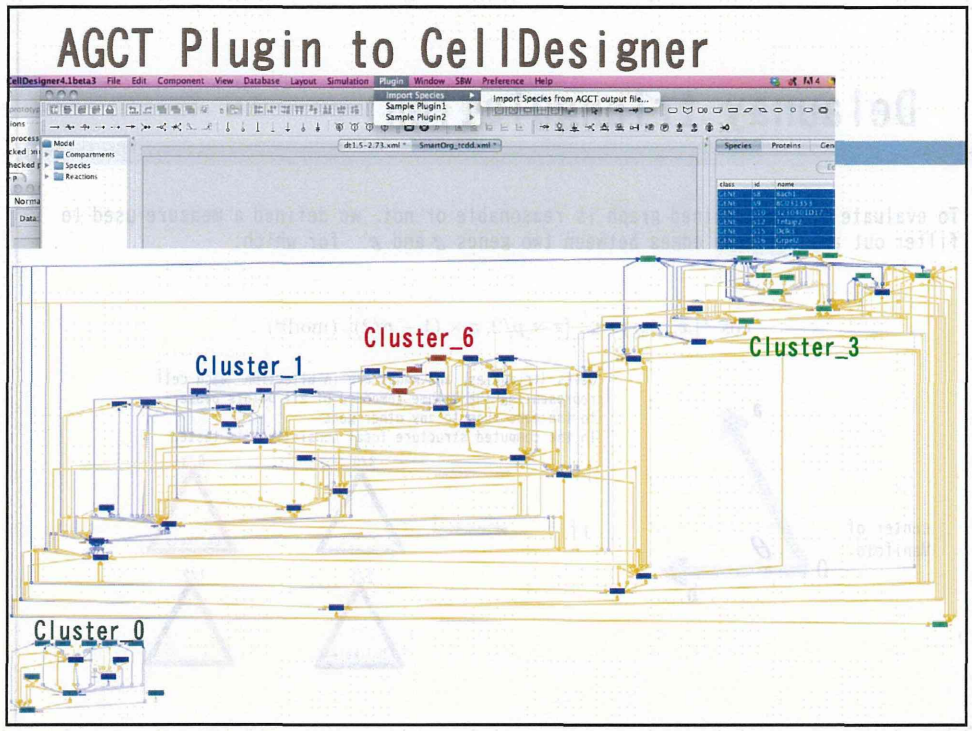
9

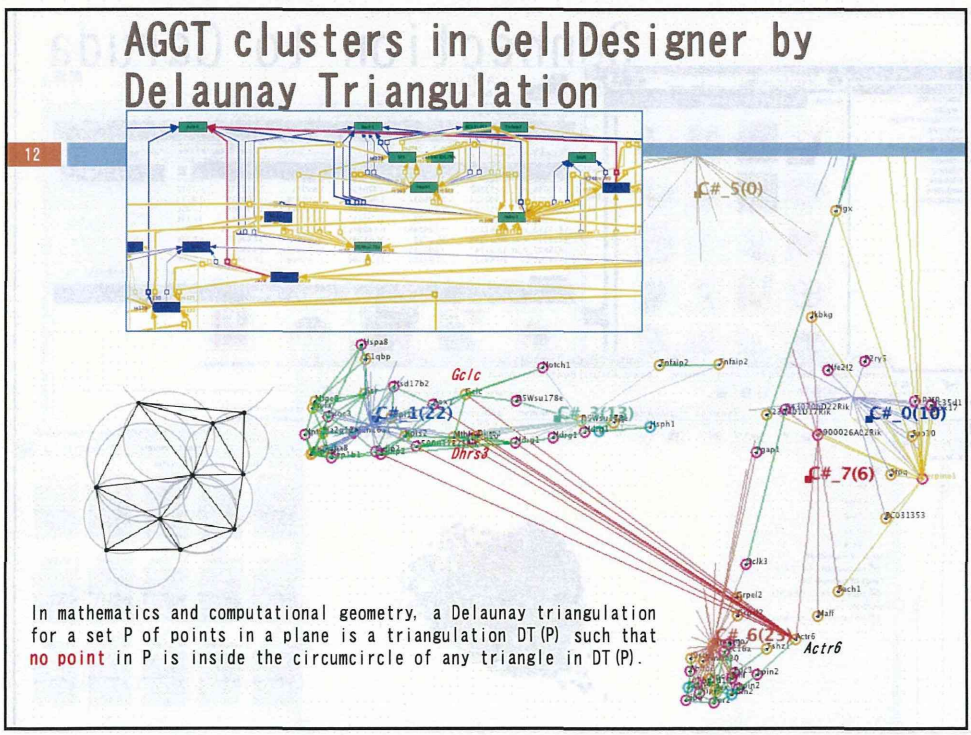


Selection of cluster number and visualization

10




AGCT Plugin to CellDesigner
Import Species
Sample Plugin1
Sample Plugin2
Import Species from AGCT output file...
Cluster_6
Cluster_1
Cluster_3
Cluster_0

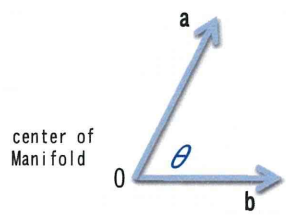

AGCT clusters in CellDesigner by Delaunay Triangulation
12
C#_5(0)
C#_1(22)
C#_3(13)
C#_0(10)
C#_7(6)
C#_6(23)
Gclc
Dhrs3
Actr6
In mathematics and computational geometry, a Delaunay triangulation for a set P of points in a plane is a triangulation DT(P) such that no point in P is inside the circumcircle of any triangle in DT(P).

Delaunay triangulation

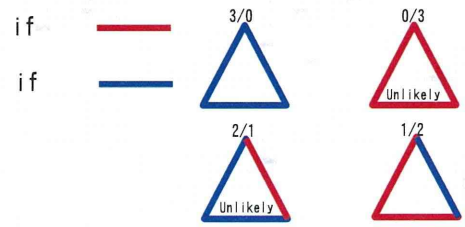
13

To evaluate if the obtained graph is reasonable or not, we defined a measure used to filter out any Delaunay edges between two genes g and g' for which:

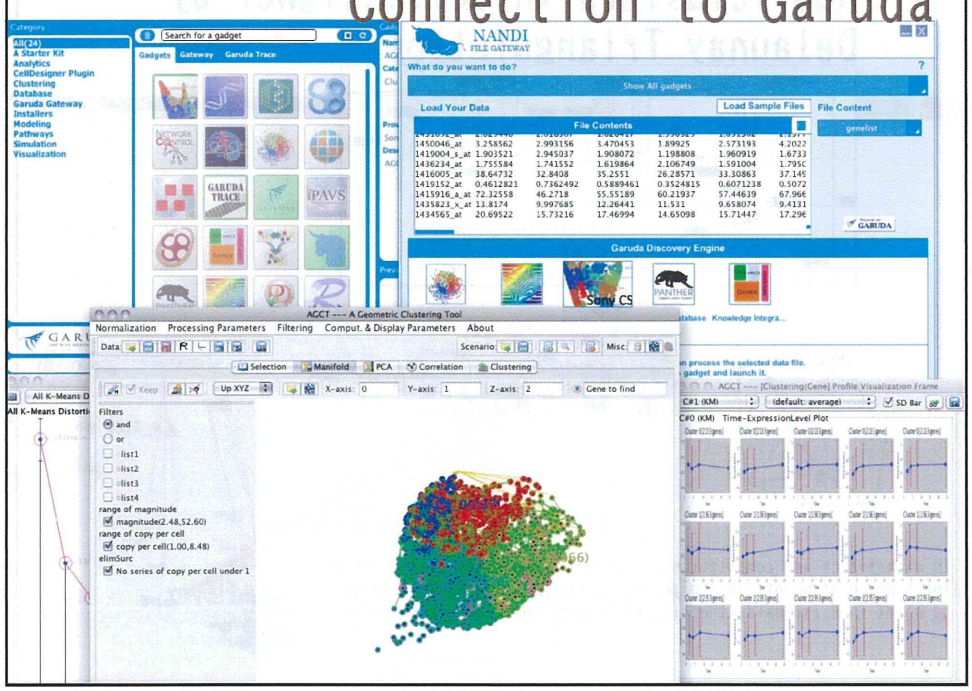
$$\cos^{-1}(\boldsymbol{x}_g, \boldsymbol{x}_{g'}) \in [\pi \times p/2, \pi \times (1-p/2)] \ (\mathrm{mod}\pi) \ .$$



Cells (triangles) and genes are in bijection, each cell representing the volume composed of all points closer to the gene than to any other gene.
In the computed structure local consistency is tested.



Connection to Garuda



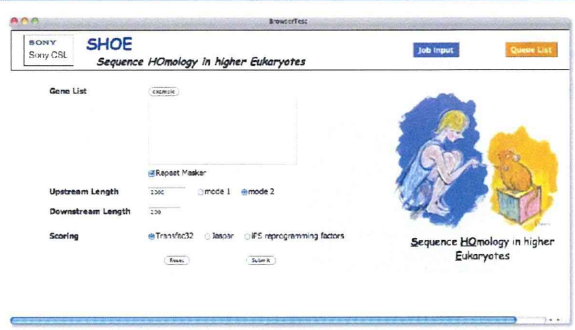
15

SHOE

- Sequence Homology in Higher Eukaryotes

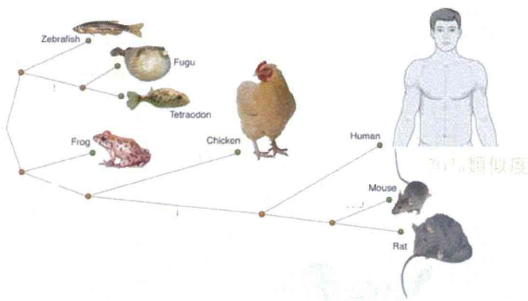
16

SHOE Interface



Sony CSL Server

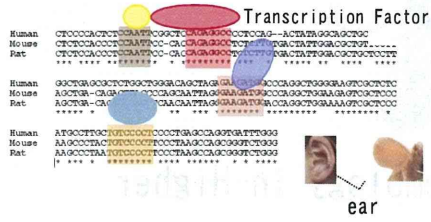
Phylogenetic footprinting finds evidence of functionality



・TFs
・Brain-specific
(Suzuki et. al, 2004)

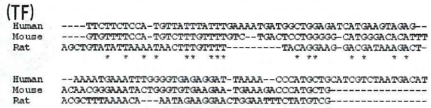
Regulation conserved !

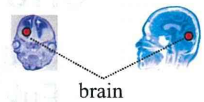
Transcription Factor (TF)



ear

Regulation not conserved



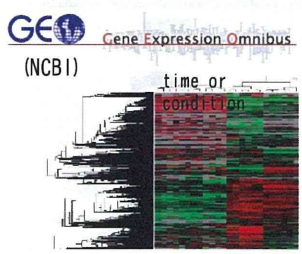


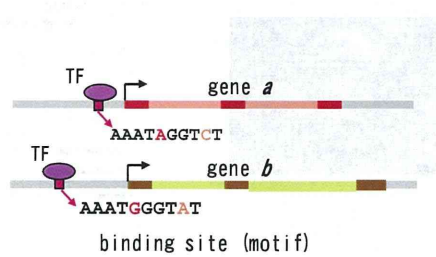
brain

17

Co-regulated genes

Similar way of expression suggests **regulation** by the same Transcription Factor (**TF**)





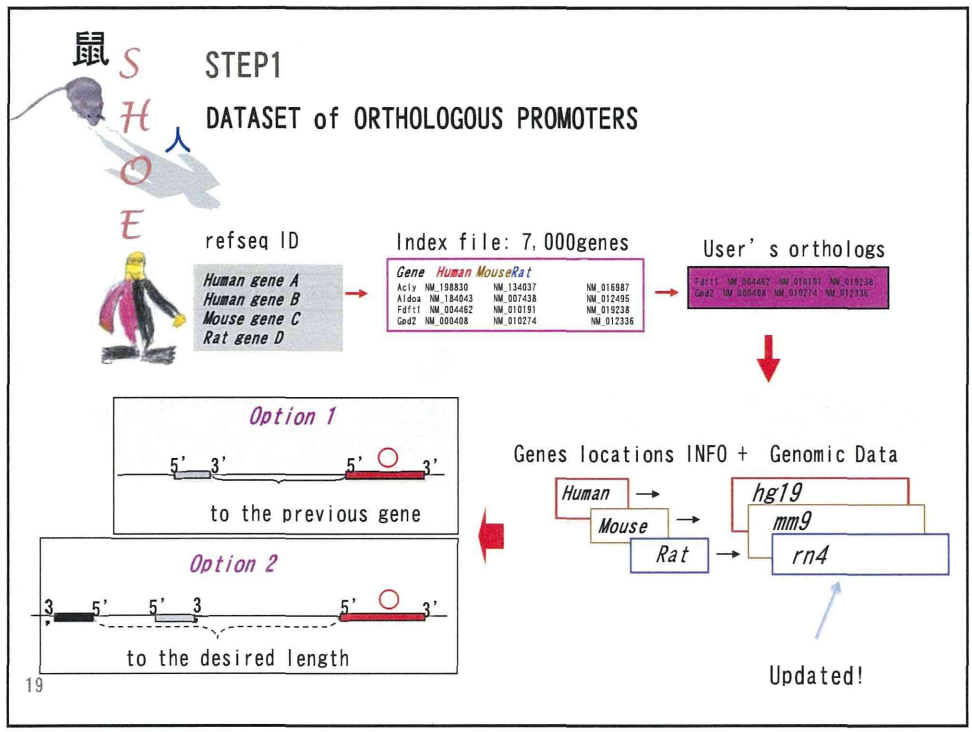
Motif discovery tools

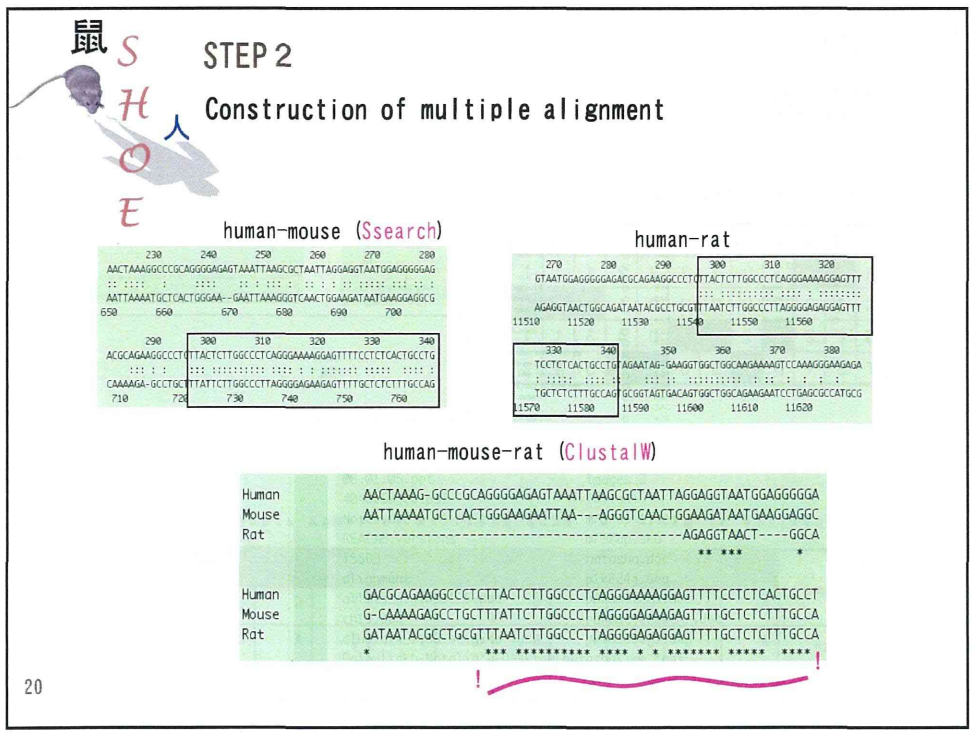
MEME (Bailey *et.al.*)
Consensus (Stormo *et.al.*)
Gibbs sampler (Lawrence *et.al.*)
Yebis (Yada *et.al.*)....

&

Comparative genomics tools

MONKEY (Moses et.al)
FootPrinter (Tompa et.al)
PhyMe (Sinha et.al)
PhyloGibbs (Siddharthan et.al) ....


鼠 SHOE 人
STEP1
DATASET of ORTHOLOGOUS PROMOTERS
refseq ID
Human gene A
Human gene B
Mouse gene C
Rat gene D
Index file: 7,000genes
Gene Human Mouse Rat
Acly NM_198830 NM_134037 NM_016987
Aldoa NM_184043 NM_007438 NM_012495
Fdft1 NM_004462 NM_010191 NM_019238
Gpd2 NM_000408 NM_010274 NM_012336
User's orthologs
Option 1
to the previous gene
Option 2
to the desired length
Genes locations INFO + Genomic Data
Human
Mouse
Rat
hg19
mm9
rn4
Updated!
19


鼠 SHOE 人
STEP 2
Construction of multiple alignment
human-mouse (Ssearch)
human-rat
human-mouse-rat (ClustalW)
Human AACTAAAG-GCCCGCAGGGGAGAGTAAATTAAGCGCTAATTAGGAGGTAATGGAGGGGGA
Mouse AATTAAAATGCTCACTGGGAAGAATTAA---AGGGTCAACTGGAAGATAATGAAGGAGGC
Rat ----------AGAGGTAACT----GGCA
Human GACGCAGAAGGCCCTCTTACTCTTGGCCCTCAGGGAAAAGGAGTTTTCCTCTCACTGCCT
Mouse G-CAAAAGAGCCTGCTTTATTCTTGGCCCTTAGGGGAGAAGAGTTTTGCTCTCTTTGCCA
Rat GATAATACGCCTGCGTTTAATCTTGGCCCTTAGGGGAGAGGAGTTTTGCTCTCTTTGCCA
20

STEP 3

How good is the alignment?

Pattern frequencies tables for "good" and "random" alignments

74 patterns

A, T, G, C, - × AA, AT, AG, AC, A-, TT, TG, TC, T-, GG, GC, G-, CC, C-, --

835 orthologous alignments (238,800bp)

1260 random alignments (239,600bp)

,

21

c - probability of pattern in each column in ***good alignment*** and ***random alignment*** tables,
m - motif length.

BIOBASE
BIOLOGICAL DATABASES

TRANSFAC

Calculation of PSSM score (PMscore)

498 human-rodent matrices

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪
A	4	5	3	0	4	3	3	2	1	1	1
C	1	2	0	0	0	0	0	1	3	4	6
G	2	2	7	2	3	7	0	4	3	1	1
T	3	1	0	8	3	0	7	3	3	4	2

10

where ***pseudocount***= 1
m is a motif length

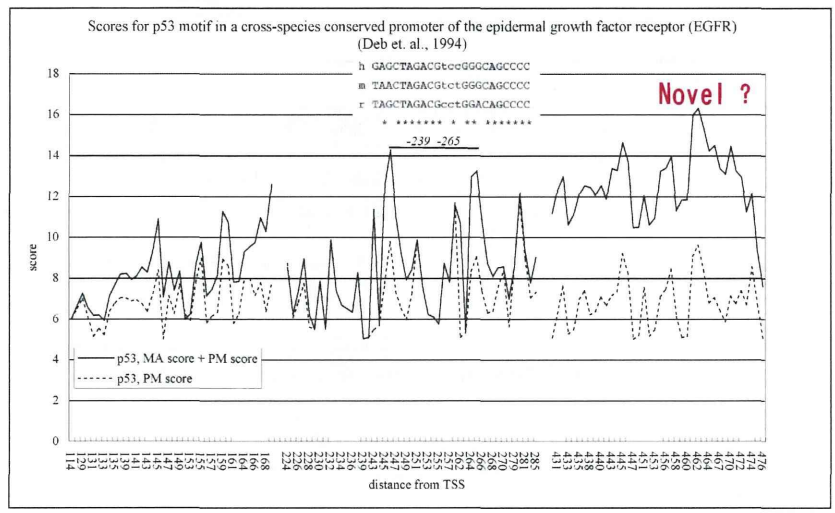
Probability of motif at each position in human sequence

	-0.48	-0.39	-0.27	-0.22	-0.57	-0.27	-0.27	-0.57	-0.57	-0.47	-0.33	-4.46
human	A	A	G	T	G	G	T	T	C	C	C	
mouse	A	T	G	T	G	G	T	C	A	C	C	
rat	A	T	C	T	G	G	T	A	A	C	C	

22

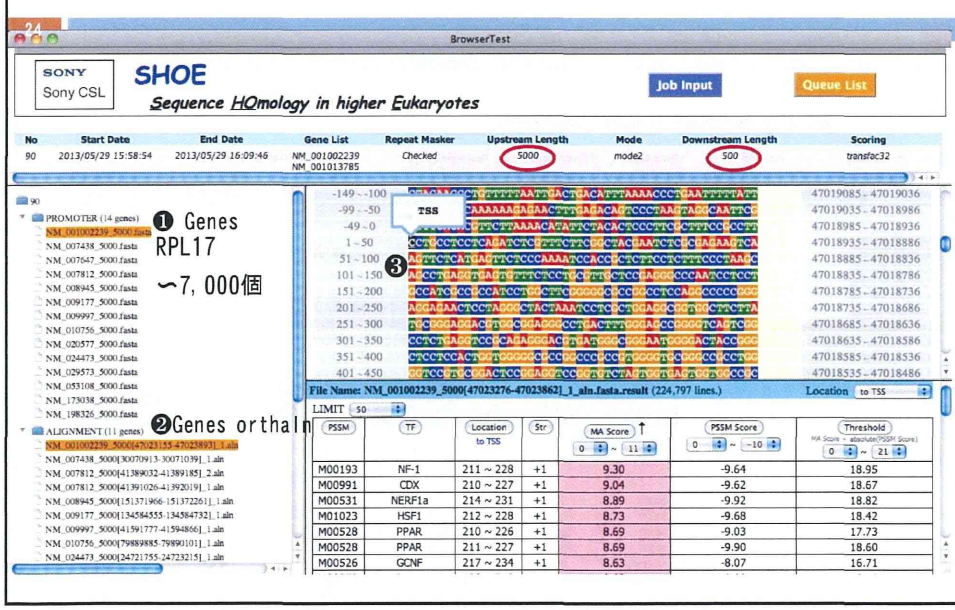
Finding motif candidates

Multiple alignment score (MA) + Transfac PSSM score (PM)

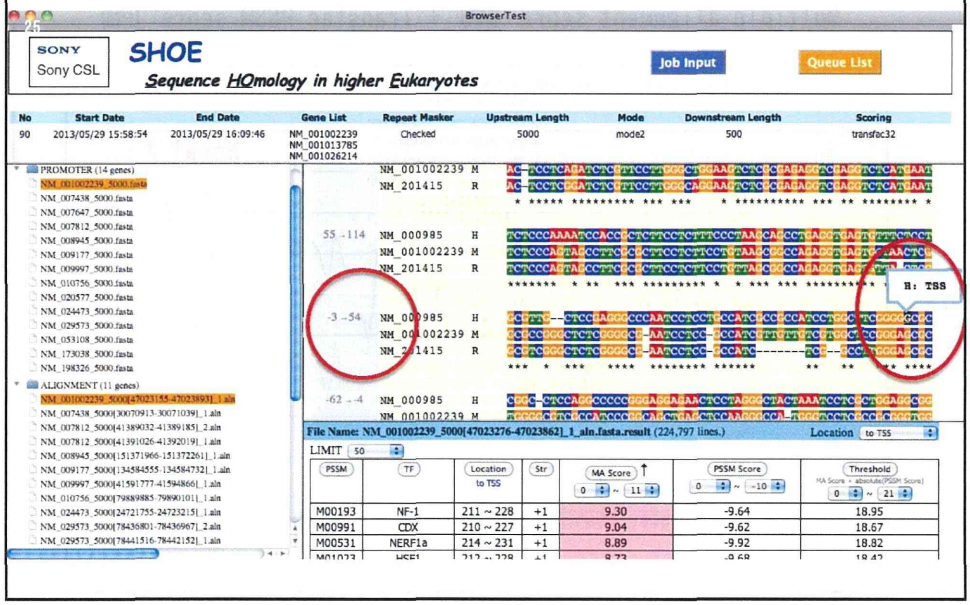


23

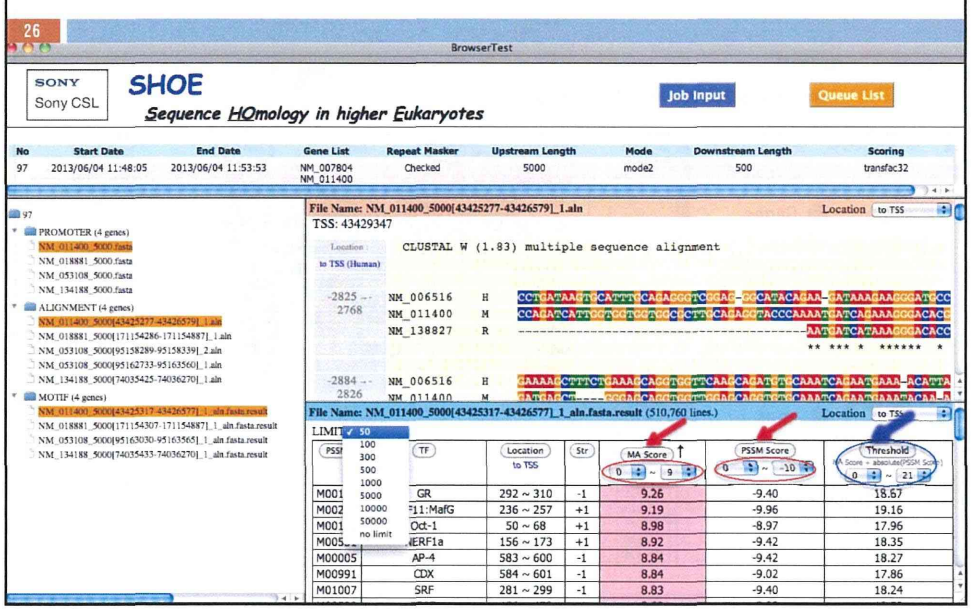
SHOE development



Alignment viewing

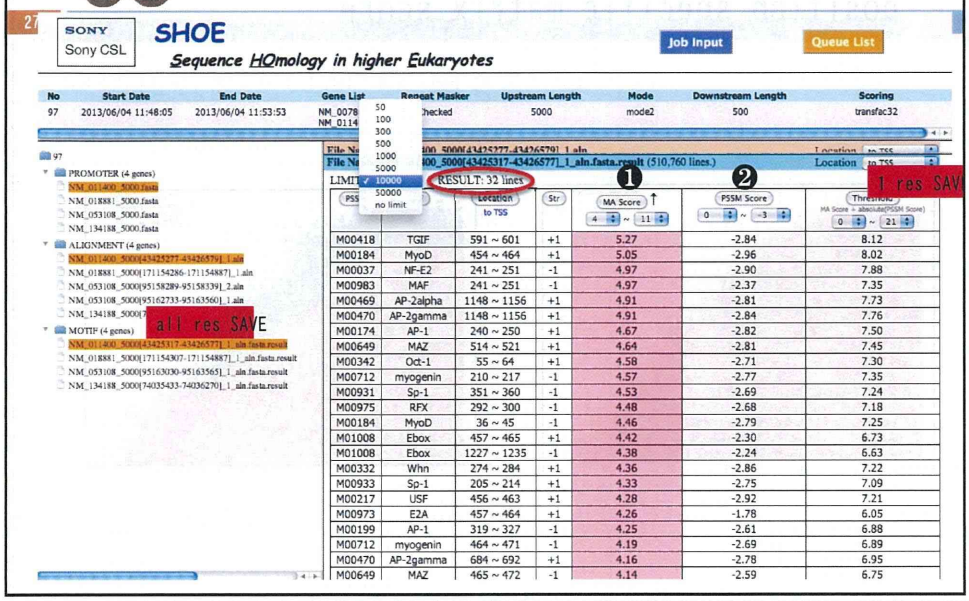


Motif result viewing



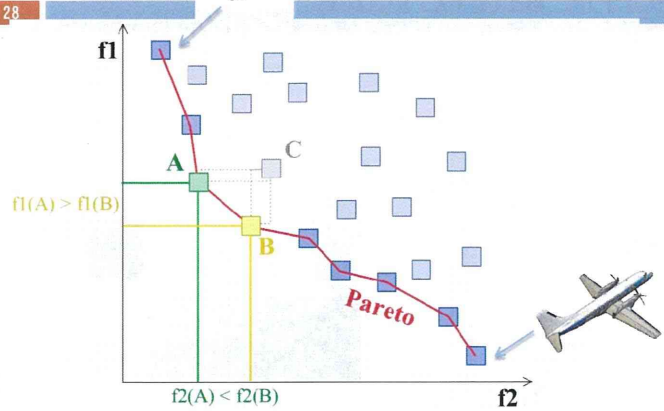
Selection of results 結果の絞り込み

❶❷



Top scoring TF motifs with Pareto Front

パレート面を用いた多目的問題最適化

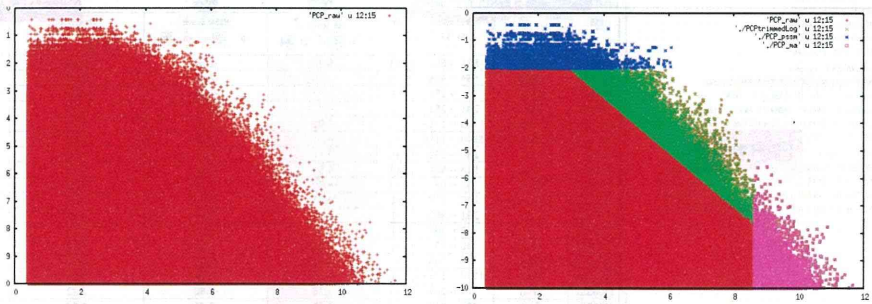


Mitsubishi Regional Jetおよび Bombardier CSeriesを支える
次世代型エンジンの設計などに用いられ、工学的な応用が盛ん

Top scoring cross-region of two distributions: multiple alignment score and position specific matrix score

29

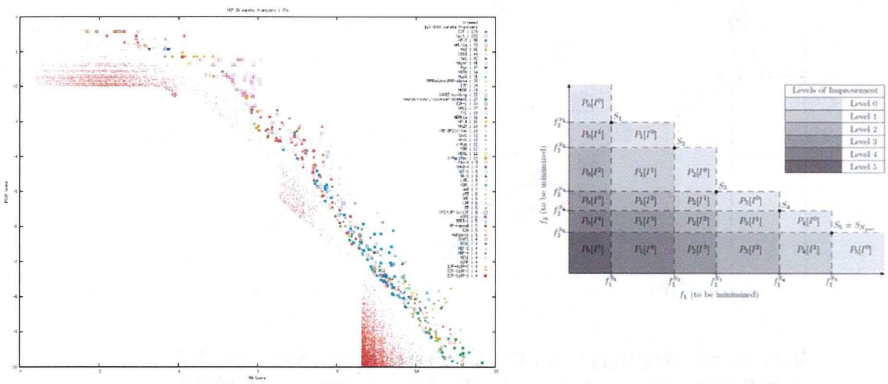
パレート面を用いた多目的問題最適化



Top scoring cross-region of two distributions: multiple alignment score and position specific matrix score

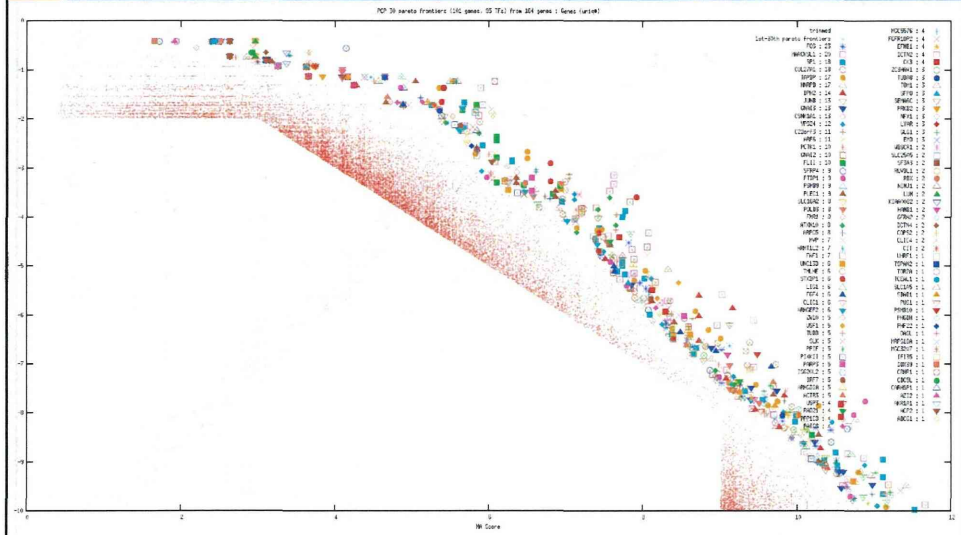
30

パレート面を用いた多目的問題最適化



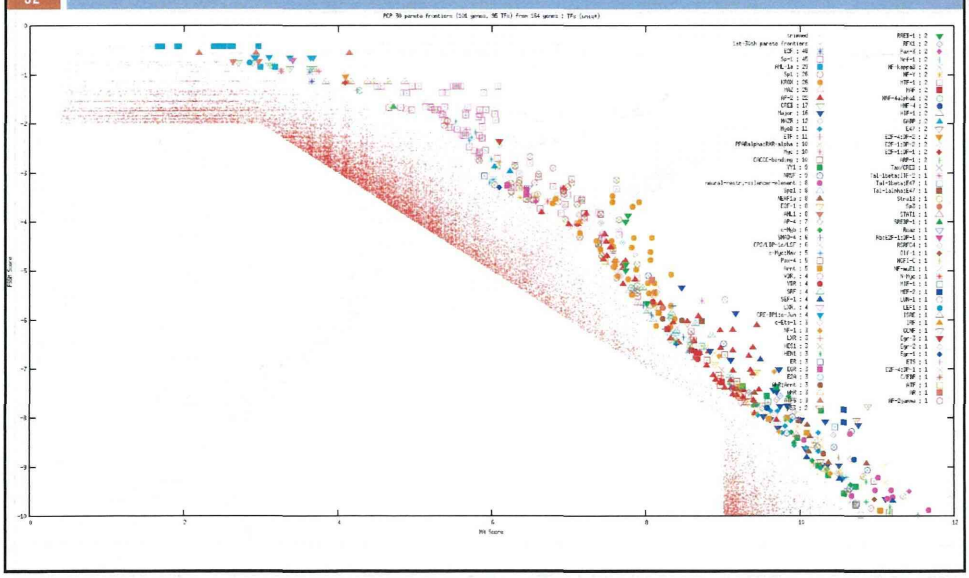
PCP network: Genes with unique number of top scoring TFs in each promoter

31



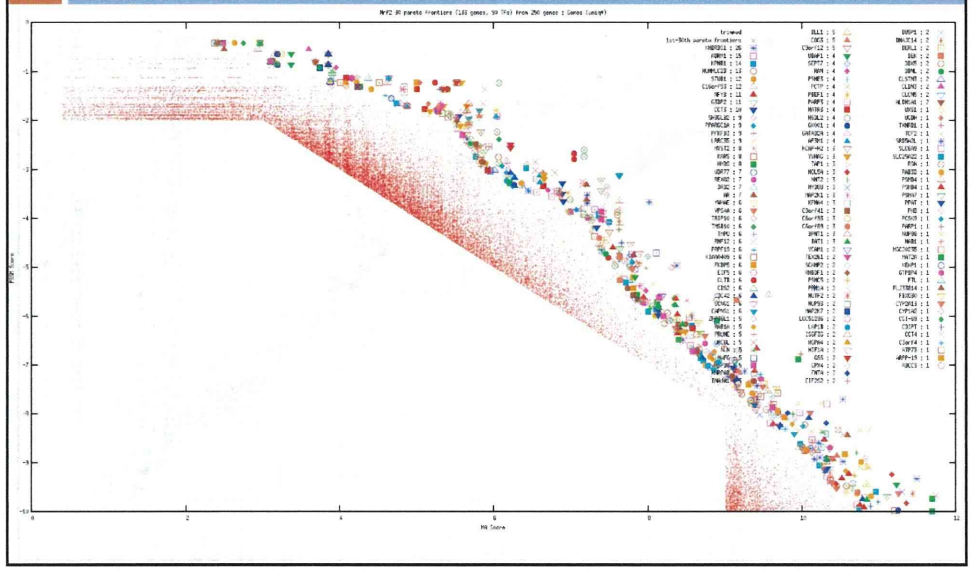
PCPnetwork: TFs with a unique number of genes sharing them

32



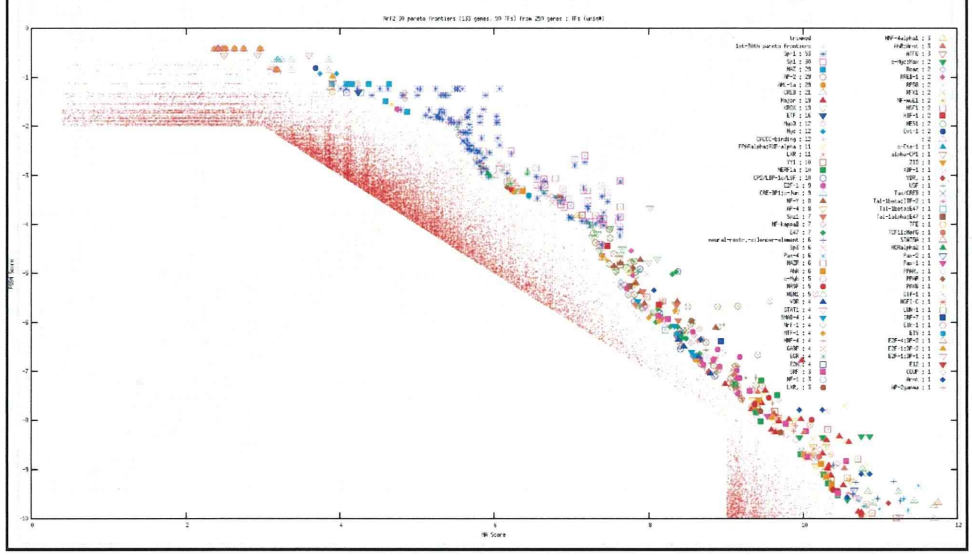
Nrf2 network: Genes with unique number of top scoring TFs in each promoter

33



Nrf2 network: TFs with a unique number of genes sharing them

34



Co-authors

35

- Hiroaki Kitano (Sony CSL)
- Frank Nielsen (Sony CSL)
- Richard Nock (University of Martinique)
- Keigo Oka (University of Tokyo)
- KoudaiTakata (Tokyo Institute of Technology
- Tomohiro Masagaki (Sony CSL)

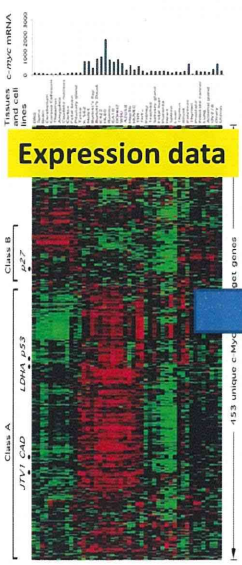
1

Systems Toxicology
Inference of gene regulatory networks from large-scale experimental data

Takeshi Hase, Samik Ghosh, and Hiroaki Kitano
The Systems Biology Institute

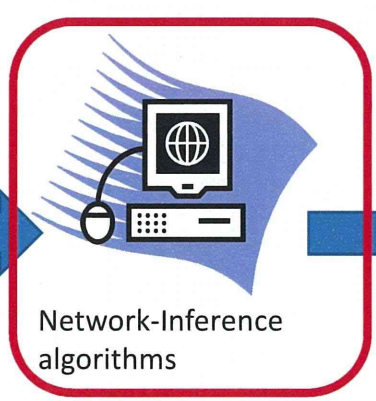
2

Inference of gene regulatory network

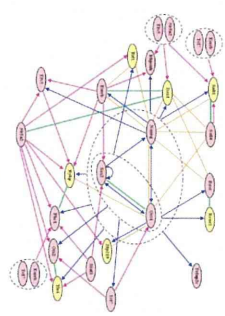


Expression data

Many network-inference algorithms have been proposed.
One of the long-standing challenge in computational biology.



Network-Inference algorithms



Gene regulatory network

3

Application of Network Reconstruction Technique to Percellome

- **Possible areas**
  - How network structure changes over time in a tissue for a specific dose of the perturbation?
  - Build tissue specific gene networks
  - How the gene interaction network changes over different dosage of the compound?

4

Inference of gene regulatory network expression dataset under pentachlorophenol

5

PCP (pentachlorophenol) dataset

- Gene expression data with PCP from mouse liver
- 4 time-points (2, 4, 8, and 24 hours) and 4 dosages (0, 10, 30, and 100 mg/kg PCP). Data of 2hour and 0 mg/kg Three replicates for each condition. 2 hours vehicle value is used for virtual 0 hour values

6

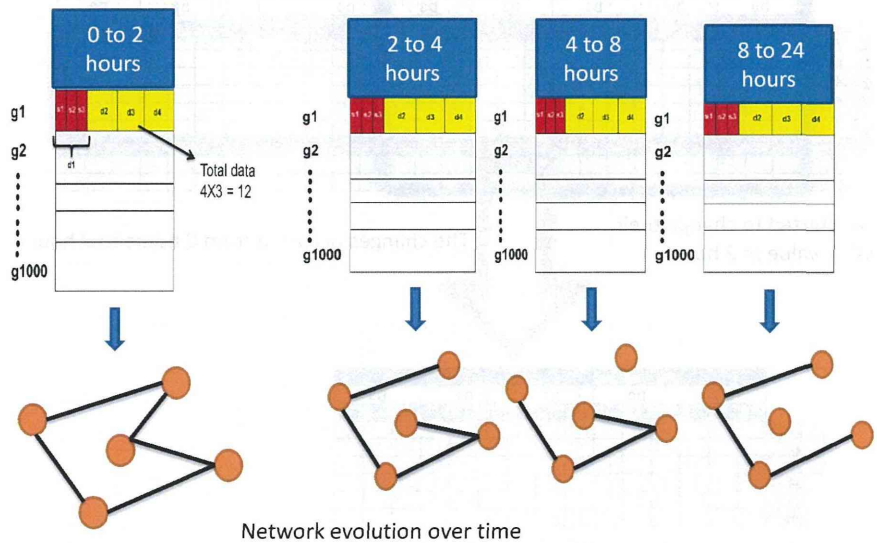
PCP (pentachlorophenol) dataset

- 98, 55, 127, and 1192 genes started to change in response to PCP treatment at 2, 4, 8, and 24 hours.
- By using the list of genes and time-course of the genes, we will generate a gene network at each time points.

Schematic Example

For a pertubagen (drug/chemical): **P**

For a tissue: **T**



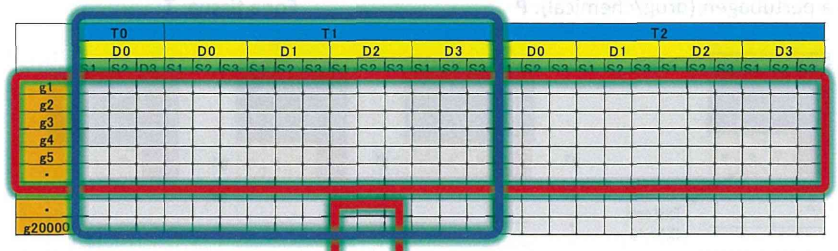
Network evolution over time

Inference of a network at 0 – 2 hours

	T0			T1												T2											
	D0			D0			D1			D2			D3			D0			D1			D2			D3		
	S1	S2	D3	S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3	S1	S2	S3
g1																											
g2																											
g3																											
g4																											
g5																											
.																											
.																											
.																											
g20000																											

9

Inference of a network at 0 – 2 hours



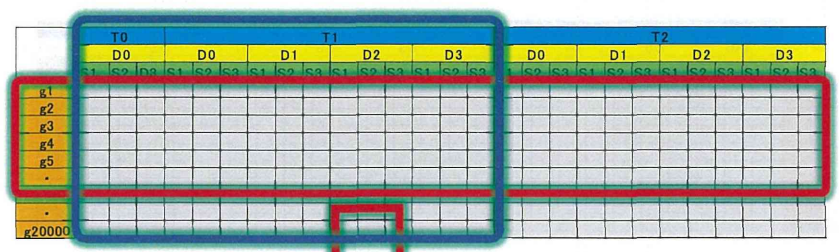
94 genes started to change their expression value at 2 hours

The changes occurred from 0 hours to 2 hours.

Expression values of 94 genes at 0 – 2 hours

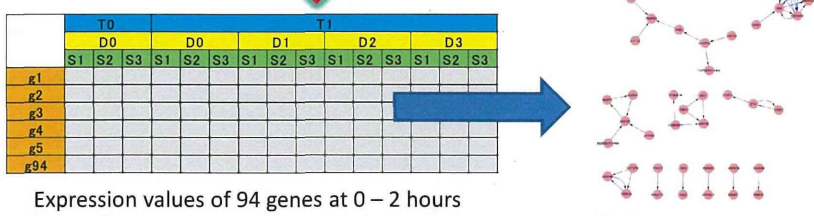
10

Inference of a network at 0 – 2 hours



94 genes started to change their expression value at 2 hours

The changes occurred from 0 hours to 2 hours.



Expression values of 94 genes at 0 – 2 hours