

associated with a kinase inhibitor using systems toxicology. *npj Systems Biology and Applications*, 1, doi:10.1038/npjbsba.2015.5, Sep. 28, 2015.

2. 学会発表

- ① Kitano, H. The Day AI wins the Nobel Prize. 1st Think X Symposium Theme: Smart Healthcare for Smart Nation, Infuse Theatre at Fusionopolis, Singapore, Apr. 9, 2015.
- ② 北野宏明. システムトキシコロジーの情報基盤としての Garuda Platform. 第42回日本毒性学会学術年会 シンポジウム:「エピゲノミクス・ゲノミクス解析の進展と適応拡大する毒性オミクス」, ホテル日航金沢, July 1, 2015. (invited)
- ③ 北野宏明. Garuda Platform: Innovations in Computational Platform for Drug Discovery and Digital Health. 第31回創薬セミナー. 大津プリンスホテル, July 22, 2015. (invited)
- ④ Kitano, H. Computational Systems Toxicology Approach Using the Garuda Platform. Workshop on Pharmacological Mechanism-Based Drug Safety Prediction: Integrated Computational Approaches for Problem Solving of Systems Comprised of Multiple Levels of Biological Organization, from Drug Exposure to Clinical Phenotype, FDA White Oak Facility, Silver Spring, USA, Aug. 6, 2015. (invited)
- ⑤ 北野宏明. システム医科学におけるオープンイノベーションを促進するガルーダ・プラットフォーム. JASIS 2015 先端診断イノベーションゾーン, 幕張メッセ, Sep. 3, 2015. (invited keynote)
- ⑥ 北野宏明. 人工知能がノーベル賞を獲る日: 科学的発見のエンジンを作る. 日本バイオインフォマティクス学会年会・生命医薬情報学連合大会, 一般公開シンポジウム: 生命科学データベースと人工知能・ロボティクスの拓く未来, 京都大学宇治キャンパス, Oct. 31, 2015. (invited)
- ⑦ 北野宏明. システム医科学におけるオープンイノベーションを促進するガルーダ・プラットフォーム. 日本薬物動態学会 第30回年会 東京, タワーホール船堀, Nov. 14, 2015. (invited)
- ⑧ Kitano, H. Recent progress in Systems Biology. ICSB 2015, Biopolis, Singapore, Nov. 23, 2015.
- ⑨ Thierry D.G.A. Mondeel, Christian Linke, Adriana Supady, Anastasia Chasapi, Wolfram Liebermeister, Mart Loog, Ioannis Xenarios, Hiroaki Kitano and Matteo Barberis. Deciphering the design principles of dynamic cell cycle control. ICSB 2015, Biopolis, Singapore, Nov. 23, 2015.
- ⑩ Huipeng Li, Elise Courtois, Debarka Sengupta, Say Li Kong, Charlene Kang, Yuliana Tan, Clarinda Chua, Chew Min Hoe, Lai Jiunn Heng, Ng Chee Yung, Dennis Koh, Tan Wah Siew, Mark Wong Te Ching, Yongli Hu, Samik Ghosh, Axel Hillmer, Hiroaki Kitano, Lawrence Wee, Daniel Tan, Iain Tan, Paul Robson and Shyam Prabhakar. Single cell RNA-seq analysis of tumor and microenvironment heterogeneity in colorectal cancer. ICSB 2015, Biopolis, Singapore, Nov. 23, 2015.
- ⑪ Akira Funahashi, Yukiko Matsuoka, Samik Ghosh, Noriko Hiroi and Hiroaki Kitano. CellDesigner: A modelling tool for biochemical networks. ICSB 2015, Biopolis, Singapore, Nov. 23, 2015.
- ⑫ Archana Bajpai, Takashi Ishii, Miyauchi Koichi, Masato Kubo and Hiroaki Kitano. Dysregulation of suppressor of cytokine signaling 3 in keratinocytes cause hyper-activation of Ap-1 related genes and develop hyperplasia like phenotype. ICSB 2015, Biopolis, Singapore, Nov. 23, 2015.
- ⑬ Kun-Yi Hsin; Yukiko Matsuoka; Tokiko Watanabe; Yoshihiro Kawaoka and Hiroaki Kitano. In-silico approach in

assessing anti-influenza agents and their targets using comprehensive pathway map (FluMap). ICSB 2015, Biopolis, Singapore, Nov. 23, 2015. (poster presentation)

- ⑭ Ghosh, S.; Matsuoka, Y.; Kitano, H. Garuda Platform and Its Applications in Biomedical Research. ICSB2015 Workshop, BioNetVisa, Yong Loo Lin School of Medicine, National University of Singapore (NUS), Singapore, Nov. 25, 2015.
- ⑮ 北野宏明. ガルーダ・プラットフォーム：創薬、デジタルヘルスのための革新的情報基盤システム. 武田薬品—京都大学医学部附属病院合同シンポジウム：日本におけるデジタルヘルス -現状と将来展望-, 京都大学医学部創立百周年記念施設芝蘭会館 稲盛ホール, 京都, Dec. 4, 2015. (invited)

H. 知的財産権の取得状況

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

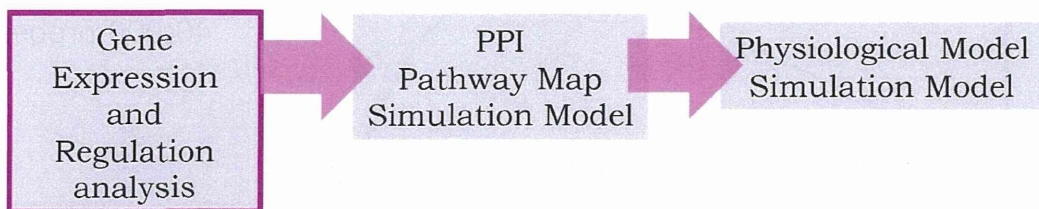
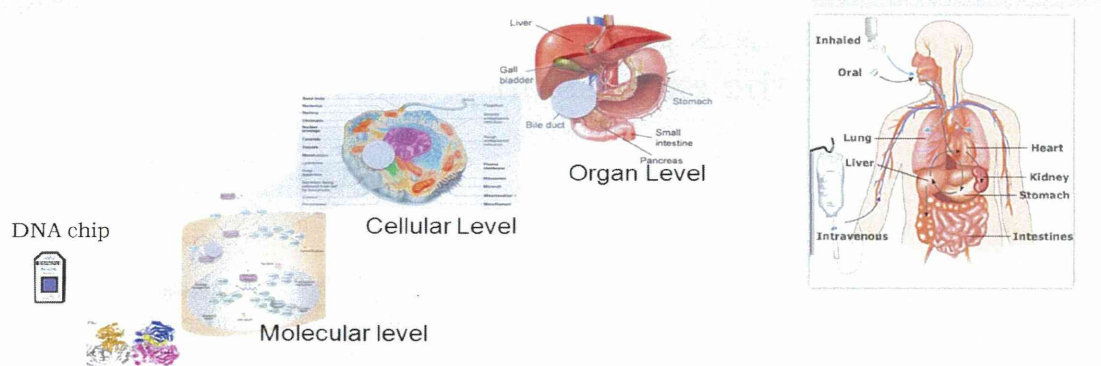
Toxicogenomics Annual Meeting

Gene expression and transcription regulation analysis
for gene network discovery

January 2016

Systems Biology Institute
Natalia Polouliakh, Hiroaki Kitano

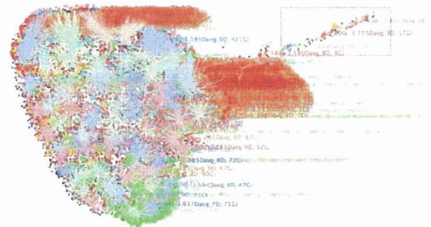
Data-driven analytic pipeline



Research Agenda

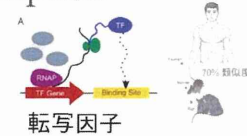
▶ **AGCT** A Geometric Clustering Tool

- ▶ Clustering Percellome data based on similarity of gene expression profile.



▶ **SHOE** Sequence Homology in Higher Eukaryote

- ▶ Phylogenetic footprinting for discovery of transcription regulation network

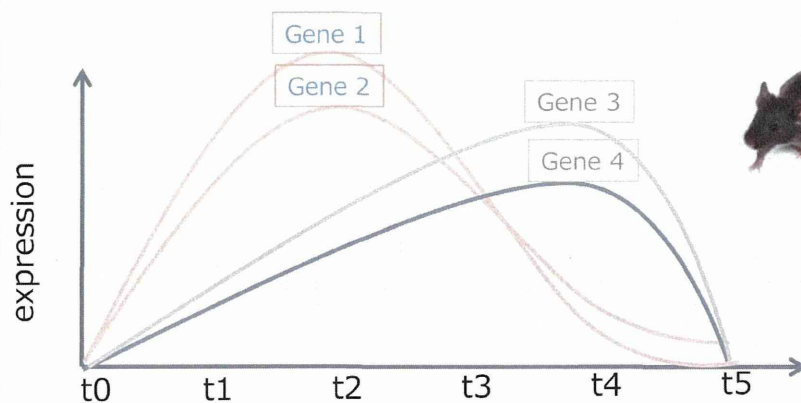


3

トキシコゲノミクス研究会平成25年

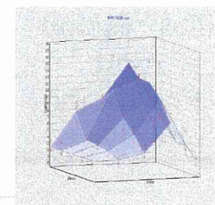
Atlas of Cell Life by AGCT

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



- chemical
- doze
- time
- expression

40,000 probes

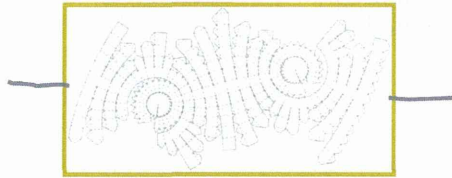
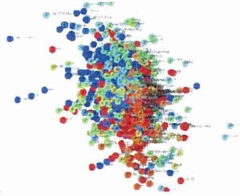


トキシコゲノミクス研究会平成25年

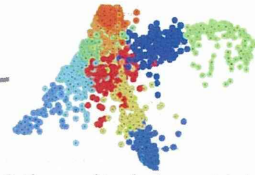
Processing data on AGCT

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 : $M \times N$ matrix



Spectral clustering:
 $M \times M$ matrix



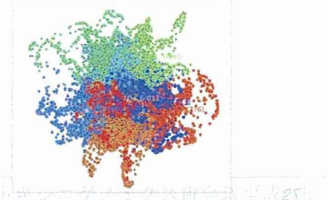
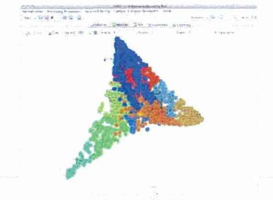
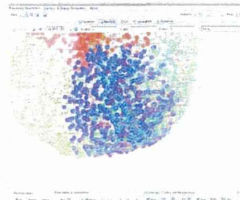
Examples of different network topologies

Orthogonal Laplacian matrix to compute
one dimension per cluster/gene

Mouse Stem cell

TCDD affected mouse liver cell

Influenza affected mouse bronchi cell



Analysis of TCDD-TCDF together data ($\approx 40,000$) Plug-in to CellDesigner to produce gene MAP

Clustering: Bergman k-means
Affinity propagation, Expectation Maximization, CP, NNMF.

The screenshot shows the AGCT (A Geometric Clustering Tool) interface. It includes several panels:

- Normalization, Processing Parameters, Iterative Filtering, Comput. & Display Parameters, About:** The top menu bar.
- Selection, Manifold, PCA, Correlation, Clustering:** The main toolbar.
- Summary:** A box on the left containing data statistics like 'mean', 'std', 'min', 'max', 'total', 'genes', 'features', 'time series', 'group', 'out', 'in', 'edge', 'total', 'features', 'time series', 'group', 'out', 'in', 'edge'.
- 1-Feature method:** A dropdown menu set to 'Pearson'.
- 2-Automatic feature selection method:** A dropdown menu set to 'None'.
- 3-User-fixed gene / time series / group selection:** A table with columns for 'Gene', 'Time series', 'Group', and 'out/in' status. It lists genes like 'gene_1427071_at', 'gene_1449037_at', etc.
- 4-Automatic prototype genes construction method:** A dropdown menu set to 'Max. distance'.
- Main Visualization:** A large scatter plot showing a complex network of points colored by cluster. A Delaunay Triangulation is overlaid on the points.
- Optimization of cluster number:** A line graph showing the relationship between the number of clusters and a metric.
- Principal Component Analysis:** A smaller scatter plot showing the first two principal components.
- Cluster centroid:** A table listing the centroid coordinates for each cluster.

Delaunay Triangulation
Surface Extraction ($p \leq 0.05$)

Optimization of cluster number

Principal Component Analysis

Cluster centroid

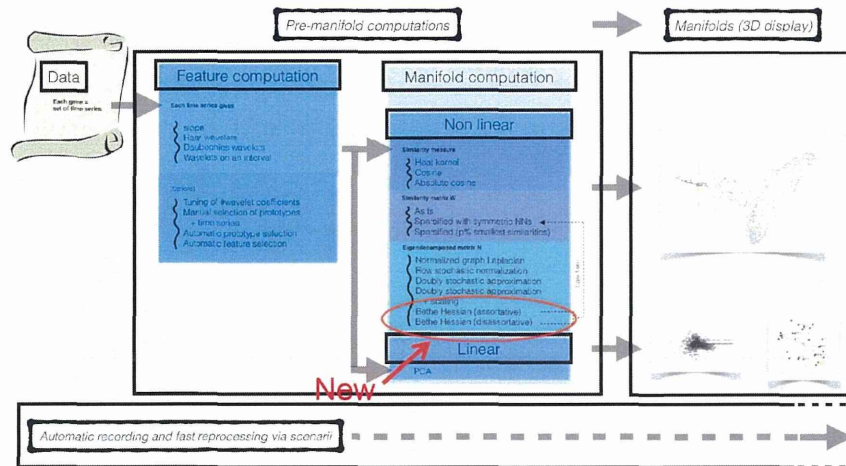
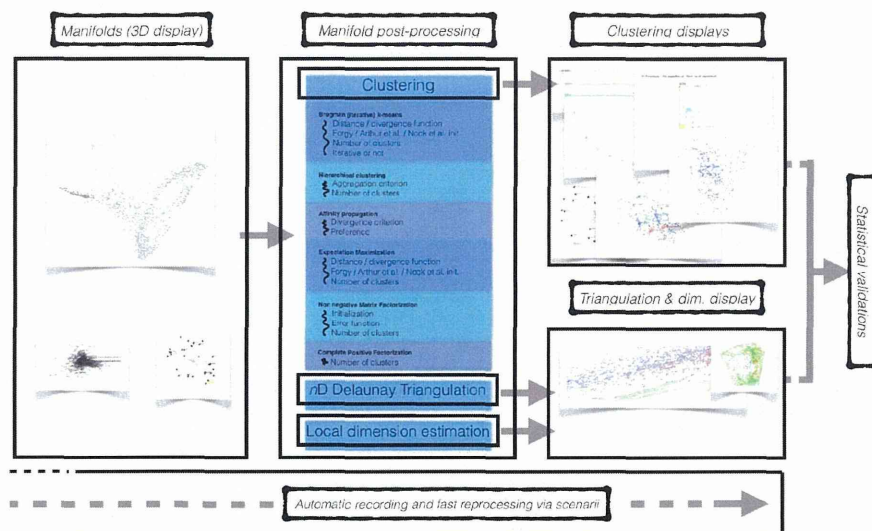


Figure 6: Summary of the workflow from loading the data to computing and displaying the manifolds. Each step can be recorded in a scenario and replayed for automated and fast reprocessing.



AGCT: A Geometric Clustering Tool for robustly unraveling the inner cluster structures of gene expressions Richard Nock, Natalia Polouliakh, Keigo Oka, Frank Nielsen, and Hiroaki Kitano (Manuscript is in submission)

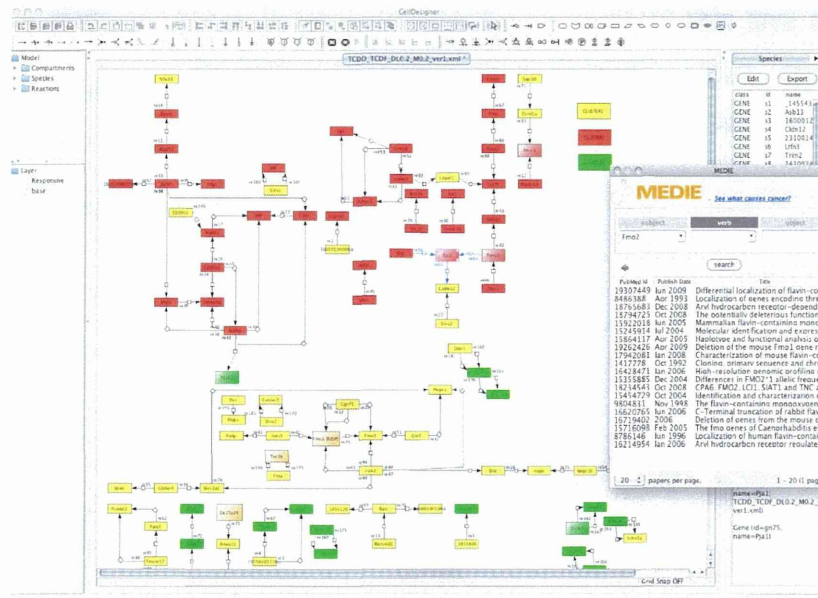
AGCT is connected with Garuda platform

The screenshot displays the Garuda platform interface. At the top, there's a search bar for gadgets and a list of categories including 'A Starter Kit', 'Analytics', 'CellDesigner Plugin', 'Clustering', 'Database', 'Garuda Gateway', 'Installers', 'Modeling', 'Pathways', 'Simulation', and 'Visualization'. The main area shows a 'Load Your Data' section with a table of file contents and a 'Garuda Discovery Engine' section with various analysis tools. A table of file contents is visible:

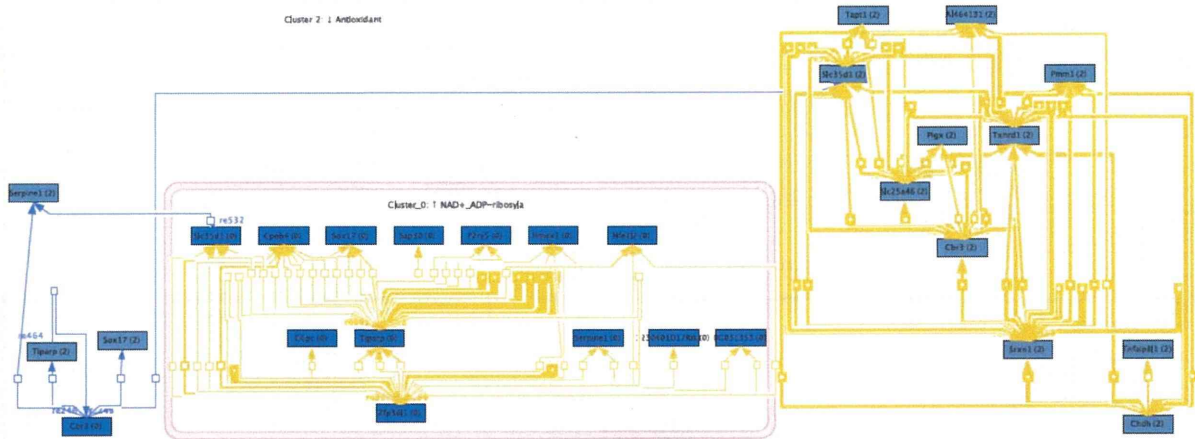
File Contents	1450046_at	1419004_s_at	1436234_at	1416005_at	1419152_at	1413916_a_at	1435823_x_at	1434565_at
	3.258562	2.993156	1.755584	38.64732	0.4612821	46.2718	9.997685	20.69522
	3.470453	1.89925	2.106749	35.2551	0.5889461	55.55189	12.26441	17.46994
	2.573193	1.198808	1.591004	26.28571	0.3524815	60.21937	11.531	14.65098
	4.2022	1.960919	1.7950	33.30863	0.6071238	57.44639	9.658074	15.71447
	1.6733	1.960919	1.7950	37.145	0.5072	67.966	9.4131	17.296

Below the table, there's a 'Garuda Discovery Engine' section with various analysis tools. The bottom part of the screenshot shows a 3D visualization of gene clusters, with a 'Gene to find' field and a 'Filters' section. The filters include 'and/or', 'list1-4', 'range of magnitude', 'range of copy per cell', 'elimSurc', and 'No series of copy per cell under 1'. A 'Time-ExpressionLevel Plot' is also visible, showing gene expression levels over time for different clusters.

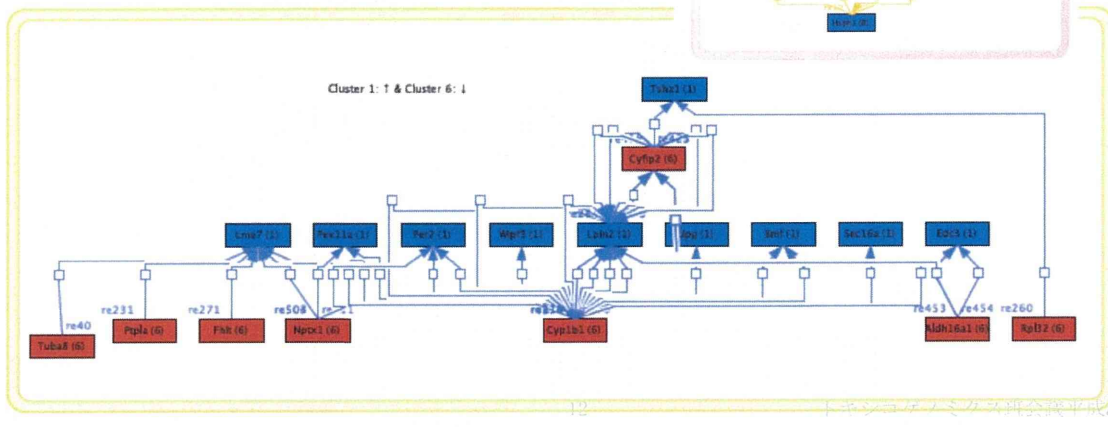
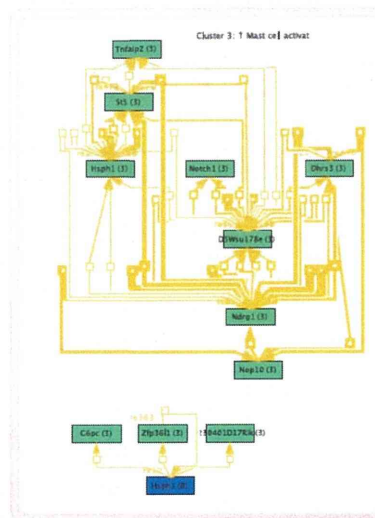
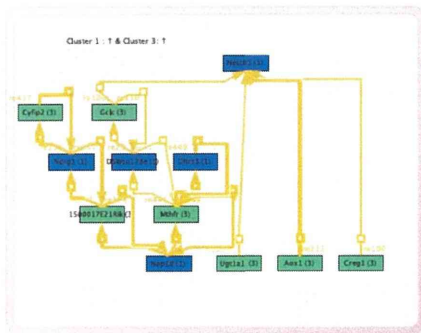
AGCT Plugin to CellDesigner -> automatically generated MAP



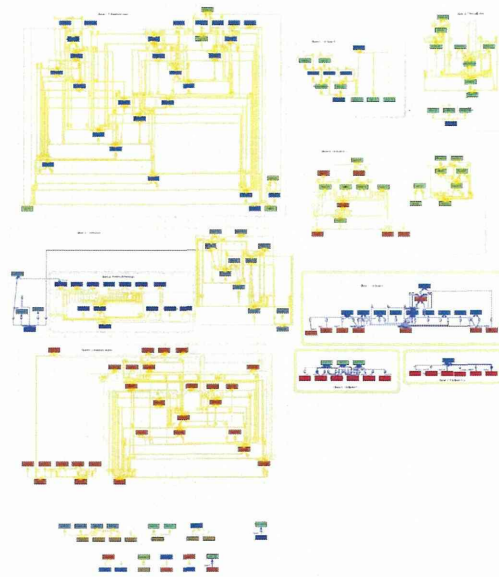
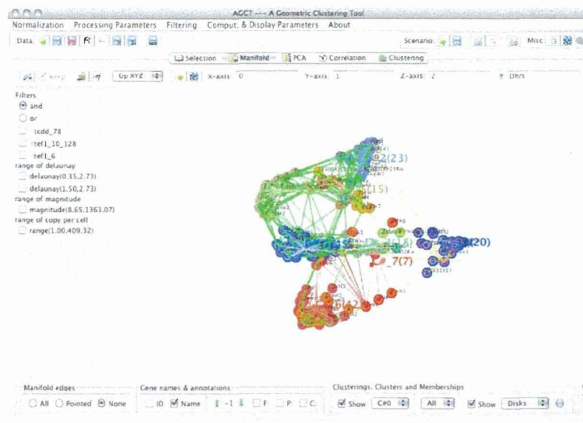
Drawn based on DT value



if $\theta > \frac{\pi}{2}$ — Negative correlation
 if $\theta < \frac{\pi}{2}$ — Positive correlation



TCDD-TCDF gene clustering automatically loaded on CellDesigner



Future plan for AGCT

- To improve the AGCT-Cell Designer MAP
 - Cut redundant edges
 - Loading speed
- Finalize AGCT 3.0

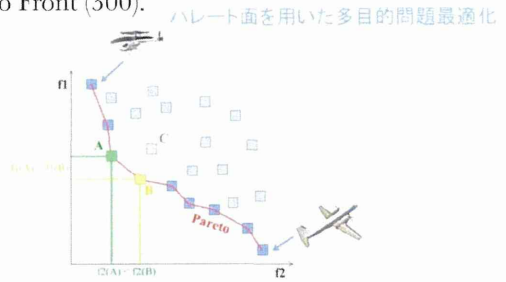
Analysis steps:

- Multiple alignment is obtained on Human-Mouse-Rat ($sim \geq 50\%$).
- Two scores Multiple Alignment score and Position Specific Scoring Matrix score are calculated.
- Top scores from two distributions **are selected** by Pareto Front (300).

File Name: NM_024207.5000134059248-124059301_1aln Location: to TSS

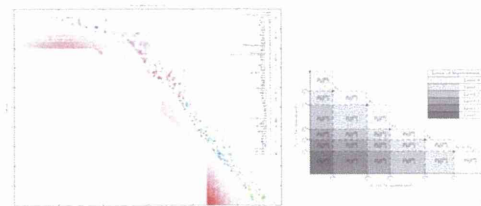
File Name: NM_024207.5000134059248-124059271_1aln.fasta.result (161,085 lines) Location: to TSS

PSM	TF	Location to TSS	Str	MA Score	PSM Score	PSM Rank	MA Rank
M00330	Major	175 ~ 193	-1	9.50	-7.38	16.89	
M00733	SMAD-4	161 ~ 175	-1	8.27	-6.86	15.14	
M00787	HIF-1	165 ~ 178	-1	7.46	-4.43	11.91	
M00647	LXR	159 ~ 176	+1	9.83	-9.95	19.59	
M00330	Major	175 ~ 193	-1	9.50	-7.38	16.89	
M00214	SEF-1	173 ~ 191	-1	9.48	-9.77	19.26	
M00806	NF-1	161 ~ 177	-1	9.47	-9.81	19.29	
M00376	Pax-8	153 ~ 170	+1	9.40	-9.59	19.09	
M00453	IRF-7	153 ~ 170	+1	9.40	-9.82	19.23	
M00929	MyoD	154 ~ 171	+1	9.37	-9.83	19.21	
M00453	IRF-7	153 ~ 170	+1	9.35	-9.95	19.31	
M00647	LXR	164 ~ 181	+1	9.17	-9.62	18.91	
M00205	AP-4	165 ~ 182	+1	9.17	-9.90	19.38	



工学では多目的最適化にあたって、ハレート面を求めてから、トレードオフを考慮して設計することが多い。様々な属性が付与された遺伝子配列群を整理。SSM ScoreおよびMA Scoreがどちらも大きくなるようにハレート面を求めた。

ハレート面を用いた多目的問題最適化



ハレート面を求めるプログラムおよび図示するところまで実装した。その結果の一部が左上の図になる。Nrf2およびPCPの遺伝子群同士で解析したところ、ハレート面の領域では、対応する転写因子数の殆どが強い相関(0.95)を示した。しかし、PCPに関してのみ、Nrf2に存在しない転写因子E2F5が最も多く現れ、差異を示した。

SONY
Sony CSL

SHOE
Sequence Homology in higher Eukaryotes

Job Input
Queue List

Job Title
Gene List

mTOR (H.sapiens)
example
HIF1A
RPS6KB1
RPS6KB2
EIF4B
EIF4EBP1
EIF4E
MAPK1
MAPK3
DDIT4
TNF
 Repeat Masker

Upstream Length 10000 mode 1 mode 2

Downstream Length 200

Scoring transfac32 ips jaspar add scoring

Max Pareto 30

Reset Submit

