Deciphering gene network regulation mechanisms through cluster and phylogenetic footprinting analysis

Natalia POLOULIAKH$^{1,2}$, Hiroaki KITANO$^{1,2}$

$^1$Sony Computer Science Laboratories, Inc., $^2$Systems Biology Institute

The 'Percellome' database is a unique source of toxicity response data for more than 100 chemicals obtained on various organs of mice (Kanno group, NIH). Originally developed “per cell” readout in mRNA copy method with a high precision in measuring gene expression permits to include each molecule to the network analysis, treating cell as an Open System.

We have developed two software tools: AGCT (A Geometric Clustering Tool) and SHOE (Sequence HOmology in higher Eukaryotes) promoter analyzer suitable for 'Percellome' data. AGCT analyses large-scale ($40,000^6$ probes) time-series data with spectral clustering manifold that represents and visualize complex data in low dimension.

Subsequently low dimension manifold is clustered with novel techniques such as Bregman K-means, Affinity Propagation (AP), Non-Negative Matrix Factorization (NMF), Expectation Maximization and others to reach the finite number of clusters with similar expression profile. Those clusters are hereafter analyzed on SHOE, which aims to predict transcription factors network basing on phylogenetic footprinting of three genomes: human, mouse and rat.

Since AGCT and SHOE are connected to Garuda platform (Systems Biology Institute), their outputs can be automatically passed to other tools on Garuda platform for further analysis or visualization.