The Ping-Pong Algorithm

Kutalik, Beckmann, Bergmann NBT 08

Some slides from Sven Bergmann
http://serverdgm.unil.ch/bergmann
The NCI-60 drug response and GE db

Scherf et al. Nat Gen 00

**R**

Response of cell line **j** to drug **i**

**E**

Exp of gene **i** in untreated cell line **j**

**Database A**

Inventory

~500,000 cmpds

60-Cell line screen

>70,000 cmpds

In vivo studies

Clinical Trials

**Database T**

Gene expression

60 cell lines

>70K cmpds

60 cell lines

>4,000,000 numbers

**T**

(9,703 genes)

(41 individually assessed targets)
Clustering the cell lines

Based on E

Based on R

Very different pictures. Can we reconcile?
Approaches

- **ISA(E)**: get transcription modules. Ignores R.
- **ISA (E·RT)**: compute drug-gene correlation across the cell lines, generate drug-gene modules by compiling info across all cell lines. Misses signals on few cell lines.
- **ISA(E)&ISA(T)**: Suppose modules are represented as row/col weights. Get modules \{\{(di,ci)\}\} in D and \{\{(gi,c*i)\}\}. Select highly correlated module pairs: \text{corr}(ci,c*j)>\alpha \text{ “late integration”}
- **Cluster (E·RT)**: Clusters the genes using HC, Set cutoff to get desired no. of clusters. Cluster the drugs too. Match gene and drug clusters by mean bic signal. “late integration”
Generalizing ISA

Gene-modules

Co-modules

Drug-modules

[AGF] [CDF] [BFC]

[AGF] [CDF] [BFC]

[AGF] [CDF] [BFC]
For a given threshold combination \((t_C: \text{condition threshold}, t_G: \text{gene threshold}, t_D: \text{drug threshold})\) the Ping-pong algorithm (PPA) is summarized in the following pseudocode:

\[ n = 0; \quad g^{(0)} = \text{random}(N_G) \in [0,1]^{NG} \text{(initial random seed)} \]

\[ \text{while } (|\hat{g}^{(n)} - \hat{g}^{(n-1)}| + |\hat{d}^{(n)} - \hat{d}^{(n-1)}| + |\hat{c}^{(n)} - \hat{c}^{(n-1)}| + |\hat{c}^{(n)} - \hat{c}^{(n)}| > \varepsilon ) \]

1. \( c = E_G^T \cdot \hat{g}^{(n)} ; \quad c_j^{(n+1)} = \begin{cases} c_j : \text{if } |c_j - \mu(c)| > t_C \sigma(c) \\ 0: \text{otherwise} \end{cases} \quad (j = 1, \ldots, N_C) \)

2. \( d = R_c \cdot \hat{c}^{(n)} ; \quad d_k^{(n+1)} = \begin{cases} d_k : \text{if } |d_k - \mu(d)| > t_D \sigma(d) \\ 0: \text{otherwise} \end{cases} \quad (k = 1, \ldots, N_D) \)

3. \( \tilde{c} = R_D^T \cdot \hat{d}^{(n)} ; \quad \tilde{c}_l^{(n+1)} = \begin{cases} \tilde{c}_l : \text{if } |\tilde{c}_l - \mu(\tilde{c})| > \tilde{t}_C \sigma(\tilde{c}) \\ 0: \text{otherwise} \end{cases} \quad (l = 1, \ldots, N_C) \)

4. \( g = E_c \cdot \tilde{c}^{(n)} ; \quad g_m^{(n+1)} = \begin{cases} g_m : \text{if } |g_m - \mu(g)| > t_G \sigma(g) \\ 0: \text{otherwise} \end{cases} \quad (m = 1, \ldots, N_G) \)

5. \( n = n + 1 \)

\[ g^* = g^{(n)} ; \quad \tilde{c}^* = \tilde{c}^{(n)} ; \quad d^* = d^{(n)} \]
Modules and Co-modules
Iteratively refine genes, cell-lines and drugs to get co-modules

The Ping-Pong algorithm!
Results on NCI-60 (859 modules)
Co-modules have predictive power for drug-gene associations
Co-modules analysis provides biological focus through data integration.


Significance of biological processes by alternative algorithm.
Analysis of large-scale expression data bears great potential to understand global transcription programs and their evolution.

Innovative analysis tools needed to extract information from such data.

ISA & Ping-Pong Algorithms:
- decomposes data into “transcription modules”
- integrates external information
- allows for interspecies comparative analysis