GSCMap – A Gene-Set-based Connectivity Map for characterizing bioactive compounds in terms of biological functional groups

Hoong-Chien Lee 李弘謙
Department of Physics, Chung Yuan Christian University
Institute of Systems Biology & Bioinformatics, National Central University
What we do

• Our interest – study complex disease, its early detection, diagnosis, and treatment

• Our methods
  • Gene-expression data of patient cohorts of complex diseases, drug and gene-set databases
  • Functional gene-set based analysis
  • Repurposed drug prediction for systems systems treatment
  • 西藥中用、舊藥新用
Systems Biology

• “Systems biology is ... integration rather than reduction, and makes heavy use of mathematical and computational models

• Advances in high-through methods, especially sequencing techniques, and powerful computers have made the practice of systems biology possible

• Philosophy of SB has similarity to Chinese traditional medicine. USFDA recently changed the term alternative medicine to integrated medicine
Gene expression data – metadata on cell activity

- Proteins are “made” by genes that are expressed
- Expressed genes appear as mRNAs in the cell
- Quantitative measures of mRNA densities in the cell samples are indirect measures of cell activity
- These can be done by microarrays experiments (lately, by “next-generation sequencing” facilities)
Functional genomics profile (FGP) and cell state

- Genes do not function individually
- Genes work in a coordinated fashion in functional groups

IGA – individual gene analysis
GSA – gene-set analysis

Microarray
NGS

cell state

functional profile

sample

functional gene groups database
The $\Delta$FGP-Disorder-Drug Trinity

- Changes in FGP $\Leftrightarrow$ change in cell state
- Disorders – change cell states
- Drugs – change cell states
GSEA – Gene Set Enhancement Analysis

GSEA assigns an enhancement score (ES) to a gene-set over a genomic profile

Subramanian A et al. PNAS 2005;102:15545-15550
### Database for ~8000 functional gene sets

**MSigDB**

**Molecular Signatures Database v4.0**

Subramanian, Tamayo, et al. (2005, PNAS 102, 15545-15550)

| c1 | positional gene sets | for each human chromosome and cytogenetic band. |
| c2 | curated gene sets | from online pathway databases, publications in PubMed, and knowledge of domain experts. |
| c3 | motif gene sets | based on conserved cis-regulatory motifs from a comparative analysis of the human, mouse, rat, and dog genomes. |
| c4 | computational gene sets | defined by mining large collections of cancer-oriented microarray data. |
| c5 | GO gene sets | consist of genes annotated by the same GO terms. |
| c6 | oncogenic signatures | defined directly from microarray gene expression data from cancer gene perturbations. |
| c7 | immunologic signatures | defined directly from microarray gene expression data from immunologic studies. |
The Connectivity Map (CMap) – Database on genomic profiles of drug effects (6,097 treatments/chips on 1,309 drugs/small molecules)

We built a local version of CMap that can be queried in batch mode.

Functional genomic studies of some complex diseases

- Colon cancer
- Brain cancer (astrocytoma)
- Psychiatric disorders (bipolar, ...)
- Type 2 diabetes
- Aging, others ...
GSCMap – A GSA-based version of Connectivity Map

Used GSEA and gene-sets from MSigDB to convert genomic profiles of drug to functional profiles

• CMap is a 6,097 x 22,283 (drug instances x probe on microarray) matrix

• GSCMap is a 1,309 x 4,883 (drugs x gene sets) matrix
CMap clusters instances by cell type
GSCMap clusters by drug effect

In Cmap, drugs cluster by cell types, not by drug effect

In GSCmap, drugs no longer cluster by cell types

Chung, Jin & Lee (2014)
Drug pairs correlate much better in GSCMap/GSA than in CMap/IGA

Test on the drug pair Trichostatin A & LY-294002
GSLHC – Gene-set local hierarchical cluster

An application of GSMap to identify properties of bioactive molecules
GSLHC protocol

1. Use the global matrix as database
2. Select drug or drugs of interest
3. Select functional gene sets (FGS) with ES scores with permutation $p < 0.005$
4. Use selected FGS to do hierarchical cluster with all drugs
5. Identify drug sub-clusters (clades) with correlation $> 0.9$, then select the clade to which drug(s) of interest belongs
Testing and first application of GSHLC – finding known and novel HDAC inhibitors
Example 2 - Identification of 0175029-0000 as a novel cyclin-dependent kinase inhibitor (CDKi)
Example 3 - Identification of CP-863187 as a potential antibiotic
• Applied GSLHC on the 106 small-molecules of unknown biofunction in C-Map using the criteria
  • at least 20 functional gene sets contain $p < 0.005$
  • Correlation $> 0.9$

• Found the putative indications of 18 among the 106 small-molecules (others failed to satisfy criteria)
The 18 C-Map perturbagens with newly discovered putative molecular target and pharmacological indication

<table>
<thead>
<tr>
<th>Test drug</th>
<th>Cor.</th>
<th>Partner drug</th>
<th>putative target *</th>
<th>Indication*</th>
</tr>
</thead>
<tbody>
<tr>
<td>5186324</td>
<td>0.99</td>
<td>neostigmine bromide</td>
<td>Acetylcholinesterase inhibitor</td>
<td>Myasthenia gravis</td>
</tr>
<tr>
<td>DL-PPMP</td>
<td>0.99</td>
<td>indoprofen</td>
<td>Cyclooxygenase-1 inhibitor</td>
<td>Non-steroidal anti-inflammatory drug</td>
</tr>
<tr>
<td>Prestwick-692</td>
<td>0.99</td>
<td>isoflupredone</td>
<td>Glucocorticoid receptor agonist</td>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>tyrphostin AG-825</td>
<td>0.99</td>
<td>camptothecin</td>
<td>DNA topoisomerase I inhibitor</td>
<td>Cancer</td>
</tr>
<tr>
<td>5248896</td>
<td>0.98</td>
<td>tyrphostin AG-825</td>
<td>human epidermal growth factor receptor (HER)-2/neu inhibitor</td>
<td></td>
</tr>
<tr>
<td>0175029-0000</td>
<td>0.98</td>
<td>GW-8510</td>
<td>Cyclin-dependent kinase 2 Inhibitor</td>
<td>Cancer</td>
</tr>
<tr>
<td>CP-863187</td>
<td>0.98</td>
<td>piperacillin</td>
<td>Sodium channel blocker</td>
<td>Anesthetic</td>
</tr>
<tr>
<td>H-7</td>
<td>0.98</td>
<td>GW-8510</td>
<td>Cyclin-dependent kinase 2 Inhibitor</td>
<td>Cancer</td>
</tr>
<tr>
<td>Prestwick-1103</td>
<td>0.98</td>
<td>pentoxifylline</td>
<td>Tumor necrosis factor antibody</td>
<td>Intermittent claudication</td>
</tr>
<tr>
<td>U0125</td>
<td>0.98</td>
<td>irinotecan</td>
<td>DNA topoisomerase I inhibitor</td>
<td>Colorectal Cancer</td>
</tr>
<tr>
<td>5109870</td>
<td>0.97</td>
<td>phenoxybenzamine</td>
<td>Alpha adrenergic receptor antagonist</td>
<td>Hypertension</td>
</tr>
<tr>
<td>MG-132</td>
<td>0.97</td>
<td>MG-262</td>
<td>Proteasome Inhibitor</td>
<td></td>
</tr>
<tr>
<td>PHA-00851261E</td>
<td>0.97</td>
<td>amrinone</td>
<td>CGMP-inhibited 3',5'-cyclic phosphodiesterase</td>
<td>Congestive heart failure</td>
</tr>
<tr>
<td>STOCK1N-35215</td>
<td>0.97</td>
<td>MS-275</td>
<td>Histone deacetylase inhibitor</td>
<td>Cancer</td>
</tr>
<tr>
<td>0297417-0002B</td>
<td>0.95</td>
<td>8-azidoguanine</td>
<td>Purine nucleoside phosphorylase Inhibitor</td>
<td>treatment of acute leukemia</td>
</tr>
<tr>
<td>F0447-0125</td>
<td>0.95</td>
<td>lomustine</td>
<td>DNA Inhibitor</td>
<td>Brain tumours</td>
</tr>
<tr>
<td>W-13</td>
<td>0.95</td>
<td>fludrocortisone</td>
<td>Mineralocorticoid receptor agonist</td>
<td>Cerebral salt-wasting syndrome</td>
</tr>
<tr>
<td>CP-944629</td>
<td>0.92</td>
<td>betulinic acid</td>
<td>DNA polymerase beta inhibitor</td>
<td>Cancer</td>
</tr>
</tbody>
</table>

Red drugs have putative anti-cancer effects
Summary

- Built local CMap that allows batch query
- Built GSCMap for gene-set-based analysis
- CMap/IGA was a drug classifier due to cell-type dominance
- GSCMap/GSA transcended cell-type and was a good drug classifier
- Built GSLHC to identify drug properties by association
- Eight "unknown" compounds in CMap were identified to have putative anti-tumor activities:
  - tyrphostin AG-825, 0175029-0000, H-7, U0125, STOCK1N-35215, F0447-0125, CP-944629, 0297417-0002B
• Work done by:
  – Dr. Feng-Hsiang Chung 鍾豐翔, PDF
  – ZH Jin 金鎮華, MSc students
  – Dr. Chih-Hao Chen 陳志浩, PDF

• Supported by
  – Ministry of Education
  – National Research Council/Ministry of Science & Technology
  – National Central University-Cathay General Hospital Joint Research Center
  – Center for Dynamical Biomarkers and Translational Medicine, National Central University
Thank you for your attention

謝謝聆聽 請多指教