BITS9

9th Cross-Strait Workshop on Biology Inspired Theoretical Science
June 26-28, Chung Yuan University, Zhongli, Taiwan

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

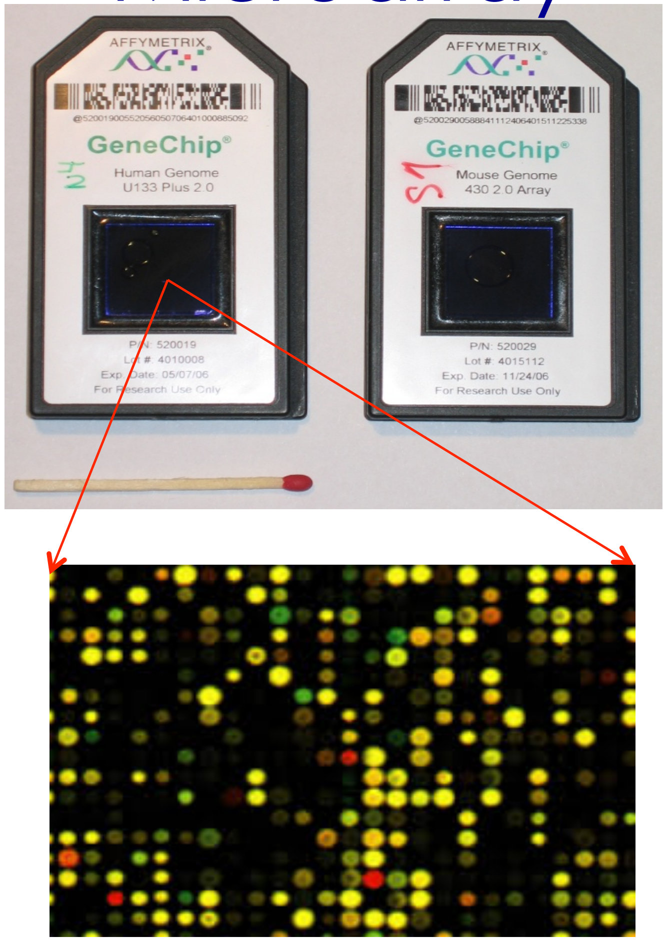
Systems Biology

- “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

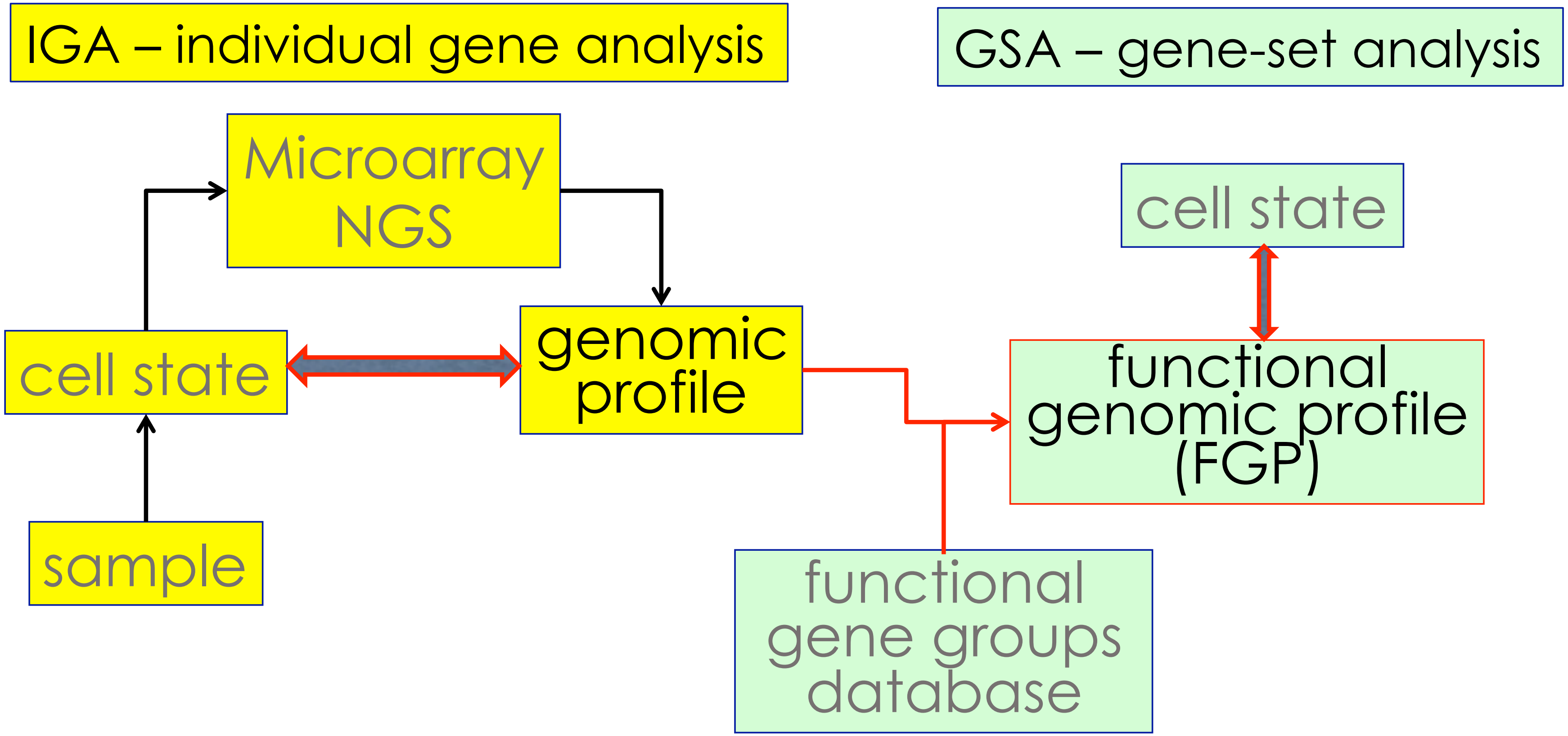
- Proteins are "made" by genes that are expressed
- Expressed genes appear as mRNAs in the cell
- Quantitative measures of mRNA densities in the cell samples are indirect measures of cell activity
- These can be done by microarrays experiments (lately, by "next-generation sequencing" facilities)

Microarray



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

genomic profile

sample

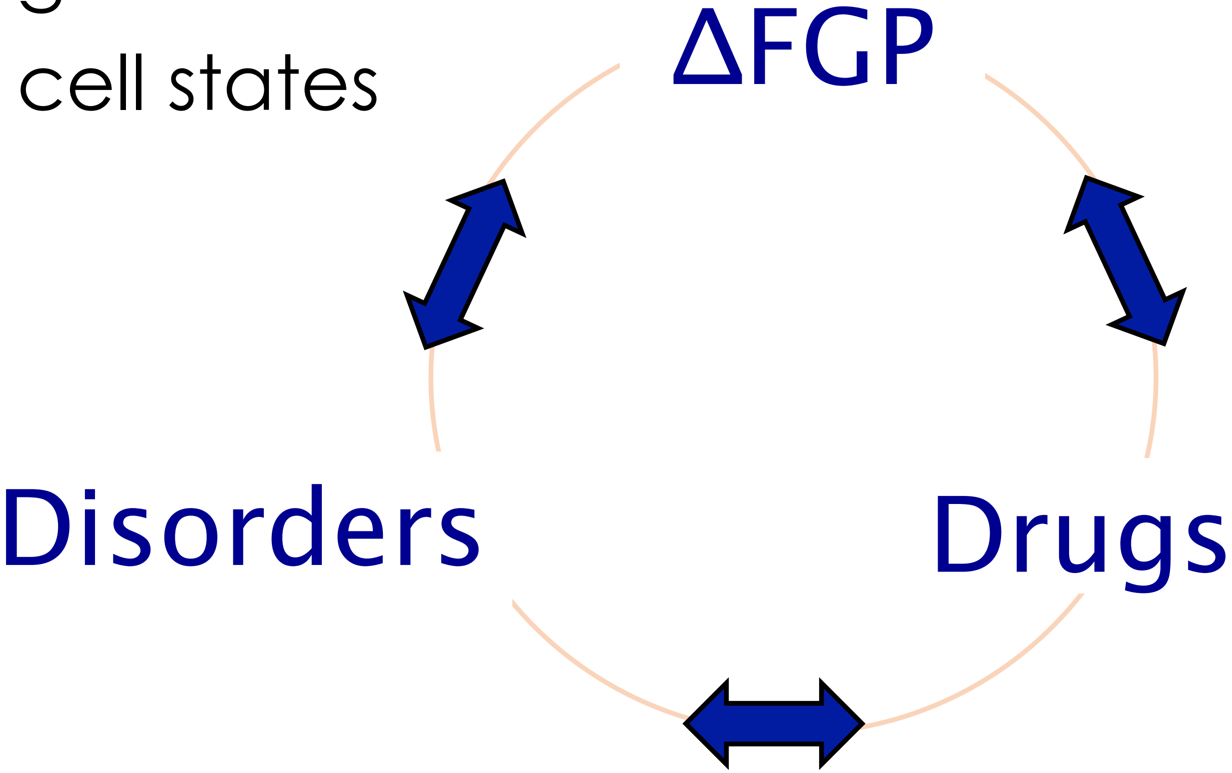
cell state

functional genomic profile (FGP)

functional gene groups database

The ΔFGP-Disorder-Drug Trinity

- Changes in FGP ⇔ change in cell state
- Disorders – change cell states
- Drugs – change cell states



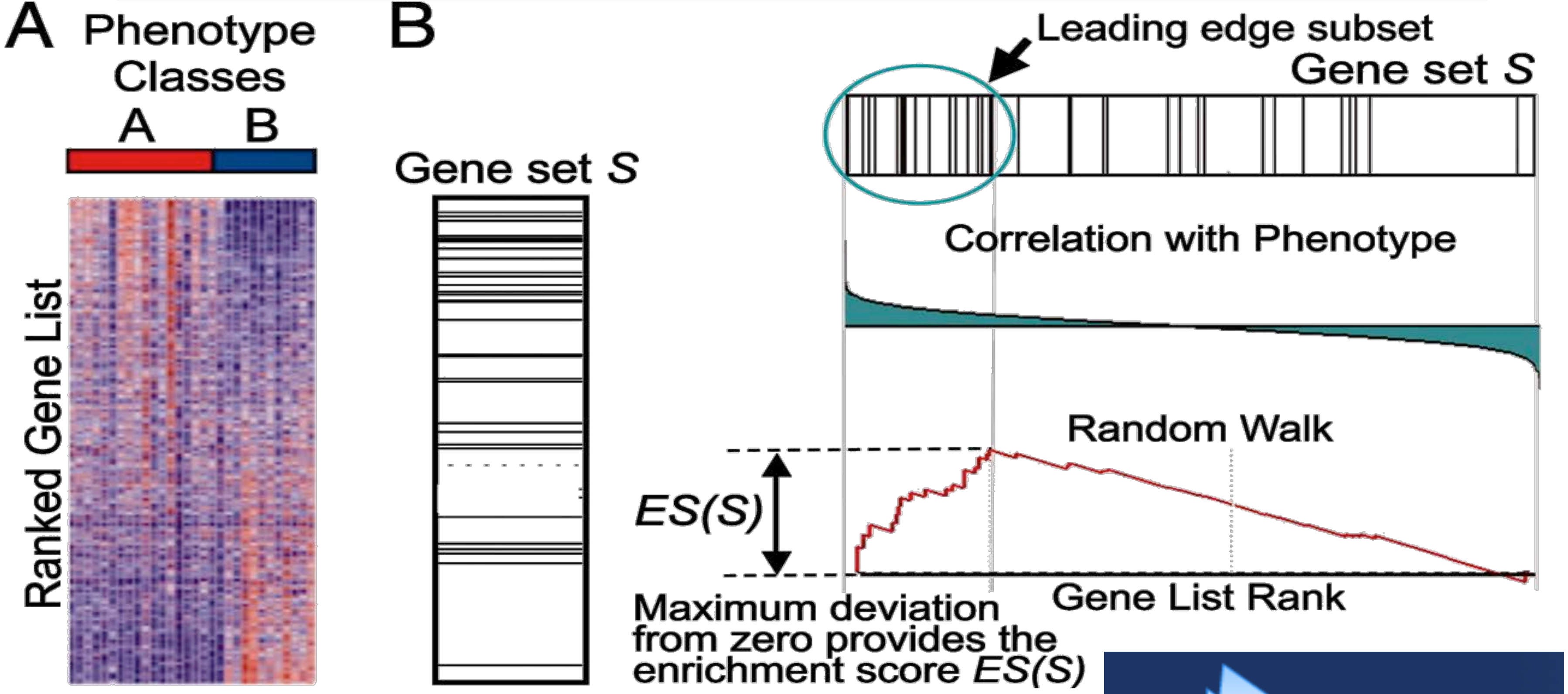
ΔFGP

Disorders

Drugs

GSEA – Gene Set Enhancement Analysis)

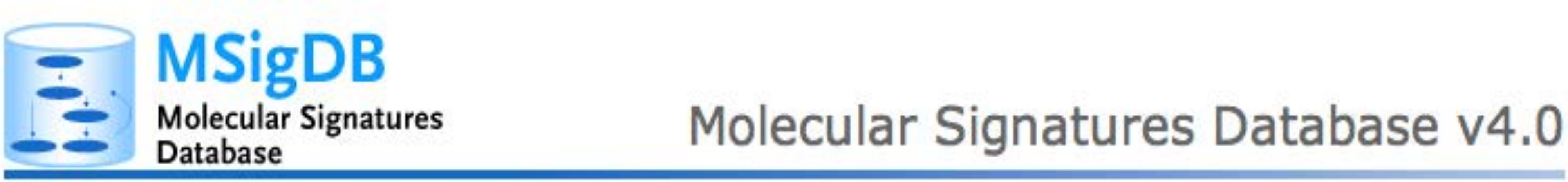
GSEA assigns an enhancement score (ES) to a gene-set over a genomic profile



Subramanian A et al. PNAS 2005;102:15545-15550

GSEA
Gene Set Enrichment Analysis

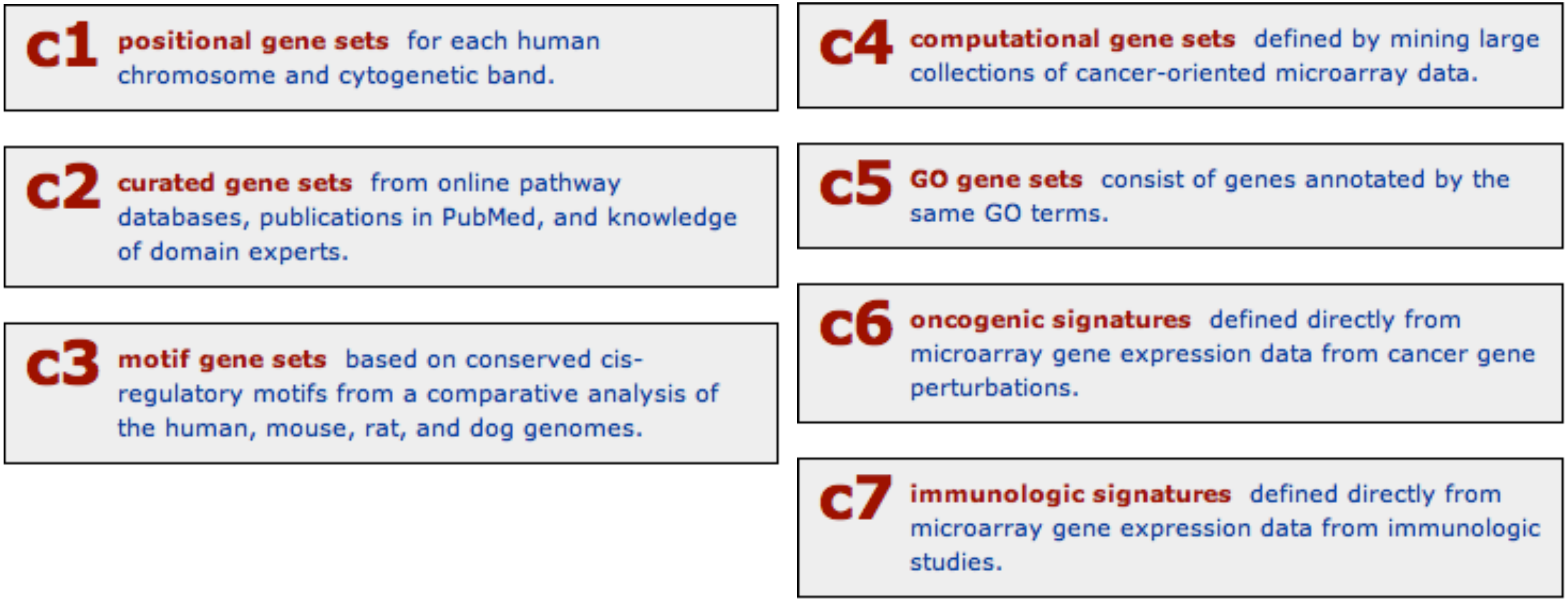
Database for ~8000 functional gene sets



MSigDB
Molecular Signatures Database

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.

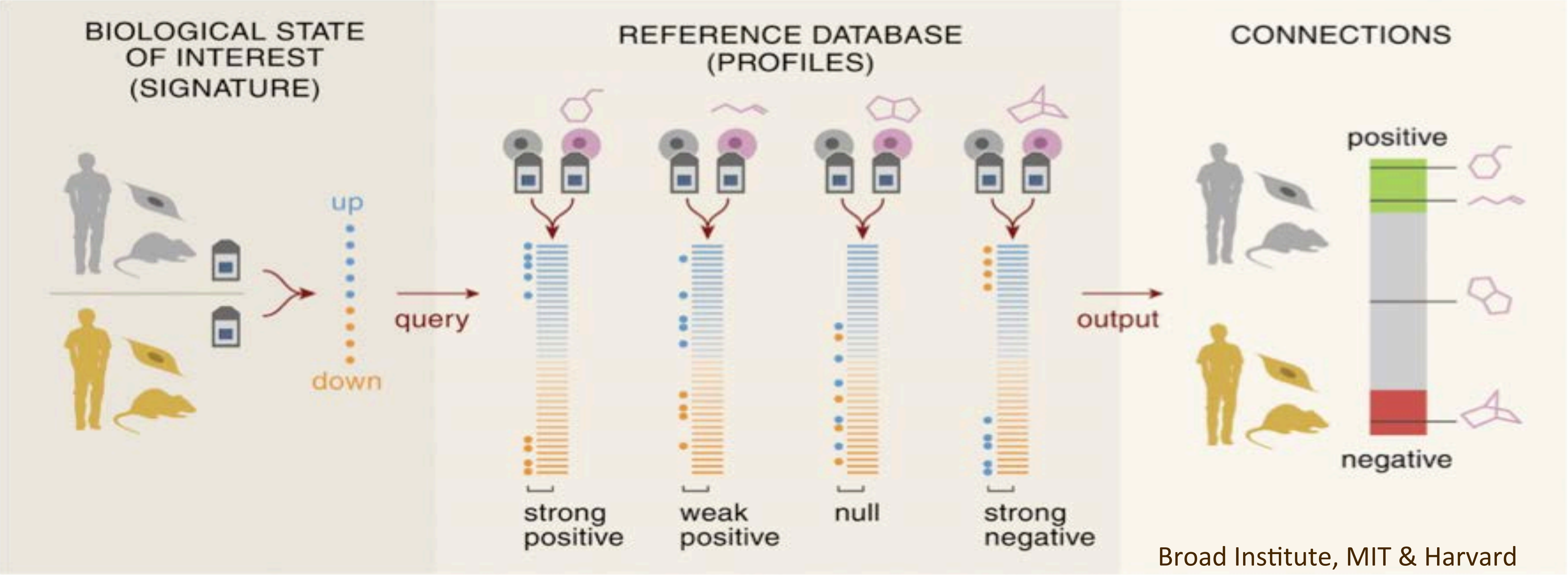
c5 **GO gene sets** consist of genes annotated by the same GO terms.

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)



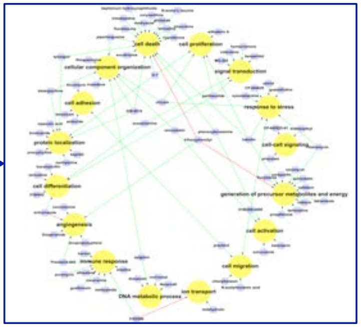
Broad Institute, MIT & Harvard

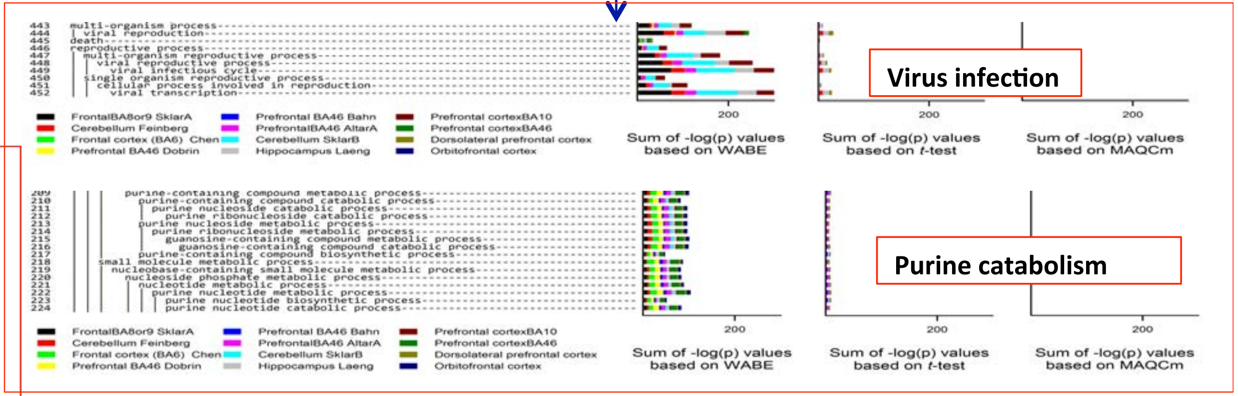
The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. 2006 Sep 29;**313**(5795):1929-35.

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

Functional genomic studies of some complex diseases

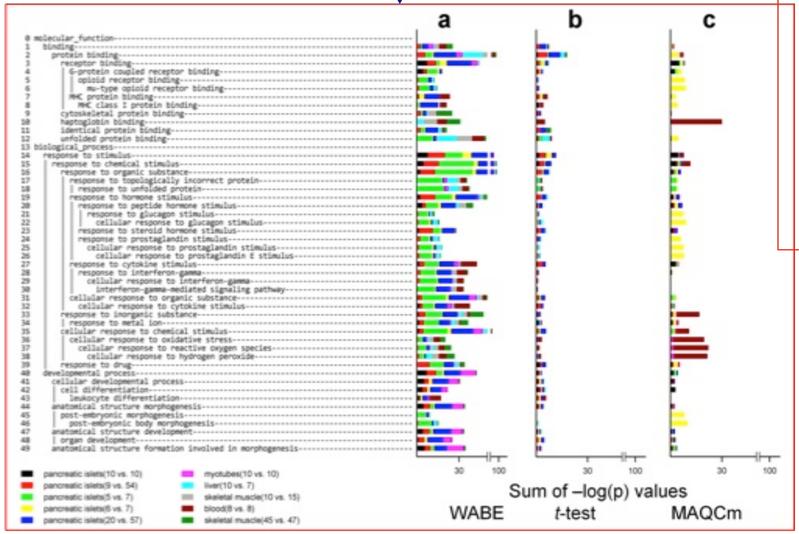
- Colon cancer
- Brain cancer (astrocytoma)
- Psychiatric disorders (bipolar, …)
- Type 2 diabetes
- Aging, others …





Virus infection

Purine catabolism

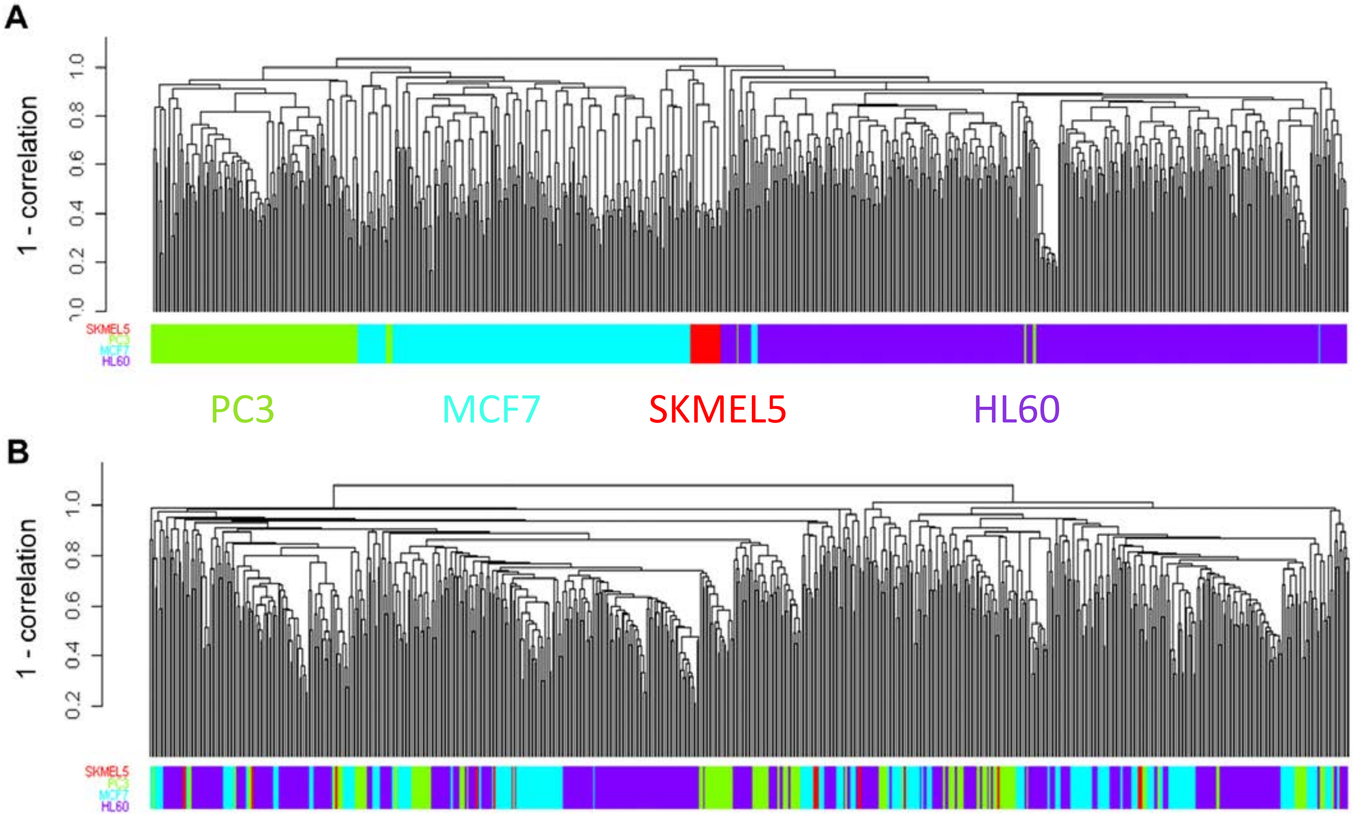


GSCMap – A GSA-based version of Connectivity Map

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



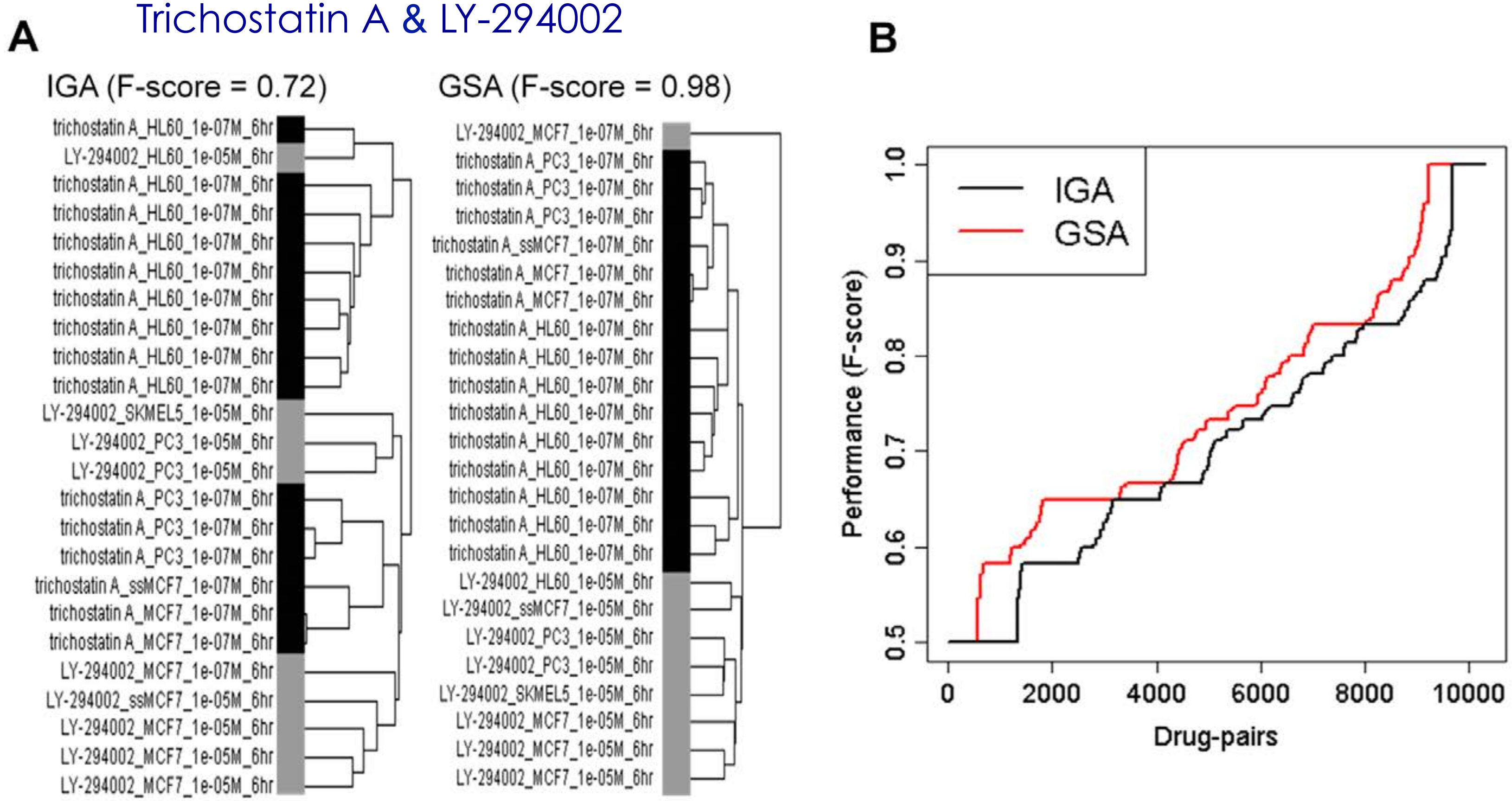
In Cmap, drugs cluster by cell types, not by drug effect

In GSCmap, drugs no longer cluster by cell types

Chung, Jin & Lee (2014)

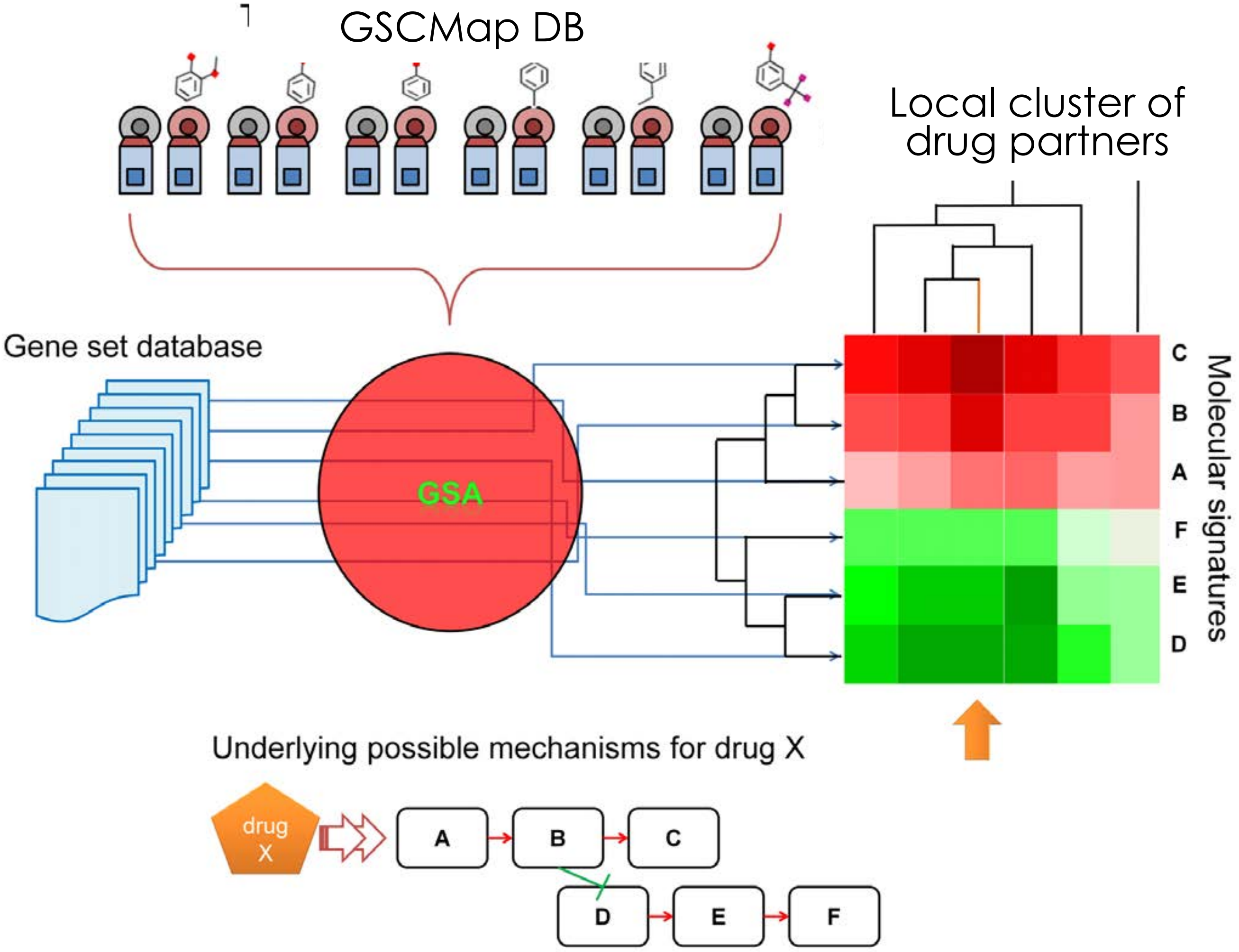
Drug pairs correlate much better in GSCMap/GSA than in CMap/IGA

Test on the drug pair
Trichostatin A & LY-294002



GSLHC – Gene-set local hierarchical cluster

An application of GSMap to identify properties of bio-active molecules



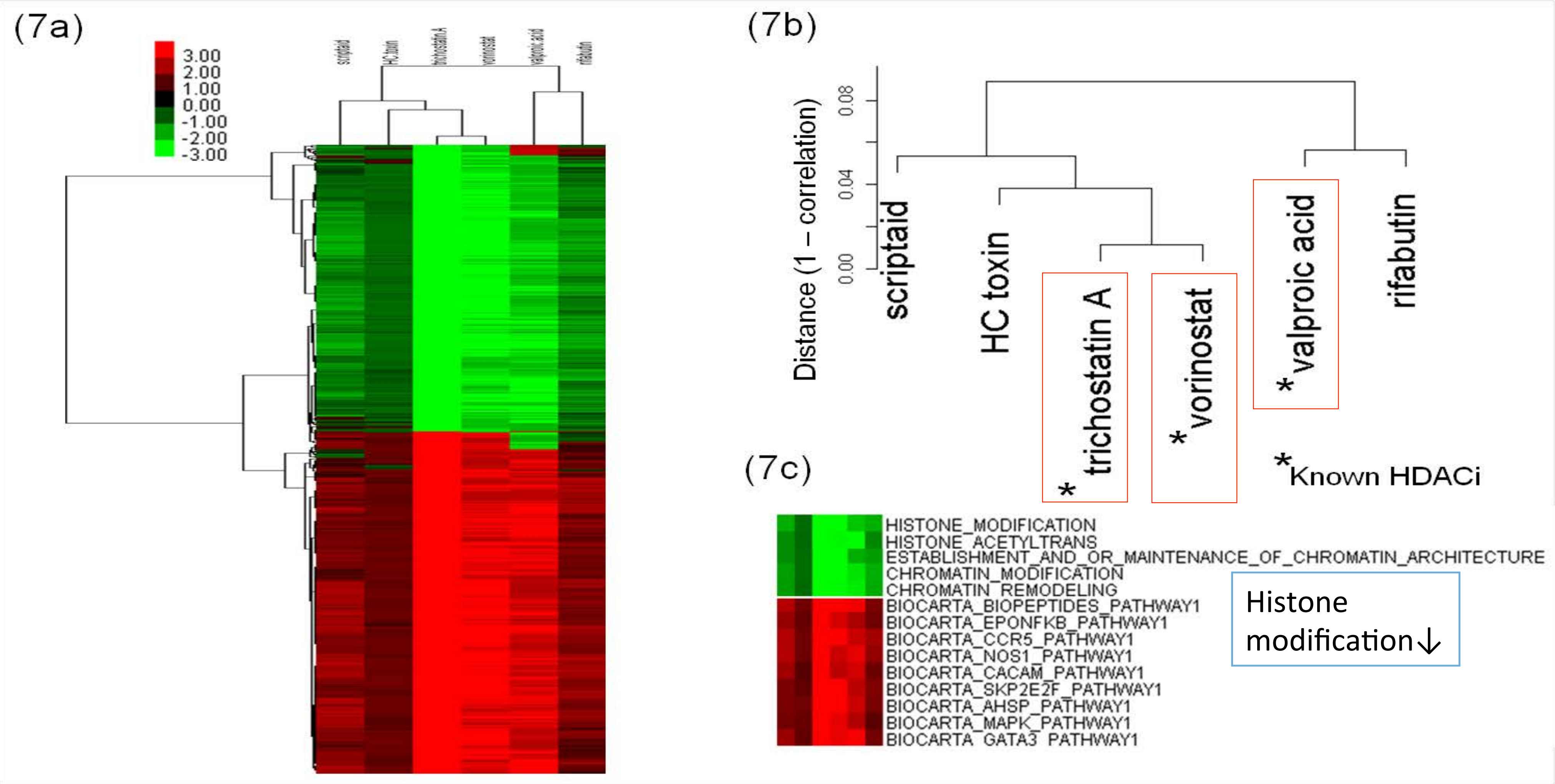
GSCMap DB

Local cluster of drug partners

GSLHC protocol

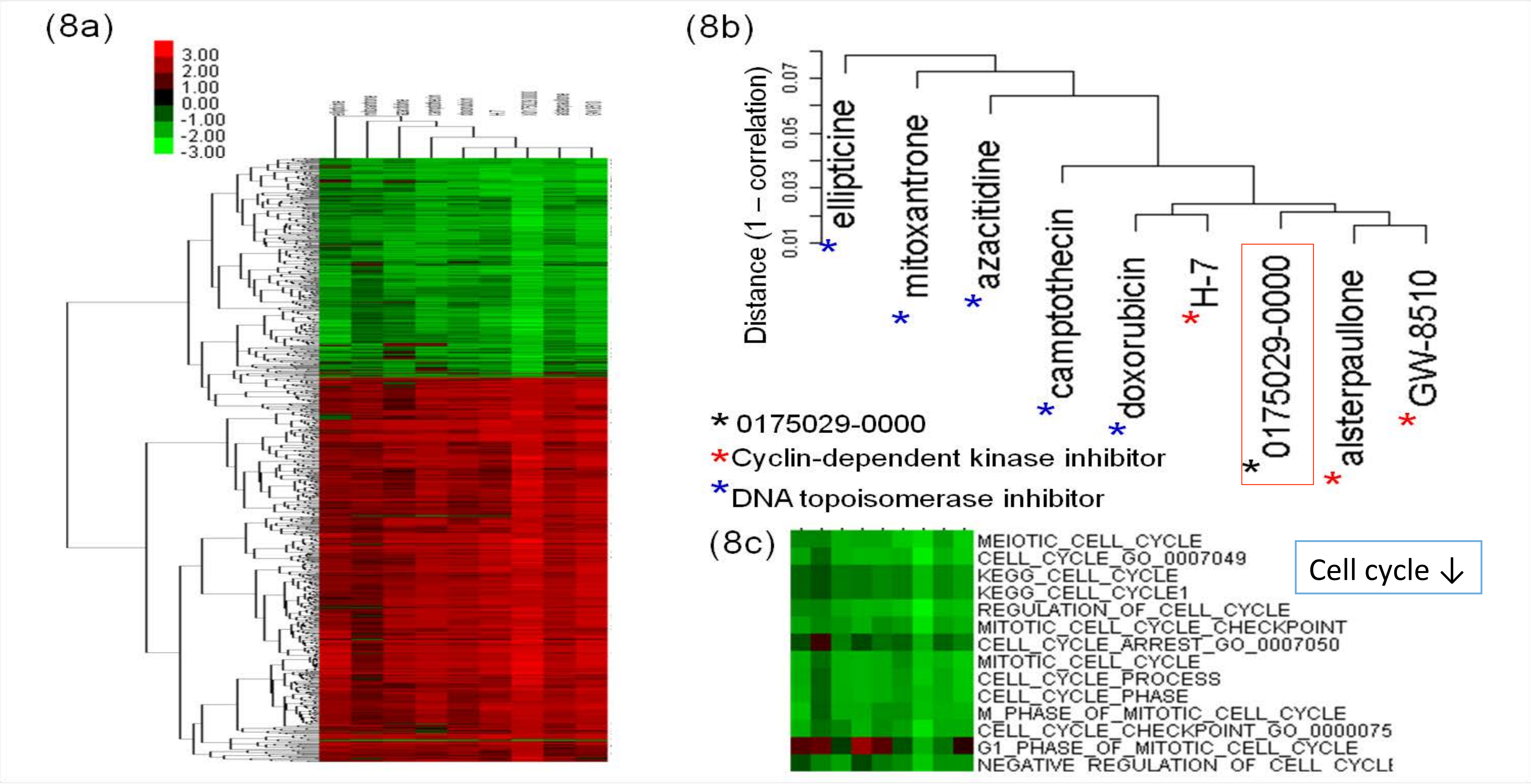
1. Use the global matrix as database
2. Select drug or drugs of interest
3. Select functional gene sets (FGS) with ES scores with permutation $p < 0.005$
4. Use selected FGS to do hierarchical cluster with all drugs
5. Identify drug sub-clusters (clades) with correlation > 0.9, then select the clade to which drug(s) of interest belongs

Testing and first application of GSHLC – finding known and novel HDAC inhibitors



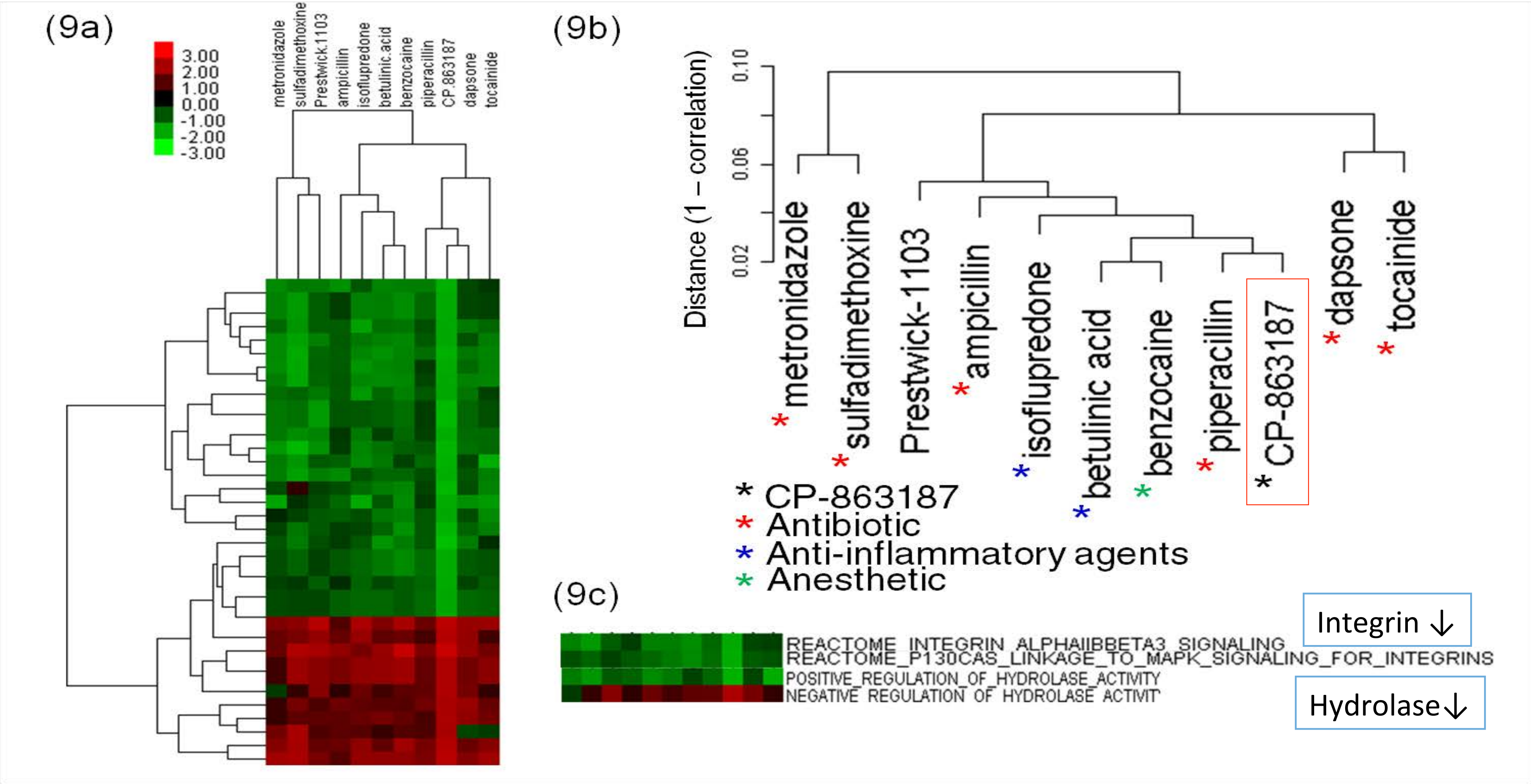
Histone modification↓

Example 2 - Identification of 0175029-0000 as a novel cyclin-dependent kinase inhibitor (CDKi)



Cell cycle ↓

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)

The18 C-Map perturbagens with newly discovered putative molecular target and pharmacological indication

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

Red drugs have putative anti0cancer 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 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

SYBBI
系統生物與生物資訊研究所
Institute of Systems Biology and Bioinformatics

- 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