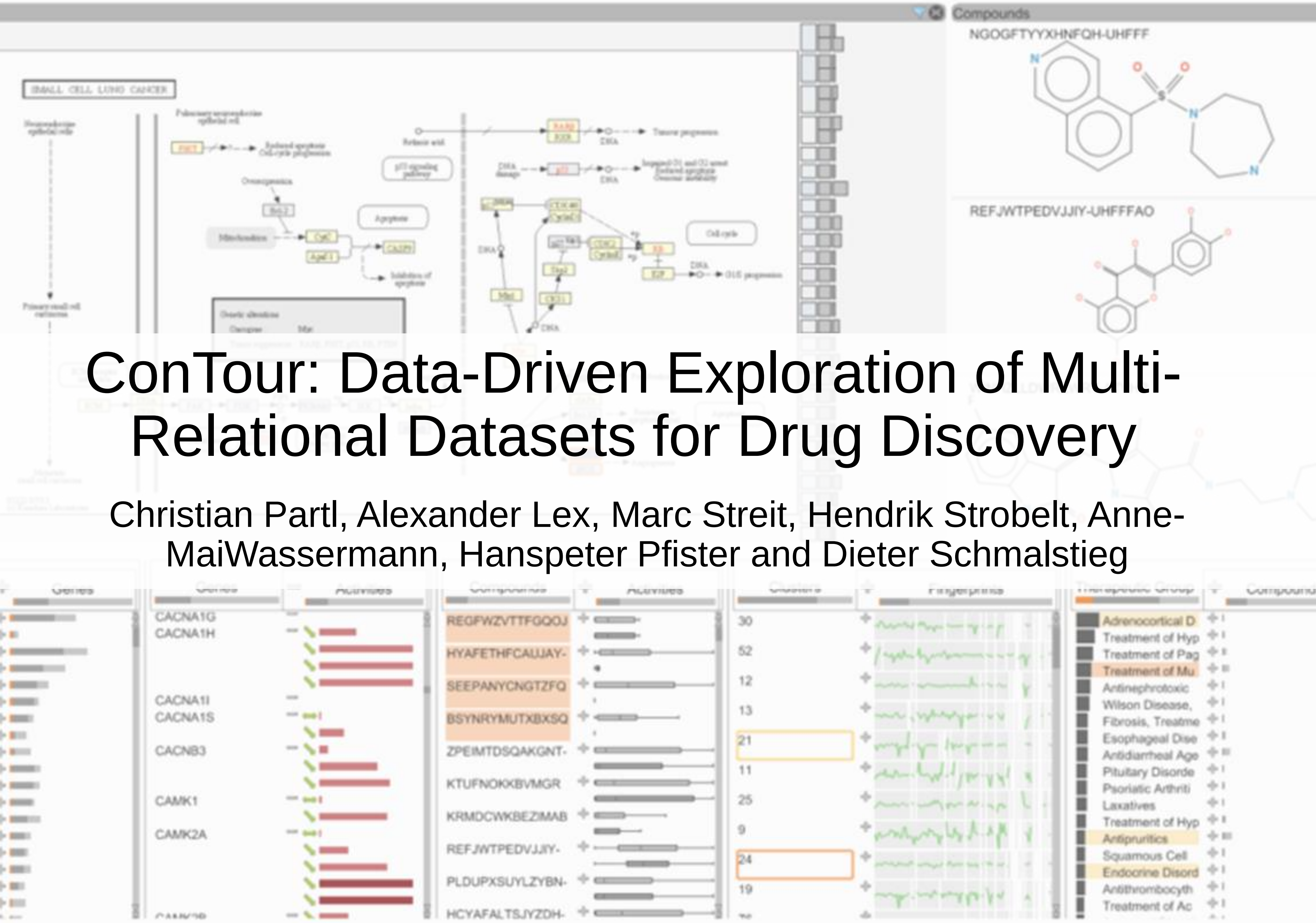


ConTour: Data-Driven Exploration of Multi-Relational Datasets for Drug Discovery

Christian Partl, Alexander Lex, Marc Streit, Hendrik Strobelt, Anne-MaiWassermann, Hanspeter Pfister and Dieter Schmalstieg

Domain Problem

UNDERSTANDING DRUG DISCOVERY

Scenario 1:
Targeted interaction, understood mechanism, desired outcome

🟦 = biological receptor

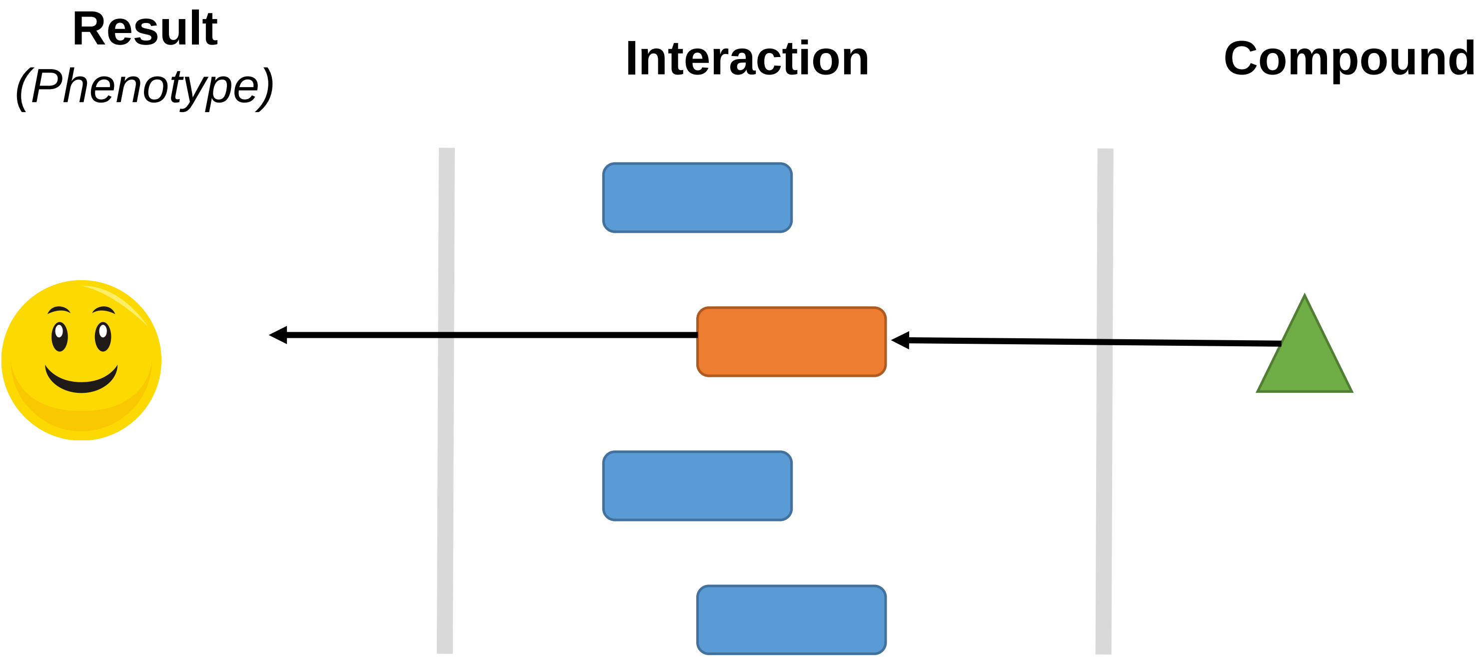
🟧 = biological target

▲ = chemical compound

╌╌╌╌ = potential target

—— = direct target

🟥╌╌ = inhibits target



Result
(Phenotype)

Interaction

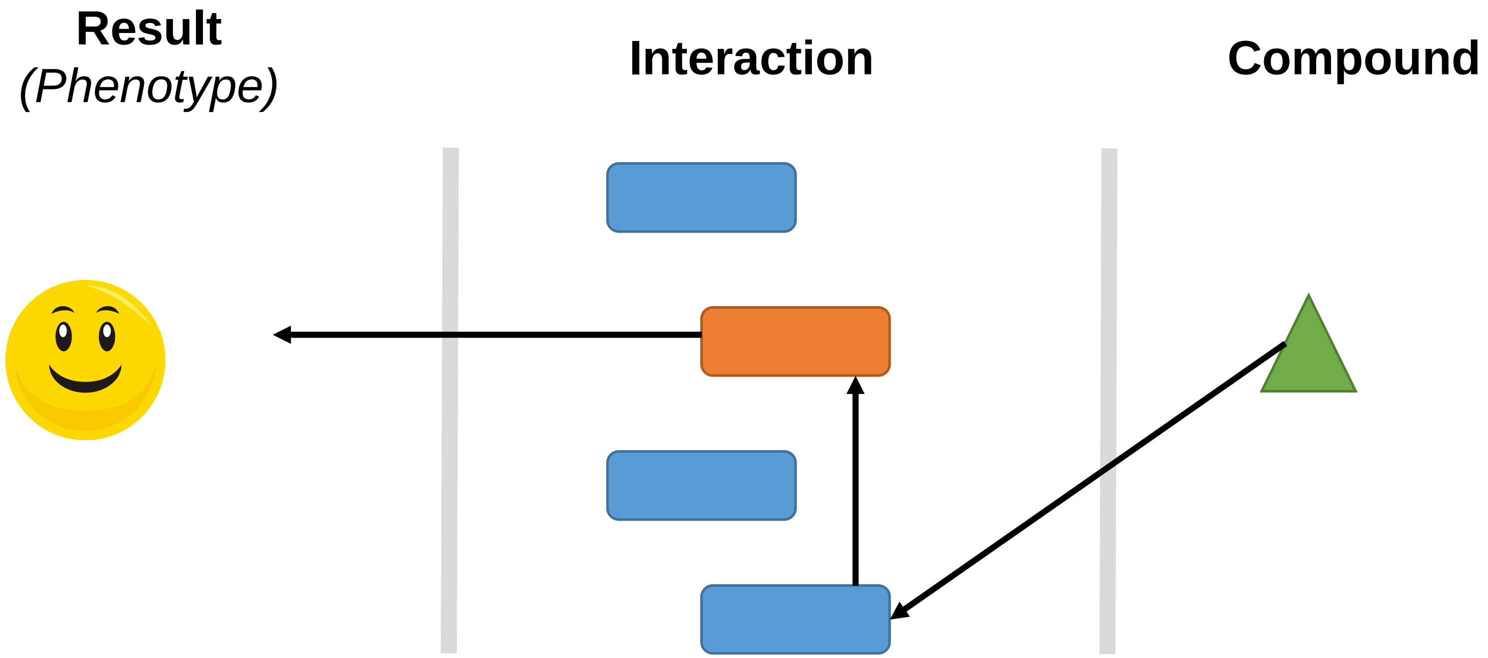
Compound

Domain Problem

UNDERSTANDING DRUG DISCOVERY

Scenario 2:
Indirect interaction, understood mechanism, desired outcome

🟦 = biological receptor
🟧 = biological target
▲ = chemical compound
---- = potential target
— = direct target
🟥 = inhibits target



Result
(Phenotype)

Interaction

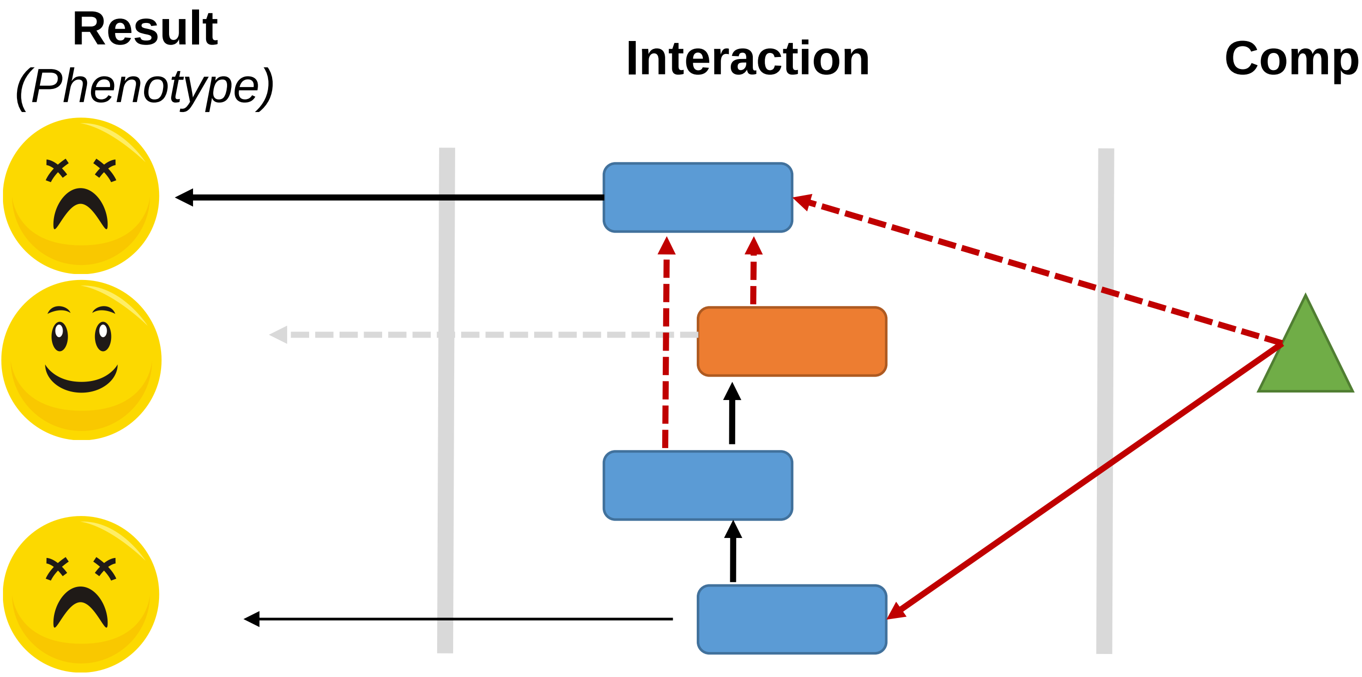
Compound

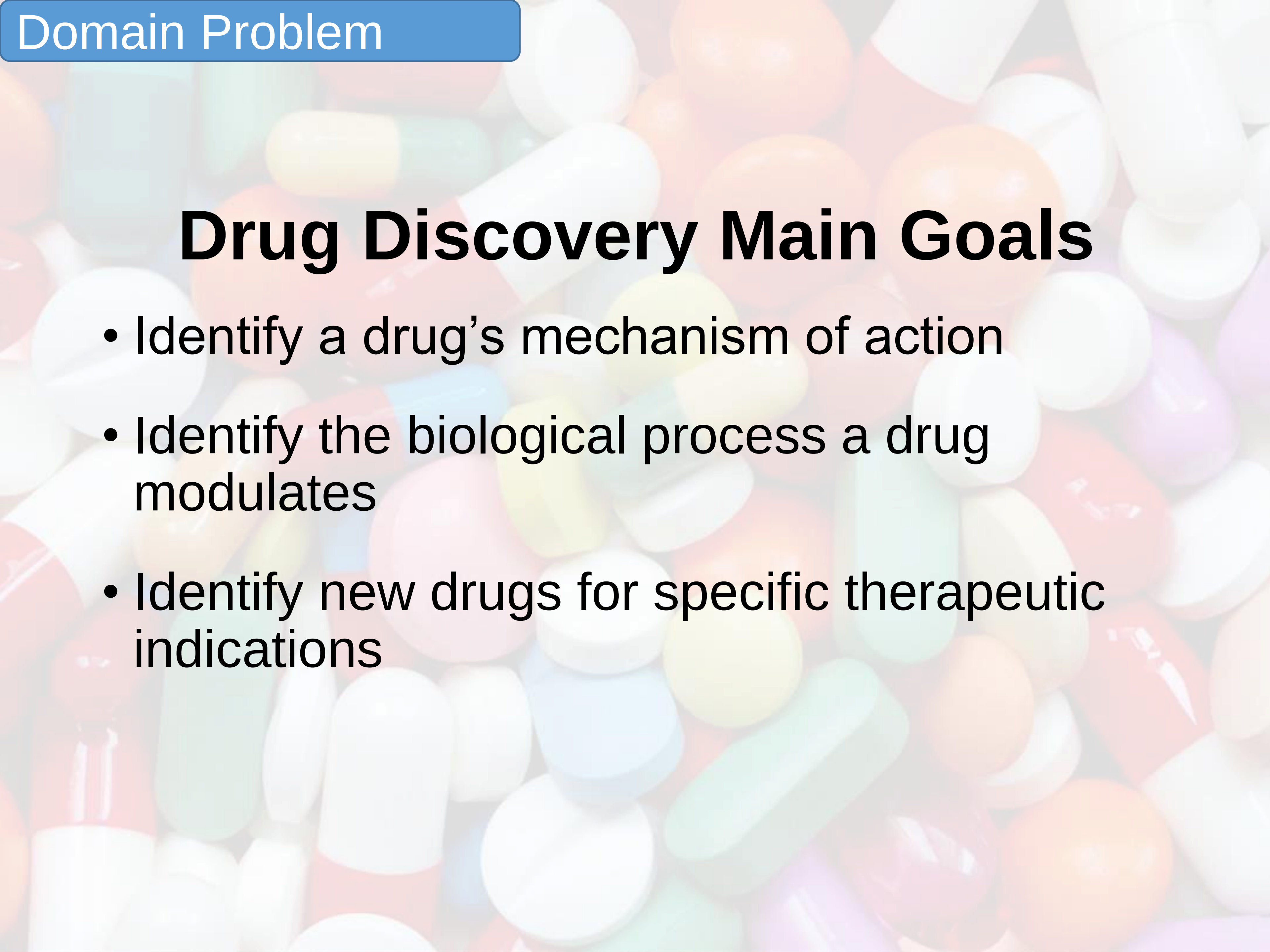
Domain Problem

UNDERSTANDING DRUG DISCOVERY

Scenario 3:
Complex interactions, mechanism
poorly understood, multiple outcomes

🟦 = biological receptor
🟧 = biological target
▲ = chemical compound
- - - - = potential target
—— = direct target
▭ = inhibits target

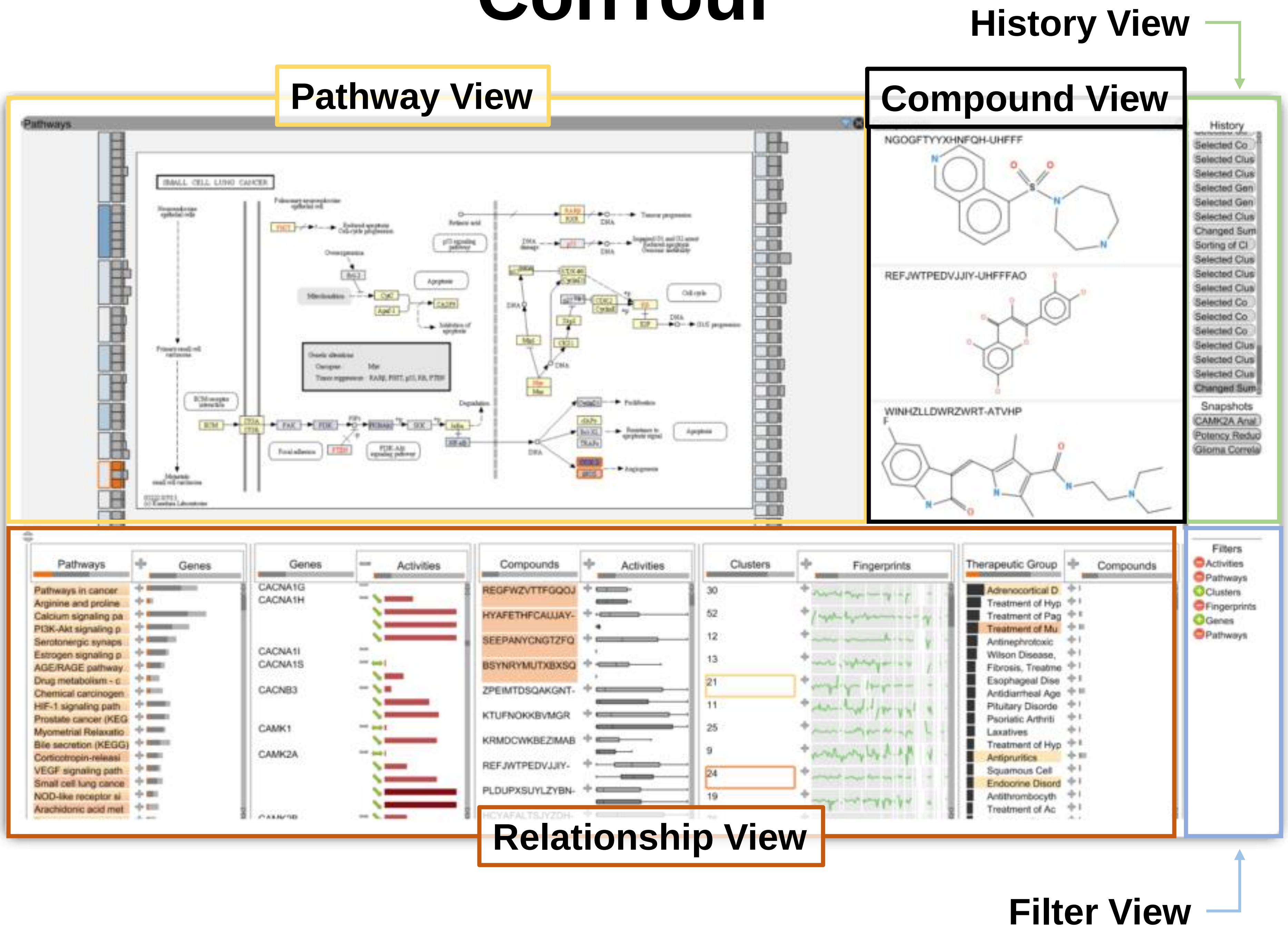




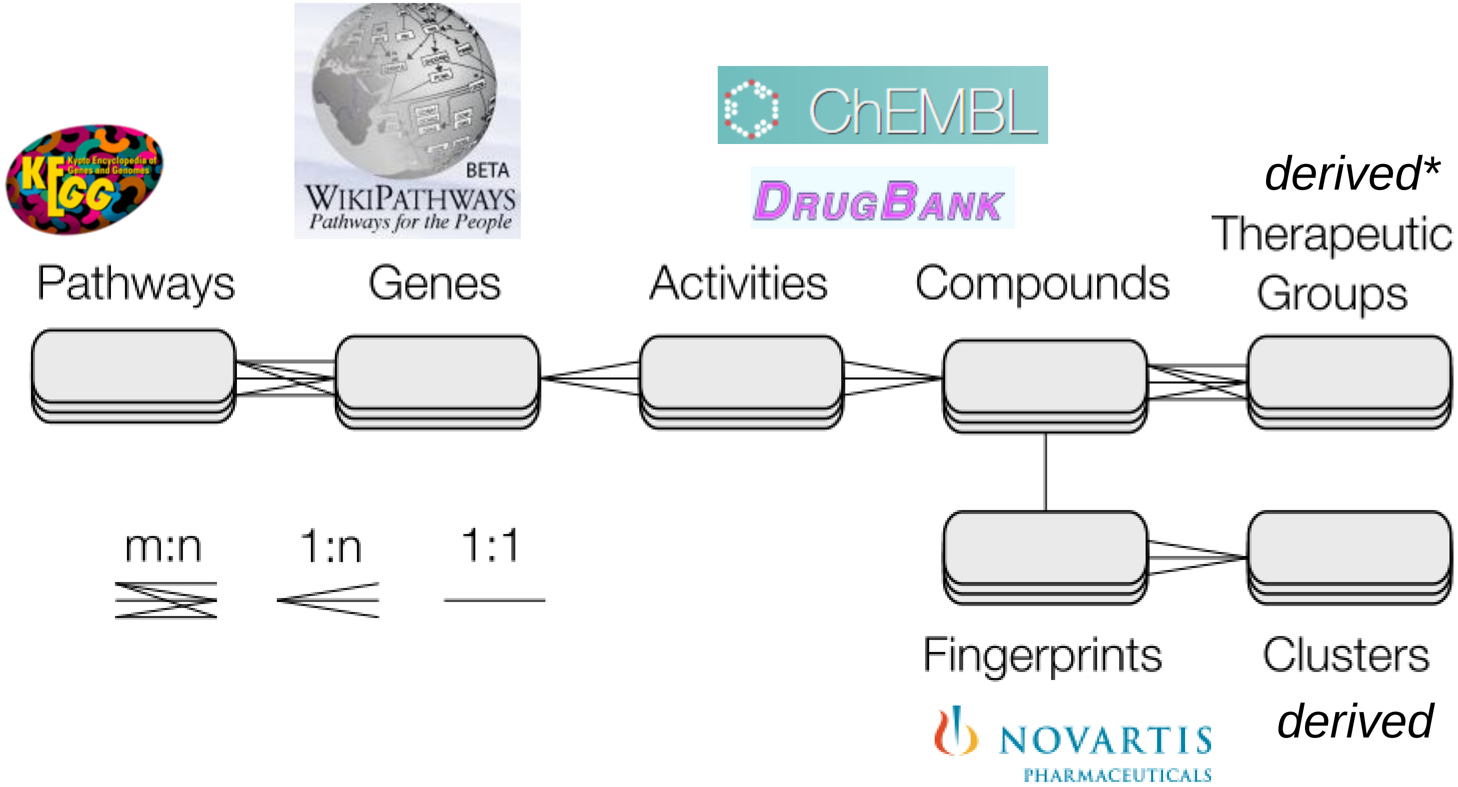
Drug Discovery Main Goals

- Identify a drug’s mechanism of action
- Identify the biological process a drug modulates
- Identify new drugs for specific therapeutic indications

ConTour

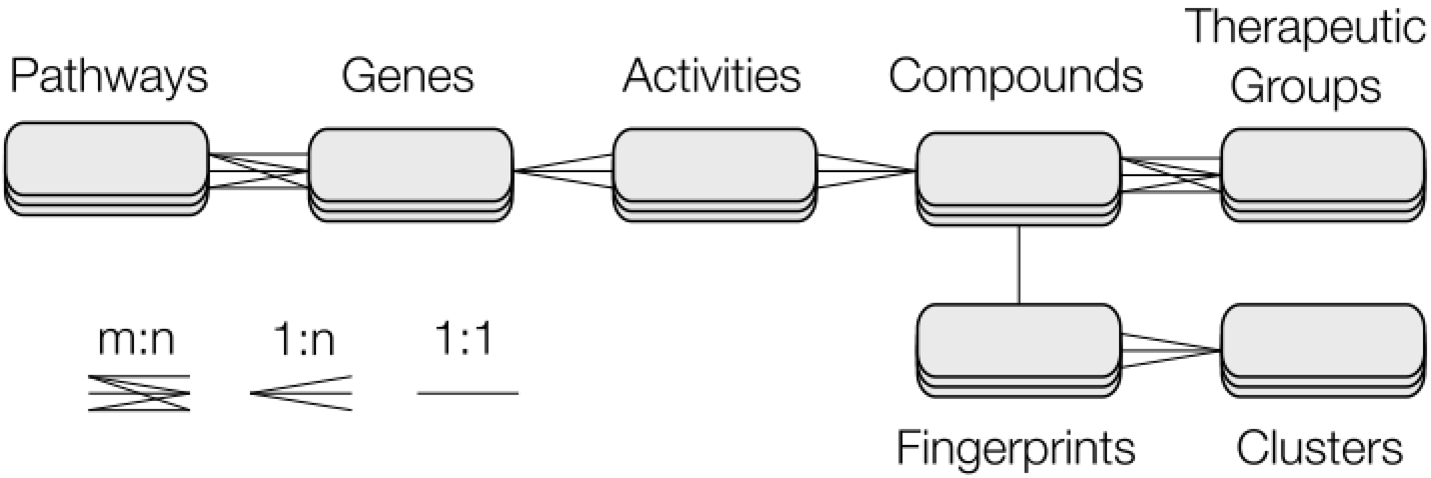

History View
Pathway View
Compound View
Pathways
SMALL CELL LUNG CANCER
NGOGFTYYXHNFQH-UHFFF
REFJWTPEDVJJIY-UHFFFAO
WINHZLLDWRZWRT-ATVHP
History
Snapshots
Pathways
Genes
Activities
Compounds
Clusters
Fingerprints
Therapeutic Group
Filters
Activities
Pathways
Clusters
Fingerprints
Genes
Pathways
Relationship View
Filter View

Data Abstraction



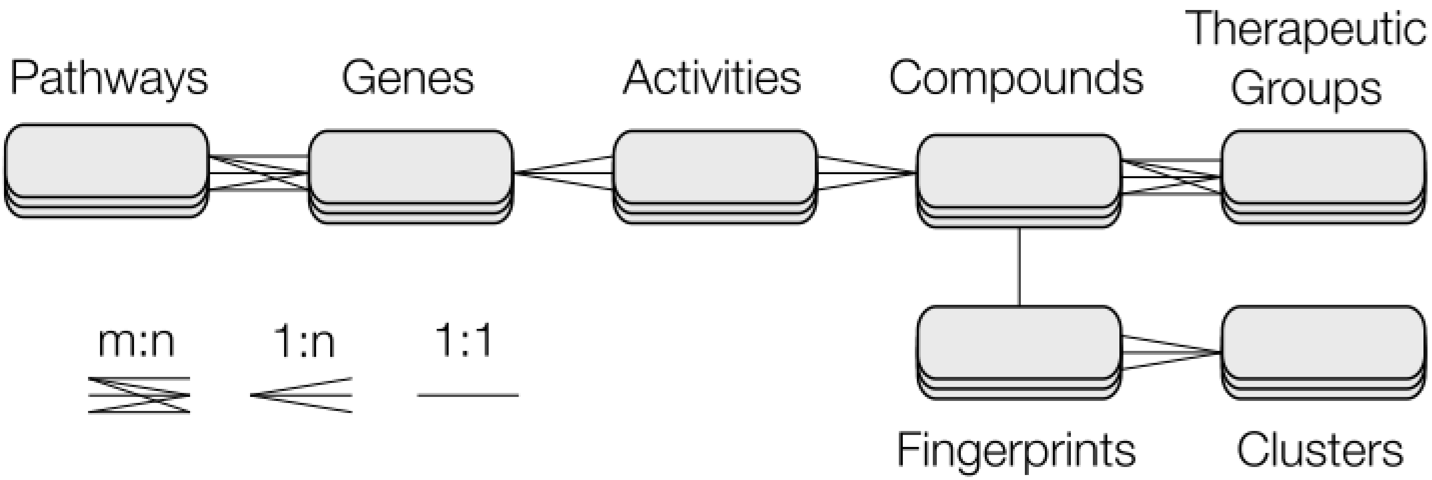
** Derived using a scheme propose by the Prous Integrity database*

Data Abstraction



“The drug discovery domain problem can be **generalized** to the problem of analysing ***multi-relational*** datasets […] Consequently, we argue that our approach is **applicable to many other problems**.”

Data Abstraction



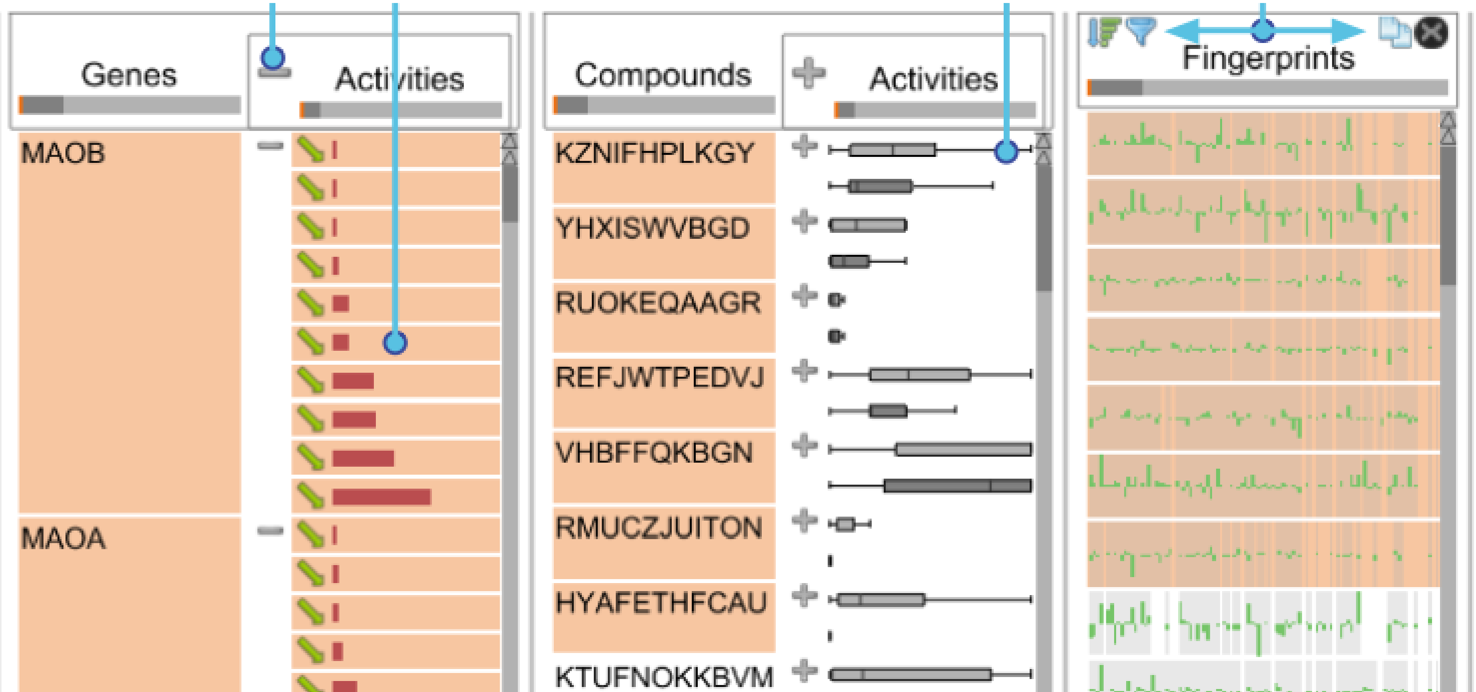
“The multi-relational data exploration problem can be interpreted as a **graph exploration problem** where each item of each dataset represents a node and the relationships between the items are the edges”

Task Analysis

- **T1**: Identify Related Items

Item selection and highlighting

Clicking, not hovering, on an item also moves all related items in columns to the top

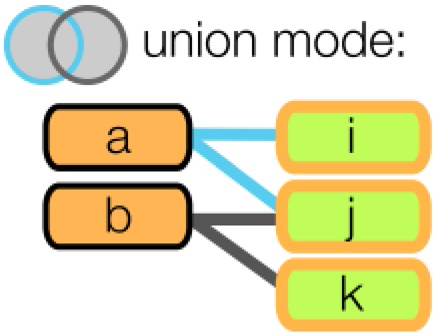


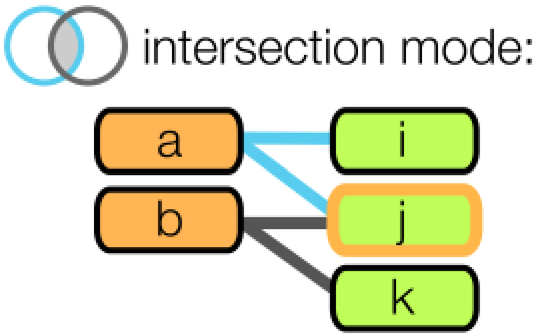
Task Analysis

- **T1**: Identify Related Items

Selection-based filters

Filter choices when multiple items are selected



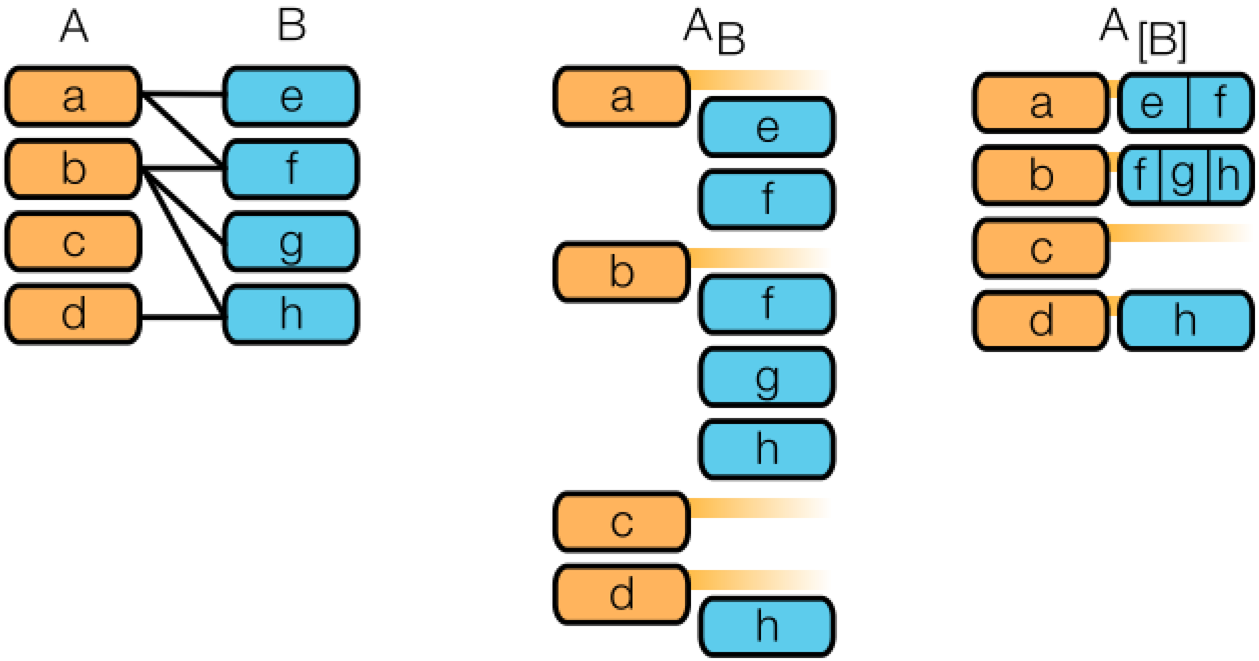


Task Analysis

- **T1**: Identify Related Items

Nesting

Simple Nesting

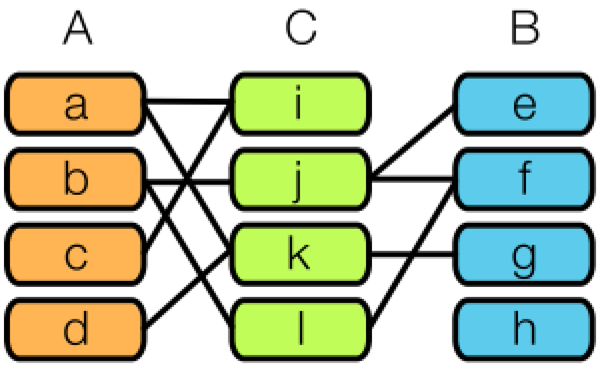


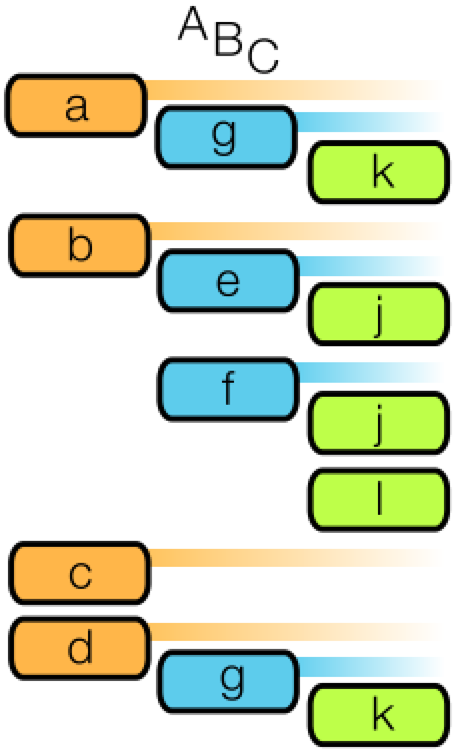
Task Analysis

- **T1**: Identify Related Items

Nesting

Recursive Nesting



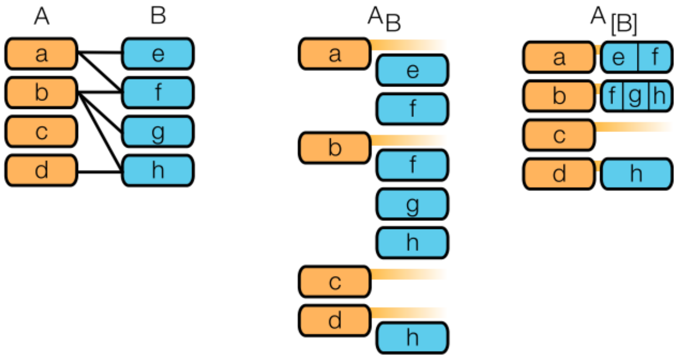


Task Analysis

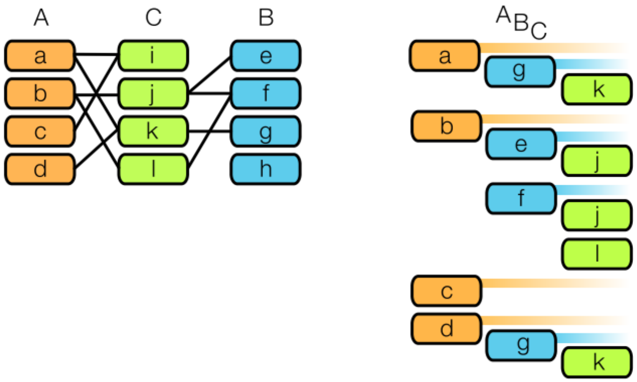
- **T2:** Identify Items that Share a Relationships with a Set of Items

Nesting

Simple Nesting



Recursive Nesting

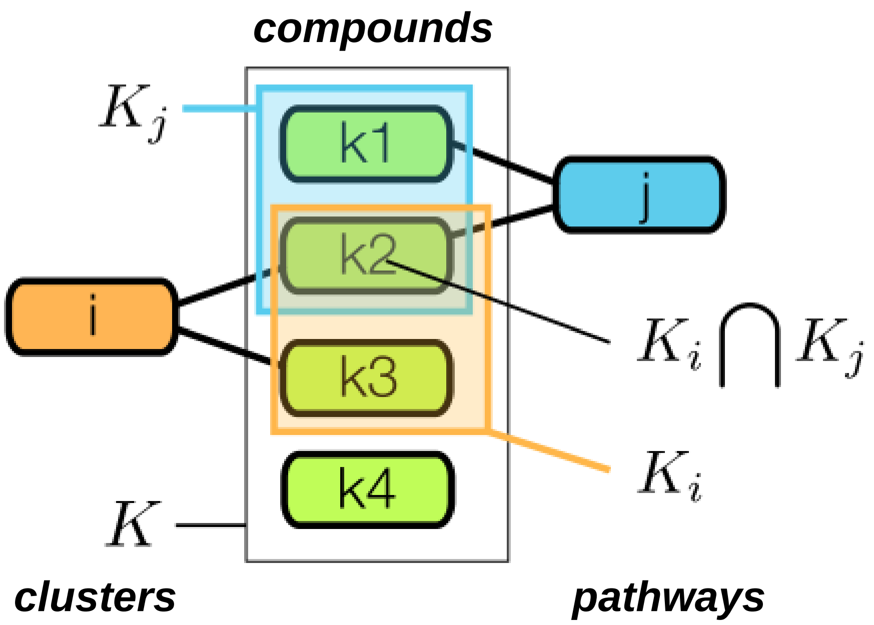


Task Analysis

- **T3:** Analyse Network Enrichment

Enrichment Score

Judging how specific two items are when compared to a third



$$s_{i,j}(K) = \frac{|K_i \bigcap K_j| / |K_j|}{|K_i| / |K|}$$

Where:

I = clusters
K = compounds
J = Pathways
S(i,j) = pair score

**I assume they take care of divide by 0?*

Task Analysis

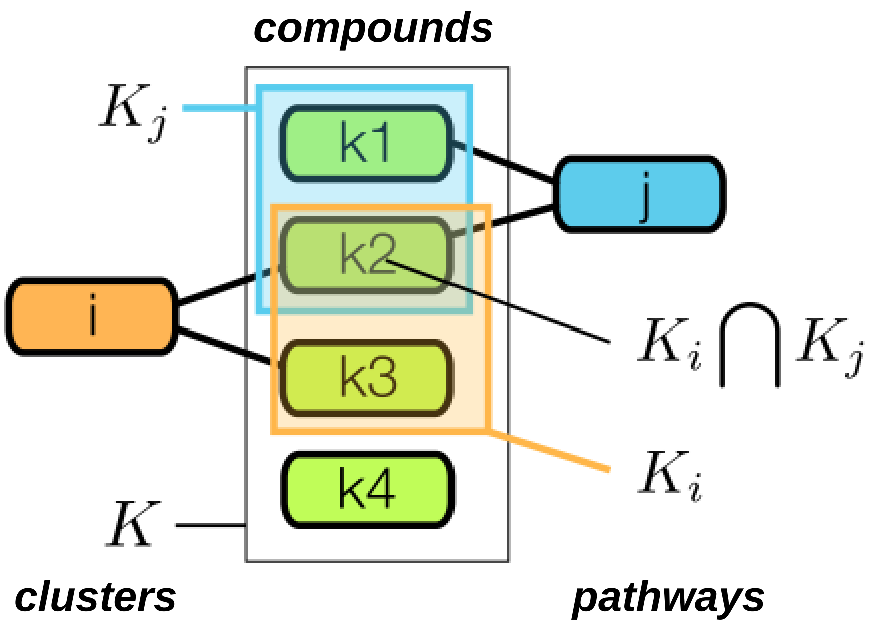
- **T4:** Rank Items

Sorting by interest

Sort alpha-numerically

Enrichment Score

Sort by enrichment score



Task Analysis

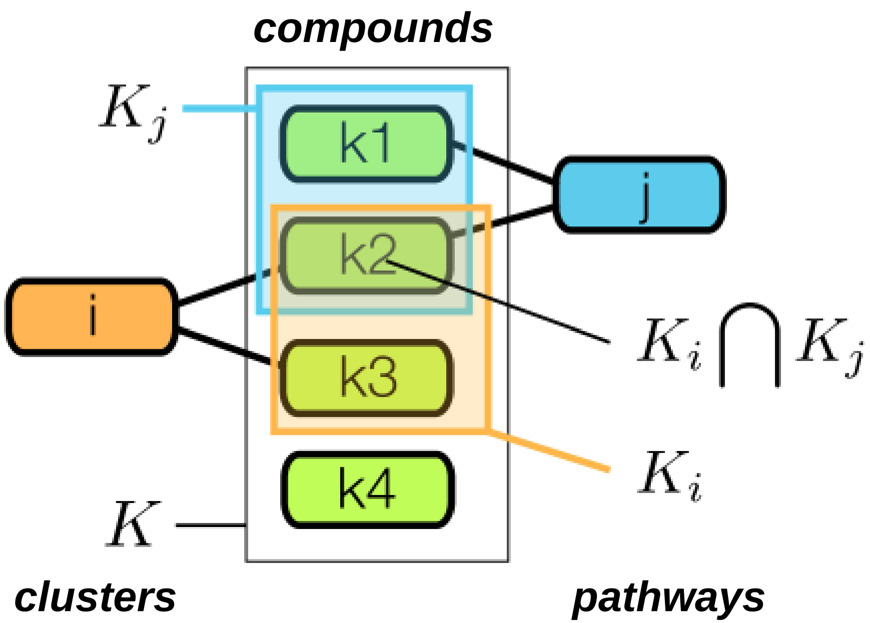
- **T4:** Rank Items

Sorting by interest

Sort alpha-numerically

Enrichment Score

Sort by enrichment score



Task Analysis

- **T5:** Filter Items

  Depends on tasks 1 and 2

Navigation

Local Filter : filter within a specific column

Global Filter: remove items that are not connected to the source column

Task Analysis

- **T5:** Filter Items

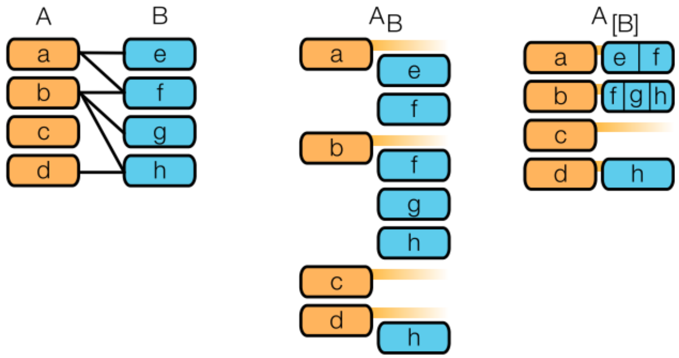
  Depends on tasks 1 and 2

Navigation

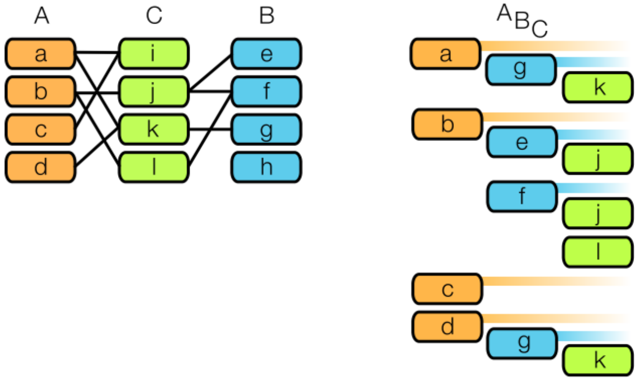
Selection-based filters

Nesting

Simple Nesting



Recursive Nesting

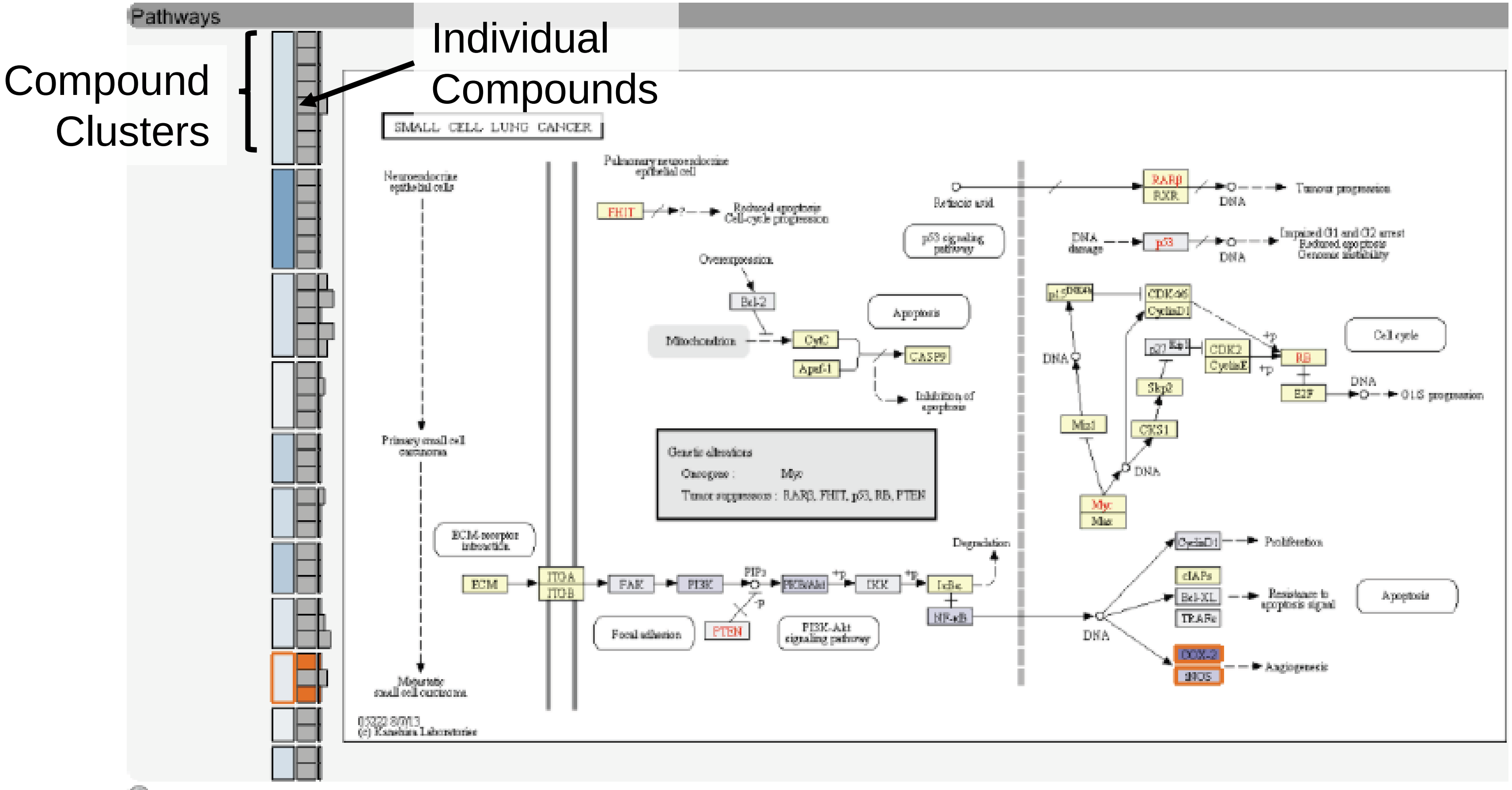


Task Analysis

- **T6**: View items in detail

Pathway View

Marks



Compound Clusters

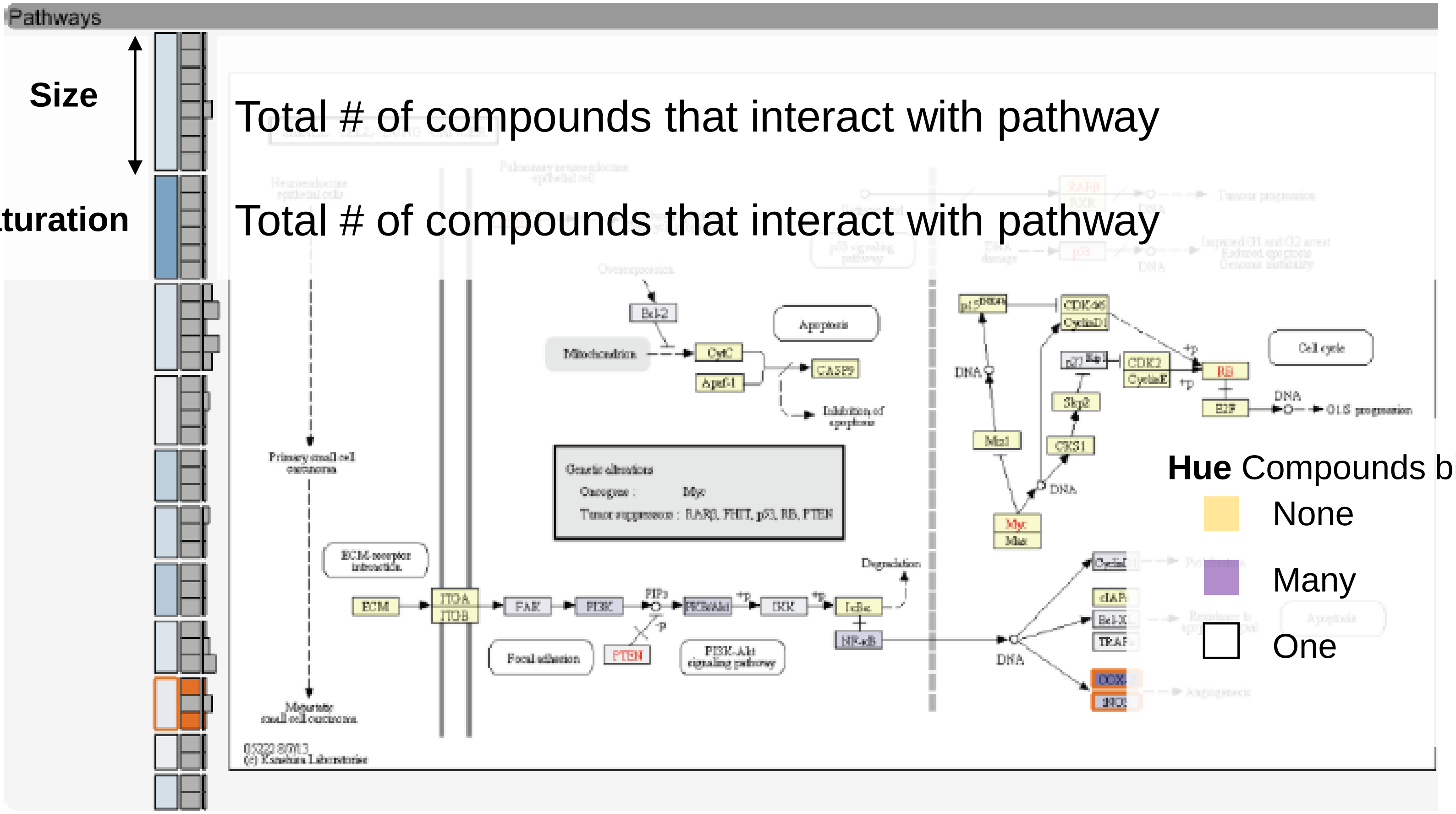
Individual Compounds

Task Analysis

- **T6**: View items in detail

Pathway View

Channels



Size

Total # of compounds that interact with pathway

Saturation

Total # of compounds that interact with pathway

Hue Compounds binding

🟨 None

🟪 Many

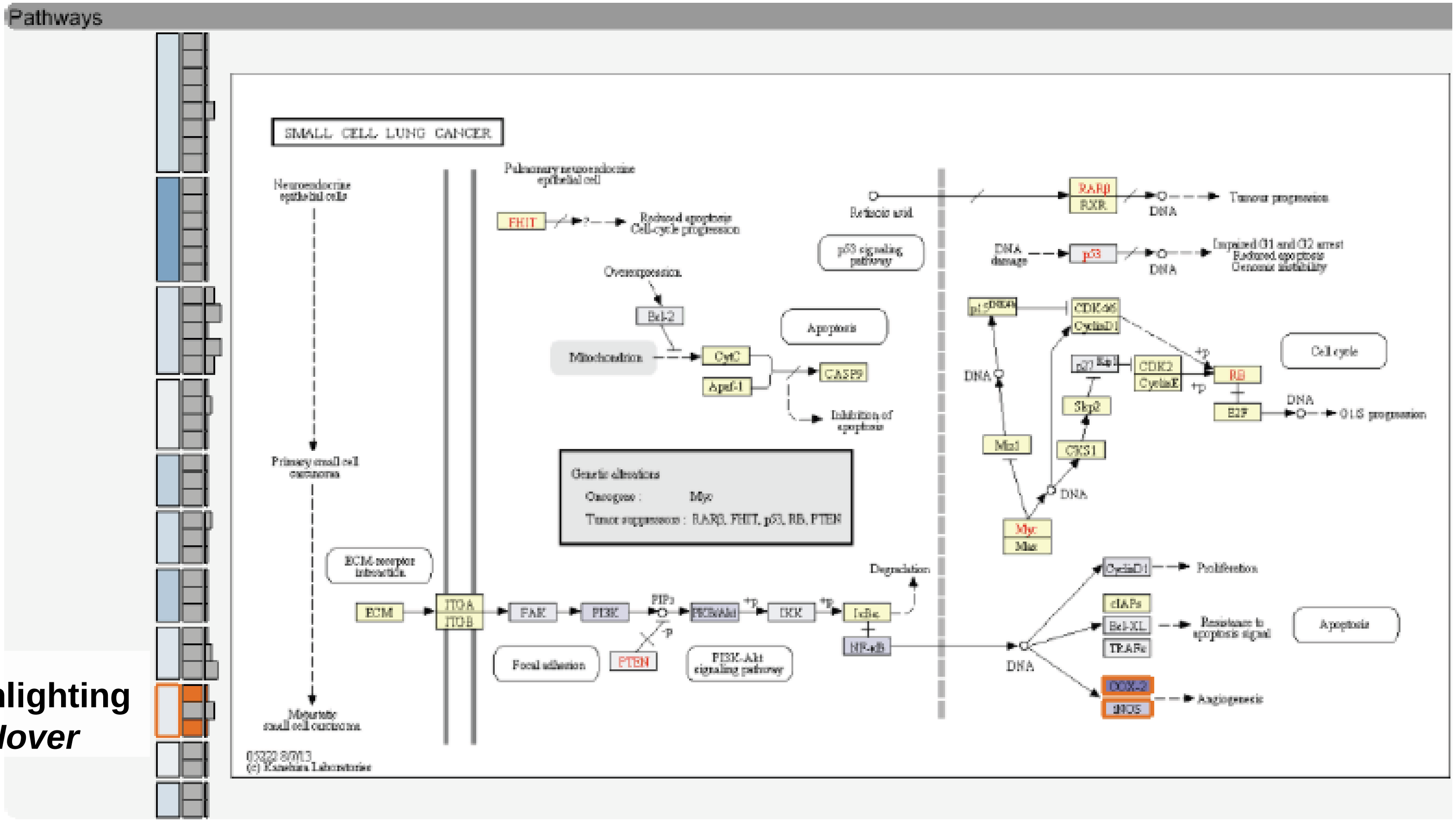
☐ One

Task Analysis

- **T6**: View items in detail

Pathway View

Linked Views



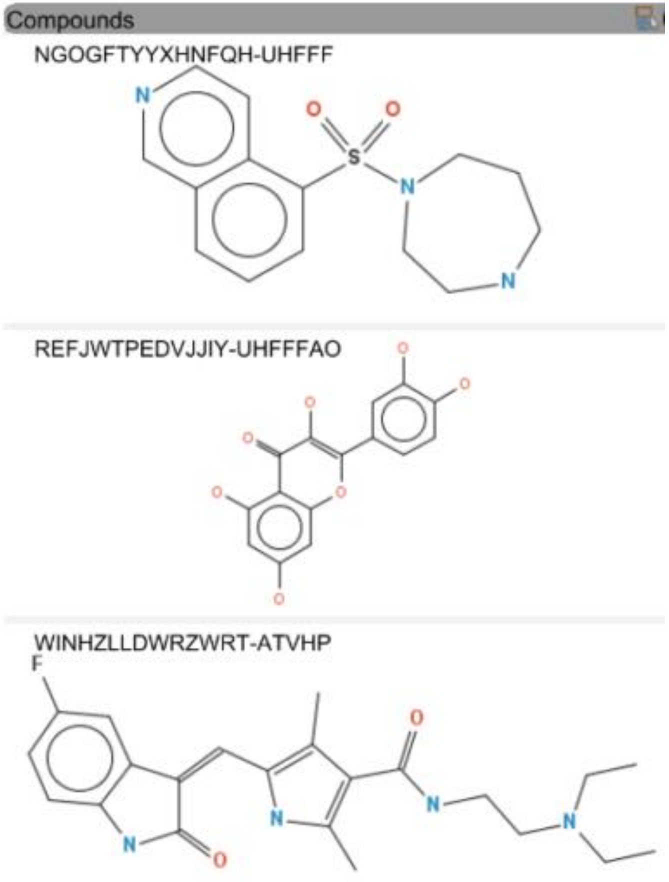
Highlighting

Hover

Task Analysis

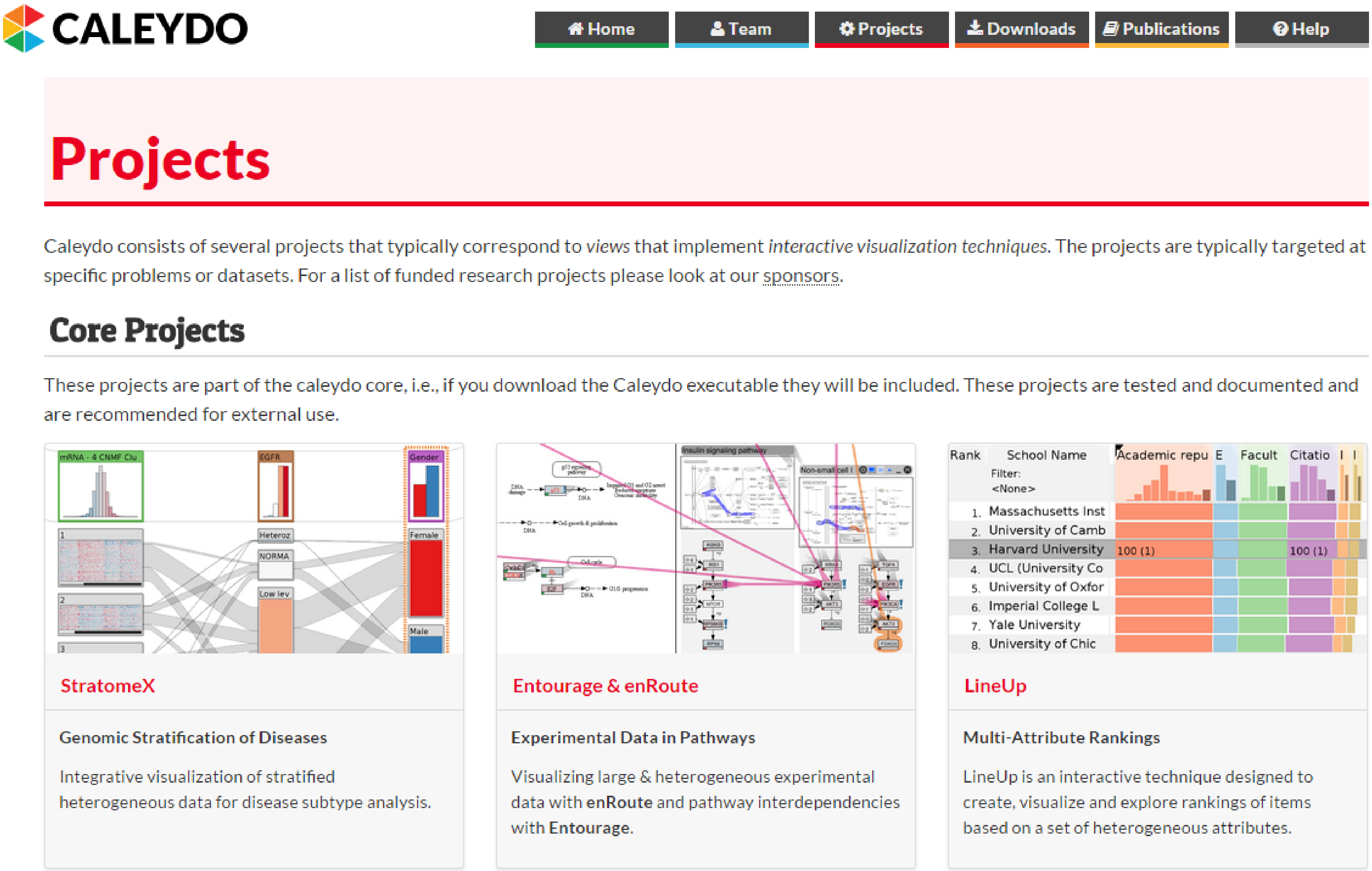
- **T6**: View items in detail

Compound View



Hue : Elements

Implementation Details



Projects

Caleydo consists of several projects that typically correspond to *views* that implement *interactive visualization techniques*. The projects are typically targeted at specific problems or datasets. For a list of funded research projects please look at our sponsors.

Core Projects

These projects are part of the caleydo core, i.e., if you download the Caleydo executable they will be included. These projects are tested and documented and are recommended for external use.

StratomeX

Genomic Stratification of Diseases

Integrative visualization of stratified heterogeneous data for disease subtype analysis.

Entourage & enRoute

Experimental Data in Pathways

Visualizing large & heterogeneous experimental data with **enRoute** and pathway interdependencies with **Entourage**.

Rank	School Name	Academic repu	E	Facult	Citatio	I	I
	Filter: <None>						
1.	Massachusetts Inst						
2.	University of Camb						
3.	Harvard University	100 (1)			100 (1)		
4.	UCL (University Co						
5.	University of Oxfor						
6.	Imperial College L						
7.	Yale University						
8.	University of Chic						

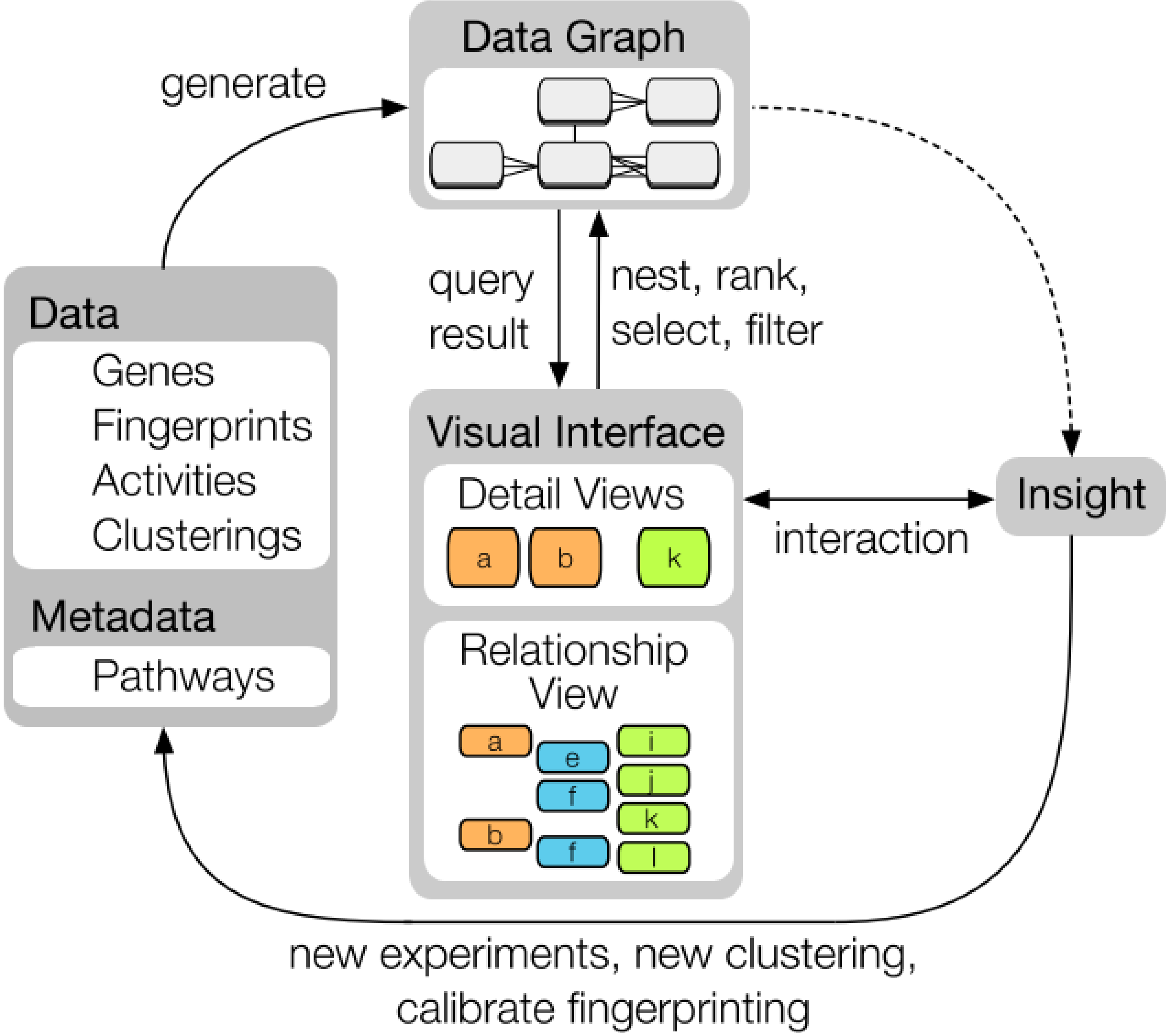
LineUp

Multi-Attribute Rankings

LineUp is an interactive technique designed to create, visualize and explore rankings of items based on a set of heterogeneous attributes.

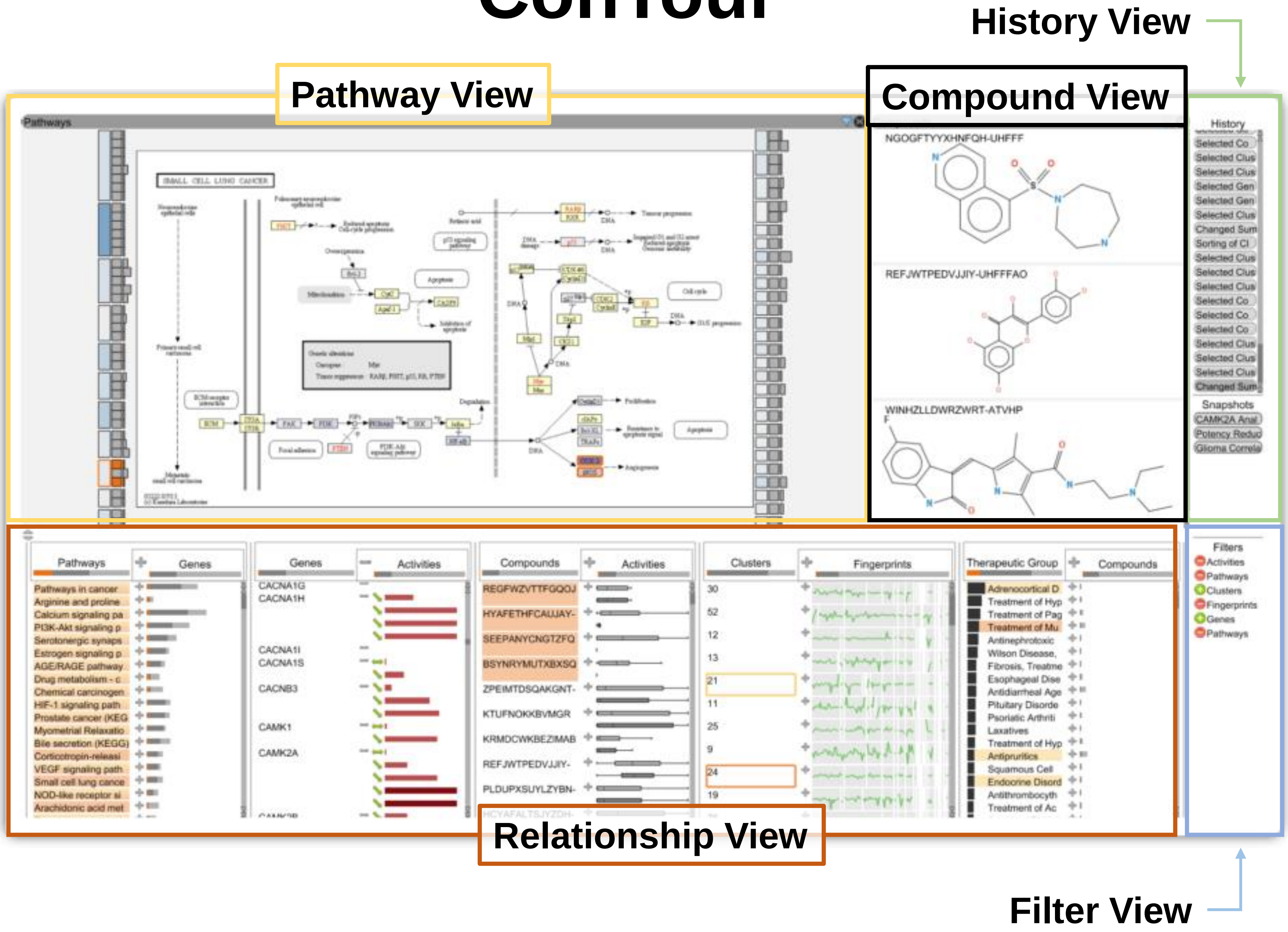
source code: https://github.com/Caleydo/

ConTour



Algorithm Design

ConTour



History:
Selected Co
Selected Clus
Selected Clus
Selected Gen
Selected Gen
Selected Clus
Changed Sum
Sorting of Cl
Selected Clus
Selected Clus
Selected Clus
Selected Co
Selected Co
Selected Co
Selected Clus
Selected Clus
Selected Clus
Changed Sum

Pathways	Genes
Pathways in cancer	
Arginine and proline	
Calcium signaling pa	
PI3K-Akt signaling p	
Serotonergic synaps	
Estrogen signaling p	
AGE/RAGE pathway	
Drug metabolism - c	
Chemical carcinogen	
HIF-1 signaling path	
Prostate cancer (KEG	
Myometrial Relaxatio	
Bile secretion (KEGG)	
Corticotropin-releasi	
VEGF signaling path	
Small cell lung cance	
NOD-like receptor si	
Arachidonic acid met	

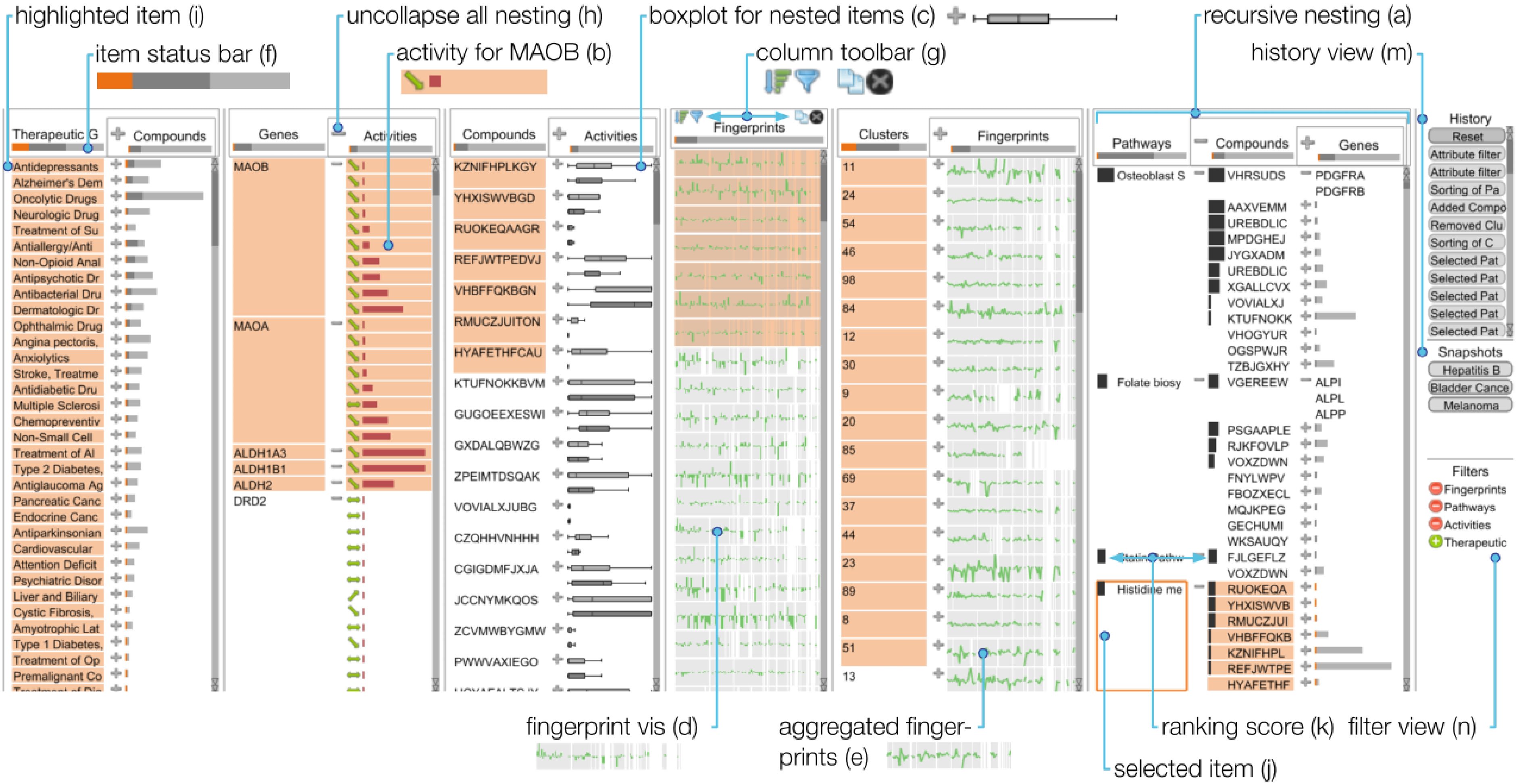
Genes	Activities
CACNA1G CACNA1H	
CACNA1I CACNA1S	
CACNB3	
CAMK1	
CAMK2A	

Compounds	Activities
REGFWZVTTFGQOJ	
HYAFETHFCALJAY-	
SEEPANYCNGTZFQ	
BSYNRYMUTXBXSQ	
ZPEIMTDSQAKGNT-	
KTUFNOKKBVMGR	
KRMDCWKBEZIMAB	
REFJWTPEDVJJIY-	
PLDUPXSUYLZYBN-	
HCYAFALTSJYZDH-	

Clusters	Fingerprints
30	
52	
12	
13	
21	
11	
25	
9	
24	
19	

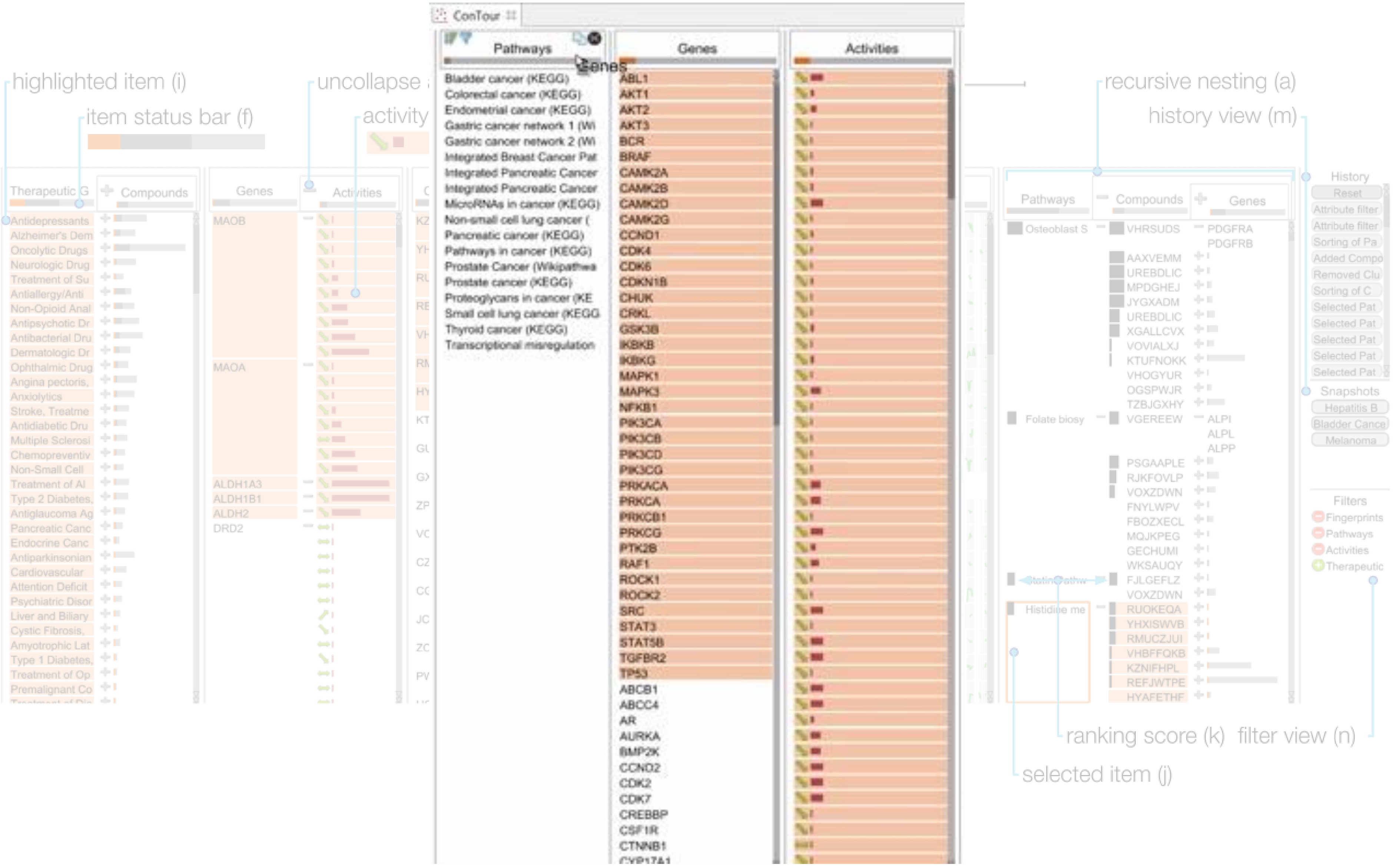
Therapeutic Group	Compounds
Adrenocortical D	
Treatment of Hyp	
Treatment of Pag	
Treatment of Mu	
Antinephrotoxic	
Wilson Disease,	
Fibrosis, Treatme	
Esophageal Dise	
Antidiarrheal Age	
Pituitary Disorde	
Psoriatic Arthriti	
Laxatives	
Treatment of Hyp	
Antipruritics	
Squamous Cell	
Endocrine Disord	
Antithrombocyth	
Treatment of Ac	

Relationship View



- There are facets within the relationship view
- Combination of tabular data and plots

Relationship View



Bladder cancer (KEGG)
Colorectal cancer (KEGG)
Endometrial cancer (KEGG)
Gastric cancer network 1 (Wi
Gastric cancer network 2 (Wi
Integrated Breast Cancer Pat
Integrated Pancreatic Cancer
Integrated Pancreatic Cancer
MicroRNAs in cancer (KEGG)
Non-small cell lung cancer (
Pancreatic cancer (KEGG)
Pathways in cancer (KEGG)
Prostate Cancer (Wikipathwa
Prostate cancer (KEGG)
Proteoglycans in cancer (KE
Small cell lung cancer (KEGG
Thyroid cancer (KEGG)
Transcriptional misregulation

ABL1
AKT1
AKT2
AKT3
BCR
BRAF
CAMK2A
CAMK2B
CAMK2D
CAMK2G
CCND1
CDK4
CDK6
CDKN1B
CHUK
CRKL
GSK3B
IKBKB
IKBKG
MAPK1
MAPK3
NFKB1
PIK3CA
PIK3CB
PIK3CD
PIK3CG
PRKACA
PRKCA
PRKCB1
PRKCG
PTK2B
RAF1
ROCK1
ROCK2
SRC
STAT3
STAT5B
TGFBR2
TP53
ABCB1
ABCC4
AR
AURKA
BMP2K
CCND2
CDK2
CDK7
CREBBP
CSF1R
CTNNB1

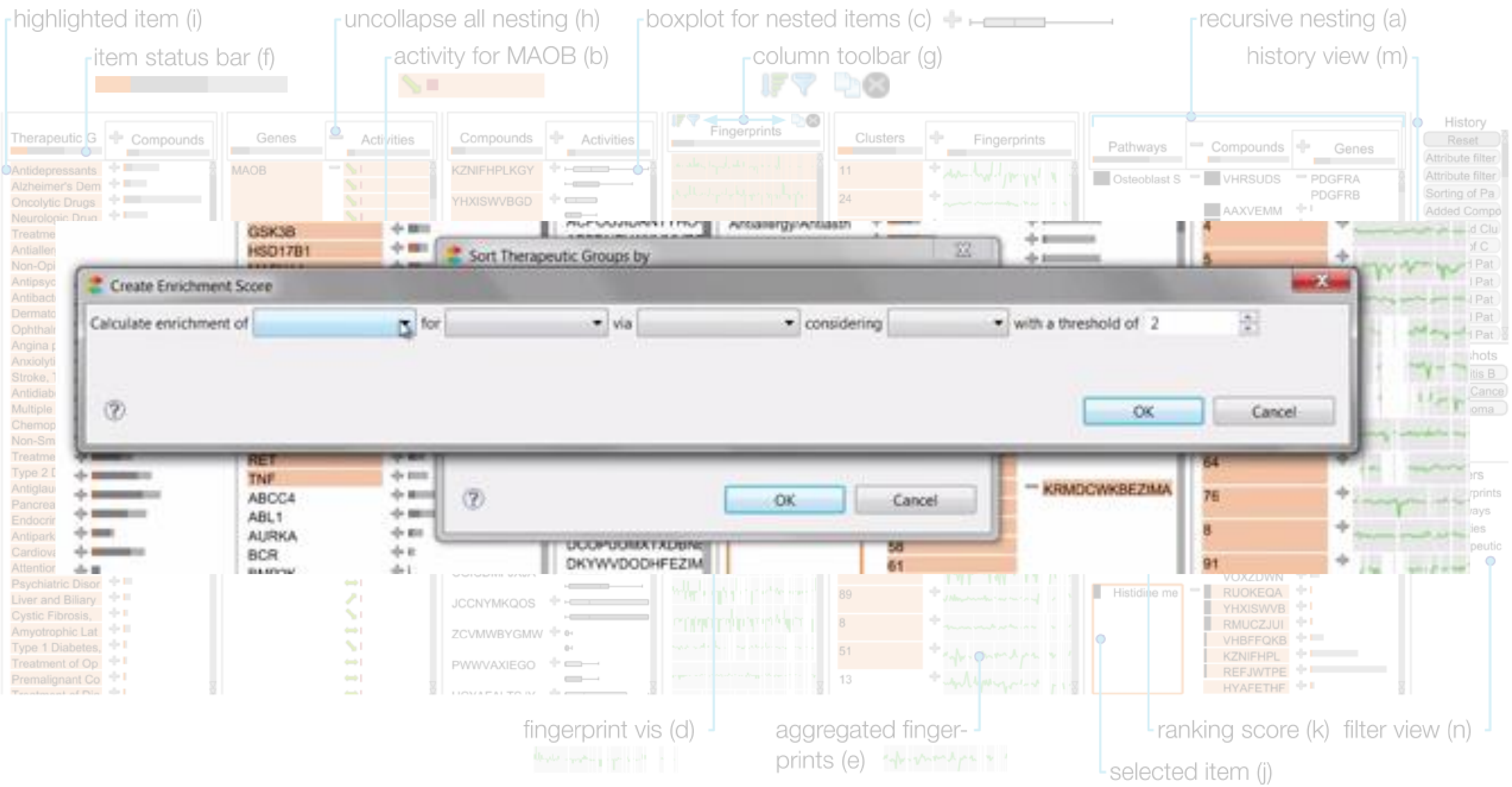
Relationship View

Pathways | Genes

Bladder cancer (KEGG)
Colorectal cancer (KE
Endometrial cancer (KE
Gastric cancer network 1
Gastric cancer network 2
Integrated Breast Cance
Integrated Pancreatic C
Integrated Pancreatic C
MicroRNAs in cancer (K
Non-small cell lung canc
Pancreatic cancer (KEG
Pathways in cancer (KE
Prostate Cancer (Wikipa
Prostate cancer (KEGG)
Proteoglycans in cancer
Small cell lung cancer (
Thyroid cancer (KEGG)
Transcriptional misregul

Bladder cancer (KEGG)

Relationship View



Sort Therapeutic Groups by

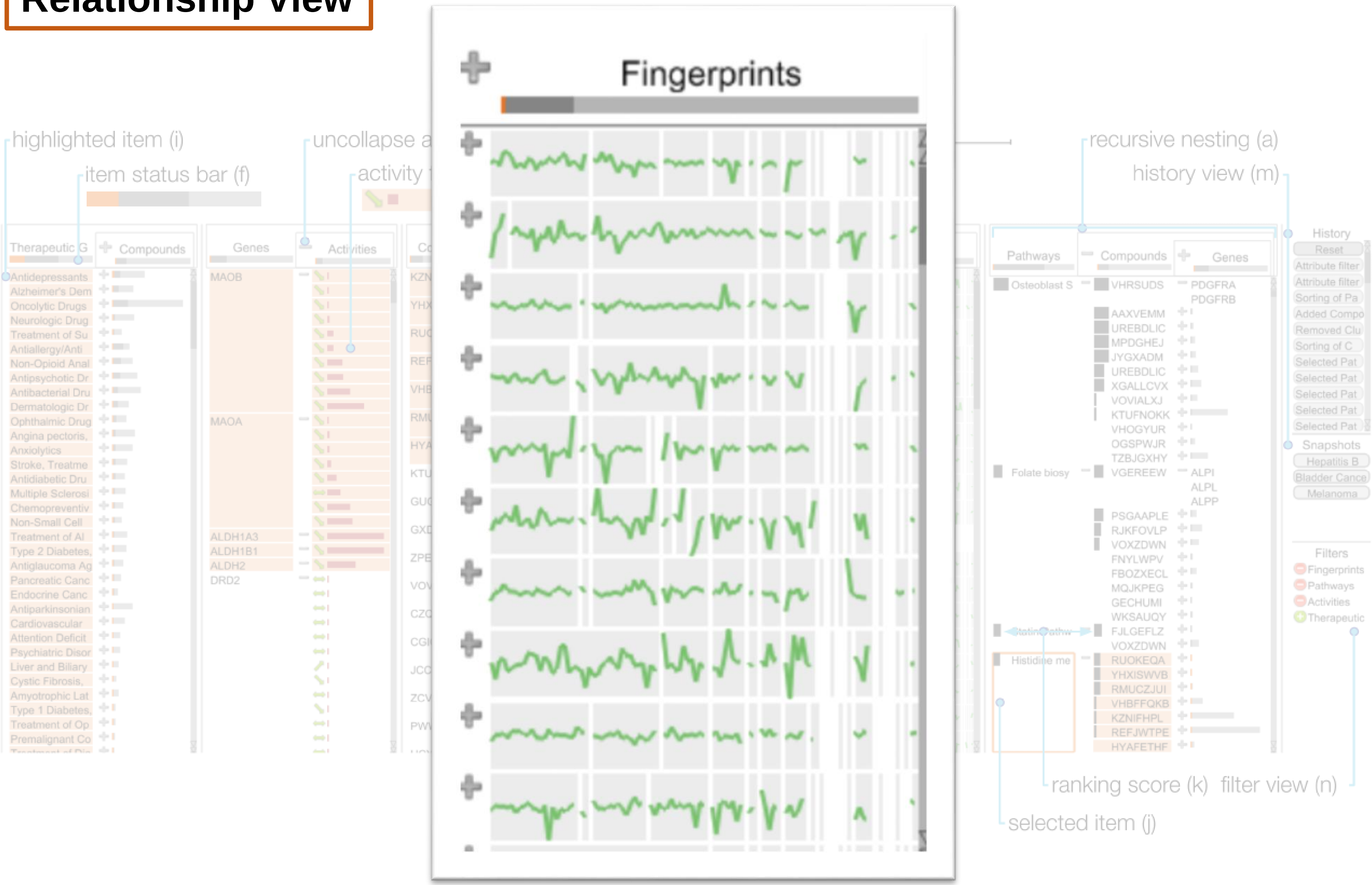
OK Cancel

Create Enrichment Score

Calculate enrichment of ▾ for ▾ via ▾ considering ▾ with a threshold of 2

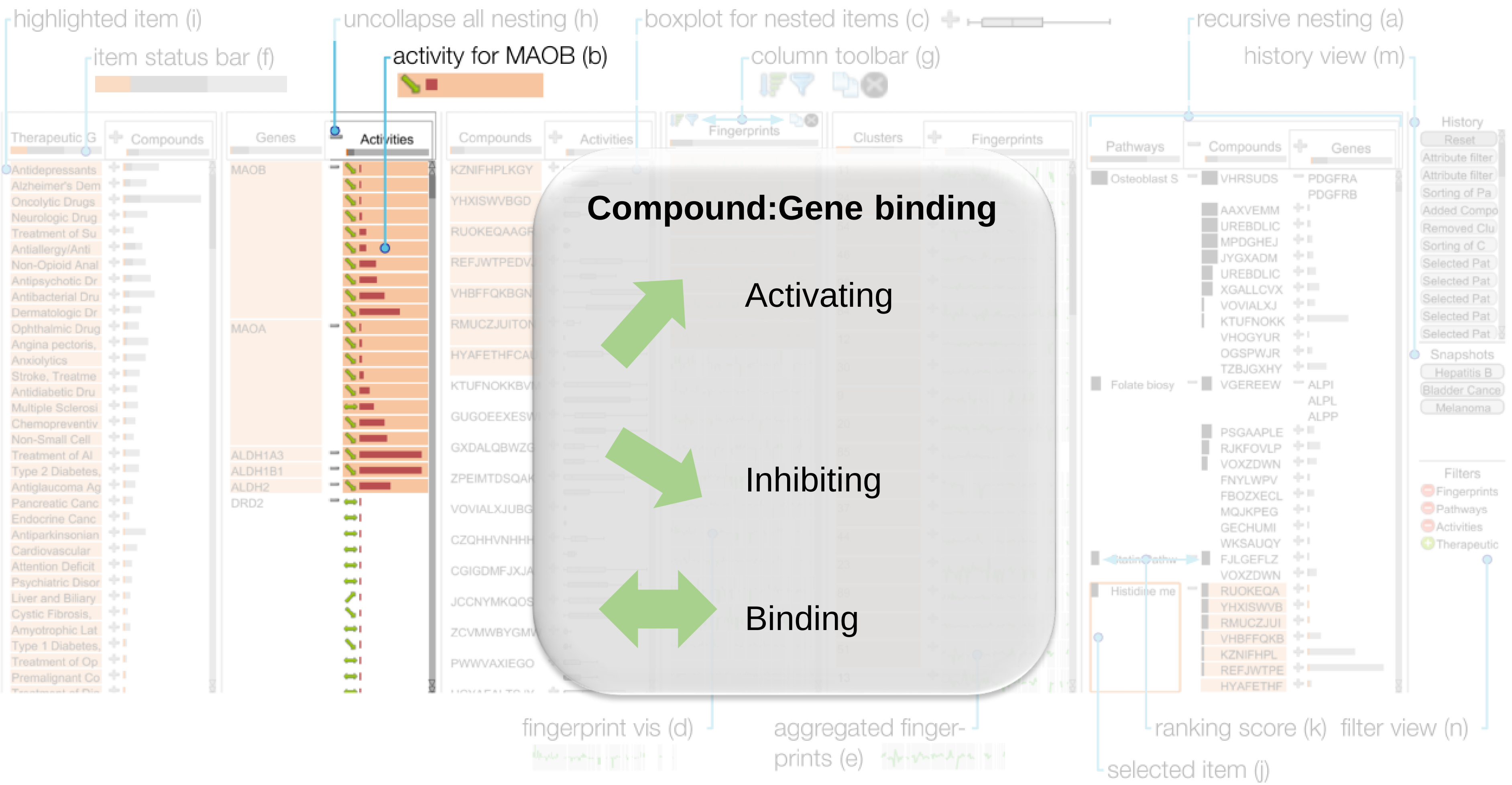
OK Cancel

Relationship View



Approximately 100 numerical values shown here

Relationship View



Compound:Gene binding

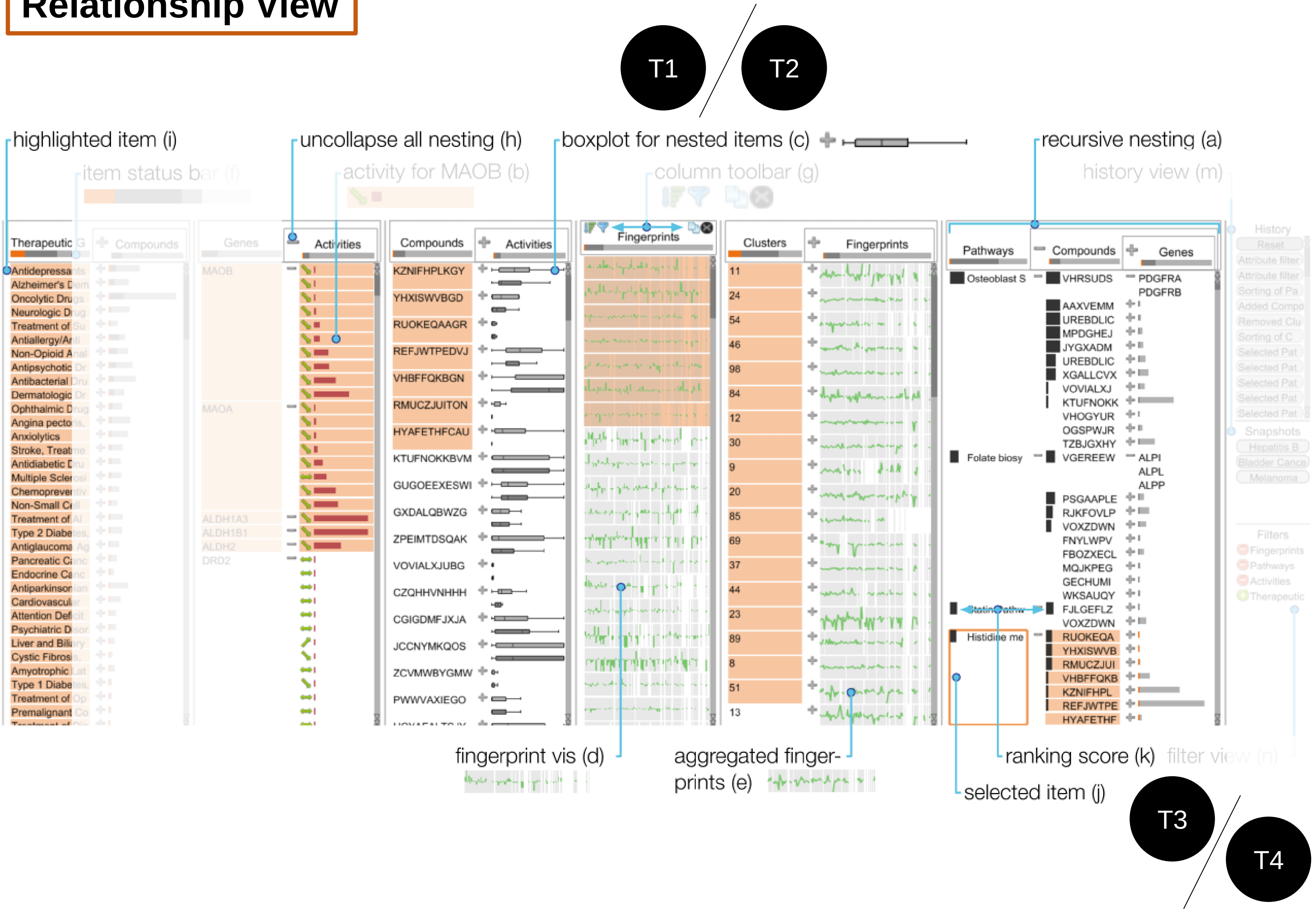
Activating

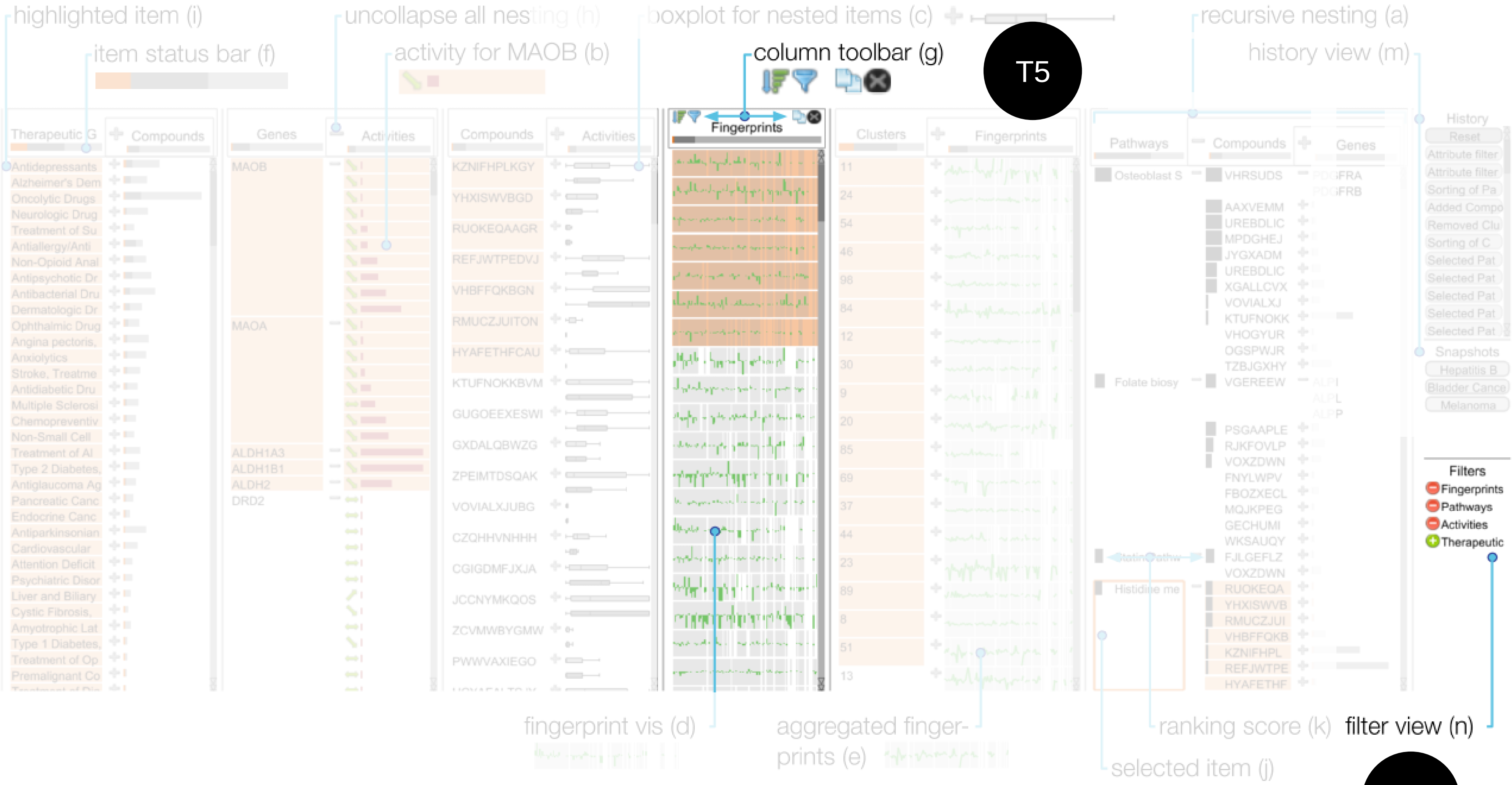
Inhibiting

Binding

Approximately 100 numerical values shown here

Relationship View




highlighted item (i)
item status bar (f)
uncollapse all nesting (h)
activity for MAOB (b)
boxplot for nested items (c)
column toolbar (g)
recursive nesting (a)
history view (m)
Therapeutic G
Compounds
Genes
Activities
Compounds
Activities
Fingerprints
Clusters
Fingerprints
Pathways
Compounds
Genes
History
Reset
Attribute filter
Attribute filter
Sorting of Pa
Added Compo
Removed Clu
Sorting of C
Selected Pat
Selected Pat
Selected Pat
Selected Pat
Selected Pat
Snapshots
Hepatitis B
Bladder Cance
Melanoma
Filters
Fingerprints
Pathways
Activities
Therapeutic
fingerprint vis (d)
aggregated finger-prints (e)
ranking score (k)
filter view (n)
selected item (j)

T5

T5

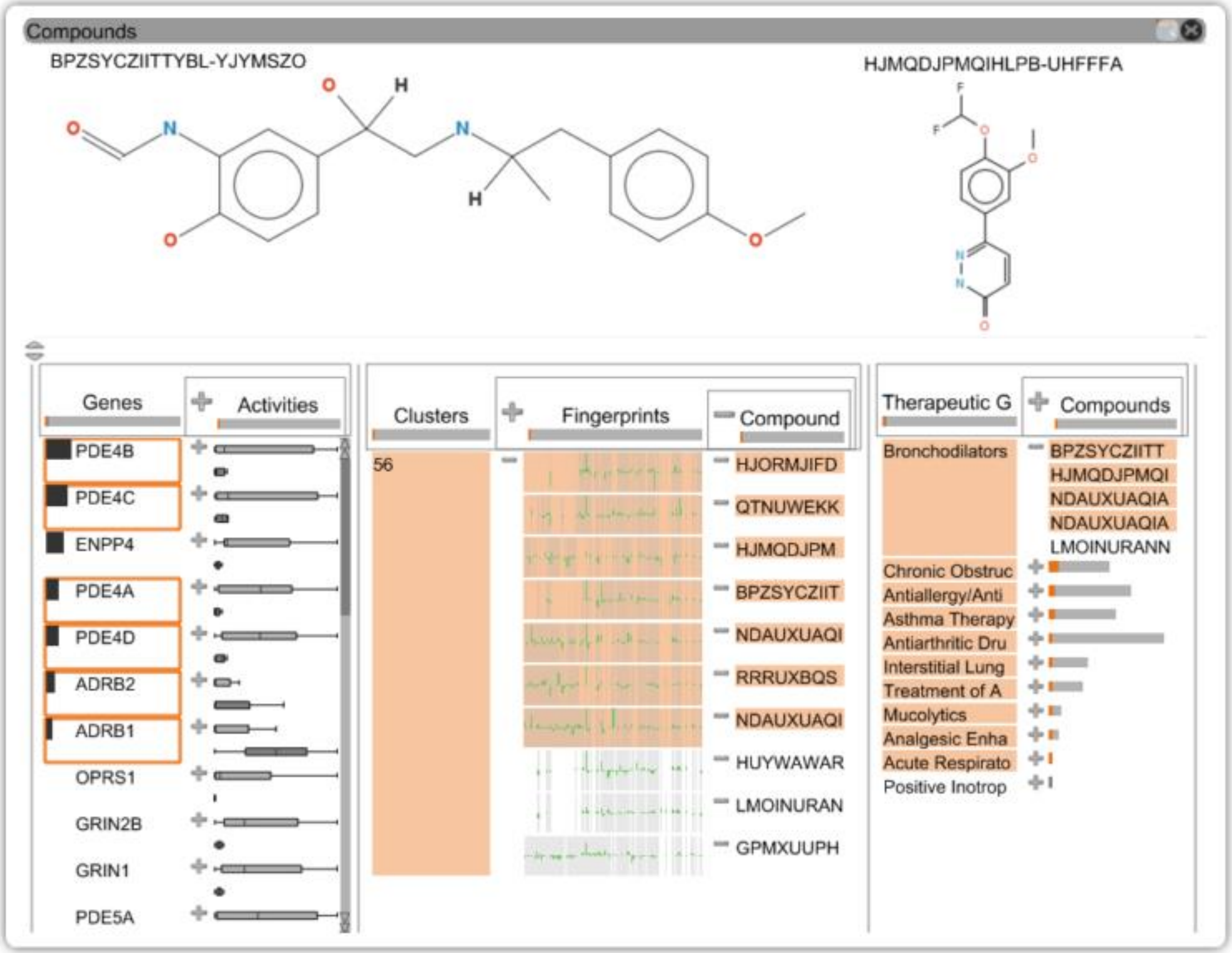
Conclusions

System	***ConTour***
What : Data	Multi-relational databases; node-link graph; clusters *(derived)*
Why : Tasks	Discovery; drill down; highlight relationships
How: Multiple Views	Relationship view; pathway view; compound view; history and filters
How: Facet	Side-by-side linked views, containing tabular data, bar plots, glyphs
How: Selection & Highlighting	Linked highlighting across facets; automatic sorting
How: Filtering	Drag and drop (nesting); user control (navigation)
How: Ranking & sorting	Enrichment score; highlight; user control (navigation)
How: Encode	Simple marks with manipulation of hue and saturation (pathway view)
Scale:	Dozens of columns; upper limit on HD display appears to about 20. Thousands of data items. Up to 8 simultaneous views for compounds; only 1 for pathways

Concluding Thoughts

- Seems like a very good tool for use on structured datasets
- When there are indirect (inferred) relationships, it would be good to highlight this with some uncertainty
- What about incomplete relationships?

Backup



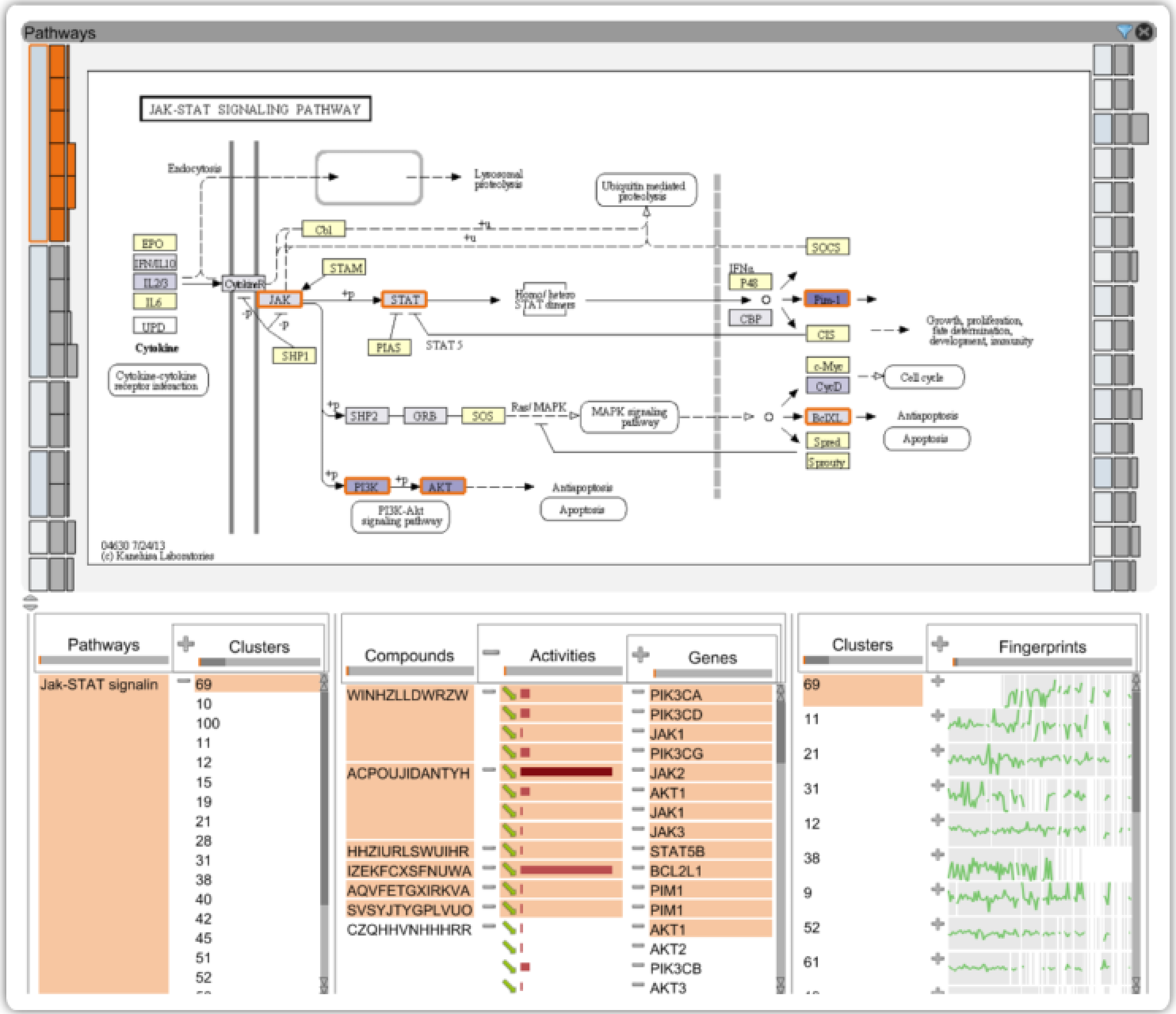
BPZSYCZIITTYBL-YJYMSZO

HJMQDJPMQIHLPB-UHFFFA

Genes	Activities
PDE4B	
PDE4C	
ENPP4	
PDE4A	
PDE4D	
ADRB2	
ADRB1	
OPRS1	
GRIN2B	
GRIN1	
PDE5A	

Clusters	Fingerprints	Compound
56		HJORMJIFD
		QTNUWEKK
		HJMQDJPM
		BPZSYCZIIT
		NDAUXUAQI
		RRRUXBQS
		NDAUXUAQI
		HUYWAWAR
		LMOINURAN
		GPMXUUPH

Therapeutic G	Compounds
Bronchodilators	BPZSYCZIITT HJMQDJPMQI NDAUXUAQIA NDAUXUAQIA LMOINURANN
Chronic Obstruc	
Antiallergy/Anti	
Asthma Therapy	
Antiarthritic Dru	
Interstitial Lung	
Treatment of A	
Mucolytics	
Analgesic Enha	
Acute Respirato	
Positive Inotrop	



Pathways	Clusters
Jak-STAT signalin	69
	10
	100
	11
	12
	15
	19
	21
	28
	31
	38
	40
	42
	45
	51
	52

Compounds	Genes
WINHZLLDWRZW	PIK3CA
	PIK3CD
	JAK1
	PIK3CG
ACPOUJIDANTYH	JAK2
	AKT1
	JAK1
	JAK3
HHZIURLSWUIHR	STAT5B
IZEKFCXSFNUWA	BCL2L1
AQVFETGXIRKVA	PIM1
SVSYJTYGPLVUO	PIM1
CZQHHVNHHHRR	AKT1
	AKT2
	PIK3CB
	AKT3

Clusters
69
11
21
31
12
38
9
52
61