

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

Christian Partl, Alexander Lex, Marc Streit, Hendrik Strobel, Anne-Mai Wassermann, Hanspeter Pfister and Dieter Schmalstieg





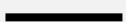



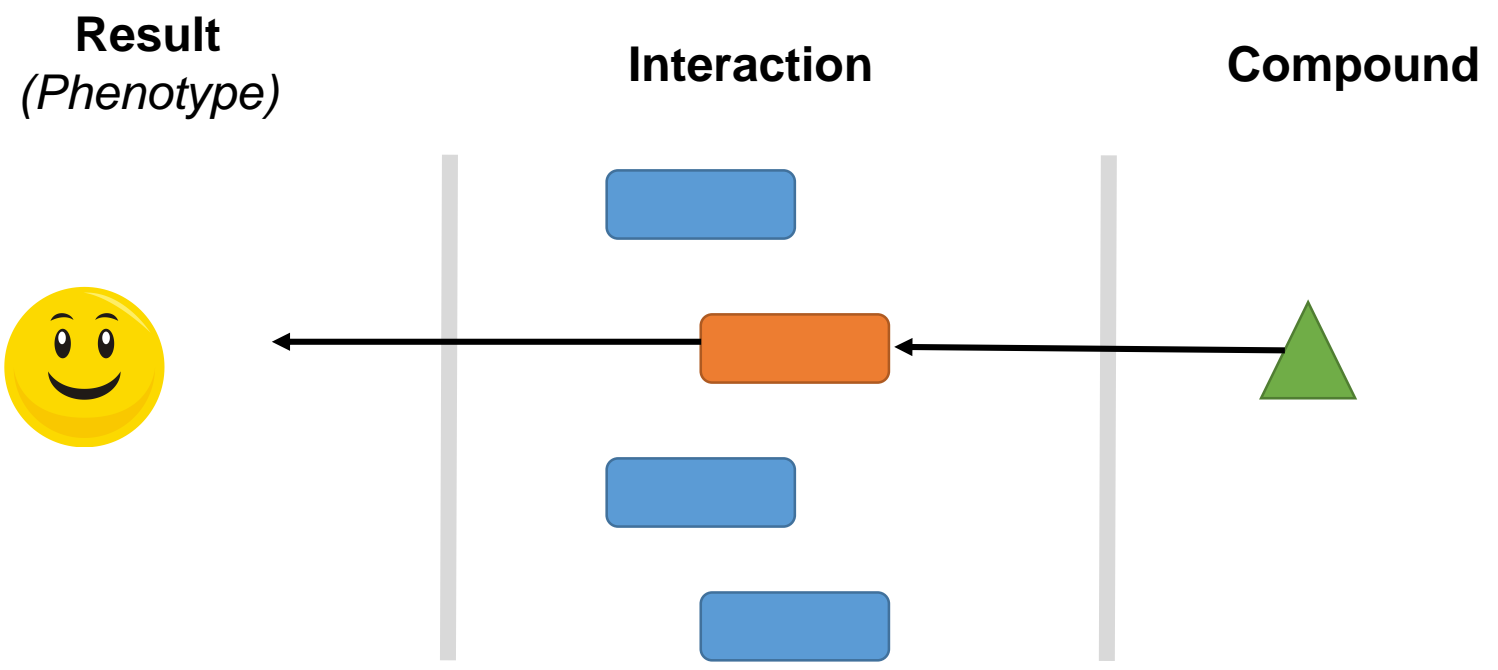
Domain Problem

UNDERSTANDING DRUG DISCOVERY

Scenario 1:

Targeted interaction, understood mechanism, desired outcome

-  = biological receptor
-  = biological target
-  = chemical compound
-  = potential target
-  = direct target
-  = inhibits target





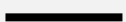



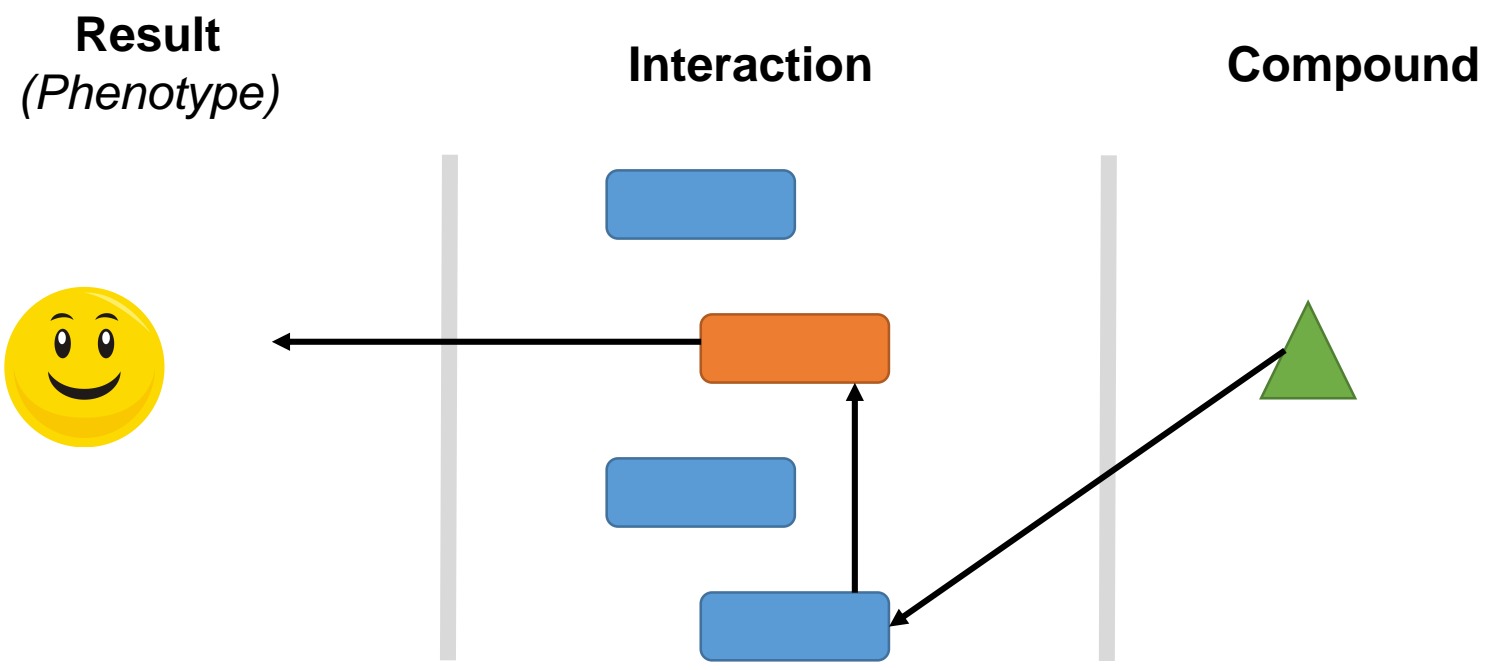
Domain Problem

UNDERSTANDING DRUG DISCOVERY

Scenario 2:

Indirect interaction, understood mechanism, desired outcome

-  = biological receptor
-  = biological target
-  = chemical compound
-  = potential target
-  = direct target
-  = inhibits target





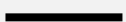



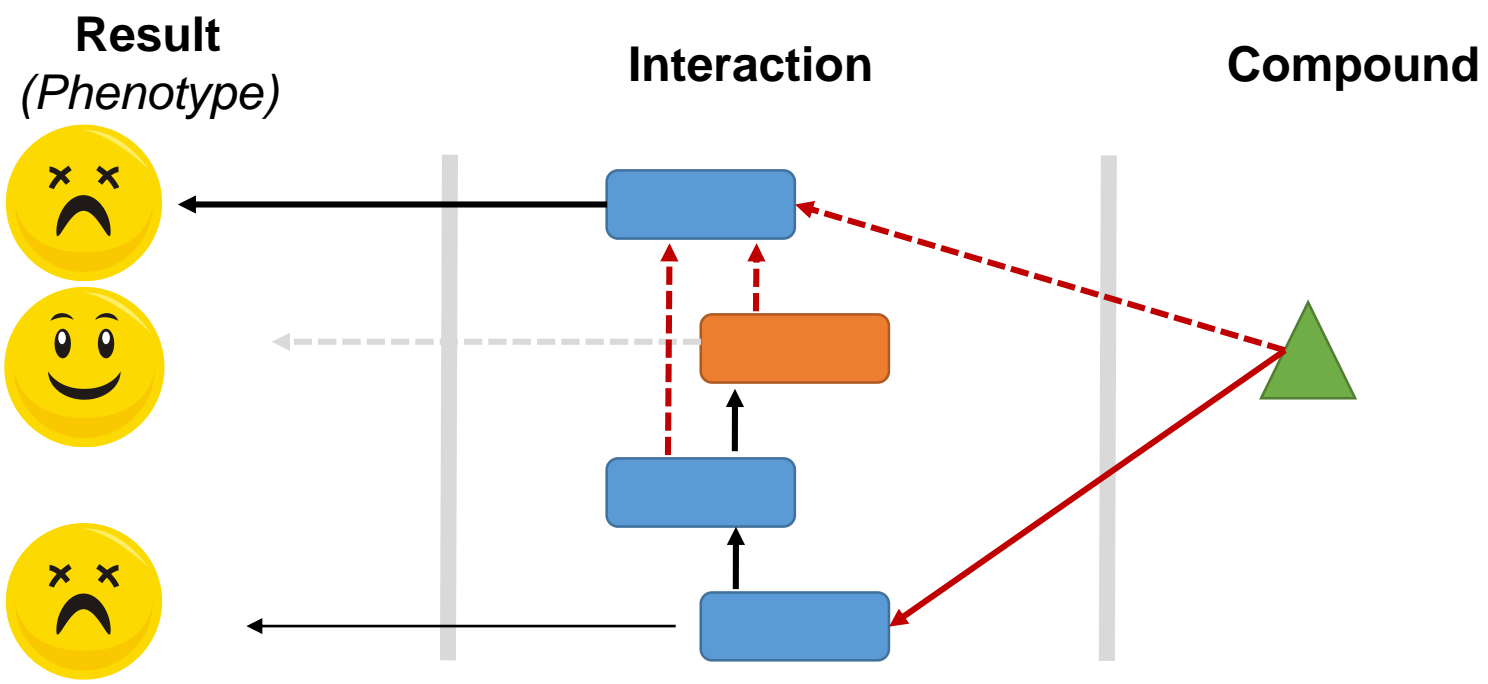
Domain Problem

UNDERSTANDING DRUG DISCOVERY

Scenario 3:

Complex interactions, mechanism poorly understood, multiple outcomes

-  = biological receptor
-  = biological target
-  = chemical compound
-  = potential target
-  = direct target
-  = inhibits target



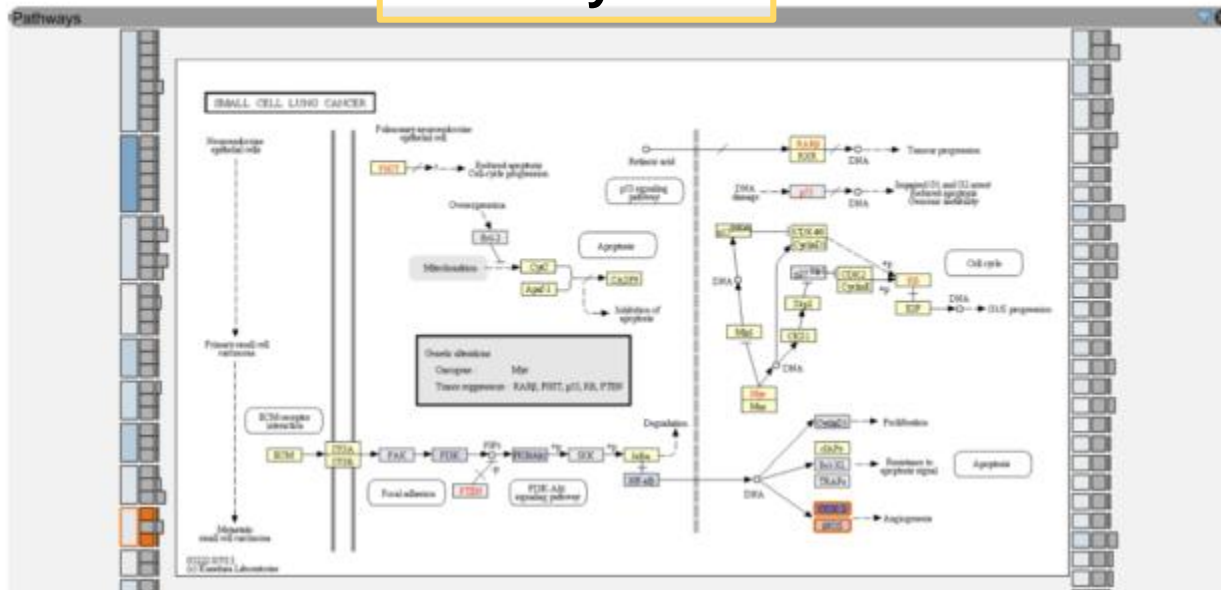
Drug Discovery Main Goals

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

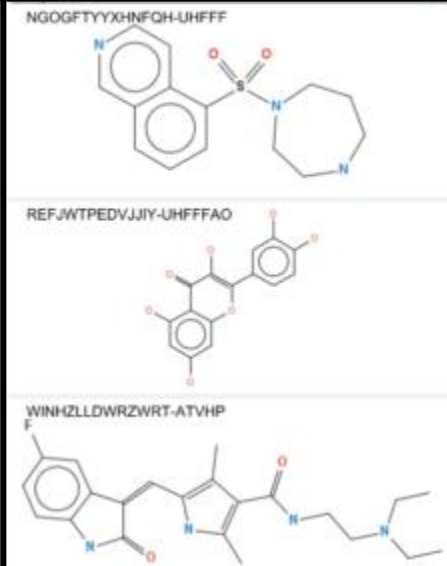
ConTour

History View

Pathway View



Compound View



- 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.
- Snapshots
- CAMK2A Anal.
- Potency Reduc.
- Glioma Correl.

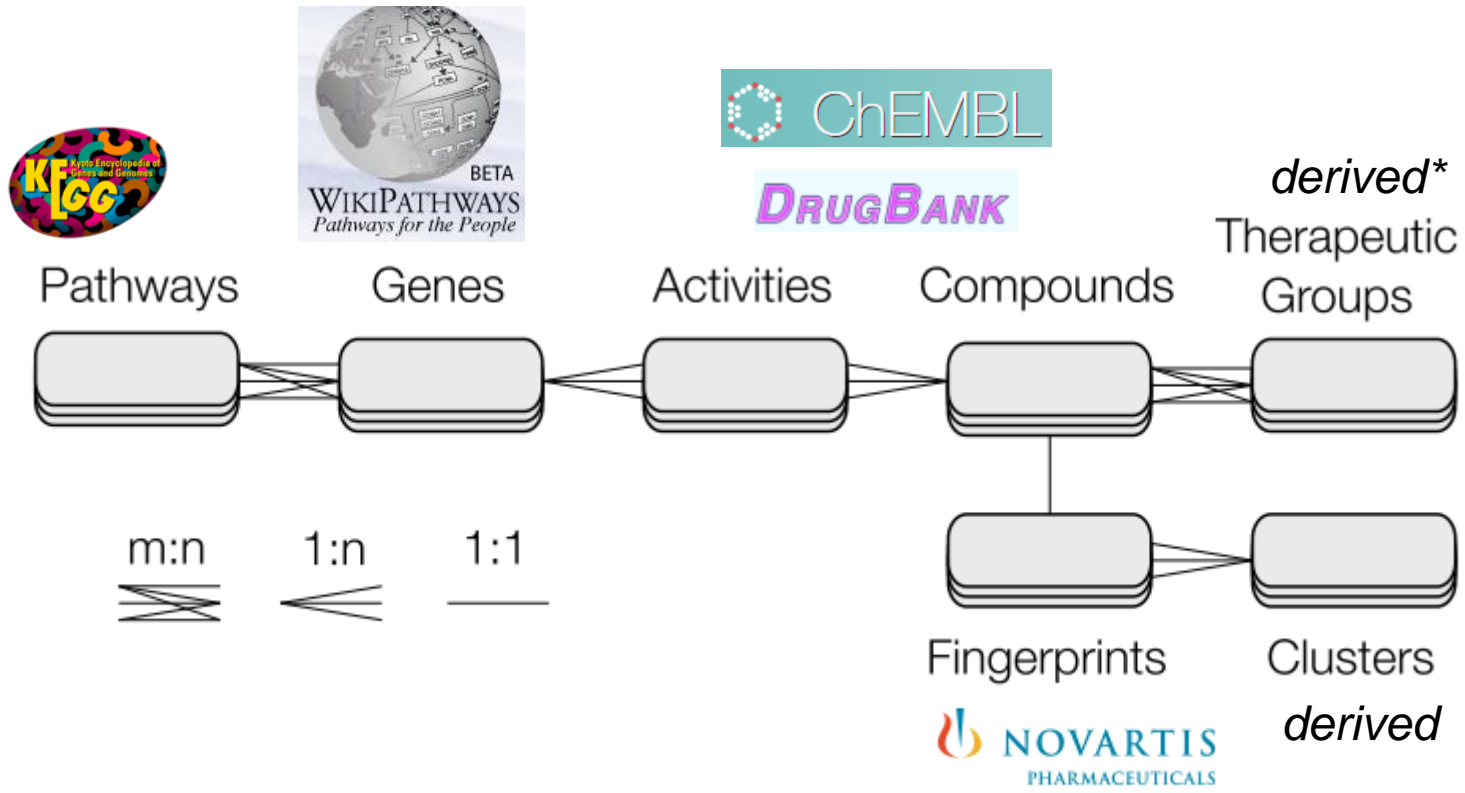
Pathways	Genes	Genes	Activities	Compounds	Activities	Clusters	Fingerprints	Therapeutic Group	Compounds
Pathways in cancer		CACNA1G		REGFWZVTFGQOJ		30		Adrenocortical D	
Arginine and proline		CACNA1H		HYAFETHFCALJAY-		52		Treatment of Hyp	
Calcium signaling pa				SEEPANYCNGTZFO		12		Treatment of Pag	
PI3K-Akt signaling p		CACNA1I		BSYNRYMUTXBSXQ		13		Treatment of Mu	
Serotonergic synaps		CACNA1S		ZPEIMTDSQAKGNT-		21		Antineoplastic	
Estrogen signaling p		CACNB3		KTUFNOKKBVMGR		11		Wilson Disease,	
AGE/RAGE pathway				KRMDCWKBEZIMAB		25		Fibrosis, Treatme	
Drug metabolism - c		CAMK1		REFJWTPEDVJJY-		9		Esophageal Dise	
Chemical carcinogen		CAMK2A		PLDUPXSUULZYBN-		19		Antidiarrheal Age	
HIF-1 signaling path								Pituitary Disorde	
Prostate cancer (KEG								Psoriatic Arthriti	
Myometrial Relaxatio								Laxatives	
Bile secretion (KEGG)								Treatment of Hyp	
Corticotropin-releasi								Antipruritics	
VEGF signaling path								Squamous Cell	
Small cell lung cance								Endocrine Disord	
NOD-like receptor si								Antithrombocty	
Arachidonic acid met								Treatment of Ac	

Relationship View

Filter View

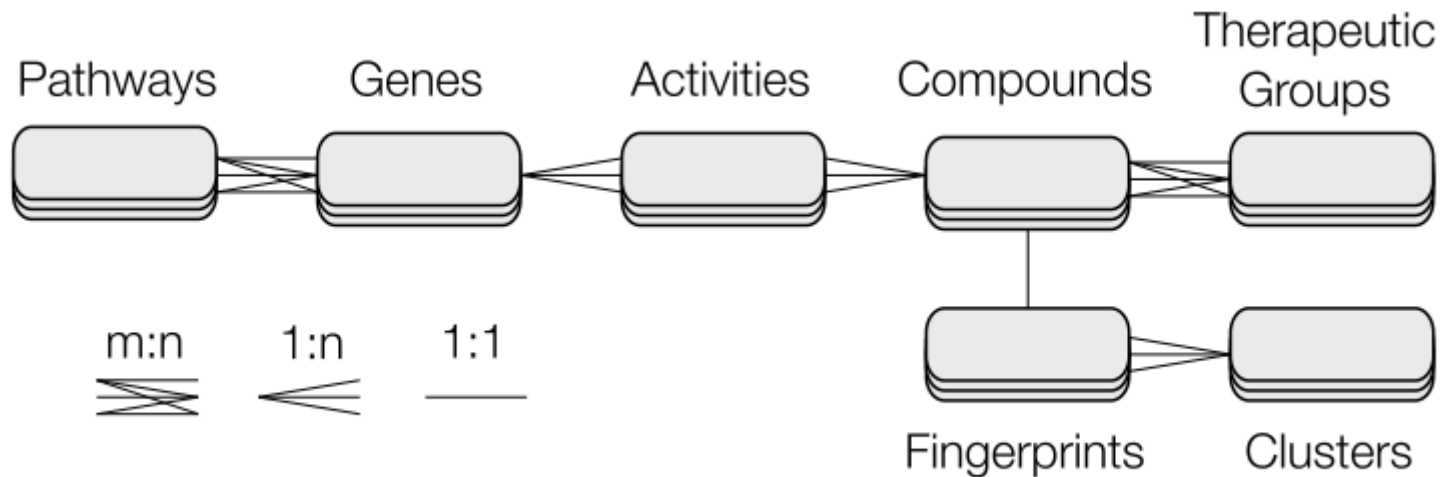
- Filters
- Activities
- Pathways
- Clusters
- Fingerprints
- Genes
- Pathways

Data Abstraction



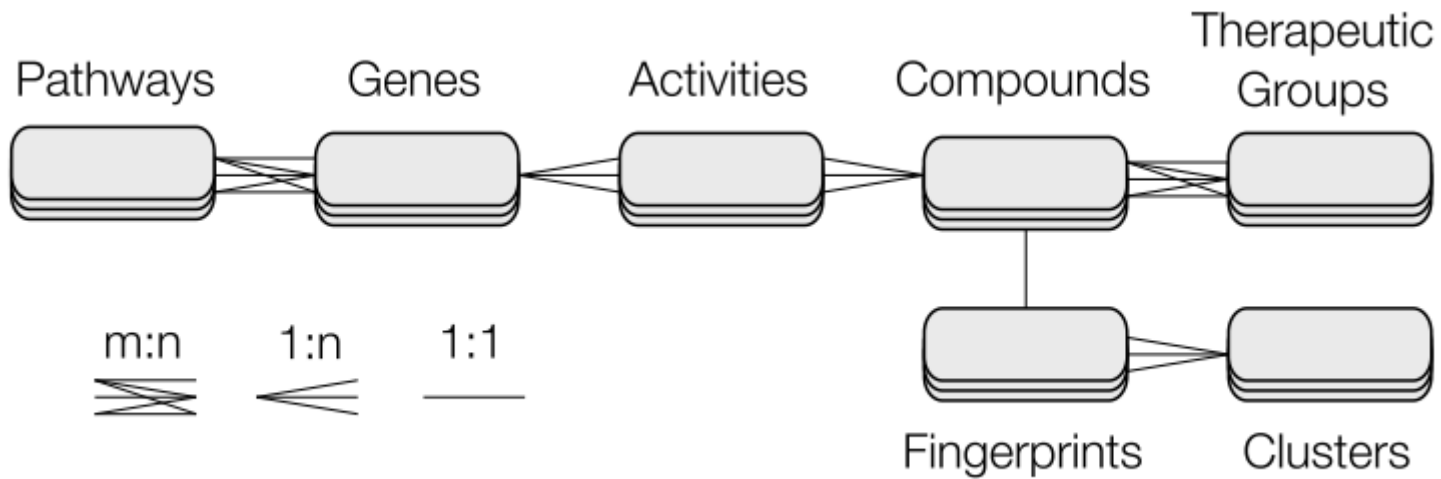
* Derived using a scheme propose by the Prous Integrity database

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

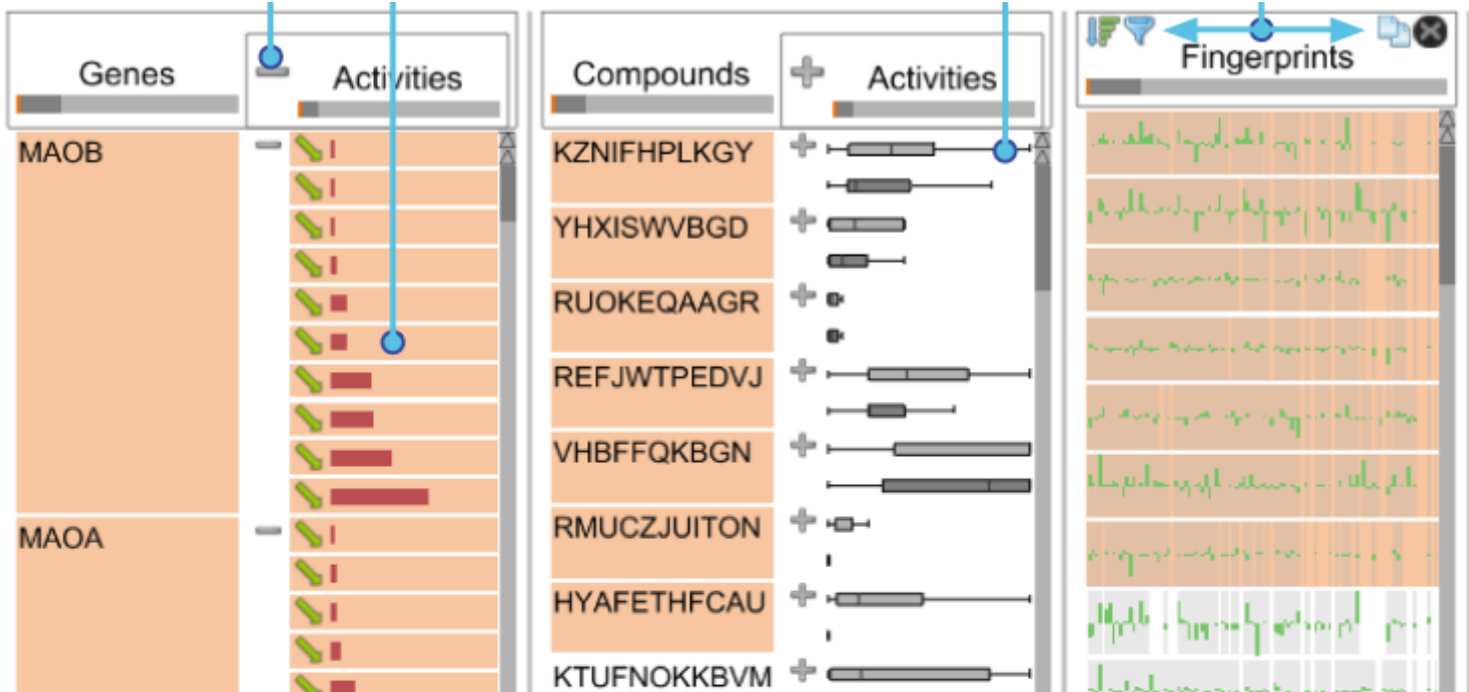


“The multi-relational data exploration problem can be interpreted as a **graph exploration problem** where each item of each dataset represents a node and the relationships between the items are the edges”

- **T1: Identify Related Items**

Item selection and highlighting

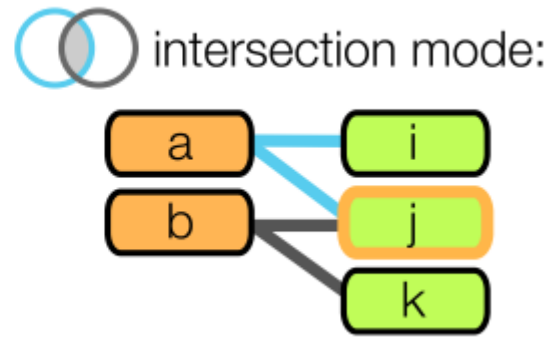
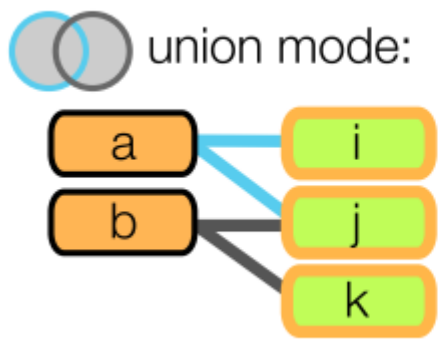
Clicking, not hovering, on an item also moves all related items in columns to the top



- **T1: Identify Related Items**

Selection-based filters

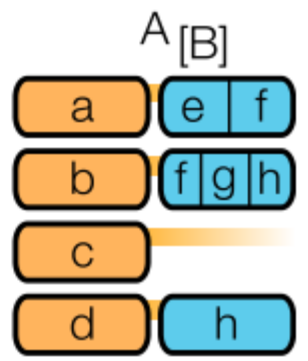
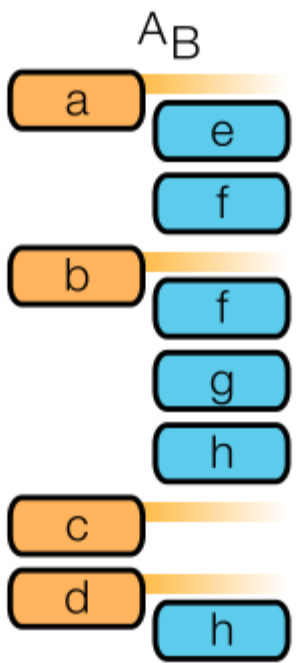
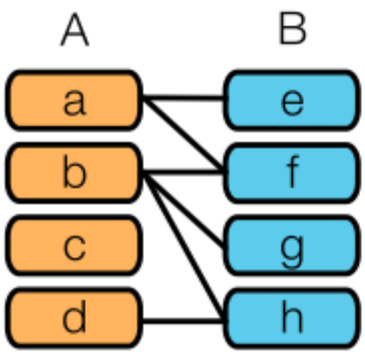
Filter choices when multiple items are selected



- **T1: Identify Related Items**

Nesting

Simple Nesting

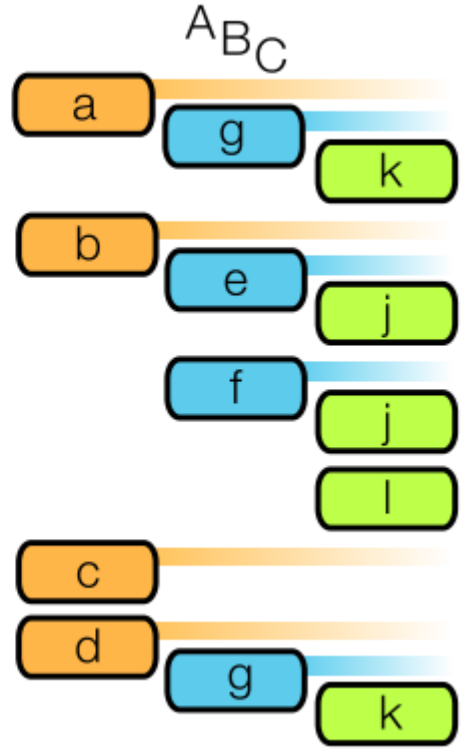
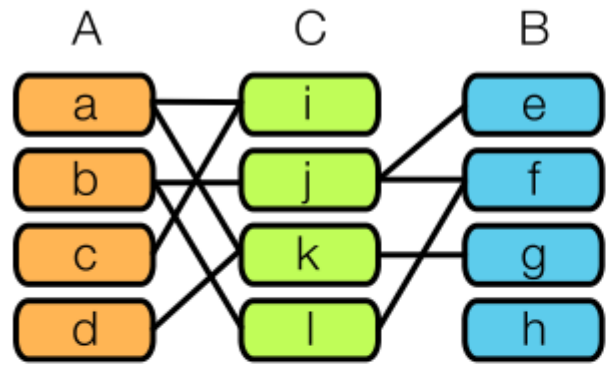


Task Analysis

- **T1: Identify Related Items**

Nesting

Recursive Nesting

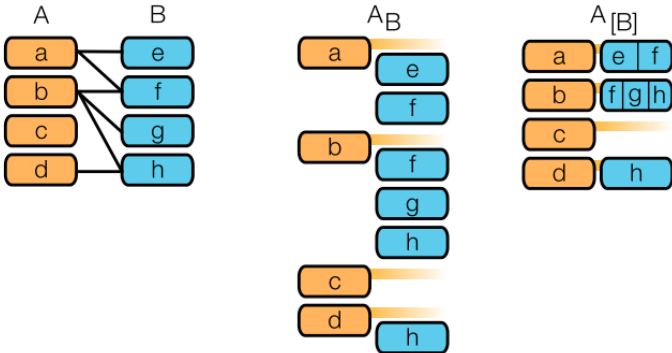


Task Analysis

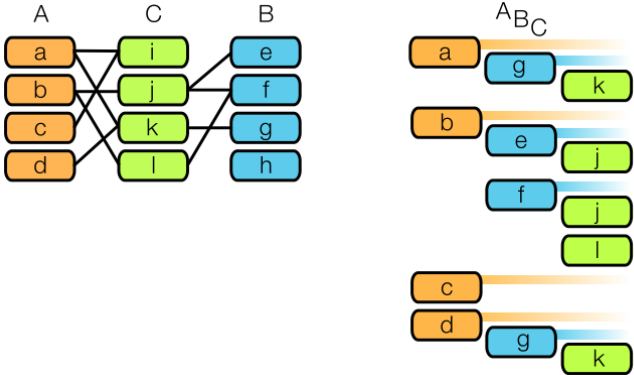
- **T2:** Identify Items that Share a Relationships with a Set of Items

Nesting

Simple Nesting



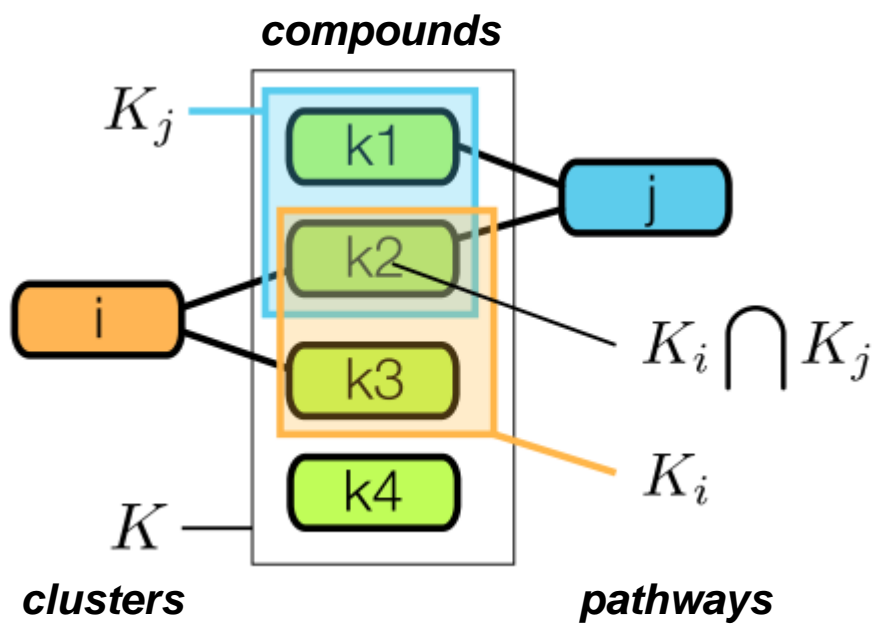
Recursive Nesting



• T3: Analyse Network Enrichment

Enrichment Score

Judging how specific two items are when compared to a third



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

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

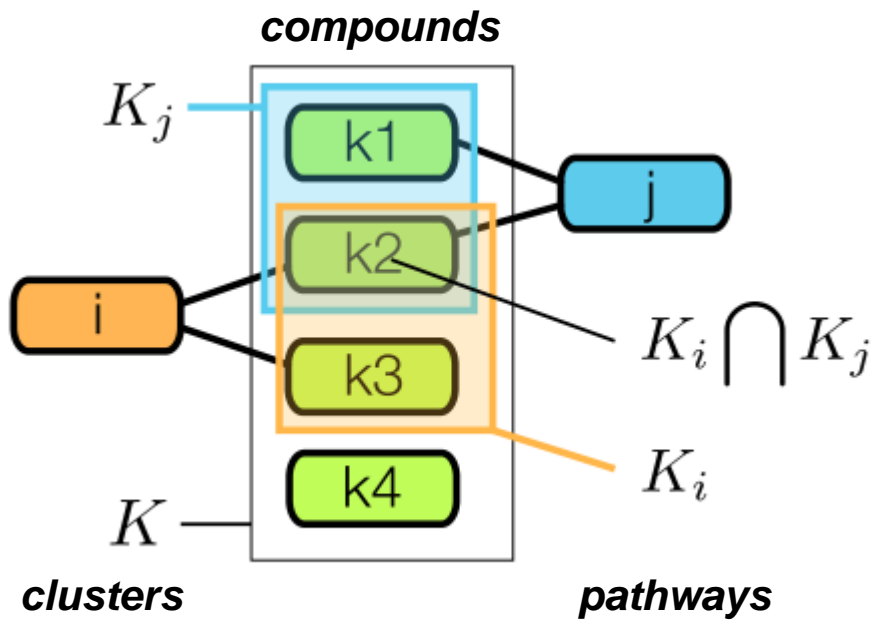
Sorting by interest

Sort alpha-numerically



Enrichment Score

Sort by enrichment score



Task Analysis

• T4: Rank Items

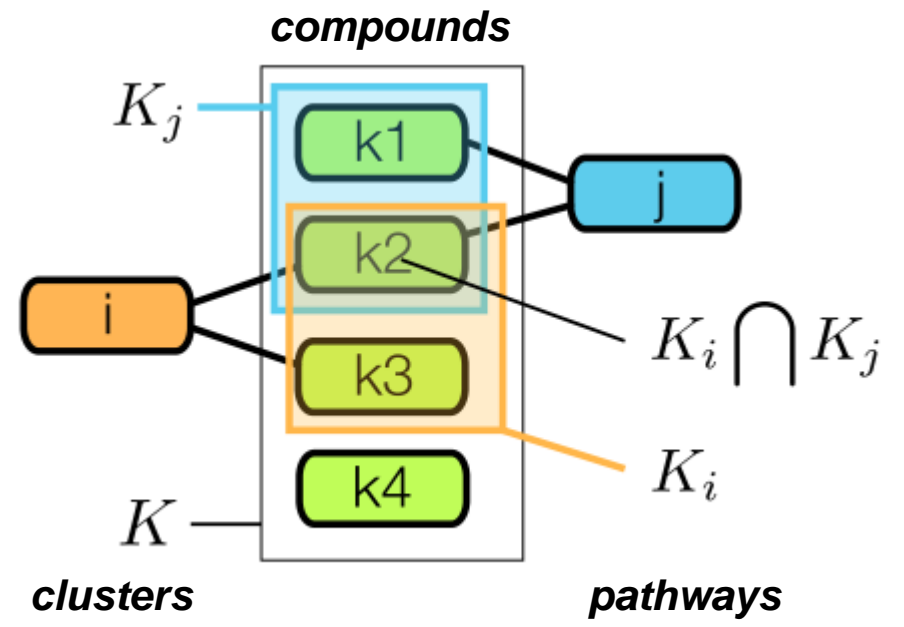
Sorting by interest

Sort alpha-numerically



Enrichment Score

Sort by enrichment score



- **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

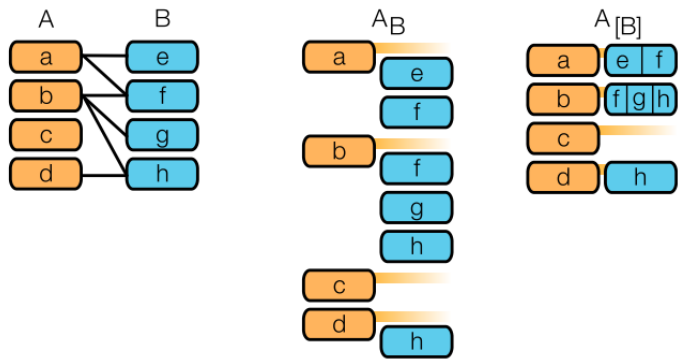


Selection-based filters

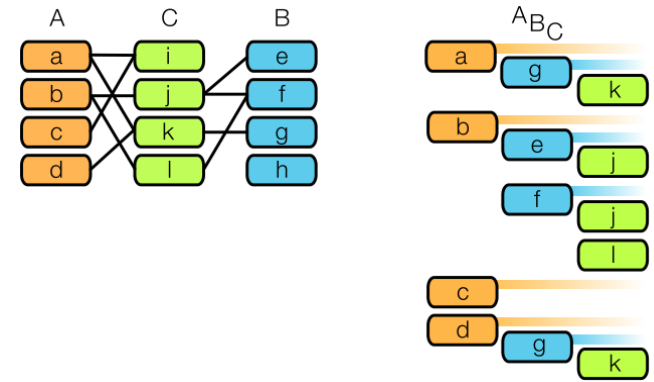


Nesting

Simple Nesting



Recursive Nesting



Task Analysis

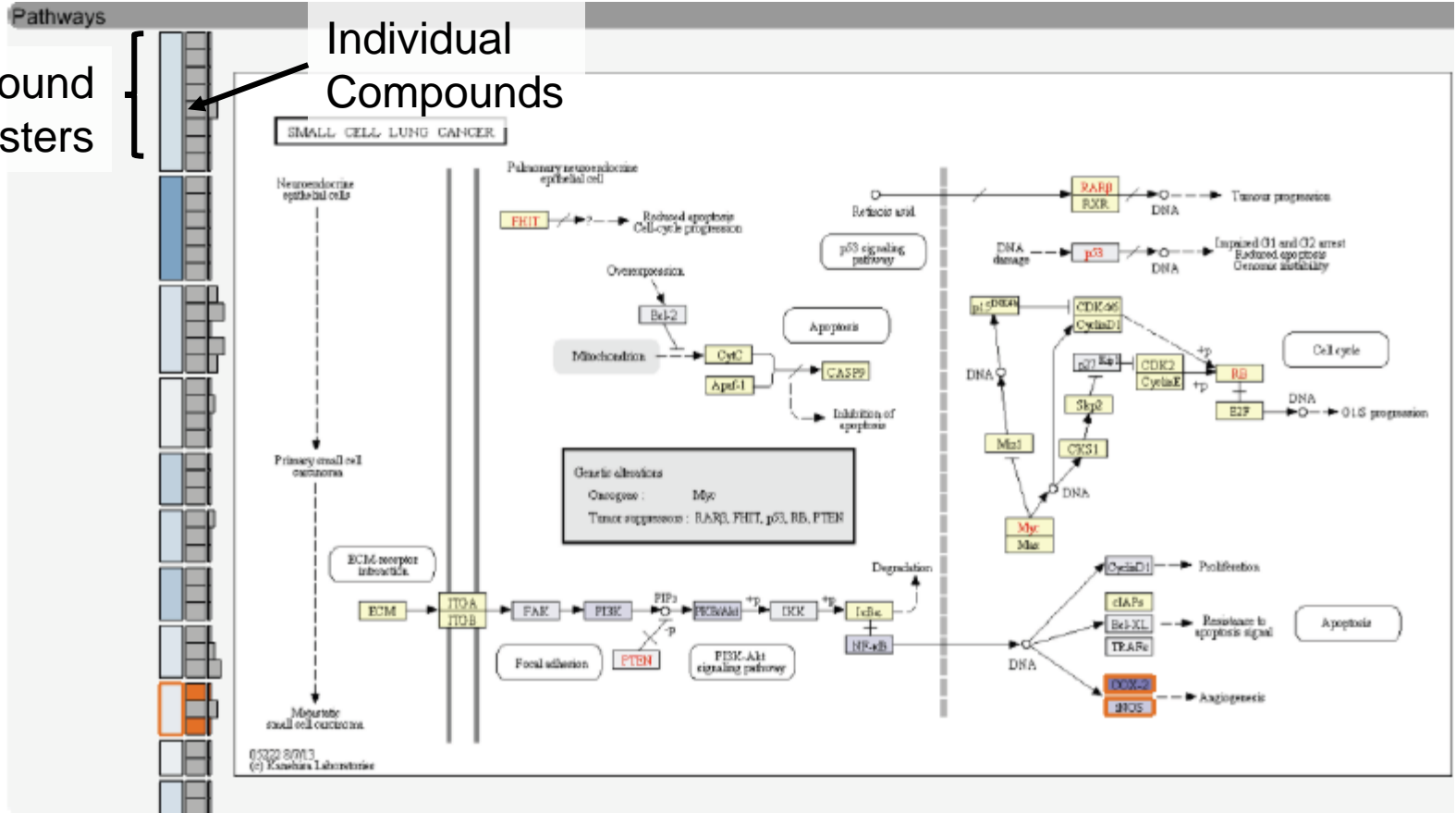
- **T6: View items in detail**

Pathway View

Marks

Compound Clusters

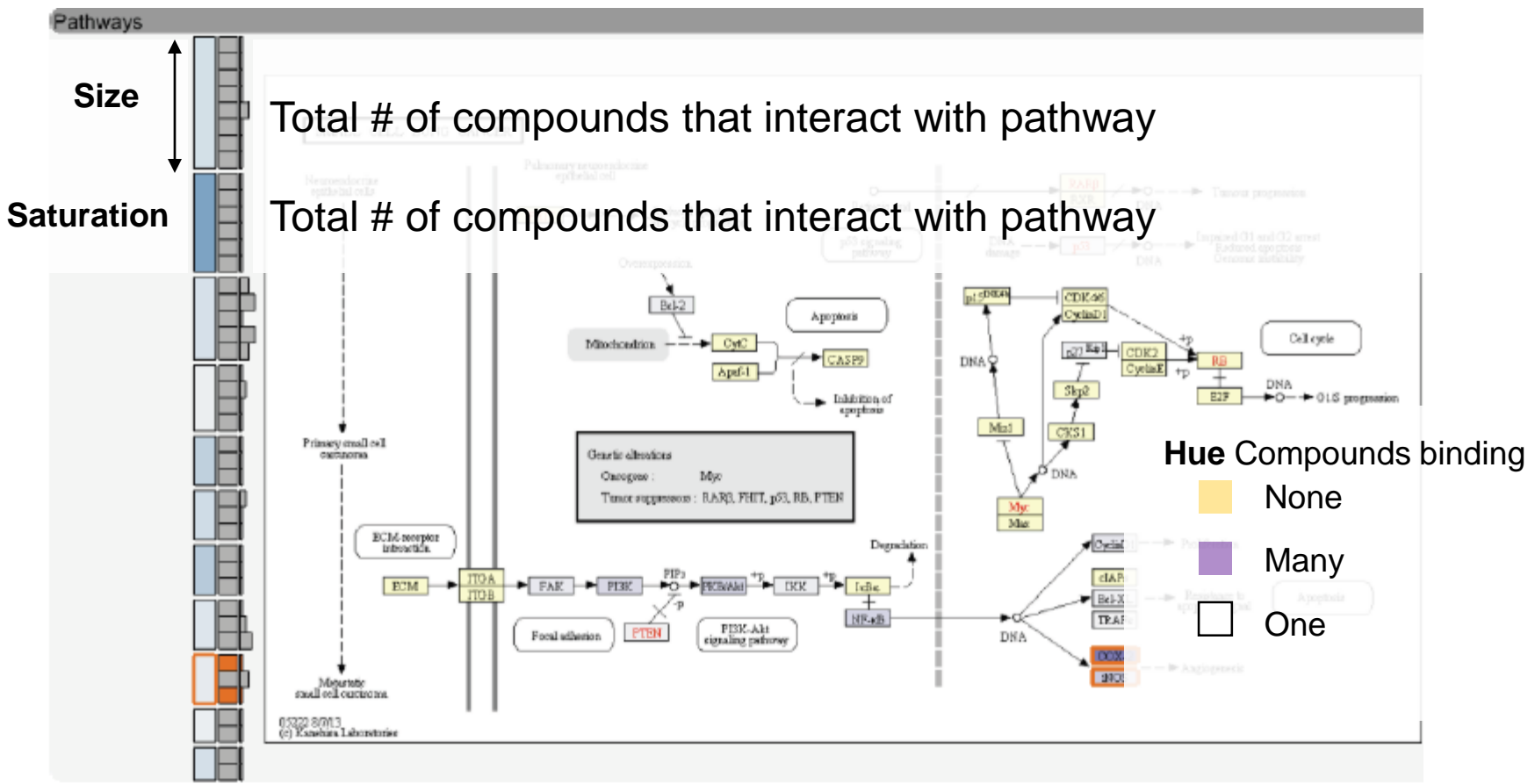
Individual Compounds



• T6: View items in detail

Pathway View

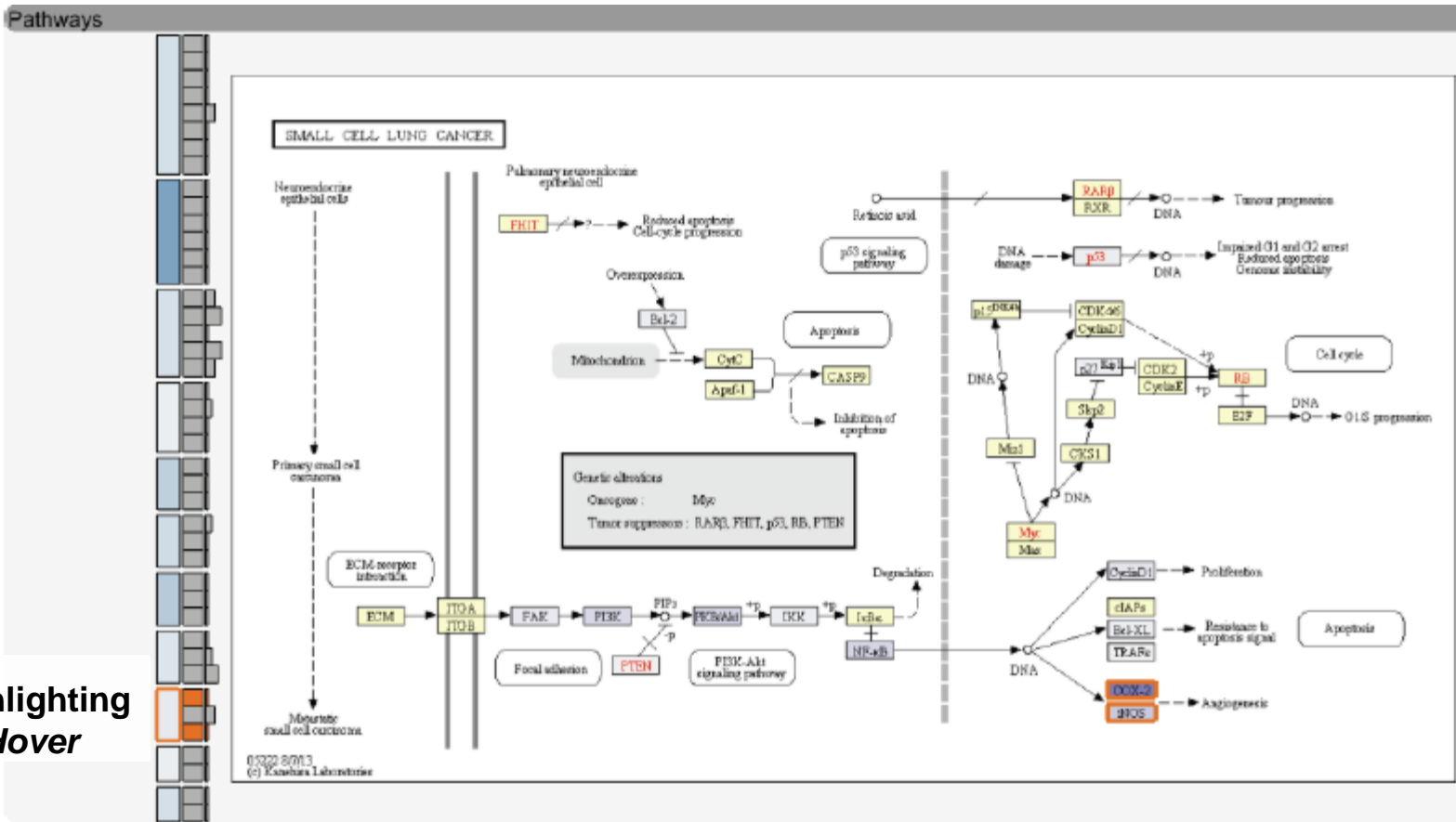
Channels



- **T6: View items in detail**

Pathway View

Linked Views



Highlighting
Hover

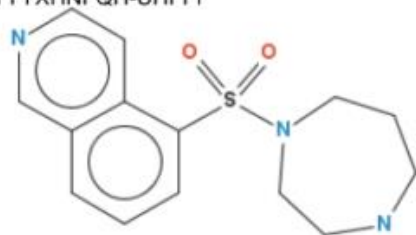
Task Analysis

- **T6: View items in detail**

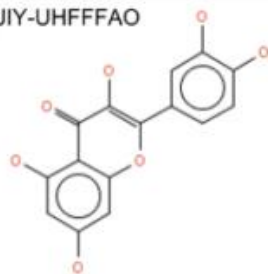
Compound View

Compounds

NGOGFTYYXHNFQH-UHFFF

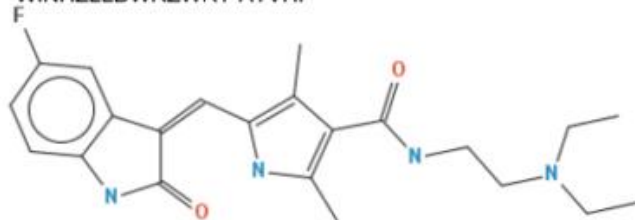


REFJWTPEDVJJIY-UHFFFAO



Hue : Elements

WINHZLLDWRZVRT-ATVHP



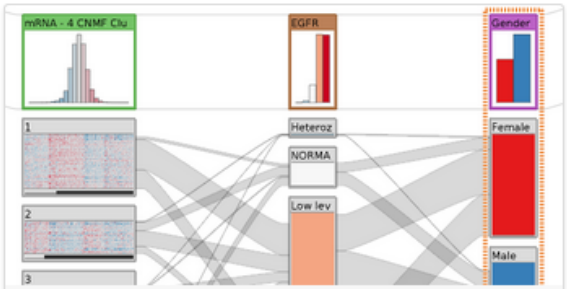
Implementation Details

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

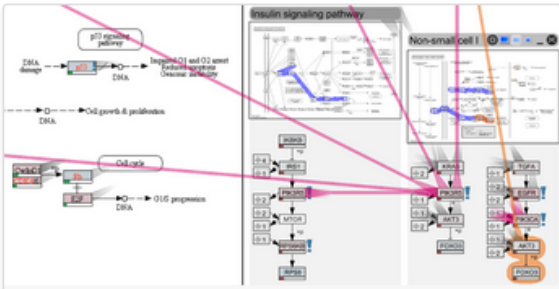
These projects are part of the caleydo core, i.e., if you download the Caleydo executable they will be included. These projects are tested and documented and are recommended for external use.



StratomeX

Genomic Stratification of Diseases

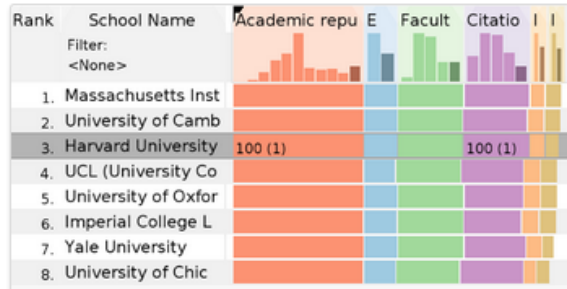
Integrative visualization of stratified heterogeneous data for disease subtype analysis.



Entourage & enRoute

Experimental Data in Pathways

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



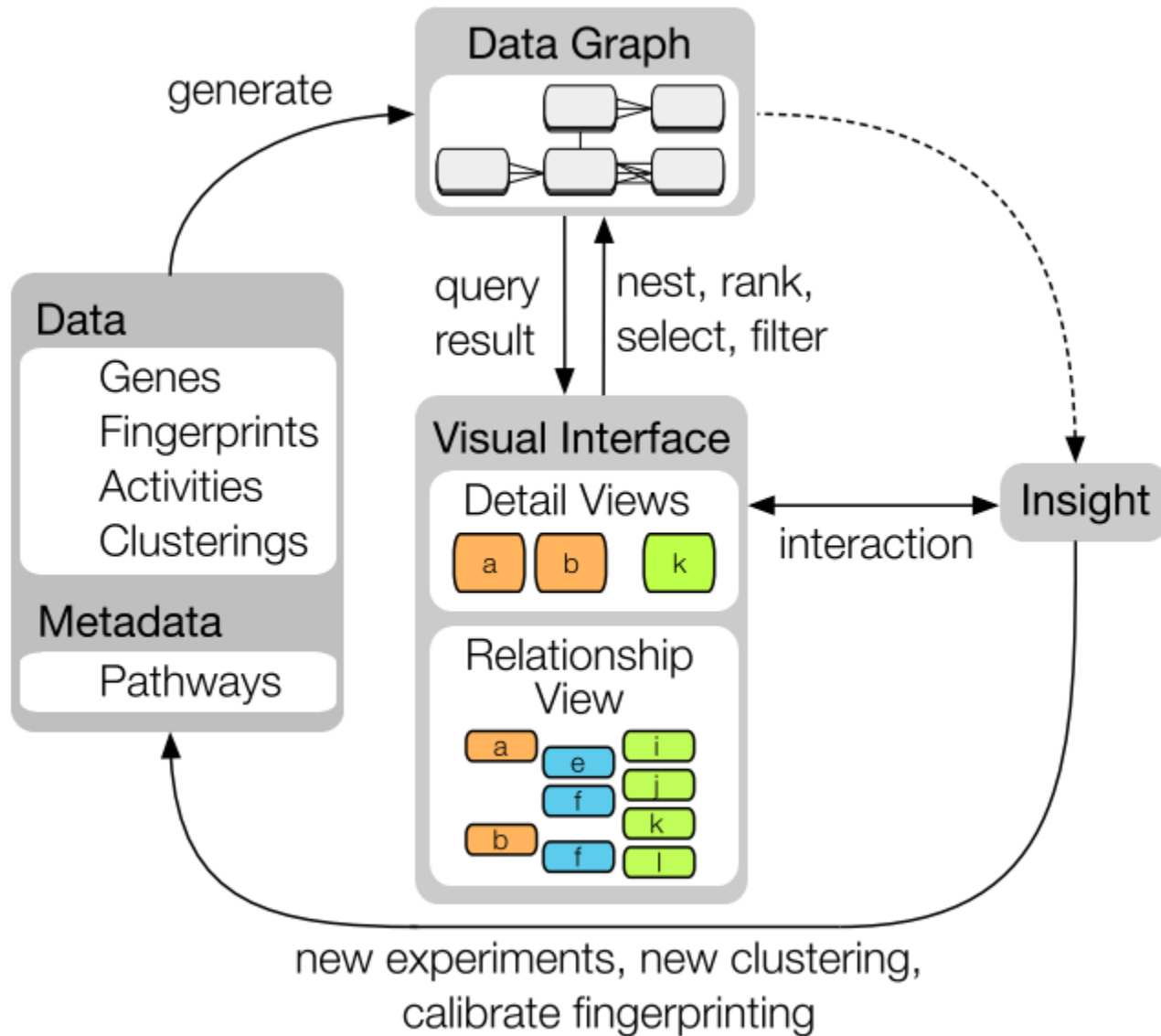
Rank	School Name	Academic repu	E	Facult	Citatio
1.	Massachusetts Inst				
2.	University of Camb				
3.	Harvard University	100 (1)			100 (1)
4.	UCL (University Co				
5.	University of Oxfor				
6.	Imperial College L				
7.	Yale University				
8.	University of Chic				

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.

ConTour

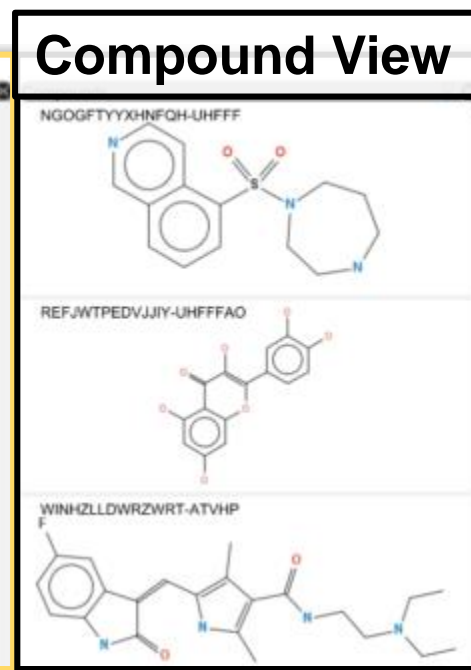
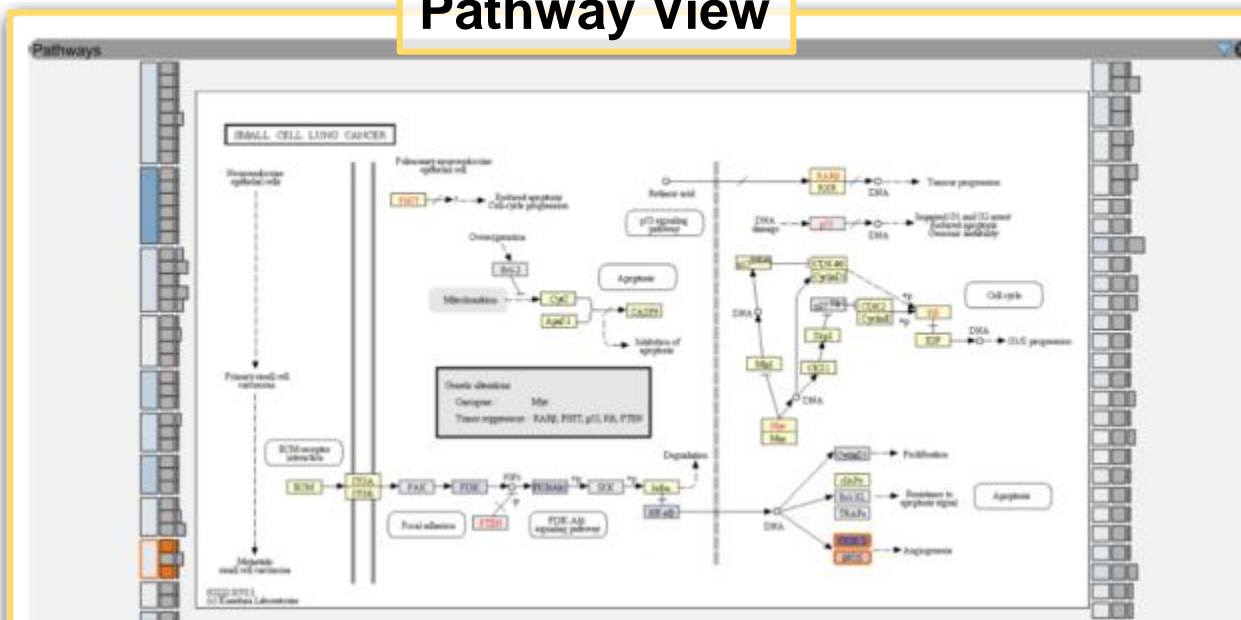


ConTour

History View

Pathway View

Compound View



- 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
- Snapshots
- CAMK2A Anal.
- Potency Reduc
- Glioma Correl

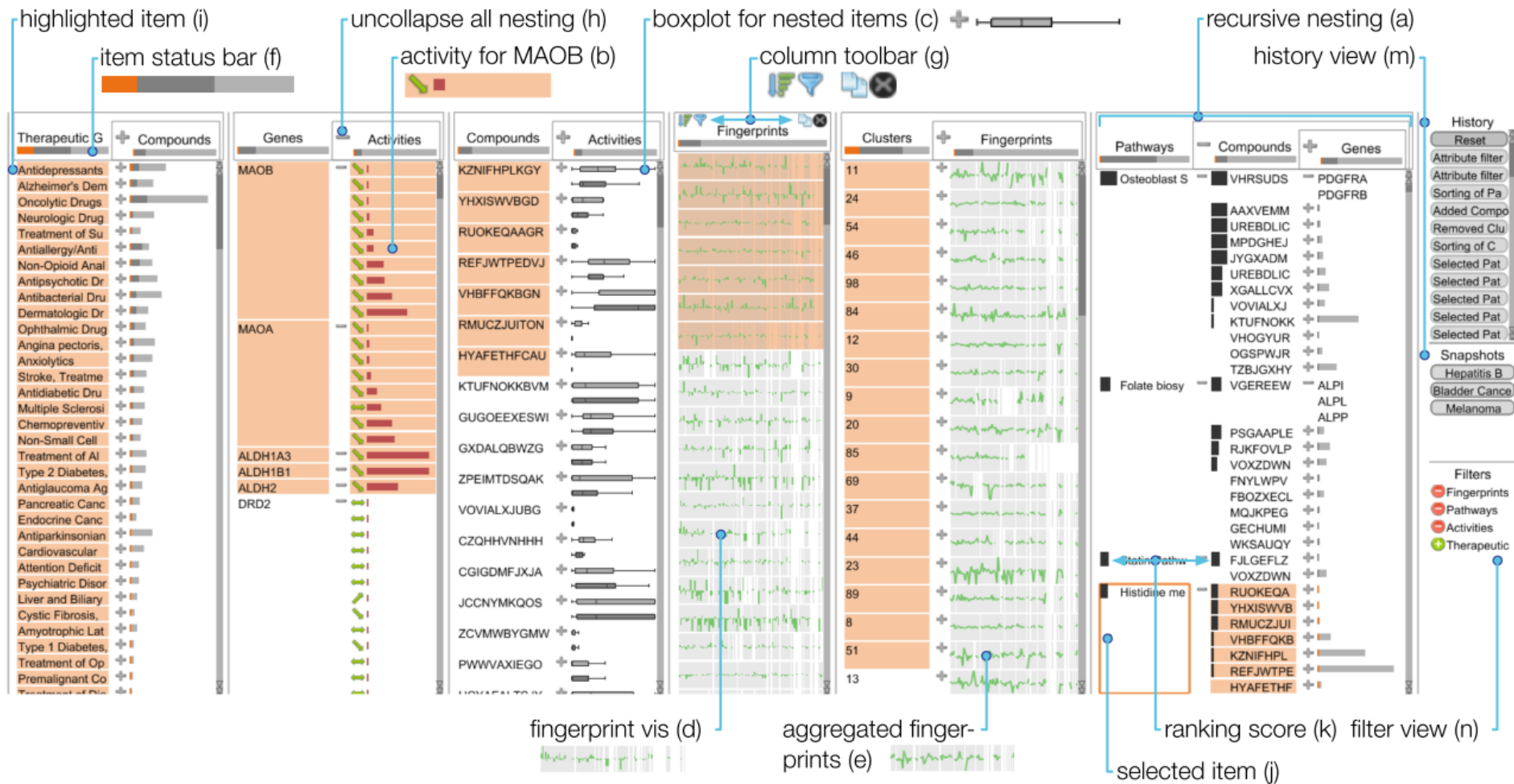
Pathways	Genes	Genes	Activities	Compounds	Activities	Clusters	Fingerprints	Therapeutic Group	Compounds
Pathways in cancer		CACNA1G		REGFWZVTFGQOJ		30		Adrenocortical D	
Arginine and proline		CACNA1H		HYAFETHFCALUAY-		52		Treatment of Hyp	
Calcium signaling pa				SEEPANYCNGTZFO		12		Treatment of Pag	
PI3K-Akt signaling p		CACNA1I		BSYNRYMUTXBSXQ		13		Treatment of Mu	
Serotonergic synaps		CACNA1S		ZPEIMTDSQAKGNT-		21		Antineoplastic	
Estrogen signaling p		CACNB3		KTUFNOKKBVMGR		11		Wilson Disease,	
AGE/RAGE pathway				KRMDCWKBEZIMAB		25		Fibrosis, Treatme	
Drug metabolism - c		CAMK1		REFJWTPEDVJJY-		9		Esophageal Dise	
Chemical carcinogen		CAMK2A		PLDUPXSUULZYBN-		24		Antidiarrheal Age	
HIF-1 signaling path						19		Pituitary Disorde	
Prostate cancer (KEG								Psoriatic Arthriti	
Myometrial Relaxatio								Laxatives	
Bile secretion (KEGG)								Treatment of Hyp	
Corticotropin-releasi								Antipruritics	
VEGF signaling path								Squamous Cell	
Small cell lung cance								Endocrine Disord	
NOD-like receptor si								Antithrombocty	
Arachidonic acid met								Treatment of Ac	

Relationship View

Filter View

- Filters
- Activities
- Pathways
- Clusters
- Fingerprints
- Genes
- Pathways

Relationship View



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

Relationship View

The screenshot displays the ConTour software interface in 'Relationship View' mode. The main content area is divided into three vertical panels: Pathways, Genes, and Activities. The Pathways panel lists various cancer-related pathways, the Genes panel lists associated genes, and the Activities panel shows activity levels for each gene across different pathways. A left sidebar contains a list of filters, and a right sidebar shows a history of actions and a filter view. Annotations with letters (i) through (m) point to specific UI elements:

- (i)** highlighted item
- (f)** item status bar
- (j)** selected item
- (k)** ranking score
- (n)** filter view
- (m)** history view
- (a)** recursive nesting
- (b)** uncollapse
- (c)** activity

The Pathways panel lists the following items:

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

The Genes panel lists the following items:

- 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
- CYP17A1

The Activities panel shows activity levels for each gene across different pathways, represented by colored bars and icons.

The left sidebar contains a list of filters, including Therapeutic G, Compounds, Genes, and Activities. The right sidebar contains a History view and a Filter view.

Relationship View

The screenshot displays the ConTour II software interface, which is divided into several functional panels:

- Top Panel:** Features a 'Pathways' and 'Genes' section with a search bar and a list of items. The 'Bladder cancer (KEGG)' item is highlighted with a blue box, and a tooltip shows its details.
- Left Panel:** Contains a 'Therapeutic G' and 'Compounds' section with a list of items and an 'item status bar (f)' below it.
- Center Panel:** Displays a list of 'Genes' and 'Activities' with associated activity bars. A blue dot indicates an 'uncollapse all n' action.
- Right Panel:** Shows a 'History' view with a list of items and a 'filter view (n)' section. A blue box highlights a 'selected item (j)'.
- Bottom Panel:** Includes a 'Ranking score (k)' section and a 'filter view (n)' section with a legend for 'Fingerprints', 'Pathways', 'Activities', and 'Therapeutic'.

Annotations with blue lines point to various elements:

- highlighted item (i)
- item status bar (f)
- uncollapse all n
- activity for
- recursive nesting (a)
- history view (m)
- ranking score (k)
- filter view (n)
- selected item (j)

Relationship View

The image displays a complex software interface for a 'Relationship View'. The interface is divided into several main sections:

- Left Panel:** A vertical list of categories such as 'Therapeutic Groups', 'Antidepressants', 'Alzheimer's Dem', 'Oncolytic Drugs', 'Neurologic Drugs', 'Treatme', 'Antialler', 'Non-Opi', 'Antipsyc', 'Antibact', 'Dermat', 'Ophthal', 'Angina p', 'Anxiolyt', 'Stroke', 'Antidiab', 'Multiple', 'Chemop', 'Non-Sm', 'Treatme', 'Type 2 D', 'Antiglau', 'Pancrea', 'Endocr', 'Antipark', 'Cardio', 'Attention', 'Psychiatric Disor', 'Liver and Biliary', 'Cystic Fibrosis', 'Amyotrophic Lat', 'Type 1 Diabetes', 'Treatment of Op', and 'Premalignant Co'. Each category has a plus sign and a progress bar.
- Top Section:** A series of panels for 'Compounds', 'Genes', and 'Activities'. The 'Genes' panel shows 'MAOB' with a sub-panel for 'Activities' containing a bar chart. The 'Compounds' panel shows 'KZNFHPLKGY' and 'YHXISWVBGD' with their own 'Activities' sub-panels.
- Center Section:** A 'Fingerprints' panel showing a boxplot for nested items (c) and a 'Clusters' panel with a table of clusters (11, 24). A 'column toolbar (g)' is located above the fingerprints panel.
- Right Section:** A 'Pathways' panel showing 'Osteoblast S', 'VHRSUDS', 'PDGFRA', 'PDGFRB', and 'AAVEMM'. A 'History' panel on the far right shows a list of items with 'Reset', 'Attribute filter', and 'Sorting of Pa' options.
- Dialog Box:** A 'Create Enrichment Score' dialog box is open in the foreground. It contains the text: 'Calculate enrichment of [dropdown] for [dropdown] via [dropdown] considering [dropdown] with a threshold of 2'. It has 'OK' and 'Cancel' buttons.
- Bottom Section:** A 'fingerprint vis (d)' panel showing individual fingerprint plots. An 'aggregated fingerprints (e)' panel shows a combined fingerprint plot. A 'ranking score (k)' panel shows a list of items with their scores (64, 76, 8, 91). A 'filter view (n)' panel shows a list of items with their filter status.
- Other Elements:** A 'selected item (j)' panel shows 'Histidine me'. A 'recursive nesting (a)' panel shows a list of items with their nesting status.

Annotations with blue lines point to various features: 'highlighted item (i)' points to a highlighted item in the left panel; 'item status bar (f)' points to a bar chart; 'uncollapse all nesting (h)' points to a button; 'activity for MAOB (b)' points to a bar chart; 'boxplot for nested items (c)' points to a boxplot; 'column toolbar (g)' points to a toolbar; 'recursive nesting (a)' points to a panel; 'history view (m)' points to the history panel; 'fingerprint vis (d)' points to a fingerprint plot; 'aggregated fingerprints (e)' points to an aggregated fingerprint plot; 'ranking score (k)' points to a list of items; 'filter view (n)' points to a filter view; and 'selected item (j)' points to a selected item.

Relationship View

highlighted item (i)

item status bar (f)

uncollapse a

activity

The sidebar contains four main sections: Therapeutic, Compounds, Genes, and Activities. Each section has a list of items with expand/collapse icons. The 'Activities' section is currently expanded, showing a list of activity names with corresponding status bars. A blue dot highlights a specific item in the 'Activities' list.



recursive nesting (a)
history view (m)

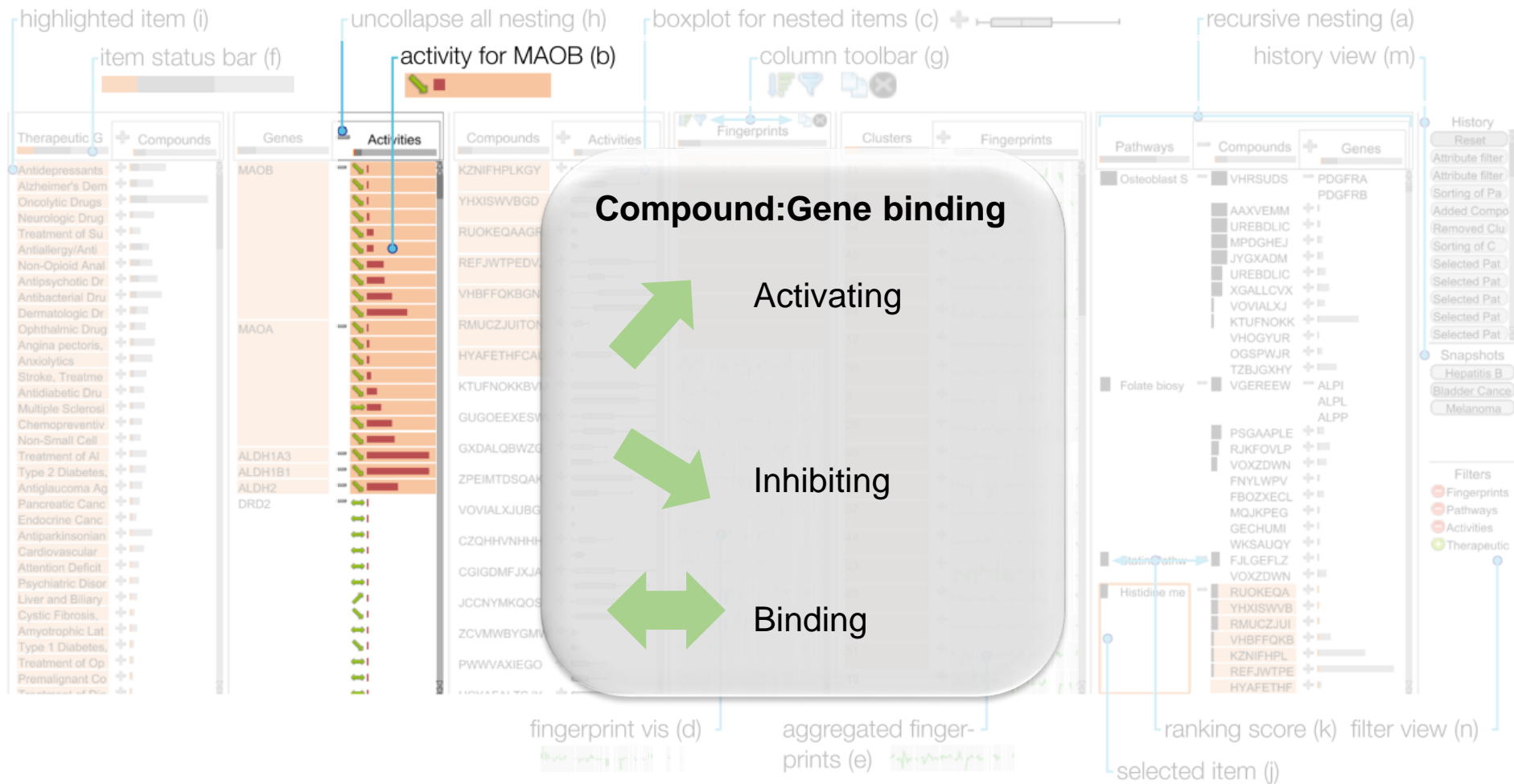
This view shows a more detailed and nested version of the sidebar. It includes sections for Pathways, Compounds, and Genes, with sub-sections for each. A 'History' panel on the right shows a list of recent actions, such as 'Reset', 'Attribute filter', and 'Sorting of Pa'. A 'Filters' panel at the bottom right shows active filters for Fingerprints, Pathways, Activities, and Therapeutic. A blue dot highlights a specific item in the 'Compounds' section.

ranking score (k) filter view (n)

selected item (j)

Approximately 100 numerical values shown here

Relationship View

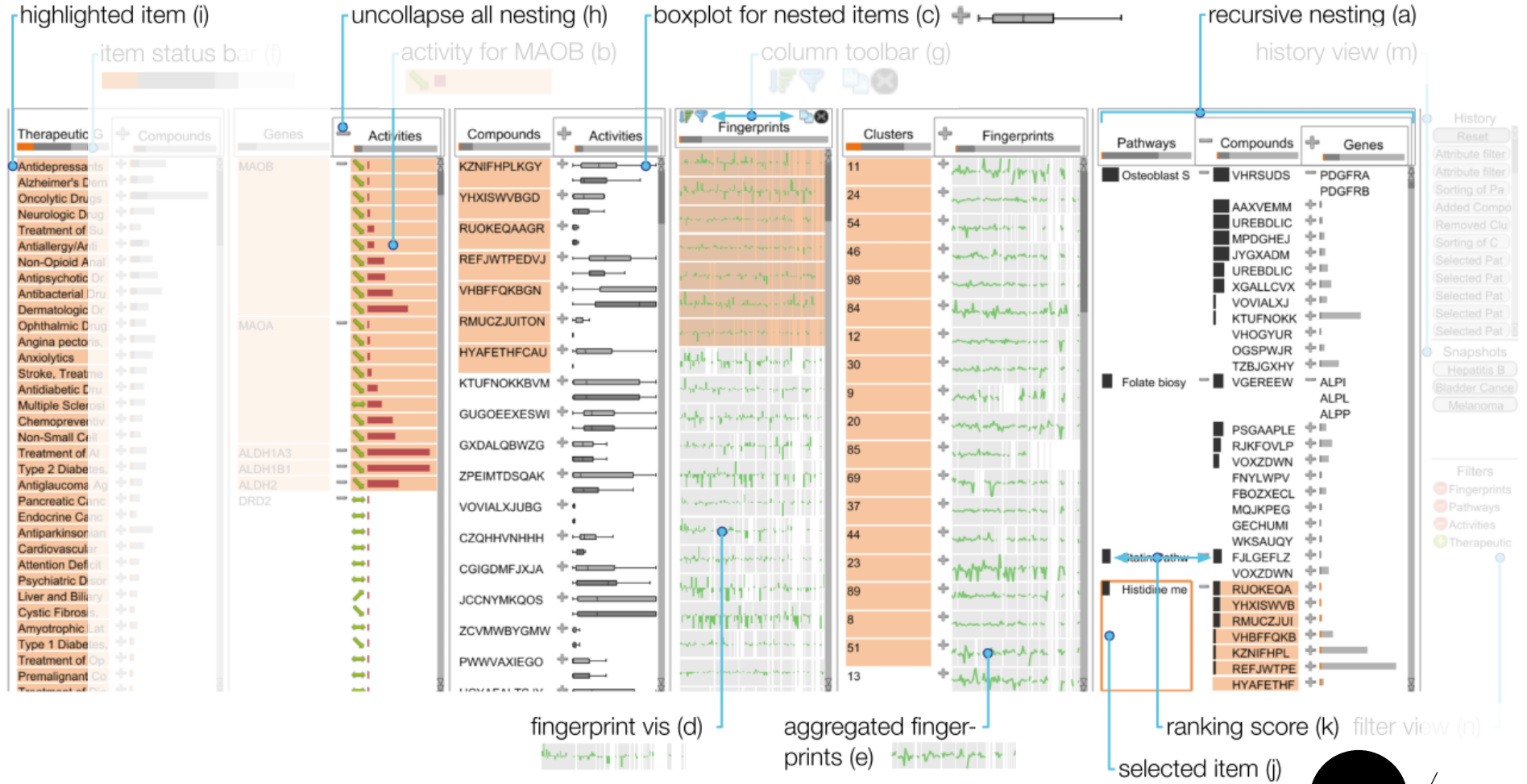


Approximately 100 numerical values shown here

Relationship View

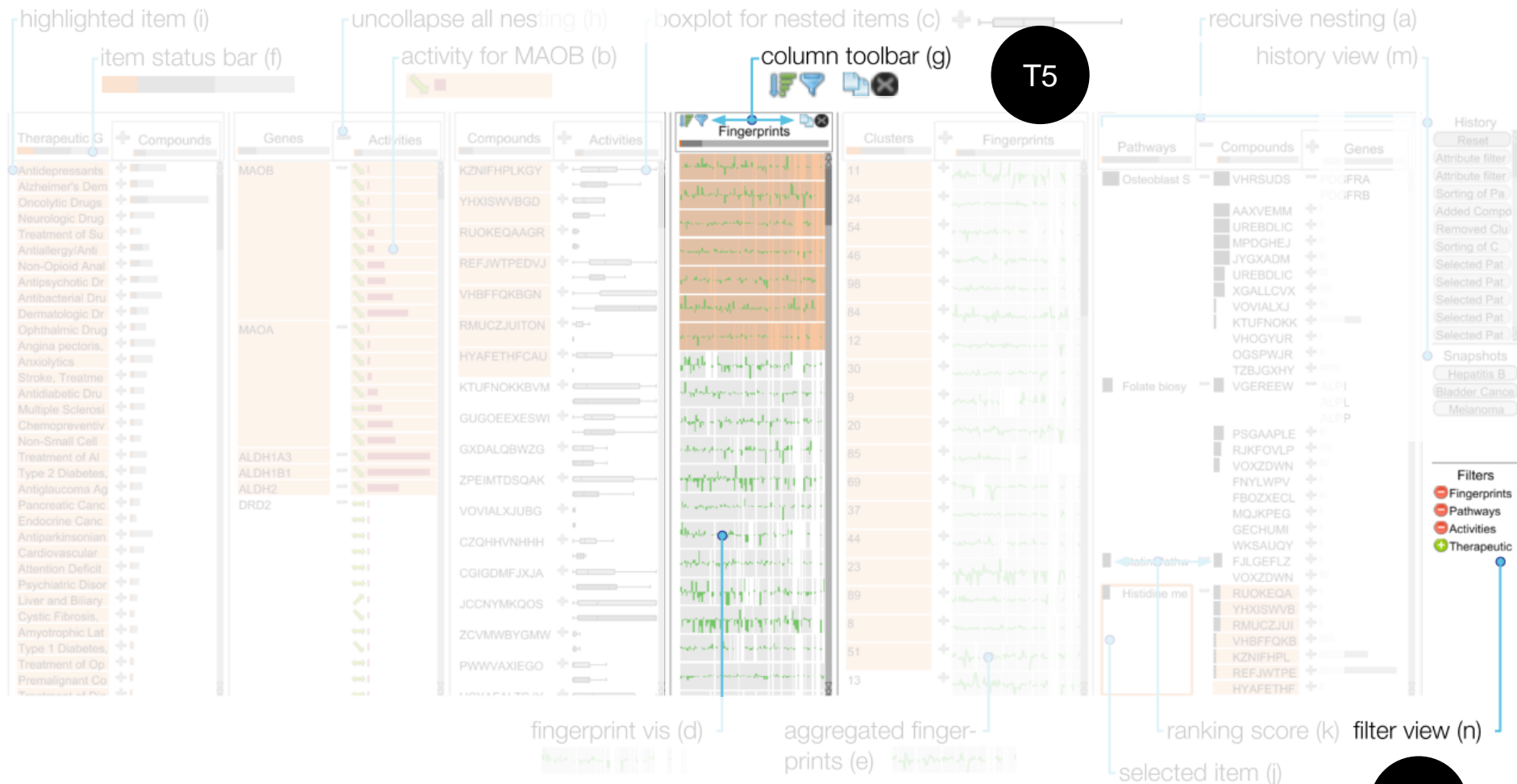
T1

T2



T3

T4



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

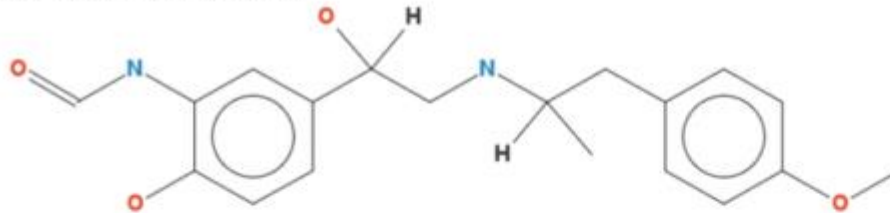
Concluding Thoughts

- 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?

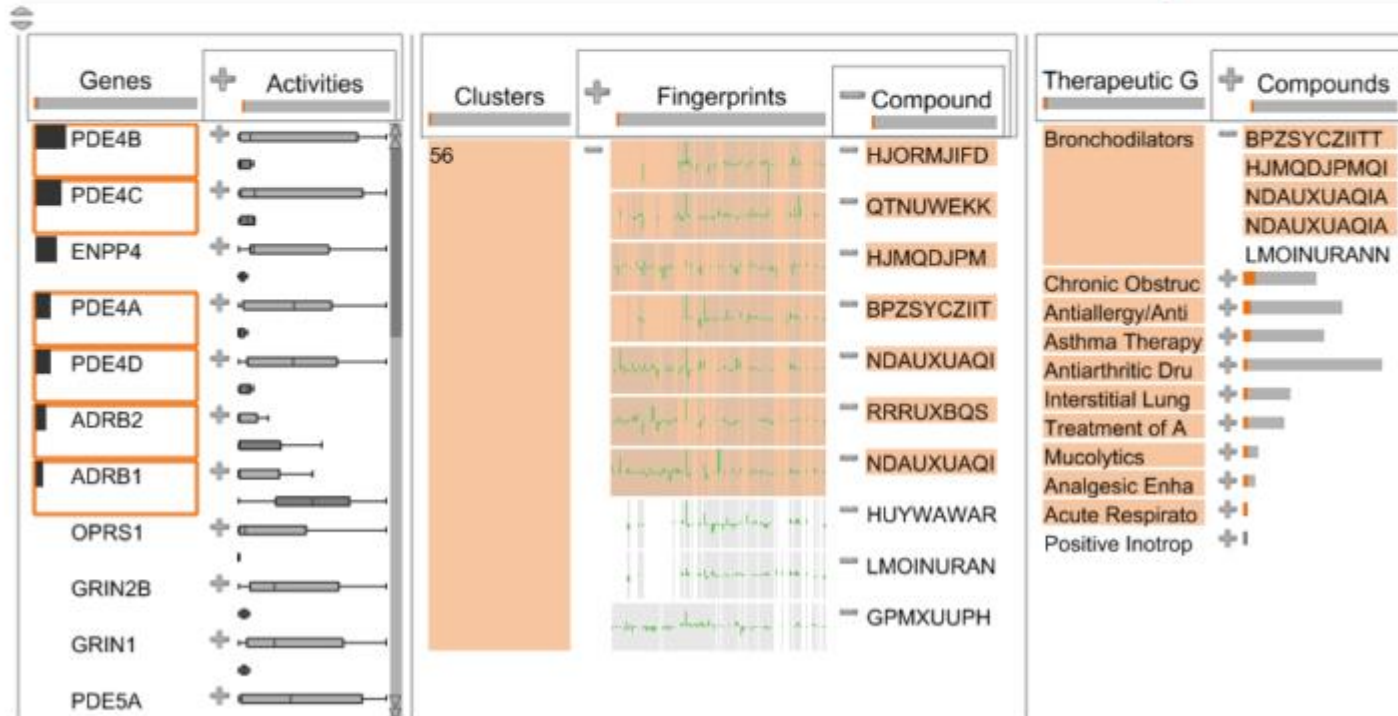
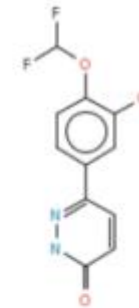
Backup

Compounds

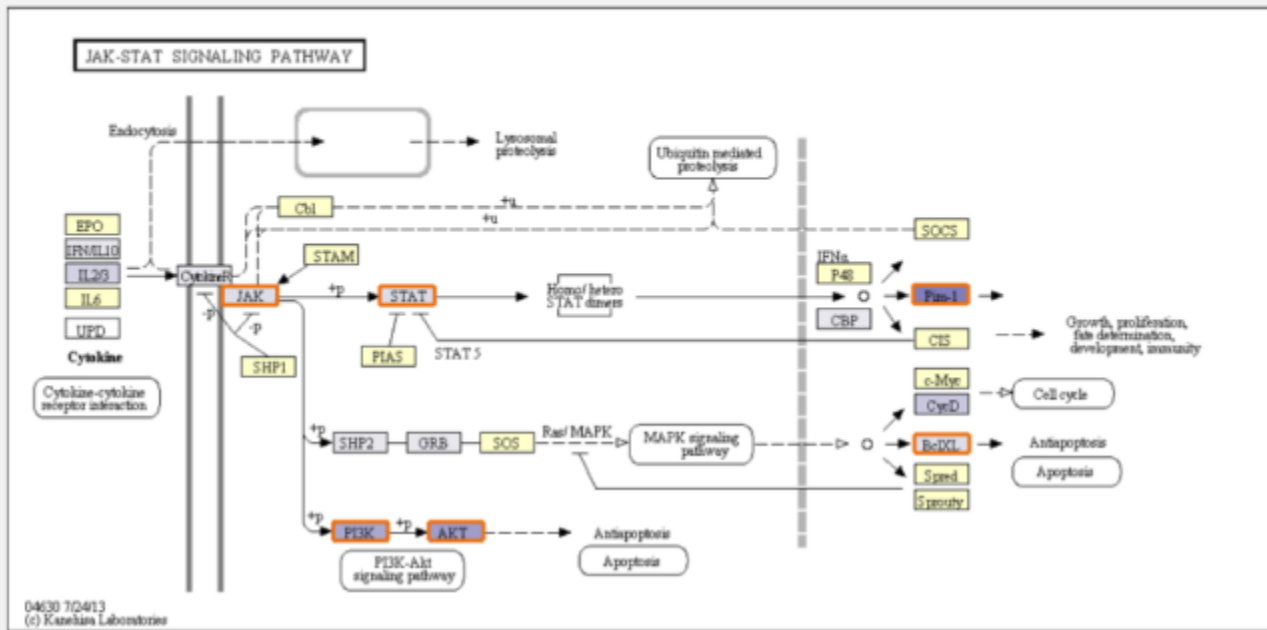
BPZSYCZIITTYBL-YJYMSZO



HJMQDJPMQIHLPB-UHFFFA



Pathways



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

Compounds	Activities	Genes
WINHZLLDWRZW	PI3K, AKT	PIK3CA, PIK3CD, JAK1, PIK3CG
ACPOUJIDANTYH	JAK, AKT	JAK2, AKT1, JAK1, JAK3
HHZIUURLSWUIHR	STAT	STAT5B
IZEKFCXSFNUWA	BCL2	BCL2L1
AQVFETGXIRKVA	PIM	PIM1
SVSJTYGPLVUO	PIM	PIM1
CZQHHVNHHR	AKT	AKT1, AKT2, PIK3CB, AKT3

Clusters	Fingerprints
69	[Fingerprint]
11	[Fingerprint]
21	[Fingerprint]
31	[Fingerprint]
12	[Fingerprint]
38	[Fingerprint]
9	[Fingerprint]
52	[Fingerprint]
61	[Fingerprint]
--	[Fingerprint]