



ConTour:

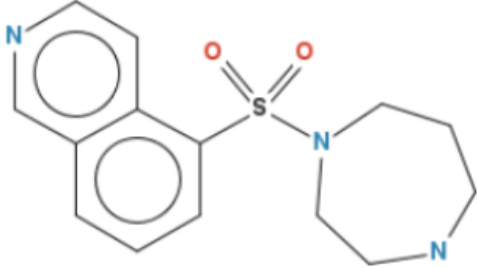
Data-Driven Exploration of Multi-Relational Datasets for Drug Discovery

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

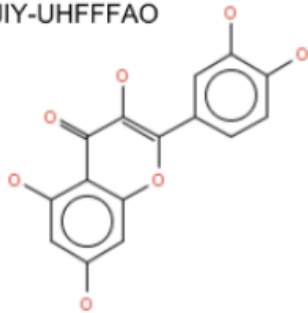
VIS 2014

Why? – Search for “magic bullet”

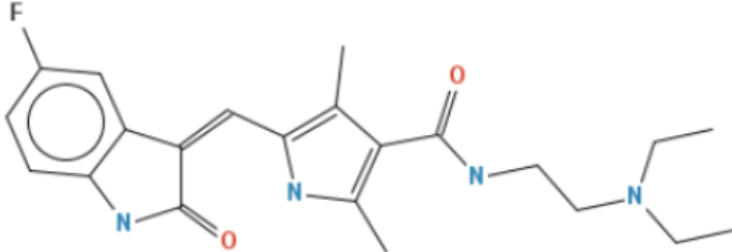
NGOGFTYYXHNFAQH-UHFFF



REFJWTPEDVJJIY-UHFFFAO

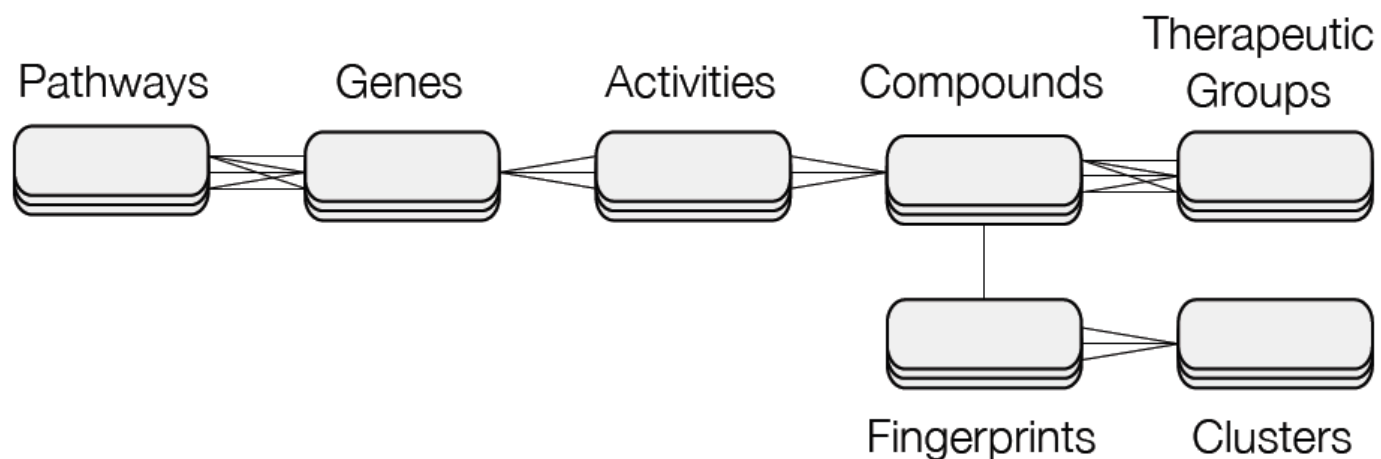


WINHZLLDWRZVRT-ATVHP



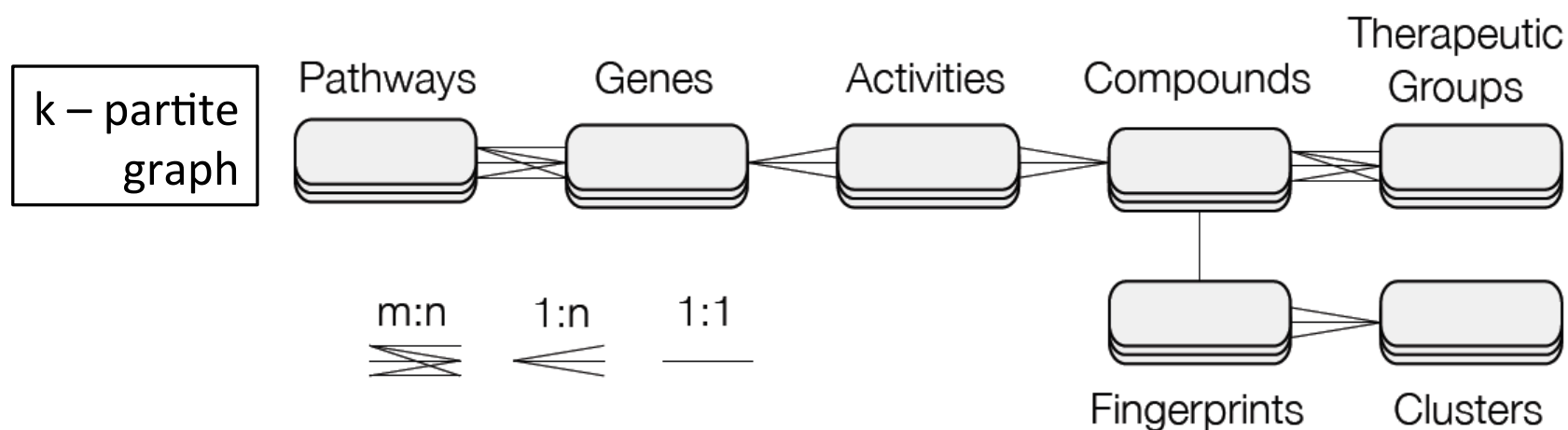
Key word definitions

- Compounds: drugs or drug candidates
- Biological fingerprints: activity of a compound across several experiments of cellular processes
- Pathways: series of action that leads to a change in the cell; regulation of genes and transmission of signals
- Therapeutic groups: treatment that compounds induces



What? – The data set

- Dataset from public bioactivity databases *ChEMBL* [Gaulton et al. *Nucleic acids research* 2012] and *DrugBank* [Law et al. *Nucleic acids research* 2014]
- Drug dataset consists of about 1100 compounds
- Compounds have been profiled in at least 50 different cell-based screens [Petrone et al. *ACS Chemical Biology* 2012]
- Correlation-based similarity measures [Wassermann et al. *J Chem Inf Model* 2013] yielded 100 distinct clusters



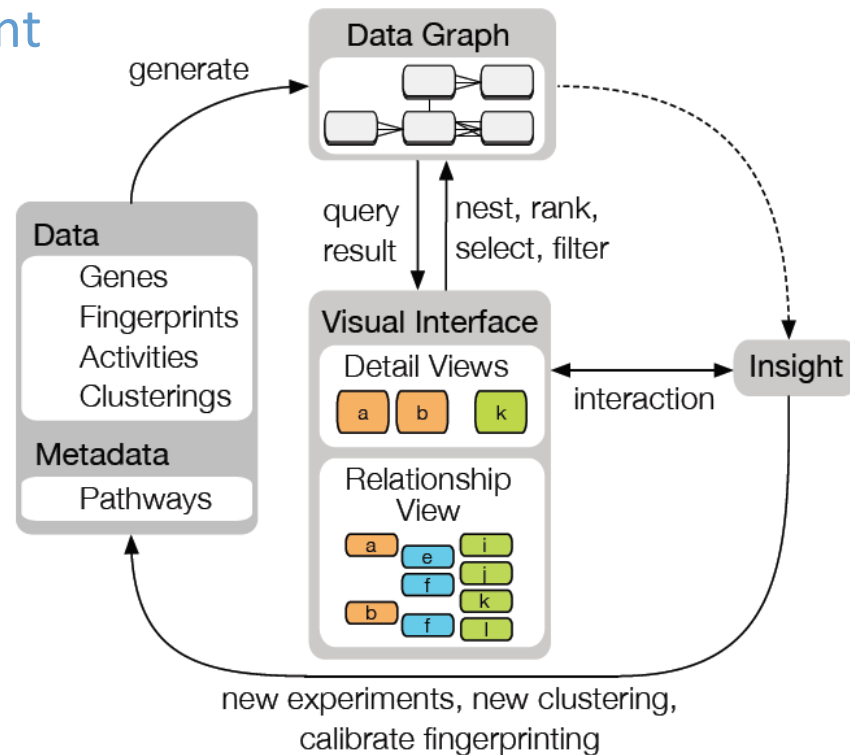
Analysis Goals

1. Identify a drug's mechanism of action:
Drugs in the same cluster are likely to have the same protein target
2. Identify the biological process a drug modulates:
Compounds binding to different target that are clustered together are likely involved in same biological processes
3. Identify new drugs for specific therapeutic indications:
Compounds clustering with drugs for particular therapeutic indication could be a novel candidate for this therapy

Task analysis

Tasks the analyst needs to perform to achieve previous goals:

1. Identify related items
2. Identify items that share a relationship with a set of items
3. Analyze network enrichment
4. Rank items
5. Filter items
6. View items in detail



ConTour for drug discovery

The screenshot displays the ConTour software interface, which is used for drug discovery. The interface is divided into several main sections:

- Therapeutic G**: A list of therapeutic categories with checkboxes, such as Antidepressants, Alzheimer's Dem, and Oncolytic Drugs.
- Compounds**: A list of compounds with checkboxes, including MAOB, MAOA, ALDH1A3, ALDH1B1, ALDH2, and DRD2.
- Genes**: A list of genes with checkboxes, including MAOB, MAOA, ALDH1A3, ALDH1B1, ALDH2, and DRD2.
- Activities**: A list of activities with checkboxes, including KZNIHPLKGY, YHXISWVBD, RUOKEQAAGR, REFJWTPEDVJ, VHBFFQKBGN, RMUCZJUITON, HYAFETHCAU, KTUFNOKKBVM, GUGOEEESWI, GXDALQBWZG, ZPEIMDSQAK, VOVIAXLJUBG, CZQHNVNHHH, CGIGDMFJXJA, JCCNYMKQOS, ZCVMWBYGMW, and PWWVAXIEGO.
- Fingerprints**: A panel showing molecular fingerprints for various clusters, with a list of clusters on the left (11, 24, 54, 46, 98, 84, 12, 30, 9, 20, 85, 69, 37, 44, 23, 89, 8, 51, 13).
- Pathways**: A list of pathways with checkboxes, including Osteoblast S, Folate biosy, and Statin Pathw.
- Compounds**: A list of compounds with checkboxes, including VHRSUDES, AAXVEMM, UREBDLIC, MPDGHEJ, JYGXADM, UREBDLIC, XGALLCVX, VOVIAXLJ, KTUFNOKK, VHOGYUR, OGSPWJR, TZBJGXHY, VGEREEW, PSGAAPLE, RJKFOVLP, VOXZDWN, FNYLWPV, FBOZXECL, MQJKPEG, GECHUMI, WKSUAUQY, FJLGEFLZ, VOXZDWN, RUOKEQA, YHXISWV, RMUCZJUI, VHBFFQKB, KZNIHPL, REFJWTP, and HYAFETH.
- Genes**: A list of genes with checkboxes, including PDGFRA, PDGFRB, ALPL, ALPL, and ALPP.
- History**: A list of history items with checkboxes, including Reset, 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 Canc, and Melanoma.
- Filters**: A list of filters with checkboxes, including Fingerprints, Pathways, Activities, and Therapeutic.

Task validation

1. Identify related items
highlighting (hovering, clicking), selection based filtering, nesting, history view
2. Identify items that share a relationship with a set of items
recursive nesting, history view
3. Analyze network enrichment
enrichment scores
4. Rank items
ranking & sorting
5. Filter items
selection based filtering, filter view
6. View items in detail
pathway, compound & parallel coordinates view

Conclusion



- Strength

- Highly exploratory through ranking, sorting and filtering
- Integrates overview, detailed view and support views
- Simple and recursive nesting illustrates parent-child relationships
- Case study showed that ConTour is an effective tool for interactively exploring relationships in drug discovery
- Applicable to other biological and non-biological domains

- Weakness

- Scaling to higher number of columns difficult due to limited space
- Nesting approach is not very space efficient
- Relationship between items of the sets are of arbitrary cardinality
→ problematic for data graphs containing cycles