

Agile Text Mining for Knowledge Discovery

David Milward

Linguamatics

September 2011

Finding information to make decisions

- A huge and growing amount of information is captured within unstructured text
 - external scientific and patent literature
 - internal reports
- Search technology is designed to find the best set of documents
 - but this leaves the user a lot of work to do, reading through each document
- I2E text mining provides an efficient way to answer complex questions
 - Which companies are working on a particular target?
 - What kinds of chemicals/targets are associated with this disease?
- Used widely in top-50 pharma



... including 9 of the top 10

“Which, What, Who?”

- Go directly to answers, e.g. find all the genes associated with a specific disease
- “Which gene” searches for >10,000s genes and all their synonyms e.g.
 - ERBB2
 - CD340, HER-2, MLN 19, NEU, TKR1, ...
- Provides highlighted evidence and link to the document:
 - Read a sentence or a single document to convince you that the gene is relevant (or not)

In conclusion, targeting **HER2** and uPAR has a synergistic inhibitory effect on breast cancer cells.

Published data on the association between **BRCA2 N372H** polymorphism and breast cancer risk are inconclusive.

Entrez Genes		Doc
ERBB2	▶ 470	15870086
BRCA1	▶ 248	15564800
BRCA2	▶ 160	15986445
PGR	▶ 129	15272277
EGFR	▶ 101	16280056
VEGFA	▶ 76	15897560
TP53	▶ 73	15583825
INS	▶ 68	15805581
CCND1	▶ 51	16140974
AKT1	▶ 50	16619501

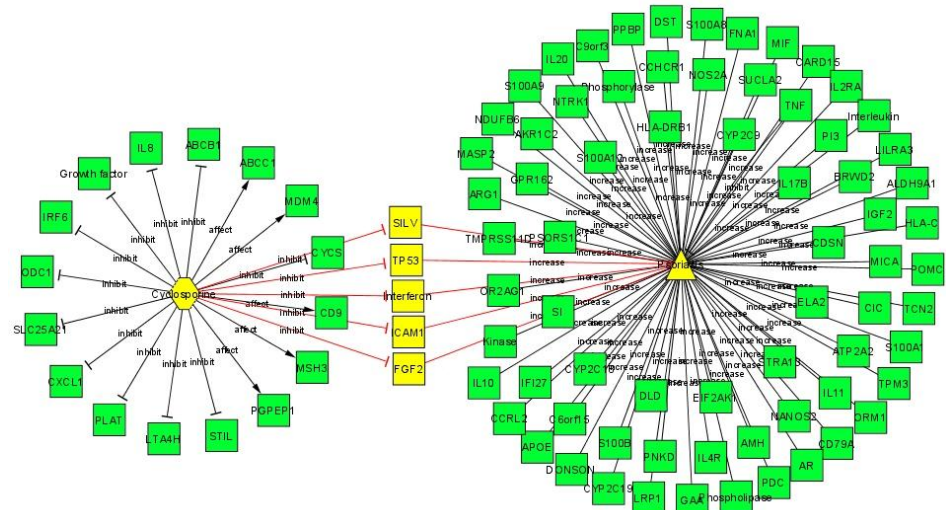
“Tell me everything about ...”

- Search can provide the most relevant documents mentioning e.g. “cyclosporine”
- Text mining can summarize the key properties of cyclosporine by pulling together different facts from different documents

Chemical	Relation	Entity	Doc	Hit
▼ Cyclosporine	▶ treat	Psoriasis	▶ 95 1401280	▶ 3 Elevation of fasting serum lipids in patients treated with low-dose cyclosporine for severe plaque-type psoriasis.
	▶ dosage	5 mg/kg/day	▶ 16 1552052	▶ 3 Alterations in renal function in psoriasis patients treated with cyclosporine, 5 mg/kg/day.
	▶ inhibit	IL8	▶ 2 9588080	1 It was found out that CsA inhibits IL-8 production by stimulated THP-1 monocyte cell line in dose dependence course.
	▶ cause	Lymphoma	1 20941955	1 Cyclosporine has been shown to increase the risk of lymphoma when used in organ transplant patients; however, studies have failed to demonstrate an increased risk of lymphoma when used at lower dermatologic doses for psoriasis.
	▶ increase	AKT1	1 20154081	1 Immunosuppressive cyclosporin A activates AKT in keratinocytes through PTEN suppression: implications in skin carcinogenesis.
	affect	CALM3	1 2277142	1 Cyclosporine binds to calmodulin with low affinity, and such binding of cyclosporine isomers does not reflect their immunosuppressive activity.
	IC50	3.0 ng/ml	1 9585801	1 We classified these patients into two groups on the basis of their PBMC sensitivity to cyclosporine with use of the median cyclosporine IC50 (3.0 ng/ml) of these patients as the cutoff point.

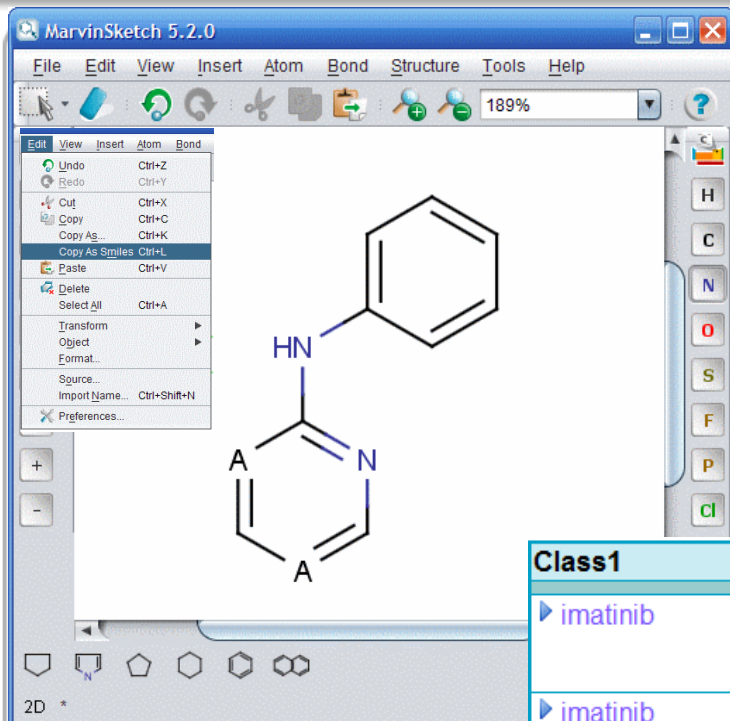
Finding Potential Mechanisms

- Combine information from separate documents
- Discover indirect associations



Cyclosporine	Compound-Gene	Entrez Genes	Gene-Disease	Psoriasis	Doc	Hit	Doc	Hit
Cyclosporine	inhibit	Interferon	affect	Psoriasis	1 1789988	1 In addition, cyclosporin blocked the interferon-gamma-induced increase in epidermal 12(S)-HETE binding.	9 9856816	1 ... helper 1-type cytokines such as interferon-gamma in psoriasis.
Cyclosporine	inhibit	IL8	affect	Psoriasis	2 9588080	1 It was found out that CsA inhibits IL-8 production by stimulated THP-1 monocyte cell ...	8 11378328	2 Interleukin-8-positive neutrophils in psoriasis.
Cyclosporine	affect	CALM3	affect	Psoriasis	1 2277142	1 Cyclosporine binds to calmodulin with low affinity, and ...	7 1879887	1 Epidermal calmodulin levels in psoriasis before & after therapy.

I2E with ChemAxon Substructure/Similarity Search



Class Properties

Chooser Advanced

molecular entity
irral product
phenylpropanoid
stilbenoid
stilbene
4,4'-bis((4-anilino-6-[bis(2-hydroxyethyl)amino]-1,3,5-triazin-2-yl)amino)stilbene-2,2'-disulfonic acid
inorganic entities
organohalogen compound
organofluorine compounds
nilotinib
ecular entities
molecular entities
ride salt
organic chloride salt
hydrochloride
pirenzepine hydrochloride

Look for: N(C1=CC=CC=C1)C1=*C=*C=N1 as: Substructure in: SMILES Search

Class1	Relation	Class2	Doc	Hit
▶ imatinib	▶ inhibit	ABL1	▶ 2 16303243	1 Imatinib , an inhibitor of BCR-ABL tyrosine kinase, also inhibits BCRP-mediated drug transport.
▶ imatinib methanesulfonate	▶ inhibit	ABL1	▶ 3 15803362	1 BACKGROUND: Imatinib mesylate is a potent inhibitor of Abl, KIT, and PDGFR tyrosine kinases.
▶ gefitinib	▶ inhibit	EGFR	▶ 2 15692759	▶ 2 The clinical benefit and safety of the EGFR tyrosine kinase inhibitor gefitinib ('Iressa') ¹ was evaluated in this Phase II, multicentre study of patients with taxane and anthracycline pretreated, metastatic breast cancer.
▶ lapatinib	▶ inhibit	ERBB2	▶ 3 16452223	▶ 2 Alternatively, inhibition of ErbB2 signaling using lapatinib (GW572016), a reversible small-molecule inhibitor of ErbB1/ErbB2 tyrosine kinases, at pharmacologically relevant concentrations, leads to marked inhibition of survivin protein with subsequent apoptosis.

Thank You!

For more information...

Please visit our table or www.linguamatics.com

Webinars:

www.linguamatics.com/welcome/events/webinars.html

Contact: Susan LeBeau

Email: susan.lebeau@linguamatics.com

Phone: 774-571-1117

Mobile: 774-277-1256