

# 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	<a href="#">15870086</a>
BRCA1	▶ 248	<a href="#">15564800</a>
BRCA2	▶ 160	<a href="#">15986445</a>
PGR	▶ 129	<a href="#">15272277</a>
EGFR	▶ 101	<a href="#">16280056</a>
VEGFA	▶ 76	<a href="#">15897560</a>
TP53	▶ 73	<a href="#">15583825</a>
INS	▶ 68	<a href="#">15805581</a>
CCND1	▶ 51	<a href="#">16140974</a>
AKT1	▶ 50	<a href="#">16619501</a>

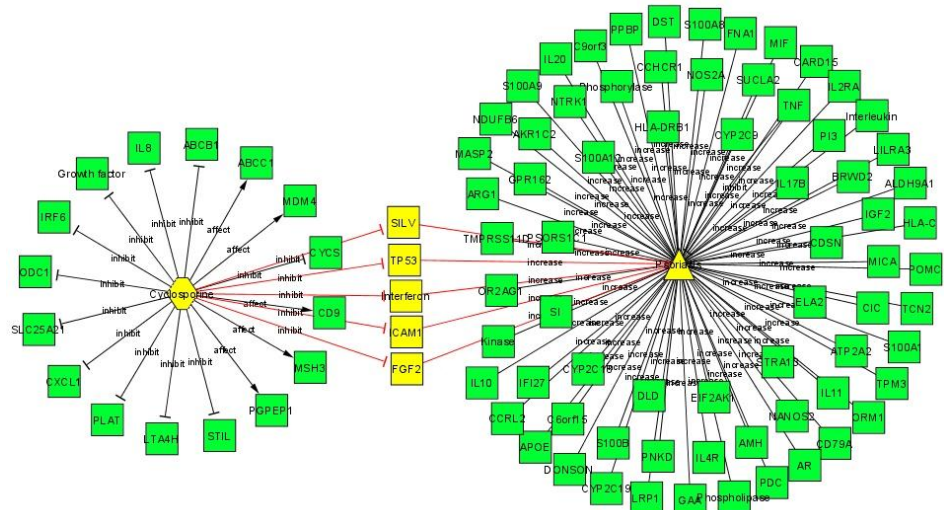
# “Tell me everything about ...”

- 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 <a href="#">1401280</a>	▶ 3 Elevation of fasting serum lipids in patients treated with low-dose cyclosporine for severe plaque-type psoriasis.
	▶ dosage	5 mg/kg/day	▶ 16 <a href="#">1552052</a>	▶ 3 Alterations in renal function in psoriasis patients treated with cyclosporine, 5 mg/kg/day.
	▶ inhibit	IL8	▶ 2 <a href="#">9588080</a>	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 <a href="#">20941955</a>	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 <a href="#">20154081</a>	1 Immunosuppressive cyclosporin A activates AKT in keratinocytes through PTEN suppression: implications in skin carcinogenesis.
	affect	CALM3	1 <a href="#">2277142</a>	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 <a href="#">9585801</a>	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.

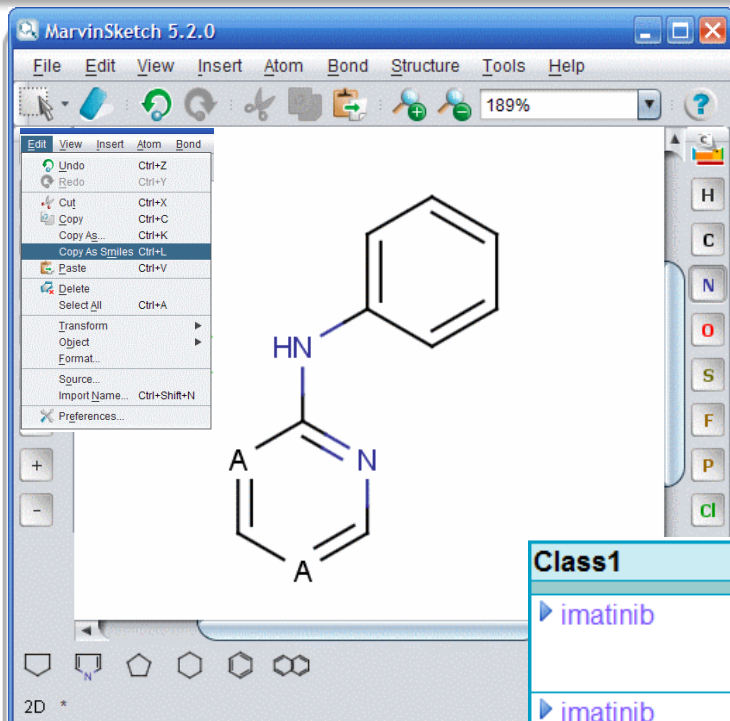
# Finding Potential Mechanisms

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

# I2E with ChemAxon Substructure/Similarity Search



Class1	Relation	Class2	Doc	Hit
▶ imatinib	▶ inhibit	ABL1	▶ 2 <a href="#">16303243</a>	1 <b>Imatinib</b> , an inhibitor of BCR-ABL tyrosine kinase, also inhibits BCRP-mediated drug transport.
▶ imatinib methanesulfonate	▶ inhibit	ABL1	▶ 3 <a href="#">15803362</a>	1 BACKGROUND: <b>Imatinib mesylate</b> is a potent inhibitor of Abl, KIT, and PDGFR tyrosine kinases.
▶ gefitinib	▶ inhibit	EGFR	▶ 2 <a href="#">15692759</a>	▶ 2 The clinical benefit and safety of the EGFR tyrosine kinase inhibitor <b>gefitinib</b> ('Iressa') <sup>1</sup> was evaluated in this Phase II, multicentre study of patients with taxane and anthracycline pretreated, metastatic breast cancer.
▶ lapatinib	▶ inhibit	ERBB2	▶ 3 <a href="#">16452223</a>	▶ 2 Alternatively, inhibition of ErbB2 signaling using <b>lapatinib</b> (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](http://www.linguamatics.com)

## **Webinars:**

[www.linguamatics.com/welcome/events/webinars.html](http://www.linguamatics.com/welcome/events/webinars.html)

**Contact:** Susan LeBeau

Email: [susan.lebeau@linguamatics.com](mailto:susan.lebeau@linguamatics.com)

Phone: 774-571-1117

Mobile: 774-277-1256