

Handling Complexity when Simulating Biological Networks

Jürgen Klenk
CEO & CTO

18 March 2004

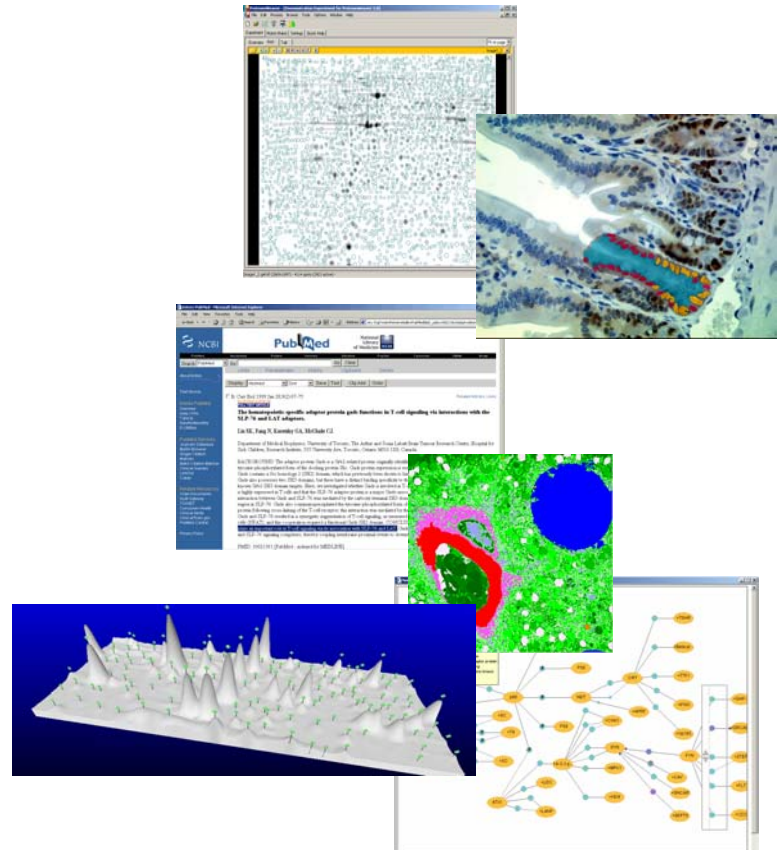
Who we are

- Founded in 1994 by Physics Nobel Laureate Prof. Dr. Gerd Binnig
- 1994 – 1999: Government-funded Think Tank
 - Developed „Cognition Network Technology“
 - First prototype for automated image analysis
- Since 2000: Commercialization of technology
 - June 2000: eCognition (Geo)
 - March 2002: Proteomweaver (LSI)
 - March 2003: Cellenger (LSI)
 - October 2003: Polymind (LSI)
- Ongoing research
 - Knowledge integration
 - Simulation



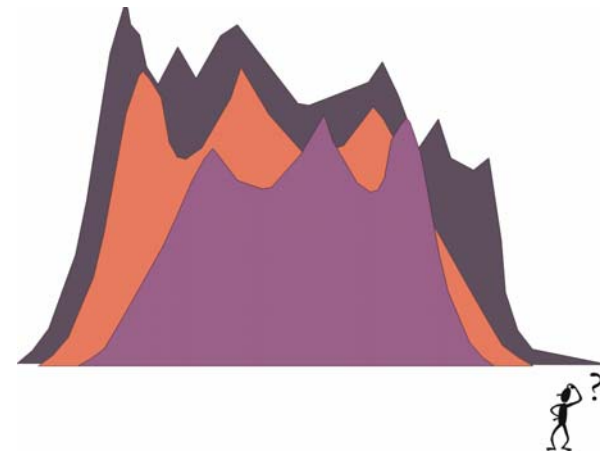
What we do

- Definiens develops and markets software for automated pattern recognition.
- Definiens software automatically finds patterns in images and texts which no other software can find.
- These patterns may be spots in gels, objects of interest in images or facts in text documents.



Cognition Machines or what?

- ❑ New technologies produce increasing amounts of data
- ❑ Key knowledge is hidden in in unstructured formats
- ❑ Images and texts are predominant sources
- ❑ Human-like cognitive skills are needed to unravel this knowledge
- ❑ Furthermore, this knowledge needs to be personalized
- ❑ **Knowledge at your fingertips**

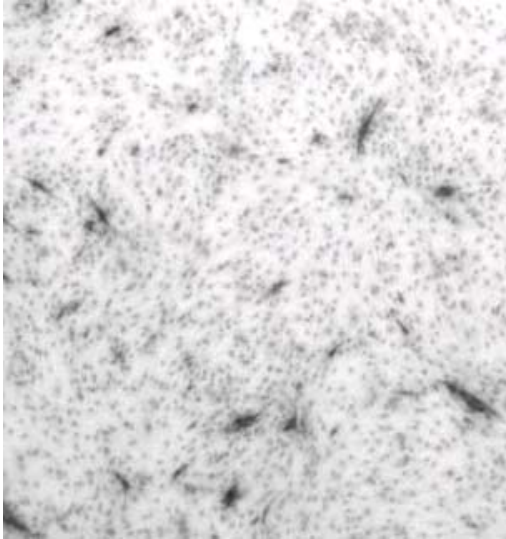


Data – Data – Data – Information – Knowledge

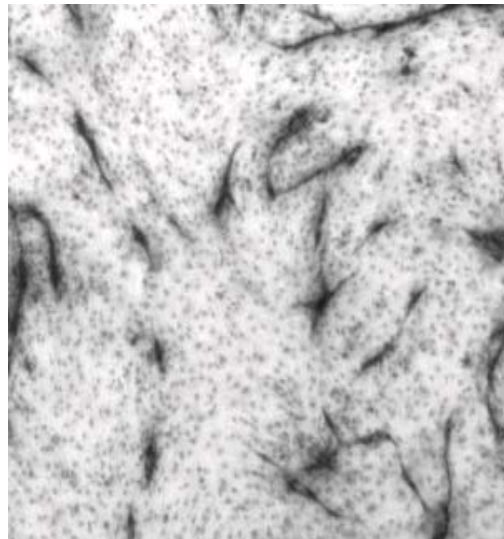
This is what
Cognition Machines
can do

Example 1: Inhibition of Vascular cell growth

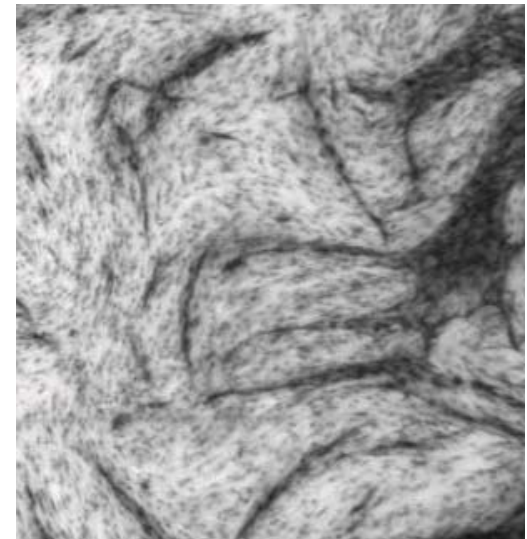
Original Images



Control



Compound 1



Compound 2

In-vitro Angiogenesis:

Multiparameter Analysis

- Sprout area/length/width
- Relational sprout and nuclei count
- Sprout branching factor

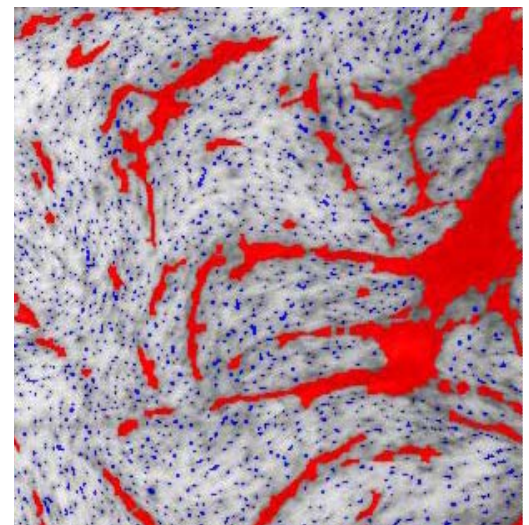
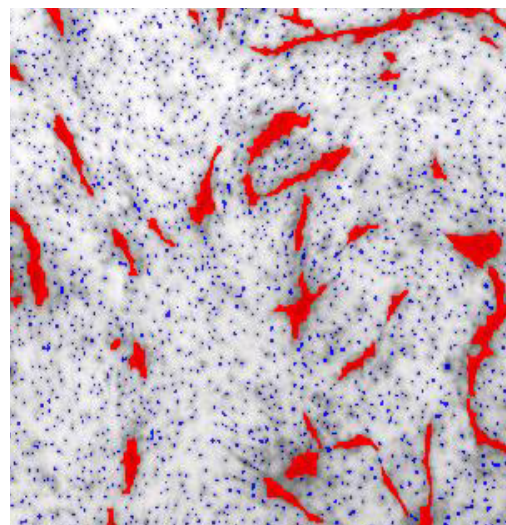
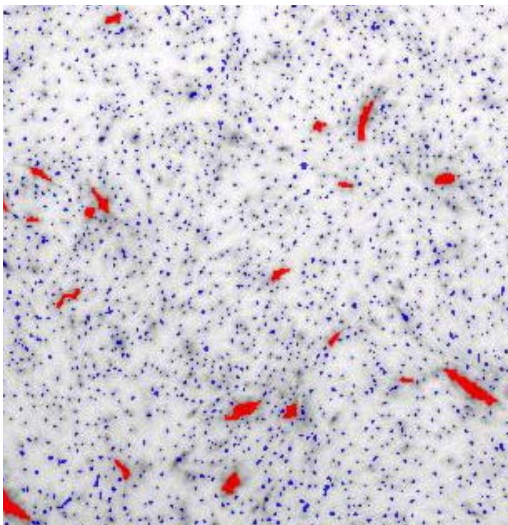
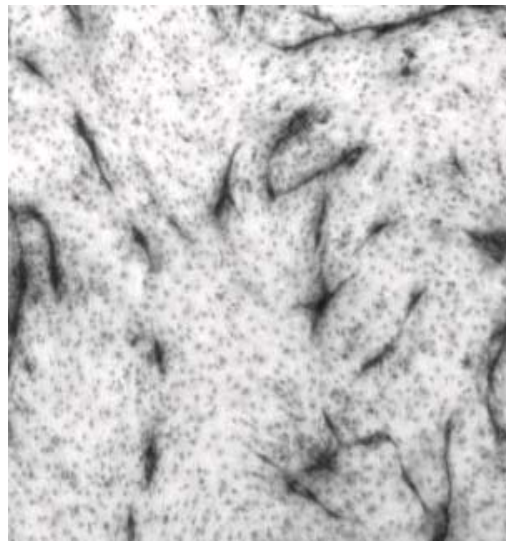
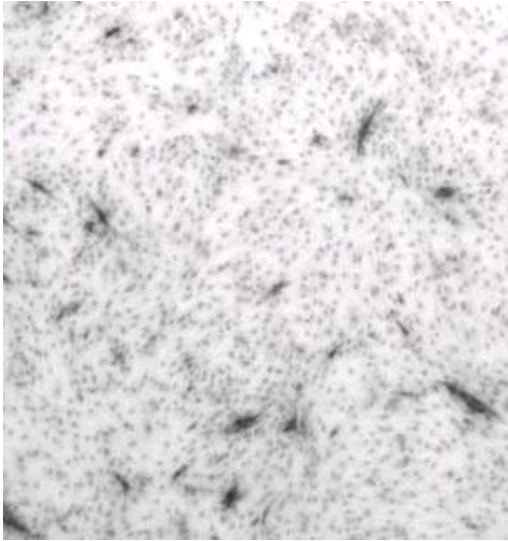
Tasks:

- Combat Diversity
- Understand Morphology
- Find sprouts and nuclei!

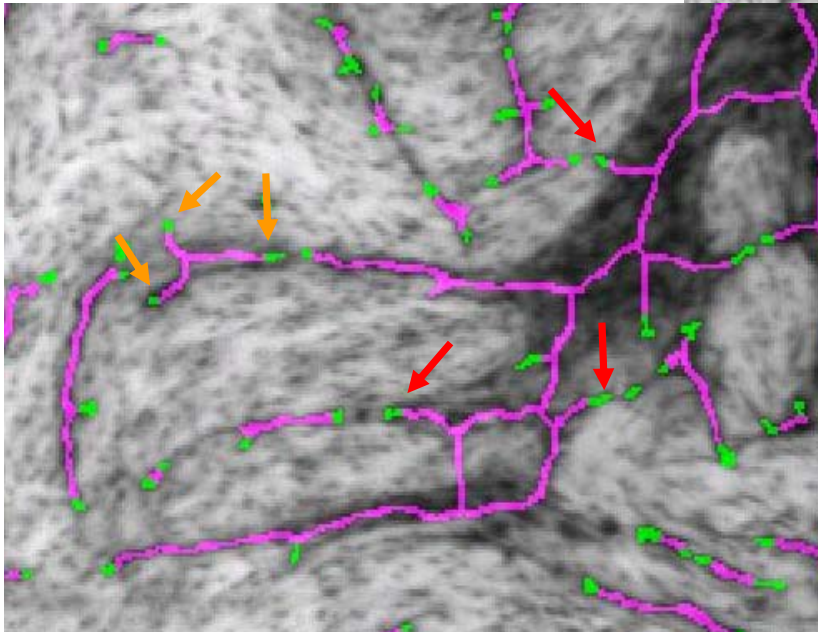
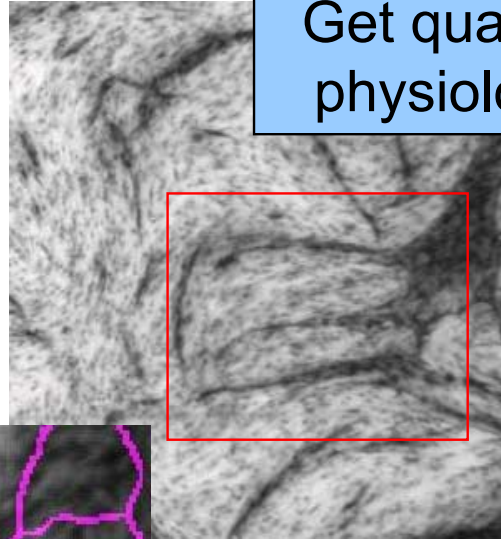
Image variety with inconsistent contrast and threshold!

Image data courtesy Novartis Pharma AG

Cellenger Result: Sprouts and Nuclei



Get quantified answers to physiological screening !

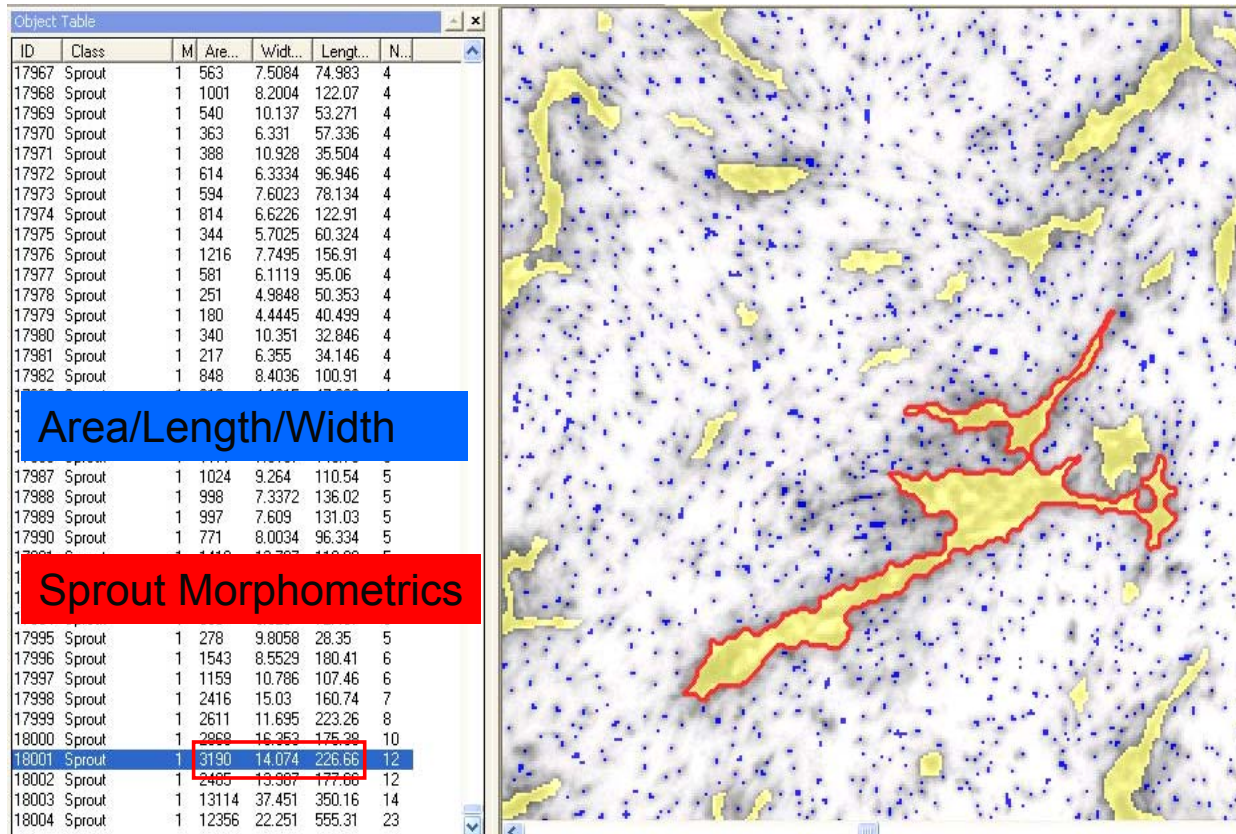


In-vitro Angiogenesis:

Multiparameter Analysis

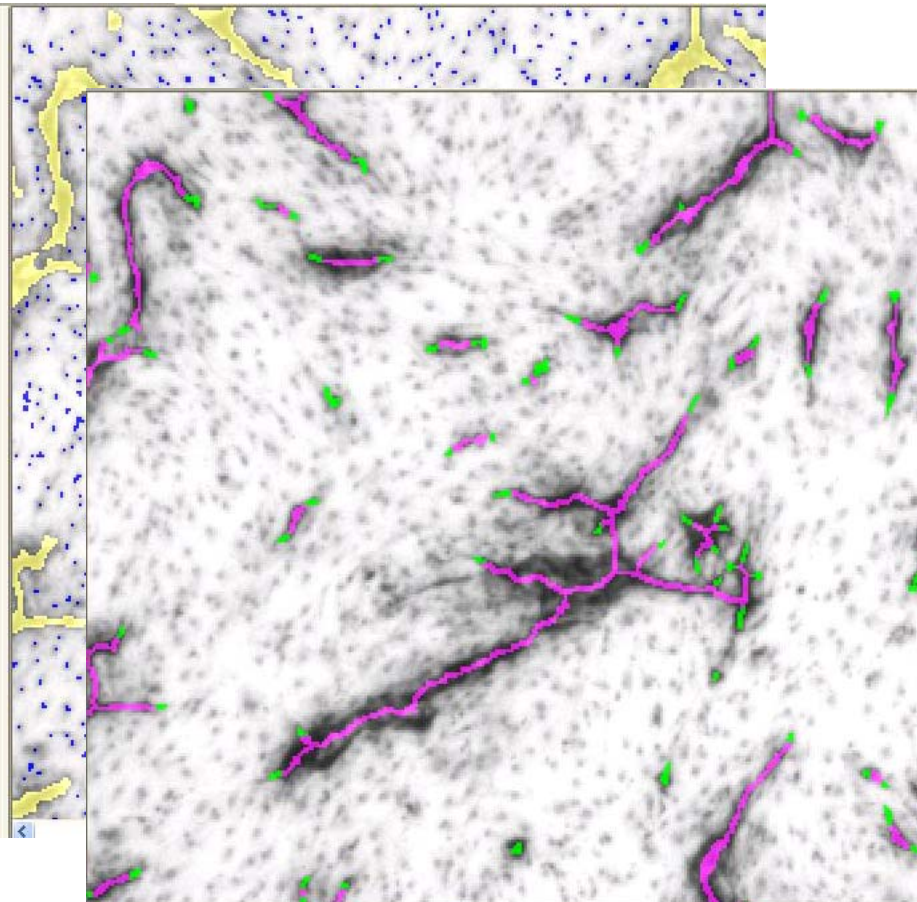
- Sprout area/length/width ✓
- Relational sprout and nuclei count ✓
- Sprout branching factor ✓

Cellenger Result: Reports & Statistics



Cellenger Result: Reports & Statistics

Object Table						
ID	Class	M	Are...	Widt...	Lengt...	N...
17967	Sprout	1	563	7.5084	74.983	4
17968	Sprout	1	1001	8.2004	122.07	4
17969	Sprout	1	540	10.137	53.271	4
17970	Sprout	1	363	6.331	57.336	4
17971	Sprout	1	388	10.328	35.504	4
17972	Sprout	1	614	6.3334	96.946	4
17973	Sprout	1	594	7.6023	78.134	4
17974	Sprout	1	814	6.6226	122.91	4
17975	Sprout	1	344	5.7025	60.324	4
17976	Sprout	1	1216	7.7495	156.91	4
17977	Sprout	1	581	6.1119	95.06	4
17978	Sprout	1	251	4.9848	50.353	4
17979	Sprout	1	180	4.4445	40.499	4
17980	Sprout	1	340	10.351	32.846	4
17981	Sprout	1	217	6.355	34.146	4
17982	Sprout	1	848	8.4036	100.91	4
Area/Length/Width						
Number of endpoints						
Sprout Morphometrics						
Sprout Branching						
17999	Sprout	1	2611	11.695	223.26	8
18000	Sprout	1	2868	16.353	175.38	10
18001	Sprout	1	3190	14.074	226.66	12
18002	Sprout	1	2485	13.587	177.66	12
18003	Sprout	1	13114	37.451	350.16	14
18004	Sprout	1	12356	22.251	555.31	23



Cellenger Result: Reports & Statistics

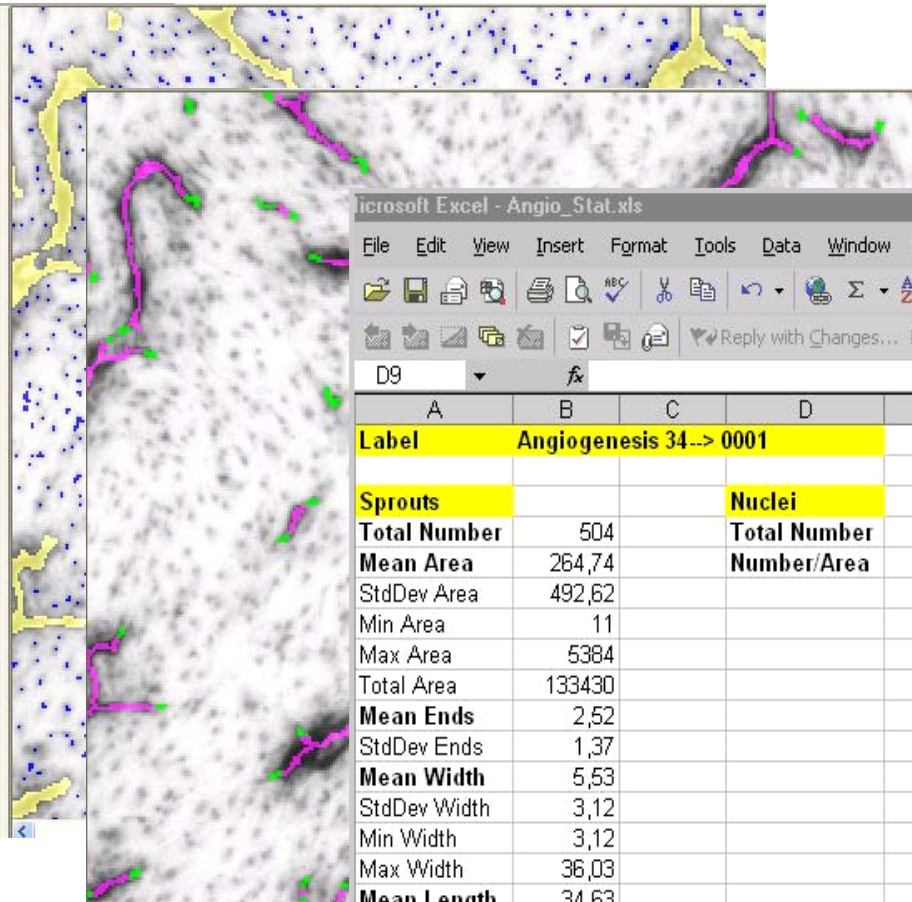
Object Table						
ID	Class	M	Are...	Widt...	Leugt...	N...
17967	Sprout	1	563	7.5084	74.983	4
17968	Sprout	1	1001	8.2004	122.07	4
17969	Sprout	1	540	10.137	53.271	4
17970	Sprout	1	363	6.331	57.336	4
17971	Sprout	1	388	10.928	35.504	4
17972	Sprout	1	614	6.3334	96.946	4
17973	Sprout	1	594	7.6023	78.134	4
17974	Sprout	1	814	6.6226	122.91	4
17975	Sprout	1	344	5.7025	60.324	4
17976	Sprout	1	1216	7.7495	156.91	4
17977	Sprout	1	581	6.1119	95.06	4
17978	Sprout	1	251	4.9848	50.353	4
17979	Sprout	1	180	4.4445	40.499	4
17980	Sprout	1	340	10.351	32.846	4
17981	Sprout	1	217	6.355	34.146	4
17982	Sprout	1	848	8.4036	100.91	4
17999	Sprout	1	2611	11.695	223.26	8
18000	Sprout	1	2868	16.353	175.28	10
18001	Sprout	1	3190	14.074	226.66	12
18002	Sprout	1	2485	13.367	177.66	12
18003	Sprout	1	13114	37.451	350.16	14
18004	Sprout	1	12356	22.251	555.31	23

Area/Length/Width

Number of endpoints

Sprout Morphometrics

Sprout Branching



Microsoft Excel - Angio_Stat.xls					
File Edit View Insert Format Tools Data Window Help					
D9					
A	B	C	D	E	F
Label Angiogenesis 34--> 0001					
Sprouts		Nuclei			
Total Number	504	Total Number	17015		
Mean Area	264,74	Number/Area	0,0146		
StdDev Area	492,62				
Min Area	11				
Max Area	5384				
Total Area	133430				
Mean Ends	2,52				
StdDev Ends	1,37				
Mean Width	5,53				
StdDev Width	3,12				
Min Width	3,12				
Max Width	36,03				
Mean Length	34,63				
StdDev Length	35,74				
Min Length	1,87				
Max Length	245,68				

Cellenger Result: Reports & Statistics

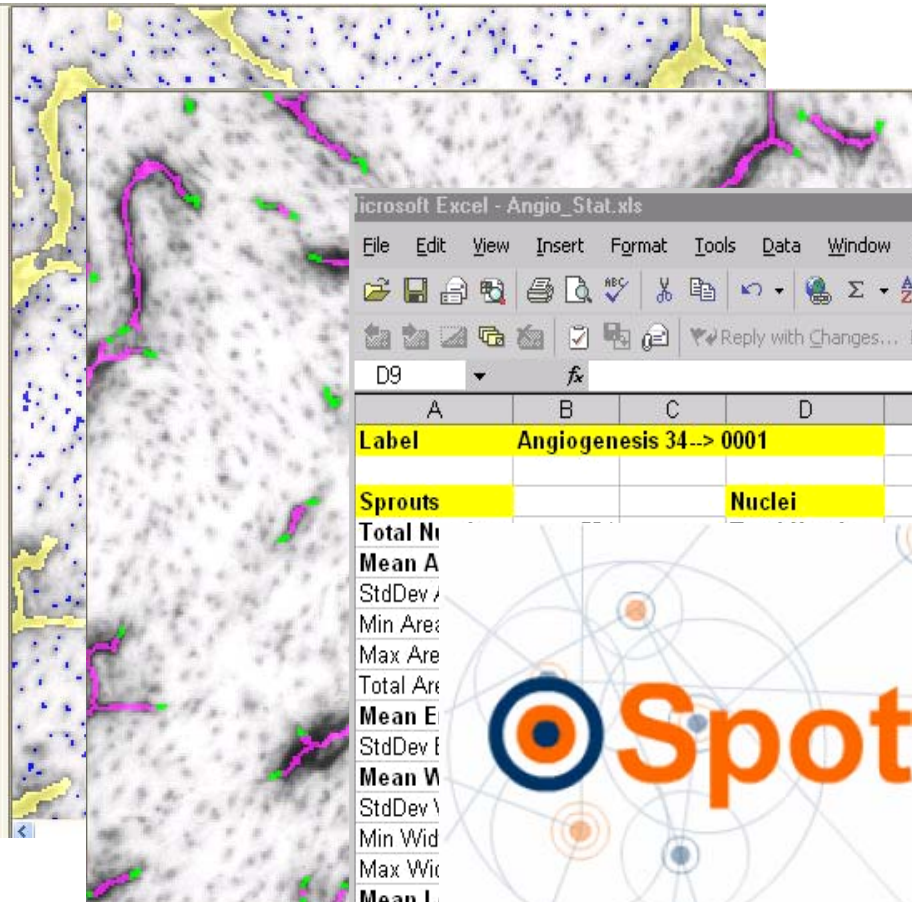
ID	Class	M	Are...	Widt...	Length...	N...
17967	Sprout	1	563	7.5084	74.983	4
17968	Sprout	1	1001	8.2004	122.07	4
17969	Sprout	1	540	10.137	53.271	4
17970	Sprout	1	363	6.331	57.336	4
17971	Sprout	1	388	10.328	35.504	4
17972	Sprout	1	614	6.3334	96.946	4
17973	Sprout	1	594	7.6023	78.134	4
17974	Sprout	1	814	6.6226	122.91	4
17975	Sprout	1	344	5.7025	60.324	4
17976	Sprout	1	1216	7.7495	156.91	4
17977	Sprout	1	581	6.1119	95.06	4
17978	Sprout	1	251	4.9848	50.353	4
17979	Sprout	1	180	4.4445	40.499	4
17980	Sprout	1	340	10.351	32.846	4
17981	Sprout	1	217	6.355	34.146	4
17982	Sprout	1	848	8.4036	100.91	4
17983	Sprout	1	553	10.137	53.271	4
17984	Sprout	1	1001	8.2004	122.07	4
17985	Sprout	1	540	10.137	53.271	4
17986	Sprout	1	363	6.331	57.336	4
17987	Sprout	1	388	10.328	35.504	4
17988	Sprout	1	614	6.3334	96.946	4
17989	Sprout	1	594	7.6023	78.134	4
17990	Sprout	1	814	6.6226	122.91	4
17991	Sprout	1	344	5.7025	60.324	4
17992	Sprout	1	1216	7.7495	156.91	4
17993	Sprout	1	581	6.1119	95.06	4
17994	Sprout	1	251	4.9848	50.353	4
17995	Sprout	1	180	4.4445	40.499	4
17996	Sprout	1	340	10.351	32.846	4
17997	Sprout	1	217	6.355	34.146	4
17998	Sprout	1	848	8.4036	100.91	4
17999	Sprout	1	2611	11.695	223.26	8
18000	Sprout	1	2868	16.353	175.38	10
18001	Sprout	1	3190	14.074	226.66	12
18002	Sprout	1	2485	13.367	177.66	12
18003	Sprout	1	13114	37.451	350.16	14
18004	Sprout	1	12356	22.251	555.31	23

Area/Length/Width

Number of endpoints

Sprout Morphometrics

Sprout Branching



Microsoft Excel - Angio_Stat.xls						
File Edit View Insert Format Tools Data Window Help						
D9						
A	B	C	D	E	F	
Label	Angiogenesis 34--> 0001					
Sprouts			Nuclei			
Total N						
Mean A						
StdDev /						
Min Area						
Max Area						
Total Area						
Mean E						
StdDev E						
Mean V						
StdDev V						
Min Width						
Max Width						
Mean L						
StdDev L						
Min Length						
Max Length	245,68					



Example 2: Proliferation Index in Jejunum

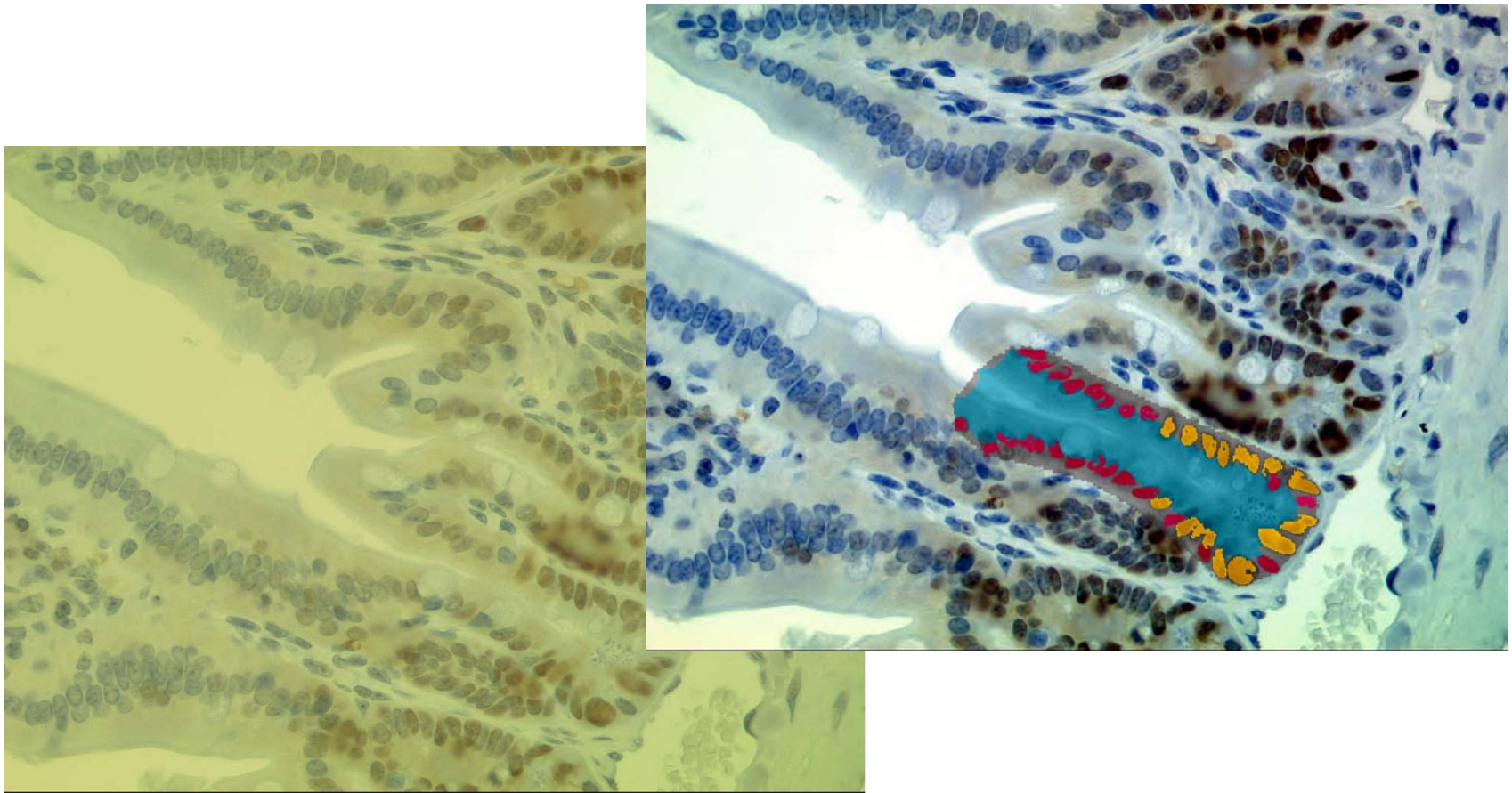
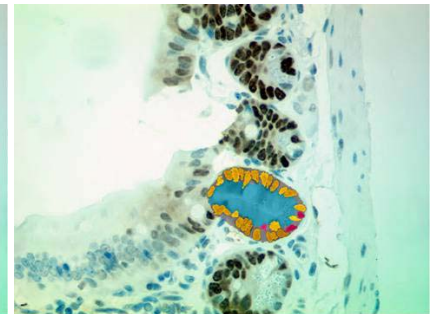
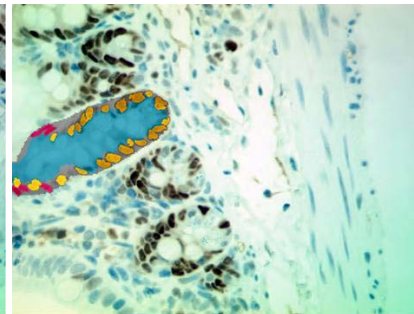
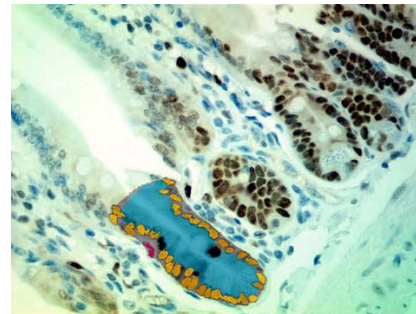
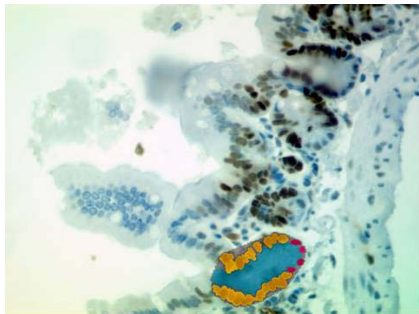
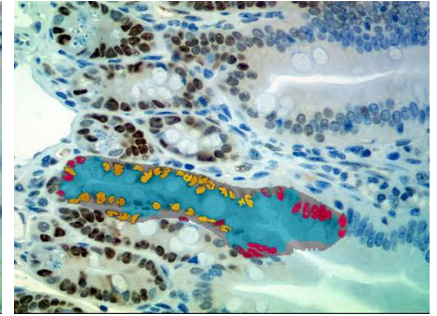
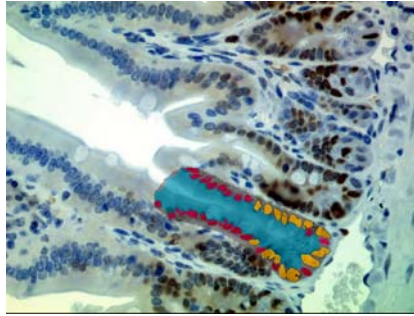
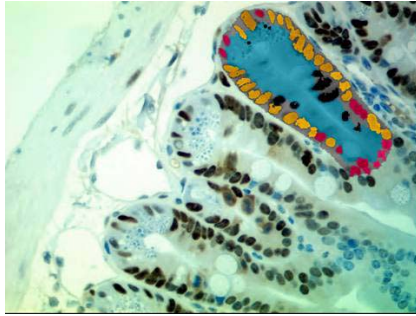
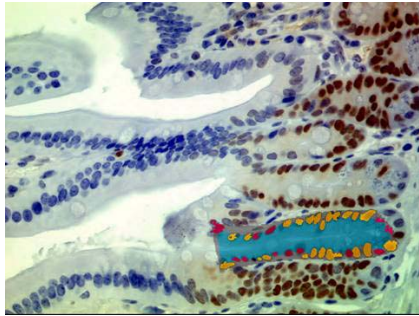
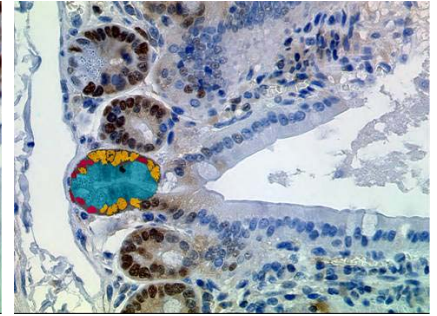
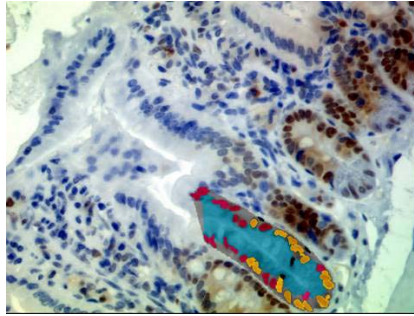
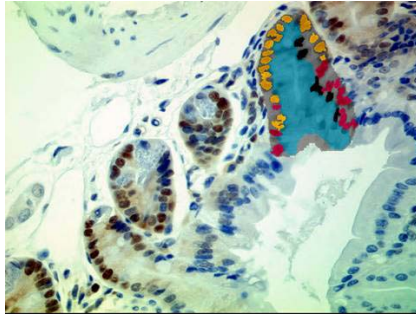
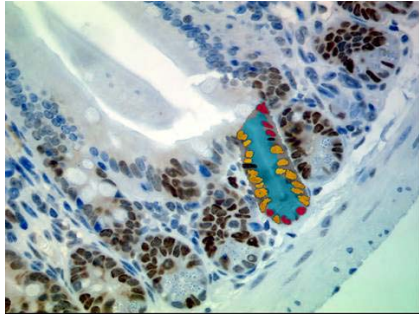


Image data courtesy Novartis Pharma AG

Pathology/Toxicology: Proliferation Index in Jejunum



Example 3: Knowledge Extraction from Literature

Entrez-PubMed - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites History Address <http://www.ncbi.nih.gov/entrez/query.fcgi> Go

NCBI PubMed National Library of Medicine NLM

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books

Search PubMed for Go Clear

Limits Preview/Index History Clipboard Details

About Entrez

Text Version

Entrez PubMed
Overview
Help | FAQ
Tutorial
New/Noteworthy
E-Utilities

PubMed Services
Journals Database
MeSH Browser
Single Citation
Matcher
Batch Citation Matcher
Clinical Queries
LinkOut
Caddy

Related Resources
Order Documents
NLM Gateway
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central

Privacy Policy

Enter one or more search terms, or click [Preview/Index](#) for advanced searching.

Enter [author names](#) as smith jc. Initials are optional.

Enter [journal titles](#) in full or as MEDLINE abbreviations. Use the [Journals Database](#) to find journal titles.

PubMed, a service of the National Library of Medicine, provides access to over 12 million MEDLINE citations back to the mid-1960's and additional life science journals. PubMed includes links to many sites providing full text articles and other related resources.

Bookshelf Additions

Developmental Biology, 6th ed.
by SF Gilbert

Surgical Treatments - Evidence Based and Problem-Oriented
edited by RA Holzheimer & JA Mannick

and NCBI's **Genes and Disease**.

These three resources are now available for interactive searches on Bookshelf.

New Journals Database

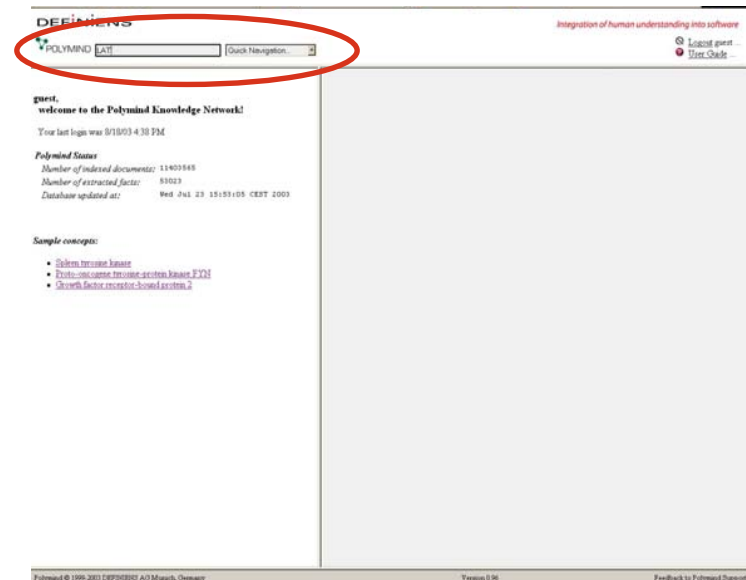
NCBI has created a new Entrez database, [Journals](#), which replaces the Journal Browser.

Journals provides additional search and display features, and includes all the journals in the Entrez databases such as PubMed, Nucleotide, and Protein.

[Write to the Help Desk](#)

Polymind PPI provides the same familiar search box...

- ❑ Automatically extracted database of protein-protein interactions
- ❑ Derived from over 12 Mio Medline abstracts
- ❑ Contains 50,000+ interactions
- ❑ Textual and graphical representation of knowledge



...and returns detailed information about the term...

The screenshot displays the DEFINIENS web interface for the term 'Son of sevenless protein'. The interface includes a search bar with 'sosl' entered, a 'Quick Navigation...' dropdown, and a 'customize GUI by setting bookmarks' link. The main content area shows the term name, a list of synonyms ('SOS'), superclasses ('DNA binding; guanyl-nucleotide exchange factor'), and a link to the SwissProt entry ('P26675'). Below this, there is an 'extensive description' of the protein's function and subunit. At the bottom, there are links to '46 Polymind-extracted Facts', 'Page [1]', and 'Show network'. Callout boxes highlight various features: 'complete name of the term' points to the term name; 'full list of known synonyms' points to the 'SOS' synonym; 'extensive description' points to the protein function text; 'access to tabular view of interactions' points to the '46 Polymind-extracted Facts' link; 'access to network view of interactions' points to the 'Show network' link; 'superclasses of the term within the ontology' points to the 'DNA binding; guanyl-nucleotide exchange factor' text; and 'link to source entry (SwissProt, NCBI Taxonomy)' points to the 'P26675' link.

complete name of the term

DEFINIENS

LYMIND Quick Navigation...

customize GUI by setting bookmarks

Son of sevenless protein [Add to / remove from favorites...](#)

Synonyms: [SOS](#)

Superclasses: [DNA binding; guanyl-nucleotide exchange factor](#)

[swissprot](#) Swissprot-ID: [P26675](#) VID=27917294

Subunit: MAY FORM A COMPLEX WITH SEVENLESS AND DRK.

Function: PROMOTES THE EXCHANGE OF RAS AND GDP BY GTP. SOS IS IMPLICATED IN NEURONAL DEVELOPMENT. CONTAINS 1 RAS-GEF DOMAIN

46 Polymind-extracted Facts

Page [1]

[Show network](#)

superclasses of the term within the ontology

link to source entry (SwissProt, NCBI Taxonomy)

access to tabular view of interactions

access to network view of interactions

...and a comprehensive list of interactions...

details about each interaction

complete list of interactions

sort list by *relevance*, *context*, *type of interaction* or *interaction partner*

link to source document at PubMed

full source reference(s), including the fact-bearing sentence

look for type of interaction, modality, role of protein, relevance, context

Polymind-extracted fact information

Extracted Fact Knowledge for Spleen tyrosine kinase

Relevant Documents

[1] The SH2 domains of Src family kinases associate with Syk.
Aoki Y, Kim YT, Sullivan R, Kim TJ, Pillai S
1995 Jux,270(26):15658-63
Syk binds directly to the SH2 domain of Btk and complexes in vivo with Lym and Btk in activated B cells.

binding (1 documents)

binder: [Spleen tyrosine kinase](#)
binder: [B lymphocyte kinase](#)
context: [B-lymphocyte](#)

binding (2 documents)

binder: [Erythropoietin receptor precursor](#)
binder: [Spleen tyrosine kinase](#)

binding (1 documents)

binder: [Spleen tyrosine kinase](#)
binder: [B lymphocyte kinase](#)
context: [B-lymphocyte](#)

ambiguous possible phosphorylation (1 documents)

modified chemical: [Squalene-hopene cyclase](#)
modifying chemical: [Spleen tyrosine kinase](#)
context: [COS cell phosphorylation; B-lymphocyte](#)

ambiguous possible phosphorylation (1 documents)

modified chemical: [SHC transforming protein](#)
modifying chemical: [Spleen tyrosine kinase](#)
context: [COS cell phosphorylation; B-lymphocyte](#)

ambiguous possible phosphorylation (1 documents)

modified chemical: [Squalene-hopene cyclase](#)
modifying chemical: [Spleen tyrosine kinase](#)
context: [phosphorylation; B-lymphocyte](#)

ambiguous binding (1 documents)

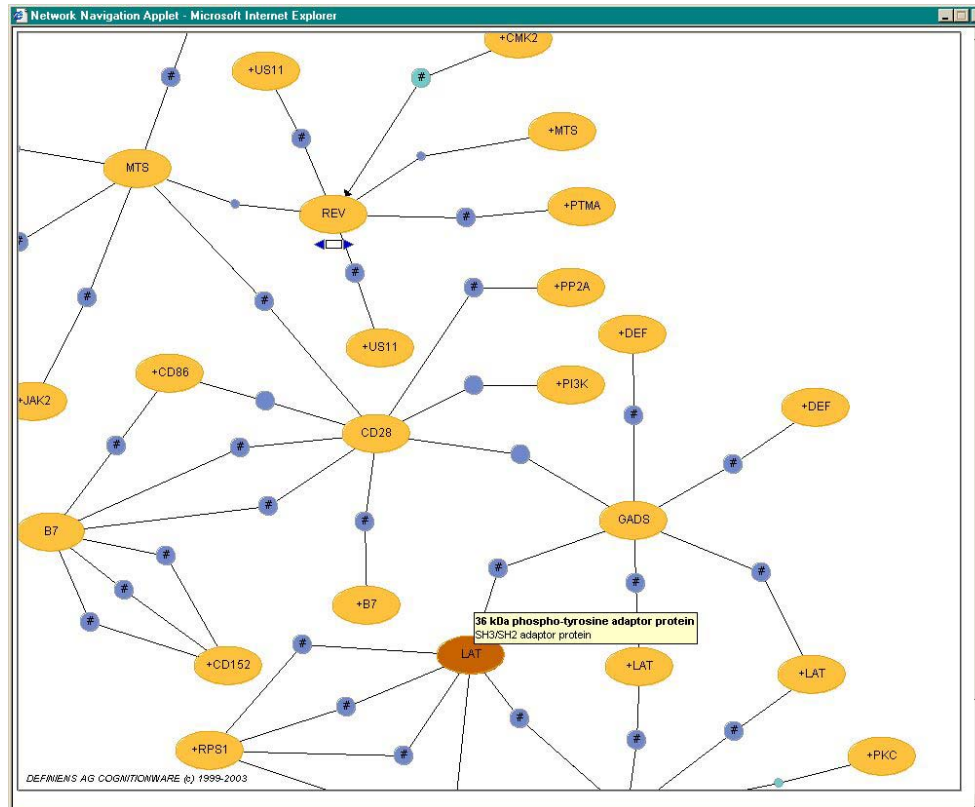
binder: [Spleen tyrosine kinase](#)
binder: [Squalene-hopene cyclase](#)
context: [phosphorylation; B-lymphocyte](#)

ambiguous binding (1 documents)

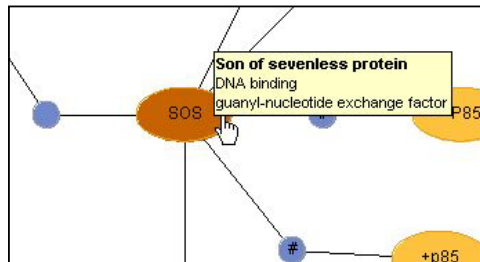
binder: [Spleen tyrosine kinase](#)

Version 0.96

...which can be displayed as a navigable network.

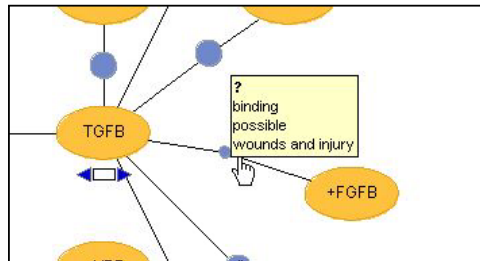


Network elements provide useful information



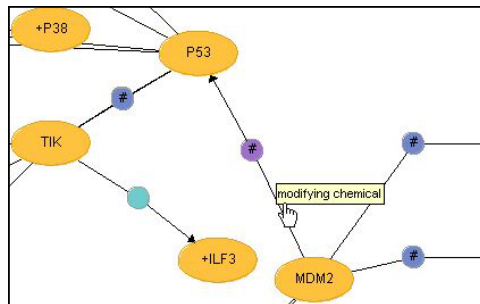
□ Concept nodes

- Full name of the selected protein node
- Names of all superclasses of the protein



□ Interaction nodes

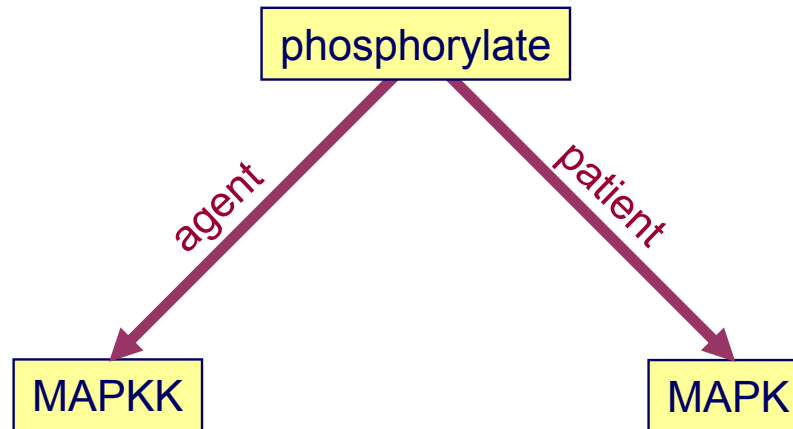
- Type of interaction (different colors for binding, phosphorylation, acetylation etc.)
- Modality of the interaction (positiv-negative, definitive-possible); different size/border
- Contextual information



□ Node connectors

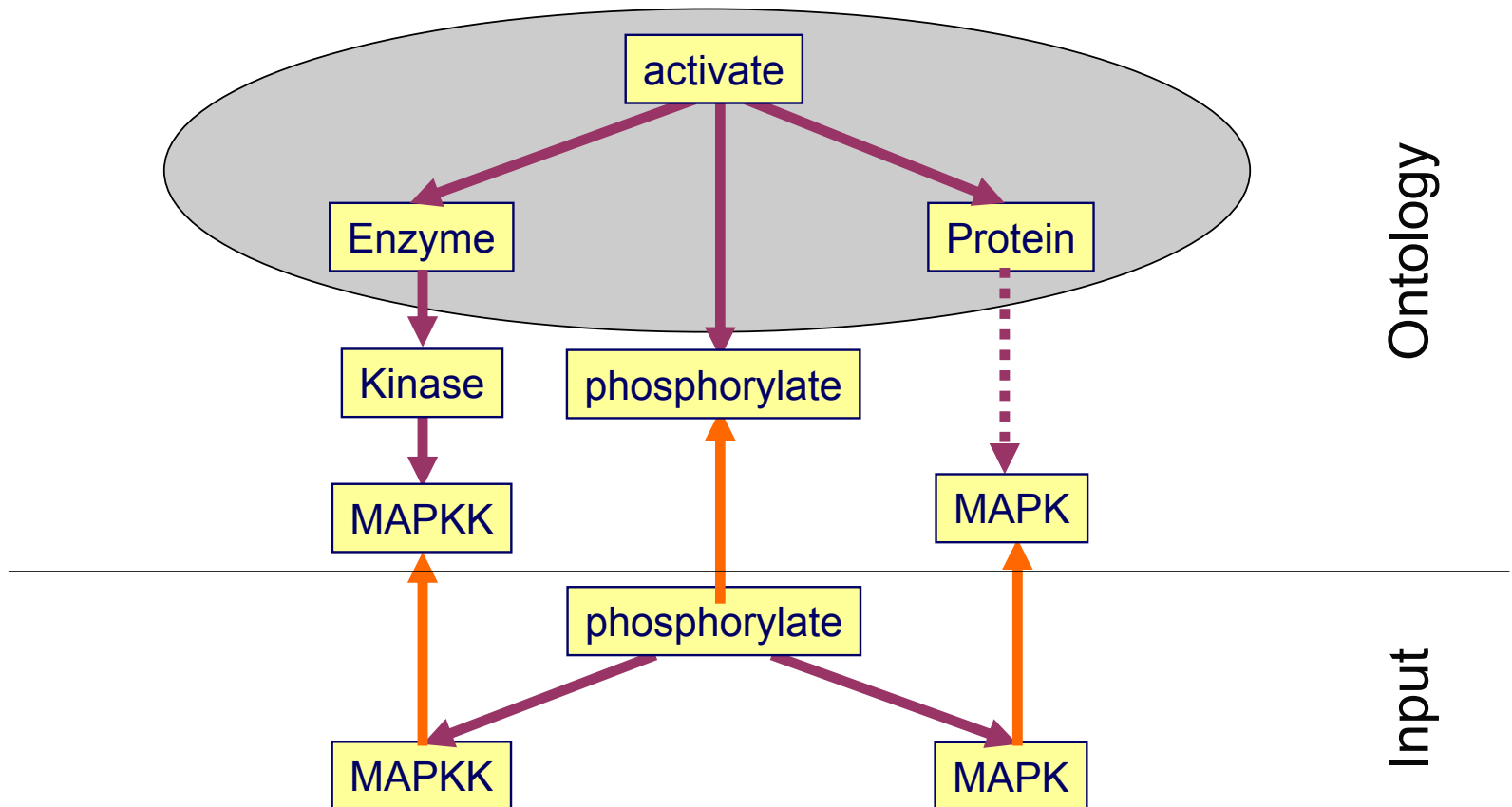
- Arrows for chemical reactions
- Role of each interaction partner

Predication: The Semantic Primitive



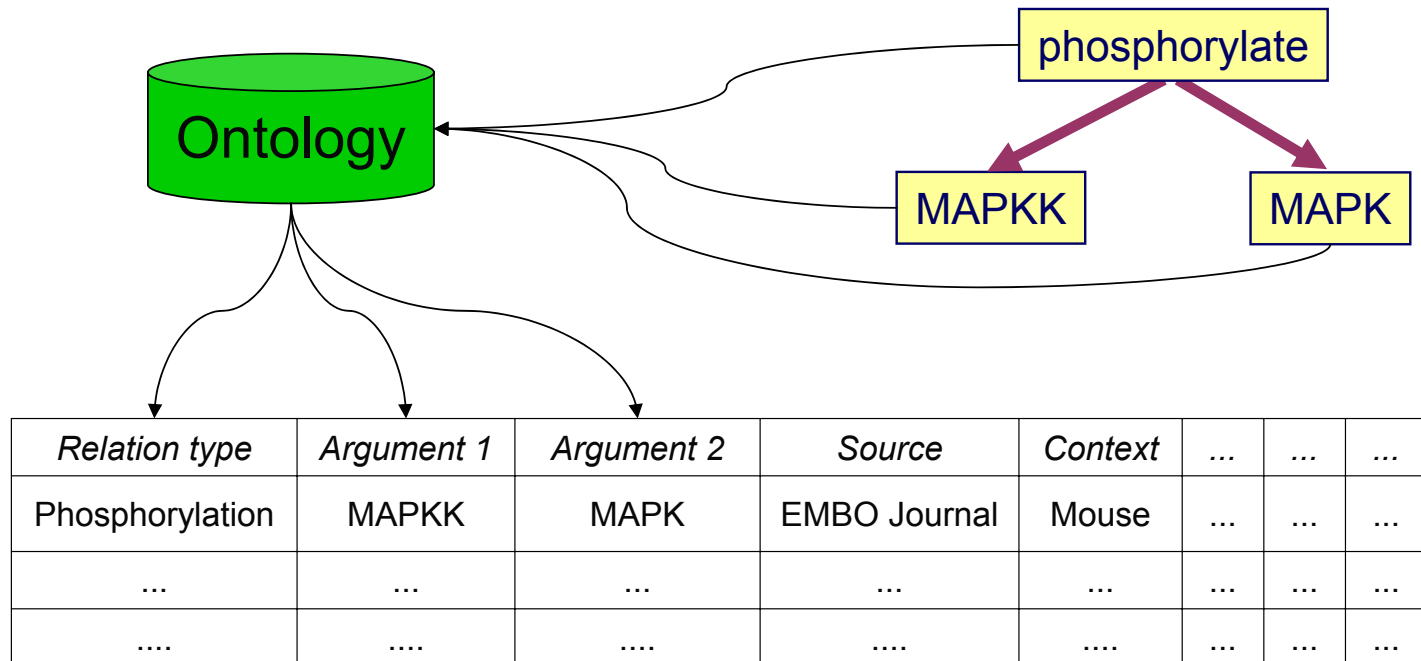
MAPKK phosphorylates MAPK.
 MAPK is phosphorylated by MAPKK.
 MAPKK-phosphorylated MAPK ...
 MAPK-phosphorylating MAPKK ...
 The phosphorylation of MAPK by MAPKK ...
 MAPK phosphorylation by MAPKK ...
 MAPKK phosphorylation of MAPK ...

From Templates to Knowledge – Match This!



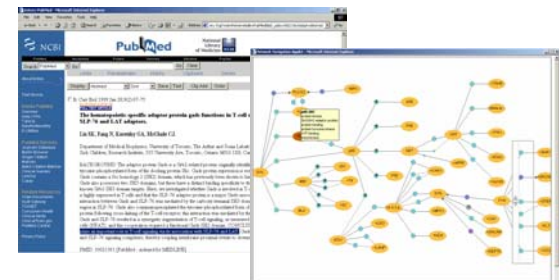
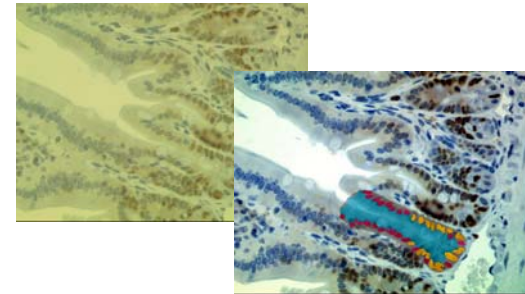
From Templates to Knowledge – Save that Fact!

References to the respective entries in the ontology
are stored in a relational fact database.



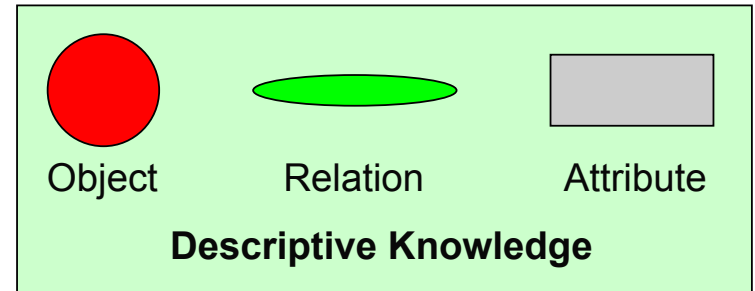
In summary...

- **Cellenger** is the **only** product capable of **automating** the analysis of complex images, such as cell-based assays or histopathological slides.
- **Polymind** automatically extracts **facts** out of text documents, for example protein-protein interactions or pathway information from Medline.

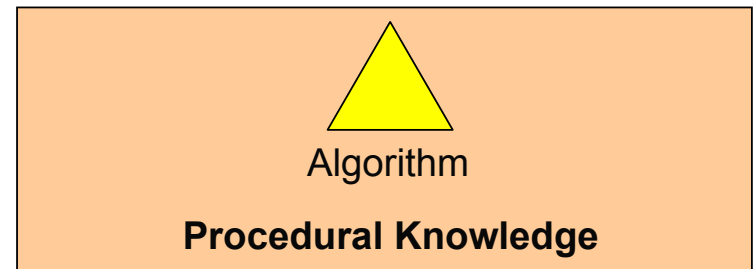


What are Cognition Networks?

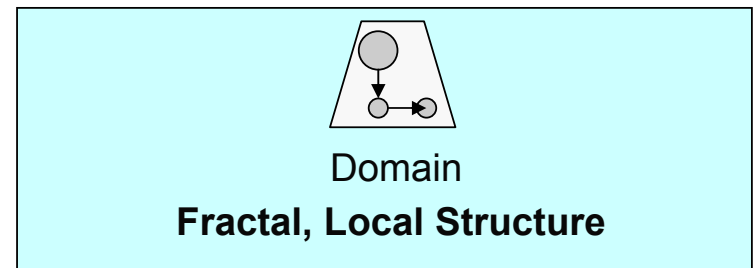
- Semantic
 - Knowledge representation



- Self-organizing
 - Cognitive understanding



- Self-similar
 - Complexity reduction

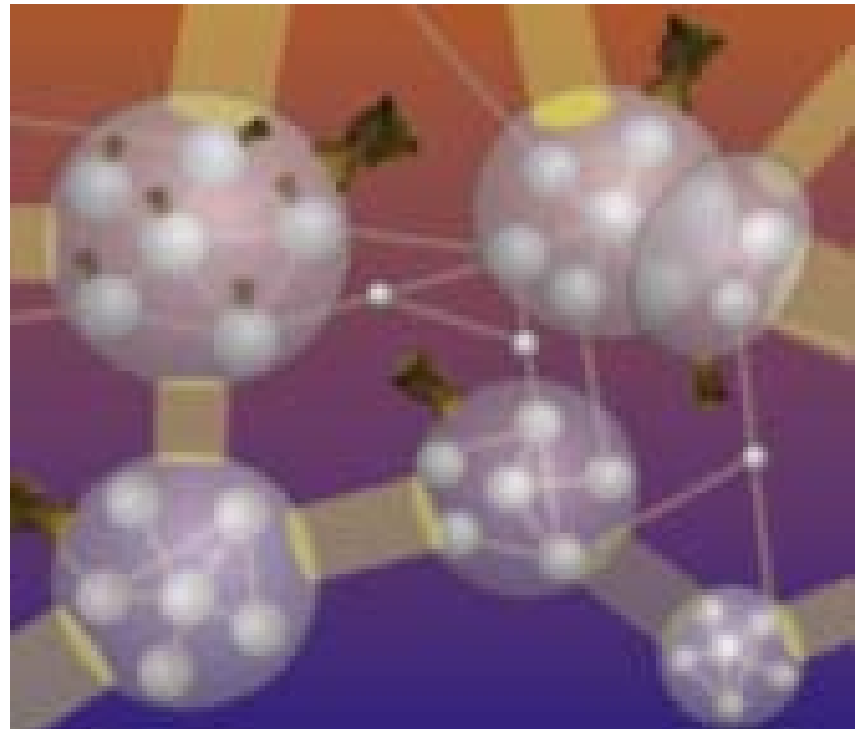


What are Cognition Machines?

- **A Network of**
 - Descriptive knowledge
 - Procedural knowledge
 - Fractal structure

- **Capable of automatically performing a cognitive task**
 - Image analysis
 - Text understanding

- **Created rapidly**
 - Next-generation programming language
 - Developer Studio



Why Cognition Machines for Simulation?

□ Data

- Initialization at $t=0$
- Extract structured data points from experiments, literature, ...
- Cellenger and Polymind provide structured format (CN)

□ Model

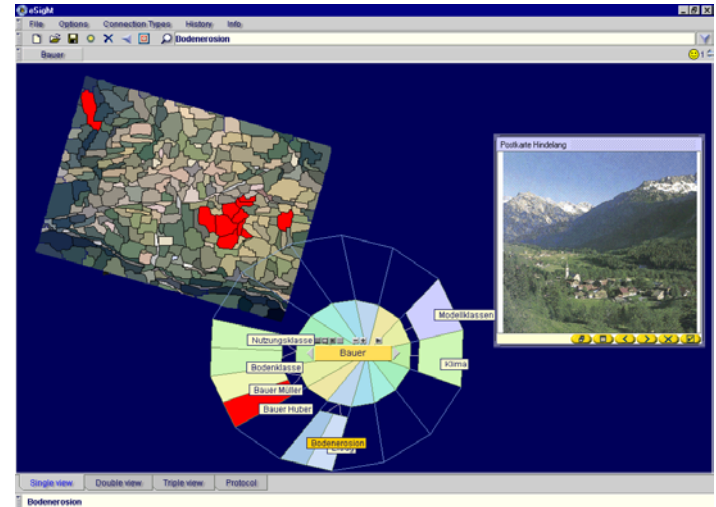
- Capture objects, relations, properties, dynamics ($t \rightarrow t+1$)
- Extract from theory, experimental evidence, literature, ...
- Integrate data into model
- Cognition Networks allow such modeling and data integration

□ Simulate

- Run and observe networked system
- Address complexity (no system of PDEs, but local dynamics)

Work done in the past - eSight

- Goal:
Modelling and simulation of complex economical and ecological systems
- Hierarchical modelling
- Object-oriented
- 2D geometry
- Automatic data acquisition
- Heterogeneous data
- User-defined functions
- Time-series



Cognition Networks and Cognition Machines: Bringing Human Like Understanding to Computers

Definiens AG

Trappentreustrasse 1

80339 Muenchen, Germany

www.definiens.com

