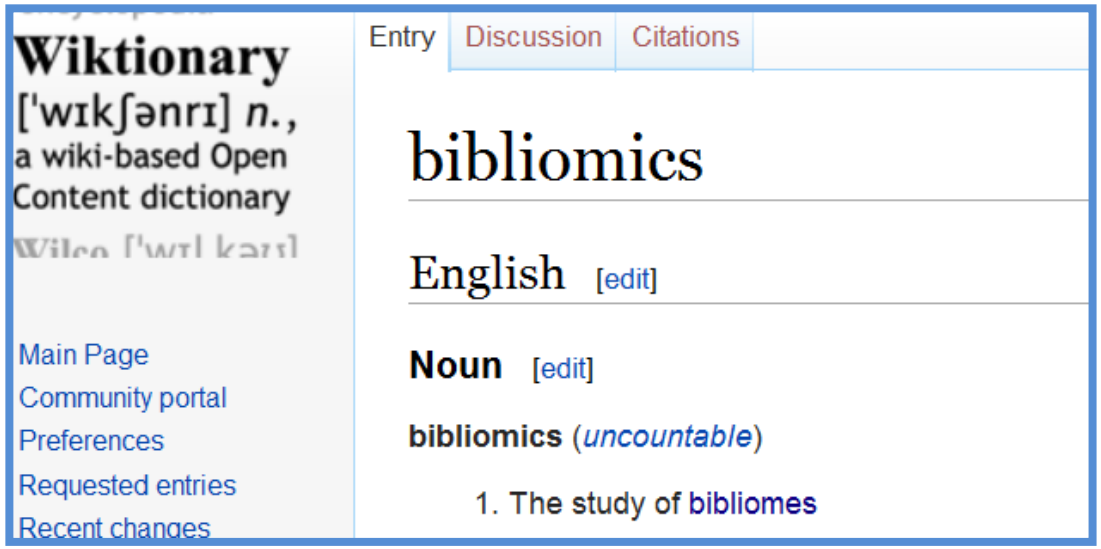
Trawling the bibliome

Wynand Alkema
NIZO Food Research
Radboud University Medical Center

Bibliomics : The study of the bibliome



Wiktionary
['wɪkʃənrɪ] *n.*,
a wiki-based Open Content dictionary

Main Page
Community portal
Preferences
Requested entries
Recent changes

Entry | Discussion | Citations

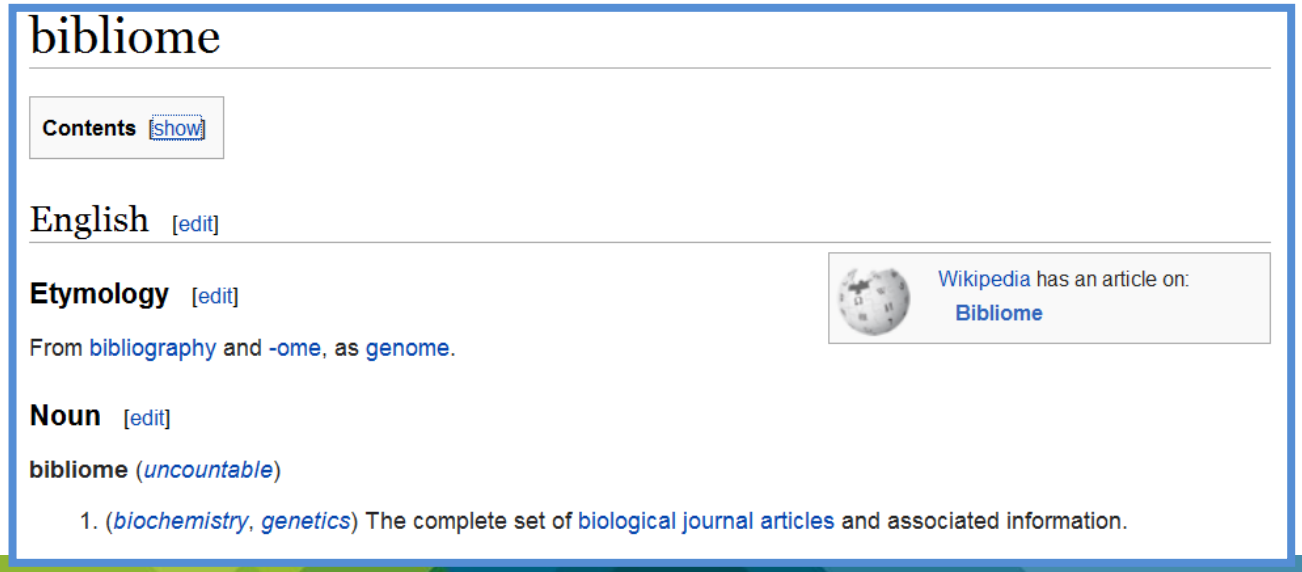
bibliomics

English [edit]

Noun [edit]

bibliomics (*uncountable*)

1. The study of bibliomes



bibliome

Contents [show]

English [edit]

Etymology [edit]

From bibliography and -ome, as genome.

Wikipedia has an article on:
Bibliome

Noun [edit]

bibliome (*uncountable*)

1. (*biochemistry*, *genetics*) The complete set of biological journal articles and associated information.

Data Flow in Research Projects

Knowledge:
New products
Life cycle extension
Intellectual Property

Literature review

Develop hypothesis

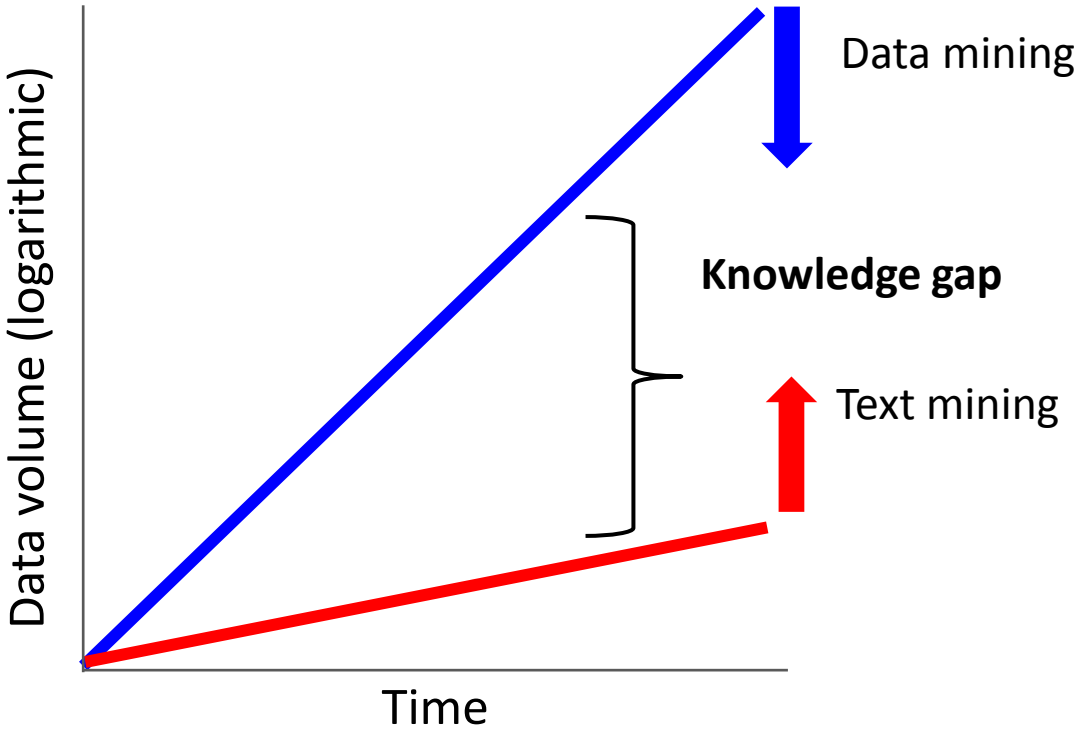
Perform experiments

Analysis interpretation

High Performance Computing

High Throughput Screening

Reduction of the knowledge gap



- Data
  - ❑ Metabolomics
  - ❑ Proteomics
  - ❑ Transcriptomics
  - ❑ Genomics
- Scientific text
  - ❑ Abstracts
  - ❑ Papers
  - ❑ Poster
  - ❑ Patents

What is text mining?

The **computer assisted** screening of **large** volumes of text to **retrieve** relevant documents, **extract** useful information and **discover** new knowledge.

A relation in the text
is a
relation in real life

An explosion of tools

Name	Input	Output	Ref	Tasks	Description
askMEDLINE	Query	Abstracts	[17, 18]	IR	Free-text, natural language (English only) query for M
Twease	Query	Ranked abstracts		IR	Indexes each word of Medline and provides features
XplorMed	Query	Ranked abstrac	[18]	IR	The system gives you the main associations between
eTBLAST	Tekst	Similar abstrac	[19]	IR	A text-similarity based search engine, using all word
MedlineRanker	Query	Ranked abstrac	[20]	IR	Ranks Medline abstracts based on user defined quer
MiSearch	Query	Ranked abstrac	[21]	IR	Ranks orders retrieved articles from PubMed based
PICO	Query	Abstracts	[22]	IR	Search engine for MEDLINE/PubMed with Spelling Ch
PubCrawler	Query	Abstracts	[23]	IR	Free "alerting" service that scans daily updates of th
PubFocus	Query	Abstracts and S	[24]	IR	Statistical analysis of the MEDLINE/PubMed search
PubGet	Query	Abstracts + Link to PDFs		IR	A comprehensive source for science PDFs, including
Pubmatrix	List of Terms	Co-occurrence r	[25]	IR	Allows simple text based mining of the NCBI literatur
PubNet	List of Terms	Graph	[26]	IR	a flexible system for visualizing literature-derived ne
GeneValorization	GeneList	Abstracts and li	[27]	IR, NER	GeneValorization gives a very clear and handful ove
DPWP	GeneList	Linked concepts	[28]	IR, NER	Disease/ Phenotype PAGE is a disease focused gene s
Anne O'Tate	Query	Abstracts and C	[29]	IR, NER	Gives an overview of the set of articles retrieved by a
Chillibot	GeneList and Te	Graph	[30]	IR, NER	Chilibot searches PubMed literature database (abstr
ProteinCorral	Protein Name	Linked proteins	[31]	IR, NER	Combines Information Retrieval and Extraction from
MEDIE	Tripple	Highligted phrases		IR, NER	Retrieve biomedical correlations from MEDLINE, bas
PubReminer	Query	Overrepresente	[32]	IR, NER	Breaks down a results of a PubmedQuery into catego
PubViz	Concepts	Hyperlinked gra	[33]	IR, NER	An Interactive Medline Search Engine Utilizing Extern
Quertle	Query + Concep	Abstracts + Ove	[34]	IR, NER	Using a combination of linguistic methods, Quertle f
Copub	Concept and Ge	Abstracts and li	[9]	IR, NER, IE	Web application with gene focussed retrieval of co-o
Coremine medical	Concept	Linked concepts		IR, NER, IE	Presents results about health, medicine and biology
Facta+	Tekst	Abstracts and li	[35]	IR, NER, IE	Finding Associated Concepts with Text Analysis
GoPubMed	Concept	Abstracts + Overrepresented co		IR, NER, IE	Uses semantic algorithms to abstracts from the MED

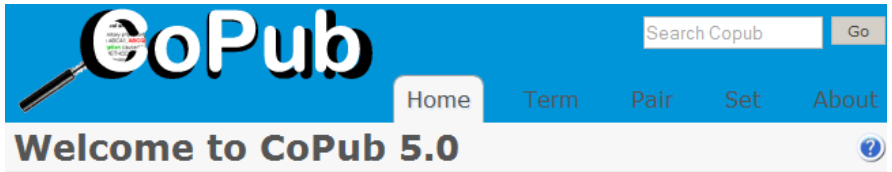
Mostly based on MEDLINE/ PubMed

Most of them co-occurrence based

Different tasks
- Document retrieval
- Information extraction
- Knowledge discovery

Fleuren and Alkema, Methods, 2015

CoPub: a database of co-citations



Nucleic Acids Research, 2011, 1–5
doi:10.1093/nar/gkr310

CoPub update: CoPub 5.0 a text mining system to answer biological questions

Wilco W. M. Fleuren[1,2,*], Stefan Verhoeven[3], Raoul Frijters[1], Bart Heupers[4], Jan Polman[3], René van Schaik[3], Jacob de Vlieg[1,3] and Wynand Alkema[3]

[1]Computational Drug Discovery (CDD), CMBI, NCMLS, Radboud University Nijmegen Medical Centre, [2]Netherlands Bioinformatics Centre (NBIC), PO Box 9101, 6500 HB Nijmegen, [3]Molecular Design and Informatics, MSD, PO Box 20, 5340 BH Oss and [4]SARA Computing and Network Services, Amsterdam, The Netherlands

Received January 27, 2011; Revised April 5, 2011; Accepted April 18, 2011

CoPub database

25 million papers

> 300 million literature relations

Statistical models for literature networks

www.copub.org

CoPub: A database of co-citations

Genes
Diseases
Metabolites
Drugs
Pathways

~450.000

PubMed abstracts

25 million

Search
Calculate co-citations

~130 million

Opposing effects of TGF-beta 2 on the Th1 cell development of naive CD4+ T cells isolated from different mouse strains.

P Hoehn;S Goedert;T Germann;S Koelsch;S Jin;N Palm;E Ruede;E Schmitt. 1995-11-22. J Immunol. 155. PMID: 7561083

The development of naive dense CD4+ T cells from different mouse strains toward Th1 cells, as monitored by measuring secondary IFN-gamma production, was affected by TGF-beta 2 in a differential way. Th1 cell development of naive CD4+ T cells from strains C57Bl/6, BALB/c, and NMRI primed by immobilized anti-CD3 mAb was strongly inhibited in the presence of TGF-beta 2. Even when the Th1 cell-inducer IL-12 was added, the same effect of TGF-beta 2 was observed. In contrast, Th1 development was substantially promoted by TGF-beta 2 with T cells from C3H/He and CBA/J mice. Further analyses using CD4+ T cells from (C57Bl/6xCBA/J)F1 hybrids or DBA/1 mice showed that Th1 development was inhibited by TGF-beta 2 if the T cells were activated by anti-CD3 mAb, but it was enhanced upon costimulation with anti-CD28 mAb. Determination of primary IL-2 production revealed that T cells from (C57Bl/6xCBA/J)F1 and DBA/1 mice produced low amounts of IL-2 following stimulation by anti-CD3 mAb alone and comparatively high amounts after coactivation by anti-CD28 mAb. In the presence of TGF-beta 2, the production of IL-2 was completely suppressed if such T cells were activated solely by anti-CD3 mAb, but it was only partially inhibited after costimulation by anti-CD28 mAb. Furthermore, TGF-beta 2-promoted Th1 development of such T cells was strongly inhibited after neutralization of endogenously produced IL-2 and completely restored by the addition of human IL-2. Thus, our results indicate that the TGF-beta 2-mediated stimulation of Th1 cell development requires the presence of relatively high concentrations of IL-2. Therefore, the opposing effect of TGF-beta 2 on the Th1 cell development of naive CD4+ T cells from different mouse strains appears to be the result of the variable potency of the respective CD4+ T cells to produce IL-2 in the presence of TGF-beta 2.

What is known about Lycopene?

CoPub

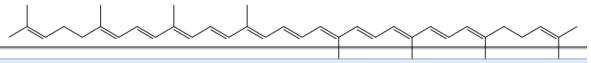
Home | Term

Term details

Preferred name	lycopene
Alternative names	(6e,8e,10e,12e,14e,16e,18e,20e,22e,24e,26e)-2,6,10,14,19,23,27,31-octamethyldotriaconta-2,6,8,10,12,14,16,18,20,22,24,26 all-trans-lycopene psi,psi-carotene
Categories	metabolite
Nr of abstracts	2297

Actions:

Retrieve abstracts in which term occurs | Add term to set | Search for indirect relations with term



Below are the most significant co-occurrences of 'lycopene' with each category. (Click in the count column to show the abstracts in which terms co-occur)

R scaled score >= 30 | Abstract count >= 3 | Order by R scaled score

disease

Cancer of prostate | Eye diseases | Gastrointestinal neoplasm | Infertility,male | Maculopathy,age-related | Myoma | Vitamin a deficiency | cardiovascular disease | colon cancer | mouth neoplasm

metabolite

7,7',8,8',11,11',12,12'-octahydro-psi,psi-carotene | alpha-carotene | beta-zeacarotene | cyclic carotene | epsilon,epsilon-carotene | gamma-carotene | lutein | neurosporene | zeaxanthin | zeta-carotene

polyphenols

5-caffeoylquinic acid | curcumin | ellagic acid | gallic acid | kaempferol | myricetin | naringenin | naringin | quercetin 3-o-rutinoside | resveratrol

GO molecular function

antioxidant | dioxygenase | glutathione peroxidase | glutathione reductase | hydratase | insulin like growth factor binding | monooxygenase | retinoic acid receptor | scavenger receptor | synthase

celltype

Adenocarcinoma cell | Breast cancer cell | Caco-2 cell | Cancer cell | Peripheral blood mononuclear cell | Premalignant cell | foam cell | peripheral blood lymphocyte | red blood cell

human gene

BXDC2 | CEL | GGT3 | GJA1 | IGFBP3 | JTV1 gene | PON1 | PTGDS | SCARB1 | XRCC1

tissue

abdominal adipose tissue | colon mucosa | endosperm | epithelial cancer | hypocotyl | mammary cancer | mycelia | prostate | sepal | uterine leiomyoma

drug

ascorbic acid | capsicum | dutasteride | finasteride | probucol | retinol acetate | retinol palmitate | tocopherol acetate | vitamin a | vitamin c

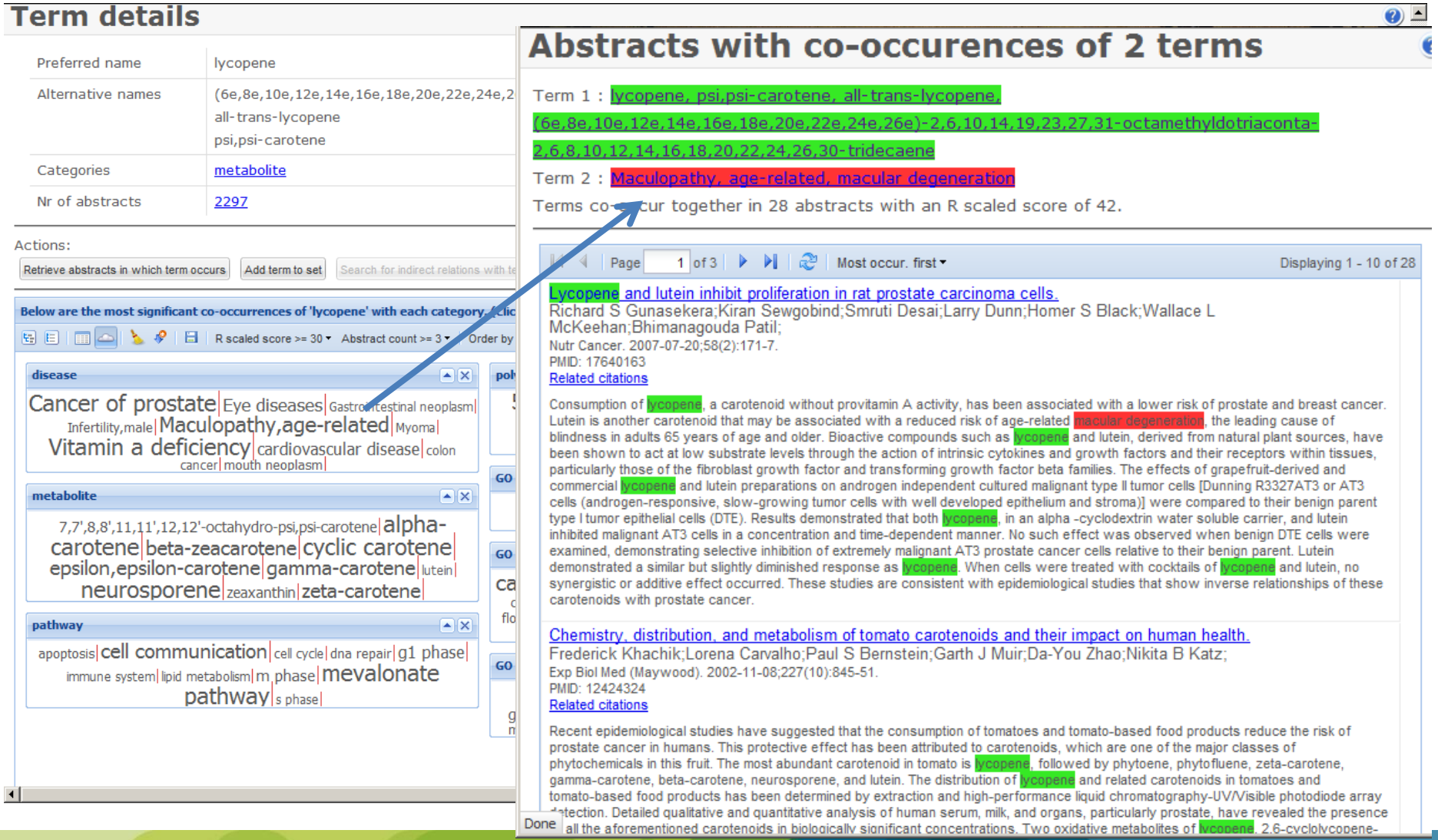
pathway

apoptosis | cell communication | cell cycle | dna repair | g1 phase | immune system | lipid metabolism | m phase | mevalonate pathway | s phase

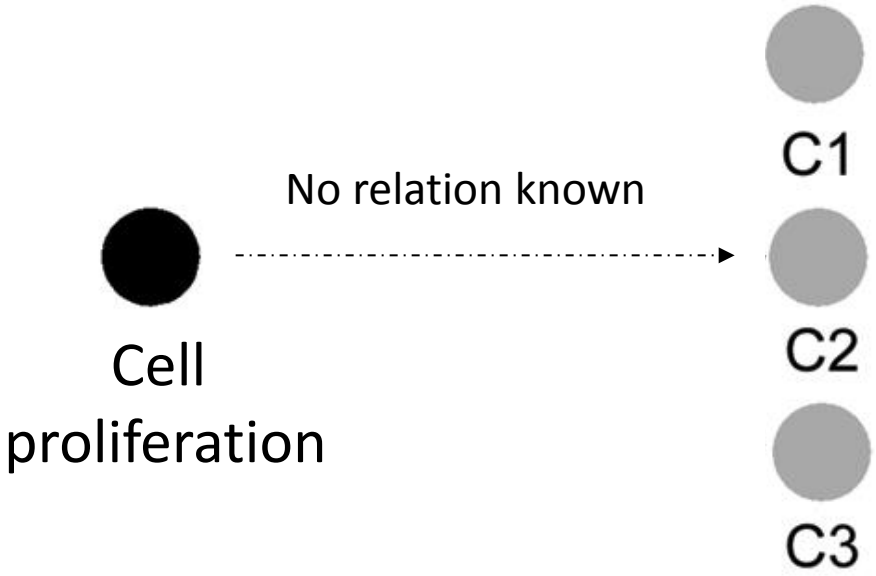
GO biological process

carotene metabolism | carotenoid metabolism | cell communication | ergosterol metabolism | ethylene metabolism | flower development | mevalonate pathway | photoprotection | ripening | xanthophyll metabolism

What is known about Lycopene?



Discovery of new knowledge



Which existing drugs can be used to inhibit or activate cell proliferation?

Existing drugs

Swanson's
ABC principle
Fish oil cures Raynaud's disease

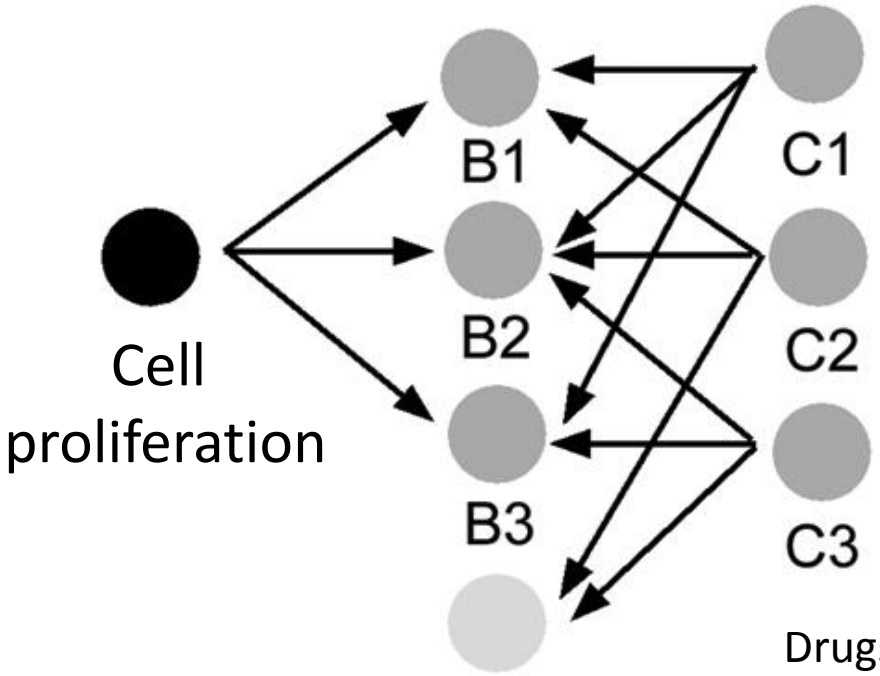
OPEN ACCESS Freely available online PLOS COMPUTATIONAL BIOLOGY

Literature Mining for the Discovery of Hidden Connections between Drugs, Genes and Diseases

Raoul Frijters[1], Marianne van Vugt[2], Ruben Smeets[2], René van Schaik[3], Jacob de Vlieg[1,3], Wynand Alkema[3]*

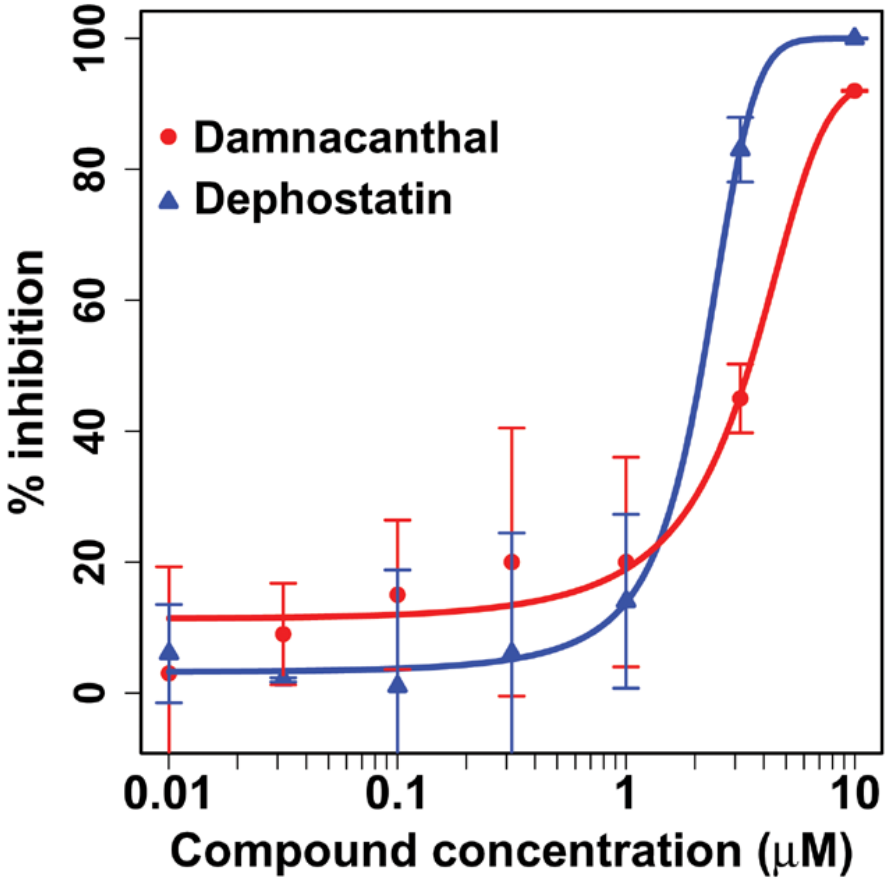
1 Computational Drug Discovery (CDD), Nijmegen Centre for Molecular Life Sciences (NCMLS), Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, 2 Department of Immune Therapeutics, Schering-Plough, Oss, The Netherlands, 3 Department of Molecular Design & Informatics, Schering-Plough, Oss, The Netherlands

Discovery of new knowledge



Drugs

Damnacanthal
Dephostatin



Experimental validation

Linking to transcriptomics

Glucocortiod regulated genes in the liver

Gene name	Symbol	Fc WT	Fc GRdim*
Insulin-like growth factor binding protein 1	*Igfbp1*	9.7	3.6
Tyrosine aminotransferase	*Tat*	4.1	1.6
6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	*Pfkfb3*	3.9	2.3
Peroxisome proliferative activated receptor, gamma, coactivator 1 alpha	*Ppargc1a*	3.2	3.2
CCAAT/enhancer binding protein (C/EBP), beta	*Cebpb*	3.1	1.6
Protein-tyrosine sulfotransferase 2	*Tpst2*	2.7	1.5
Solute carrier family 2 (facilitated glucose transporter), member 1	*Slc2a1*	2.6	1.9
Forkhead box O1	*Foxo1*	2.2	1.7
Serine dehydratase	*Sds*	2.1	1.5
Aminoadipate-semialdehyde synthase	*Aass*	2.1	1.3
Phosphoenolpyruvate carboxykinase 1, cytosolic	*Pck1*	1.9	1.3
Tryptophan 2,3-dioxygenase	*Tdo2*	1.5	1.2
Aconitase 2, mitochondrial	*Aco2*	1.4	1.3
Transketolase	*Tkt*	-1.2	-1.1
Protein kinase, AMP-activated, beta 1 non-catalytic subunit	*Prkab1*	-1.5	1
Adiponectin receptor 2	*Adipor2*	-1.6	-1.3
Purinergic receptor P2Y, G-protein coupled, 5	*P2ry5*	-1.6	-1.3
Pyruvate kinase liver and red blood cell	*Pklr*	-2.1	-1.4
CCAAT/enhancer binding protein (C/EBP), alpha	*Cebpa*	-2.3	-1.2
Sterol regulatory element binding factor 1	*Srebf1*	-3	-1.6
Insulin receptor substrate 1	*Irs1*	-4	-1.5

RESEARCH ARTICLE Open Access

Prednisolone-induced differential gene expression in mouse liver carrying wild type or a dimerization-defective glucocorticoid receptor

Linking to transcriptomics

Table 1: CoPub keyword enrichment analysis on prednisolone-regulated genes.

Biological process	# of associated genes
Metabolism	
Glucose metabolism/transport	22
Gluconeogenesis	21
Lipid metabolism/glycosylation	21
Glycolysis	14
Carbohydrate metabolism/ transport	14
Amino acid metabolism/transport	13
Cell cycle and apoptosis	
Apoptosis	77
Cell proliferation	72
Cell cycle	65
Cell growth and-or maintenance, cell growth	63
Homeostasis	58
Cell differentiation	49
Cell cycle arrest	37
Immune/Inflammatory response	
Inflammatory response	31
Acute-phase response	20

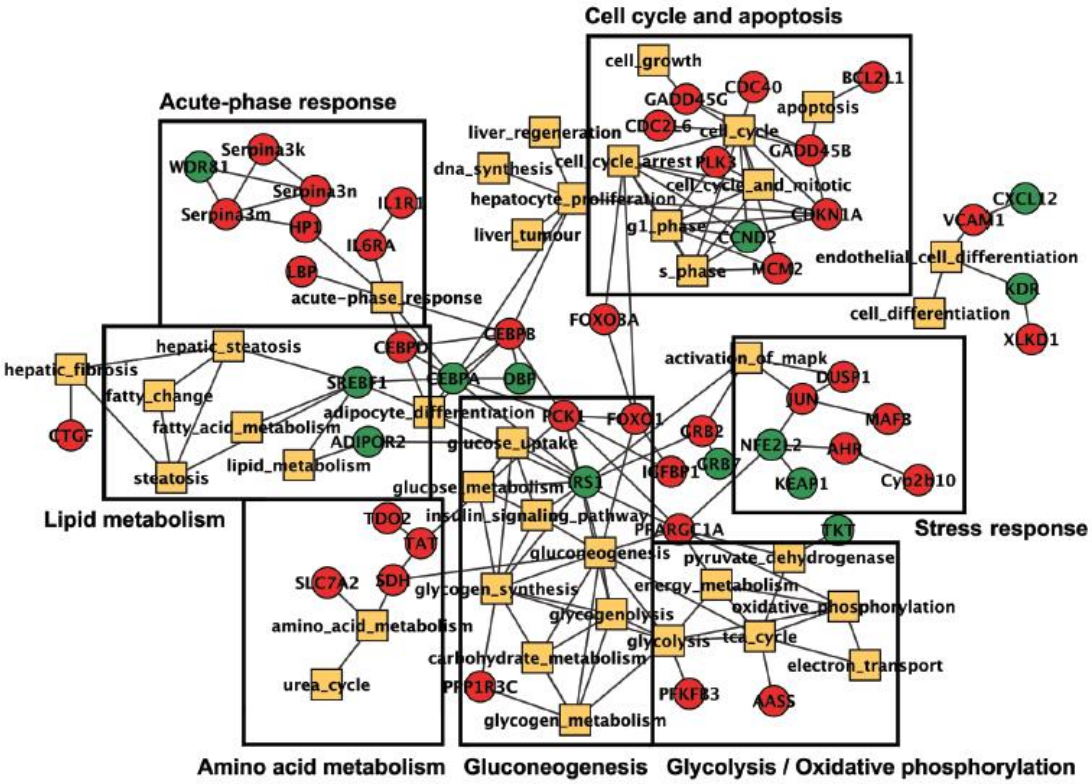
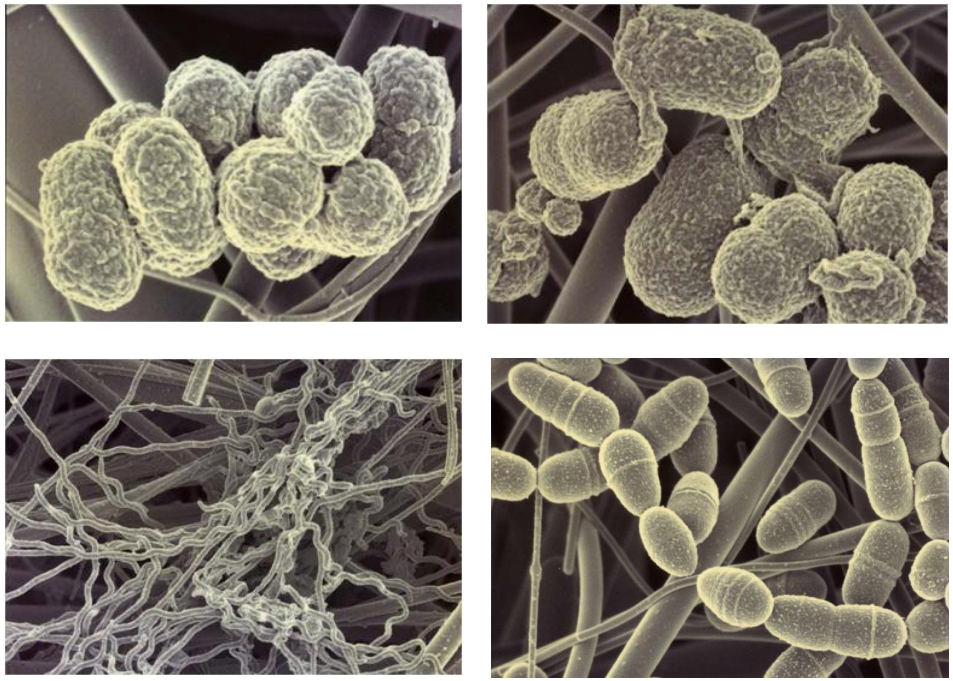


Figure 3 Literature-based network of glucocorticoid-induced effects. A network representation of the enrichment results was generated, in

You're mostly microbe

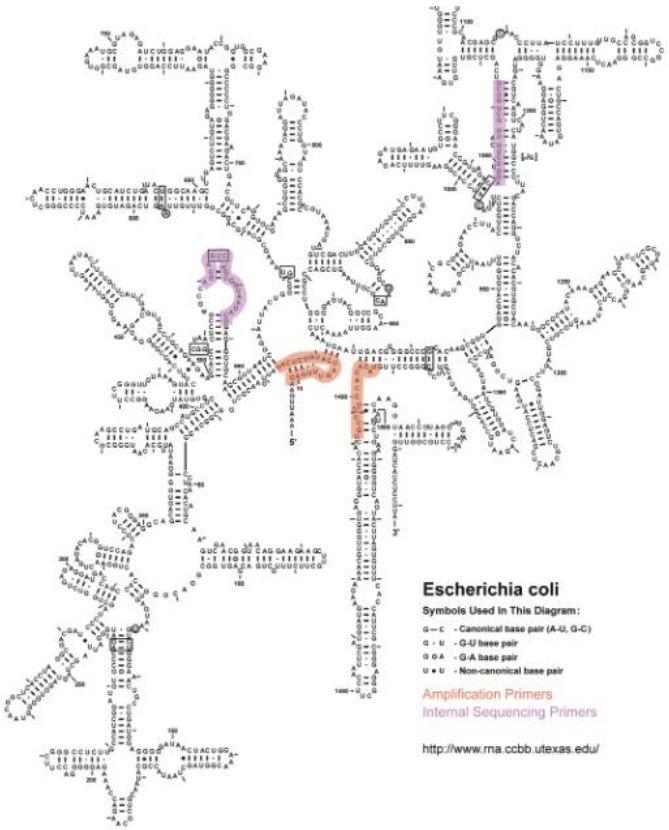


- 10^{18} cells, 10 times more than a human body
- In total more than 100 fold the gene content of a human
- Gut, skin, nasal-oral cavity
- Linked to health and disease

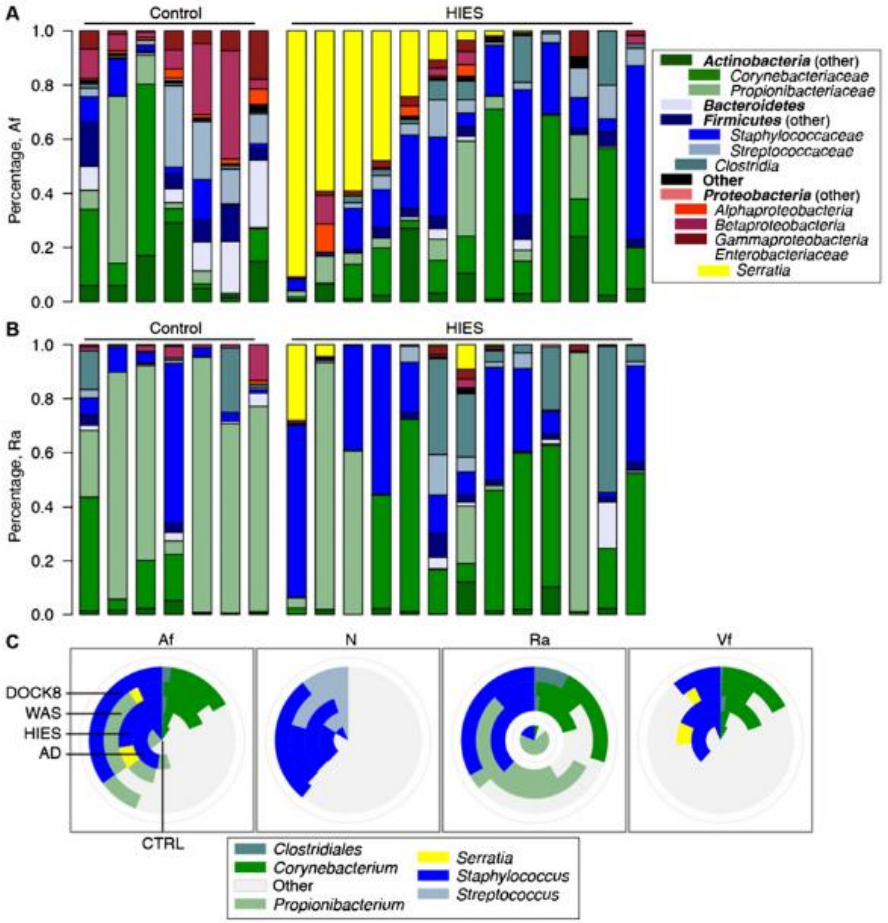
Metagenomics

"the use of genomics techniques to study communities of microbial organisms directly in their natural environments" (Chen and Pachter, 2005)

Profiling the skin microbiome : Who is there?



Secondary Structure: 16S small subunit ribosomal RNA



The altered landscape of the human skin microbiome in patients with primary immunodeficiencies

Julia Oh,[1] Alexandra F. Freeman,[2] NISC Comparative Sequencing Program,[3] Morgan Park,[3] Robert Sokolic,[1] Fabio Candotti,[1] Steven M. Holland,[2] Julia A. Segre,[1,5,6] and Heidi H. Kong[4,5,6]

[1]Genetics and Molecular Biology Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892, USA; [2]Laboratory of Clinical Infectious Diseases, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, Maryland 20892, USA; [3]NIH Intramural Sequencing Center, National Human Genome Research Institute, Bethesda, Maryland 20892, USA; [4]Dermatology Branch, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland 20892, USA

Profiling the skin microbiome : What do they do?

Table 2. Select abundant taxonomies differentially abundant between controls and STAT3-HIES individuals

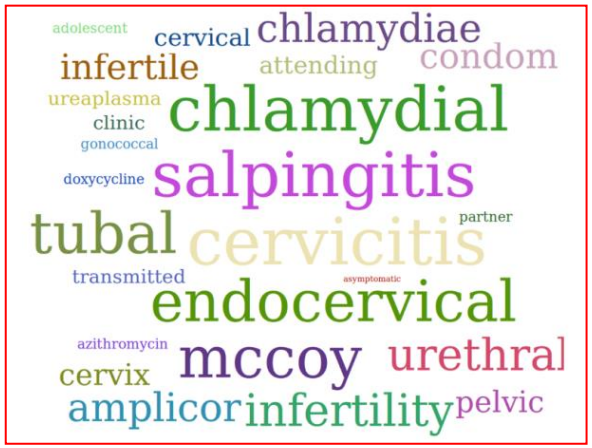
Site	Classification			CTRL (mean ± SE)	STAT3-HIES (mean ± SE)	Adjusted *P*-value
Af	*Actinobacteria*	*Actinomycetales*	*Corynebacterium*	9.1 ± 1.3%	17.2 ± 2.6%	0.04678
			Propionibacterium	24.2 ± 3.0%	10.9 ± 2.7%	0.10688
	Bacteroidetes	*Flavobacteriales*	*Porphyromonas*	0.9 ± 0.1%	0.3 ± 0.1%	0.04827
			Cloacibacterium	3.0 ± 0.7%	0.0 ± 0.0%	0.04678
	Firmicutes	*Bacilliales*	*Staphylococcus*	11.3 ± 1.7%	27.7 ± 4.0%	0.03092
			S. aureus	0.4 ± 0.2%	10.3 ± 3.4%	0.02281
			S. epidermidis	5.4 ± 1.4%	13.2 ± 2.7%	0.04893
			S. haemolyticus	0.2 ± 0.0%	1.4 ± 0.4%	0.00050
		Clostridiales	*Anaerococcus*	0.6 ± 0.1%	1.8 ± 0.4%	0.04480
			Finegoldia	0.2 ± 0.0%	1.6 ± 0.4%	0.00001
			Peptoniphilus	0.4 ± 0.1%	0.8 ± 0.2%	0.08678
	Proteobacteria	*Betaproteobacteria*	*Diaphrobacter*	2.1 ± 0.6%	0.0 ± 0.0%	0.04678
		Gammaproteobacteria	*Serratia*	0.0 ± 0.0%	8.1 ± 3.3%	0.00218
N	*Firmicutes*	*Bacilliales*	*Staphylococcus*	10.7 ± 1.3%	32.3 ± 4.3%	0.00002
			S. aureus	3.2 ± 0.9%	5.0 ± 1.9%	0.74406
			S. epidermidis	6.0 ± 0.9%	24.5 ± 4.2%	0.00000
			S. haemolyticus	0.0 ± 0.0%	0.5 ± 0.2%	0.02412
			Streptococcus	6.2 ± 1.1%	5.3 ± 1.4%	0.98904
Ra	*Actinobacteria*	*Actinomycetales*	*Corynebacterium*	4.8 ± 1.2%	19.0 ± 4.5%	0.00163
			Propionibacterium	67.5 ± 4.8%	20.8 ± 6.8%	0.00004
	Firmicutes	*Bacilliales*	*Staphylococcus*	18.9 ± 3.6%	37.8 ± 5.2%	0.00556
			S. aureus	0.0 ± 0.0%	9.1 ± 3.9%	0.00824
			S. epidermidis	13.8 ± 2.9%	24.6 ± 5.4%	0.04219
			S. haemolyticus	0.0 ± 0.0%	2.1 ± 1.6%	0.07403
			Finegoldia	0.3 ± 0.1%	2.0 ± 0.6%	0.00197
Vf	*Actinobacteria*	*Actinomycetales*	*Corynebacterium*	7.9 ± 1.4%	17.6 ± 2.6%	0.01084
	Bacteroidetes	*Flavobacteriales*	*Cloacibacterium*	2.1 ± 0.6%	0.0 ± 0.0%	0.10420
			Porphyromonas	1.2 ± 0.2%	0.5 ± 0.1%	0.07706
	Firmicutes	*Bacilliales*	*Staphylococcus*	6.3 ± 1.1%	18.3 ± 2.7%	0.02686
			S. aureus	0.6 ± 0.3%	7.5 ± 2.4%	0.05623
			S. epidermidis	1.6 ± 0.2%	7.1 ± 1.5%	0.00002
			S. haemolyticus	0.2 ± 0.1%	1.4 ± 0.4%	0.00042
		Clostridiales	*Anaerococcus*	0.5 ± 0.1%	2.0 ± 0.4%	0.00001
			Finegoldia	0.2 ± 0.0%	2.2 ± 0.5%	0.00000
			Peptoniphilus	0.3 ± 0.1%	1.2 ± 0.3%	0.00084
	Proteobacteria	*Gammaproteobacteria*	*Serratia*	0.0 ± 0.0%	8.4 ± 3.3%	0.00473

Microbes : What do they do?

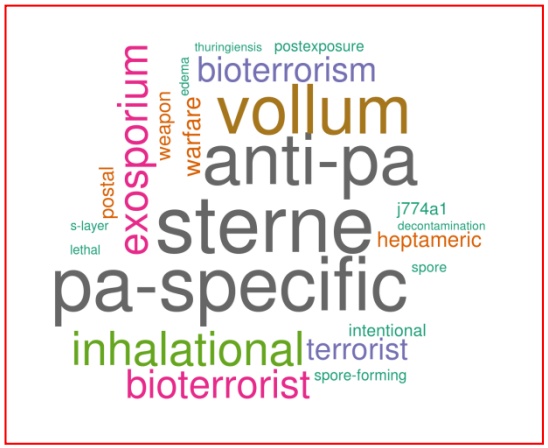
Mapped entire NCBI taxonomy (~300000 species) against total MEDLINE

Unsupervised analysis : Count the words associated with your organism

Staphylococcus aureus
> 70000 scientific papers



Chlamydia
> 1200 scientific papers

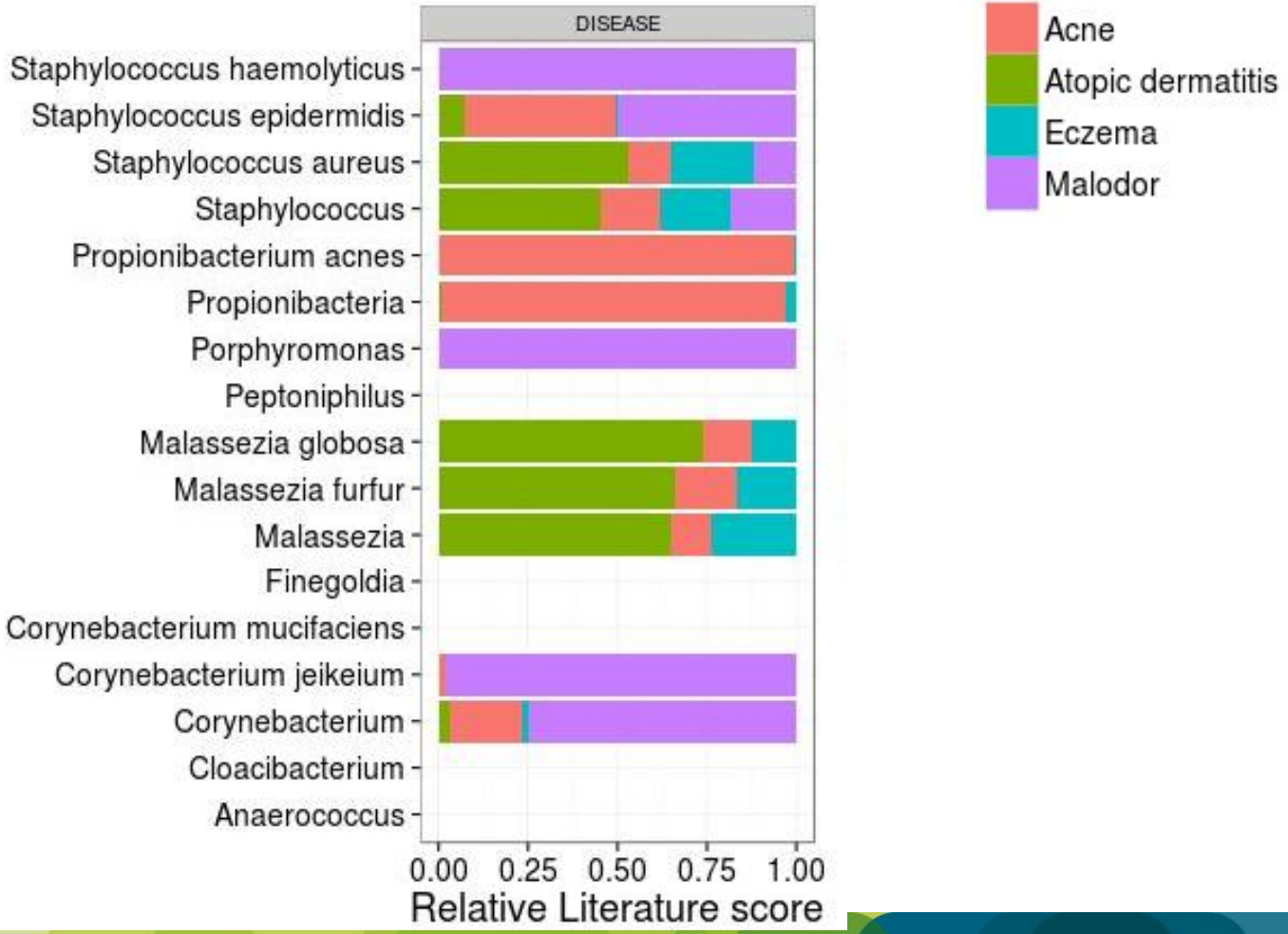


Bacillus anthracis
> 4000 scientific papers

Statistics, also when creating the word clouds

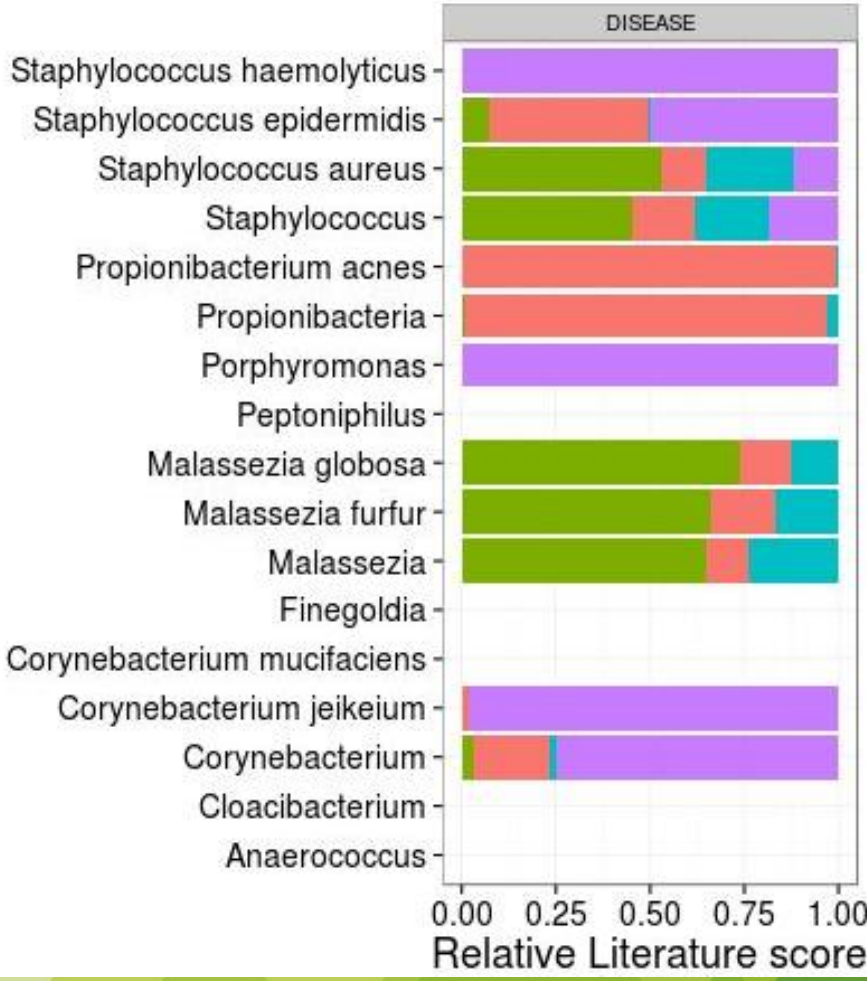
Linking the species to literature

Supervised analysis : Analyze known relations with diseases of interest



Linking the species to literature

Supervised analysis : Analyze known relations with diseases of interest



The in vivo effect of UVB radiation on skin bacteria in patients with atopic dermatitis.
Jekler;I M Bergbrant;J Faergemann;O Larkö. 1992-06-18. Acta Derm Venereol. 72. PMID: 1350138

Fourteen patients suffering from atopic dermatitis under treatment with UVB radiation were subjected to aerobic bacterial cultures in order to investigate whether this ultraviolet waveband has any in vivo germicidal effects, and, if so, whether there is a correlation with clinical improvement. Treatments were given 3 times a week for 8 weeks. Bacterial samples were collected before, midway and after the termination of therapy. On the latter two occasions, cultures were performed 30 min and 24 h post-UVB irradiation. The main bacteria found were Staphylococcus epidermidis and S. aureus. S. aureus carriage was found in 12 patients in lesional, dermatitic skin, and in 11 patients in clinically non-lesional skin. UVB radiation was found to have an antimicrobial effect primarily concerning S. aureus. Bacterial counts of this organism in lesional skin were decreased from a mean of 1.3 x 10(3) to 1.2 x 10(1) bacteria per cm2 skin at the 8-week 30-min count (p less than 0.01) and 7.5 x 10(1) at the 8-week 24-h count (p less than 0.05). The treatment yielded a statistically significant clinical improvement.

Antistaphylococcal IgE in patients with atopic dermatitis.
S Abramson;M V Dahl;G Walsh;M N Blumenthal;S D Douglas;P G Quie. 1982-10-21. J Am Acad Dermatol. 7. PMID: 7107990

Levels of IgE antibodies to Staphylococcus aureus and Staphylococcus epidermidis were determined in eleven patients with typical atopic dermatitis, with no history of furuncles or severe staphylococcal infection. Increased IgE binding to S. aureus but not to S. epidermidis was observed. Fifteen patients with hyperimmunoglobulinemia E-staphylococcal abscess syndrome had increased IgE binding not only to S. aureus but also to S. epidermidis. Other control groups of patients with elevated IgE levels or recurrent staphylococcal infection had normal IgE binding activity to both strains of staphylococci. Interaction of staphylococcal antigens from bacteria on skin with antistaphylococcal IgE antibodies on mast cells could induce mast cell release, evoke itch, and aggravate atopic dermatitis.

Comparative study of staphylococci from the skin of atopic dermatitis patients and from healthy subjects.
S Higaki;M Morohashi;T Yamagishi;Y Hasegawa. 1999-06-17. Int J Dermatol. 38. PMID: 10321941

tml/skin/start.html Bacterial infections occur frequently on the skin of atopic dermatitis (AD) patients.

Final remarks

- Technical issues in ‘bibliomics’ are largely solved
- Text mining is a versatile tool
  - Applicable to all kind of ~omics data.
  - Standardized workflows are needed!
  - Text mining on demand
- We need dedicated vocabularies that describe a set of concepts for a given topic, field of interest.
- Ontologies are good starting point but are not always practical (People do not use ontology terms in their papers)
- Algorithms for connecting concepts can be improved
  - Statistics
  - Network theory

Acknowledgements