

RetroMine,
or how to provide in-depth retrospective
studies from Medline in a glance:
the hepcidin use-case.

Bertrand Ameline de Cadeville

Olivier Loréal

Fouzia Moussouni-Marzolf



Institut national
de la santé et de la recherche médicale

Introduction

Life Science is one of the most VOLUMINOUS science

**Rapid Expansion of the biomedical literature
Available papers exploding**

**Increased demand and development of
effective text mining tools to find quickly
relevant information.**

Introduction

Biologists are reluctant to use text mining tools.

why ?

(1) PubMed Search Engine
(survey of [Eva Lee et al. 2011])

Few users review results beyond the first page.

Users seek simple interfaces :
difficulties using advanced features of PubMed

Users retrieve in-depth information
focusing on one category :
Gene Cell Disease ...

**Many pertinent articles not being pulled
in a Medline search**

Introduction

Non computational biologists are reluctant to use text mining tools.

(2) For fine-grained tools

Identification of **Entities**



In abstracts

PubMed
US National Library of Medicine
National Institutes of Health

Advanced

Abstract

J Mol Med (Berl). 2010 May;88(5):477-86. doi: 10.1007/s00109-009-0588-3. Epub 2010 Feb 19.

Natural and synthetic STAT3 inhibitors reduce hepcidin expression in differentiated mouse hepatocytes expressing the active phosphorylated STAT3 form.

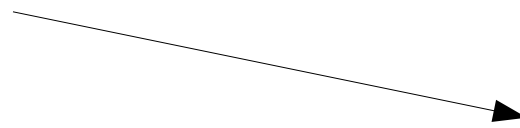
Fatih N¹, Camberlein E, Island ML, Corlu A, Abquequen E, Détiवाद L, Leroyer P, Brissot P, Loréal O.

Author information

Abstract

During the inflammatory process, hepcidin overexpression favours the development of anaemia of chronic diseases which represents the second most common form of anaemia worldwide. The identification of therapeutic agents decreasing hepcidin expression is therefore an important goal. The aim of this study was to target the STAT3 signalling involved in the development of increased hepcidin expression related to chronic inflammation. In a co-culture model associating mouse hepatocytes and rat liver epithelial cells, the mRNA levels of hepcidin1, albumin, aldolase B, Cyp3a4, Stat3, Smad4 and iron regulatory genes were measured by real-time PCR. STAT3 and phosphorylated SMAD1/5/8 proteins were analysed by Western blot. At variance of hepatocyte pure culture, co-culture provided high levels of hepcidin1 mRNA, reaching 400% of the freshly isolated hepatocyte values after 6 days of culture. Hepcidin expression was associated with the maintenance of hepatocyte phenotype, STAT3 phosphorylation and functional BMP/SMAD pathway. Stat3 siRNAs inhibited the hepcidin1 mRNA expression. STAT3 inhibitors, including curcumin, AG490 and a peptide (PpYLKTK), reduced hepcidin1 mRNA expression even when cells were additionally exposed to IL-6. Hepcidin1 mRNA was expressed at high levels by hepatocytes in the co-culture model, and STAT3 pathway activation was controlled through STAT3 inhibitors. Such inhibitors could be useful to prevent anaemia related to hepcidin overexpression during chronic inflammation.

Identification of **Events** connecting Entities



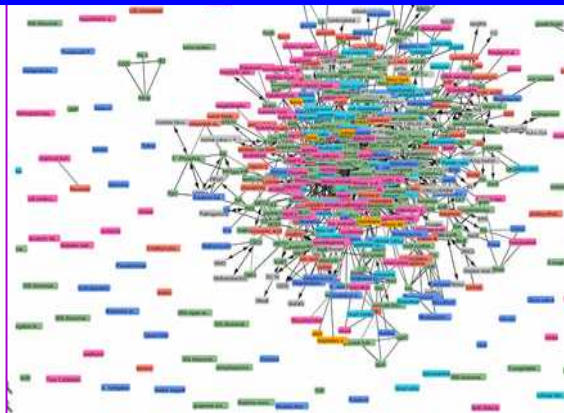
In sentences

Curcumin reduces *hepcidin1*

Introduction

These tools extract a deluge of information
Very dense data

Query « Hecpudin » using Ali-Baba [1]



For non expert
Information is dense
and unreadable

**Pertinent information is
hidden.**

**Non computational
biologists are discouraged
from using this kind of tool.**

For an expert
A considerable amount of
background knowledge

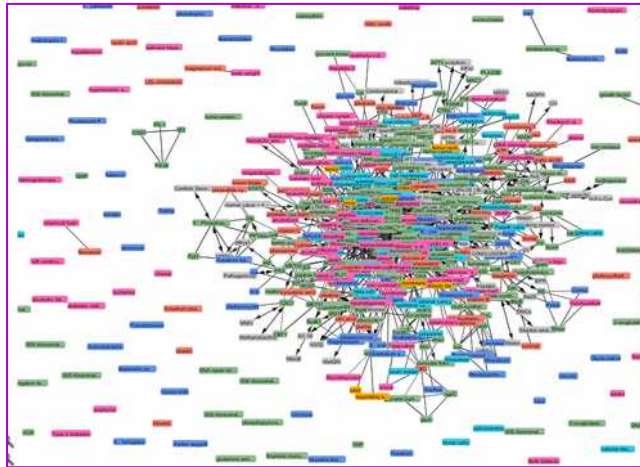
Other Event Extraction Tools

- ➡ TEES for Turku Event Extraction System of [Jari Björne *et. al*]
- ➡ EVEX web site and database
- ➡ BioNLP initiatives . . .

**Mostly devoted to the computational
biology and bioinformatics
community.**

Despite these efforts ...

few studies addressed the problem of processing these **big datasets** in order for **non-computational experts** to rapidly convert them into **meaningful patterns** over time.



?



PROPOSAL

Improve Data Selection

Identify **background** knowledge and filter it to reduce the density of information

Make time explicit : perception of events chronology enhance comprehension



Select **relevant** non background events over time and provide them to bio-investigators.

Methods

SOME BASICS

What is a Background Event ?

This concept is relative to a certain time point t

A background event is spotted and extracted repeatedly from literature at different time points

Example : IL6 transcription factor of Hepcidin published first time in April 2003

May 2005		July 2005					Dec 2005				
IL-6	express	fusionprotein	Co-occurrence	Hepcidin	16141345	2005-09-01					
interleukin6	express	GFP	expressing	hepcidin	16141345	hepcidin	exposed	Epo	16332970	2005-12-01	
IL-6	co-occu	Hemojuvelin	encoding	HAMP	15967692	Hepcidin	decreased	IL-6	16351643	2005-12-01	
Hepcidin	co-occu	Hepcidin	induced	IL-6	15886319	Hepcidin	influence	IL-6	16351643	2005-12-01	
Hepcidin	co-occu	Hepcidin	induced	interleukin6	15886319	hepcidin	decreased	interleukin-6	16351643	2005-12-01	
Ferroportin1	increa	Hepcidin	Co-occurrence	IL-6	16198622	Hepcidin	interacting	SLC40A1	16351644	2005-12-01	
Ferroportin1	is	Hepcidin	upregulation	IL-1beta	16198622						
		hepcidin	induced	IL-6	15886319						
		hepcidin	Co-occurrence	IL-6	16198622						
		IL-1beta	upregulated	Hepcidin	16198622	2005-09-01					
		IL-1beta	production	IL-6	16198622	2005-09-01					
		RGMC	encoding	HAMP	15967692	2005-09-01					

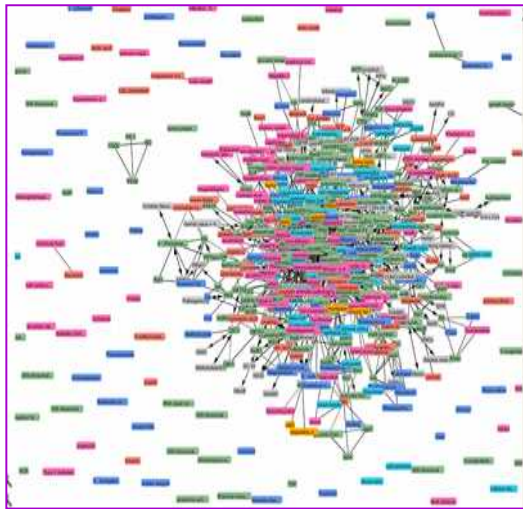
Background events are continuously returned but remain trivial to Hepcidin experts

Methods

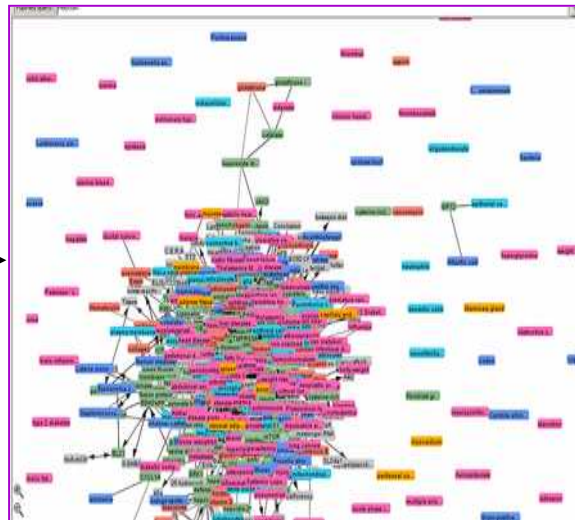
Definition 1

A recognized event e is defined as being a background if it has been spotted repeatedly in different abstracts and at different time points. In other words, when an event e is published for the first time at t , it becomes background at time $t+\Delta$.

G_0 = Graph of events
obtained for time t_0

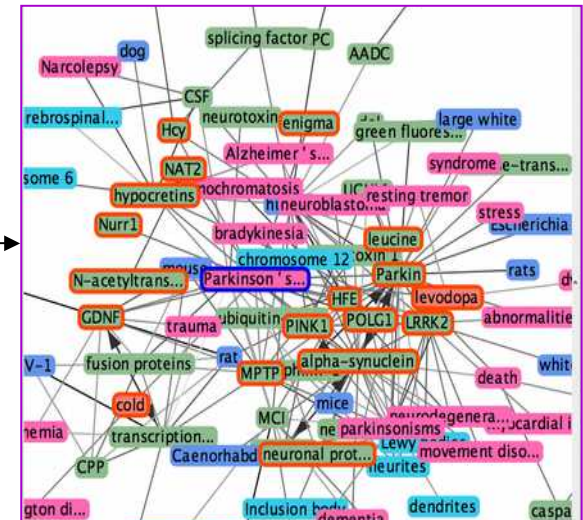


G_1 = Graph of events
 $t_1 = t_0 + p$



$(e \in G_1)$ and $(e \in G_0)$

G_2 = Graph of events
 $t_2 = t_1 + p$

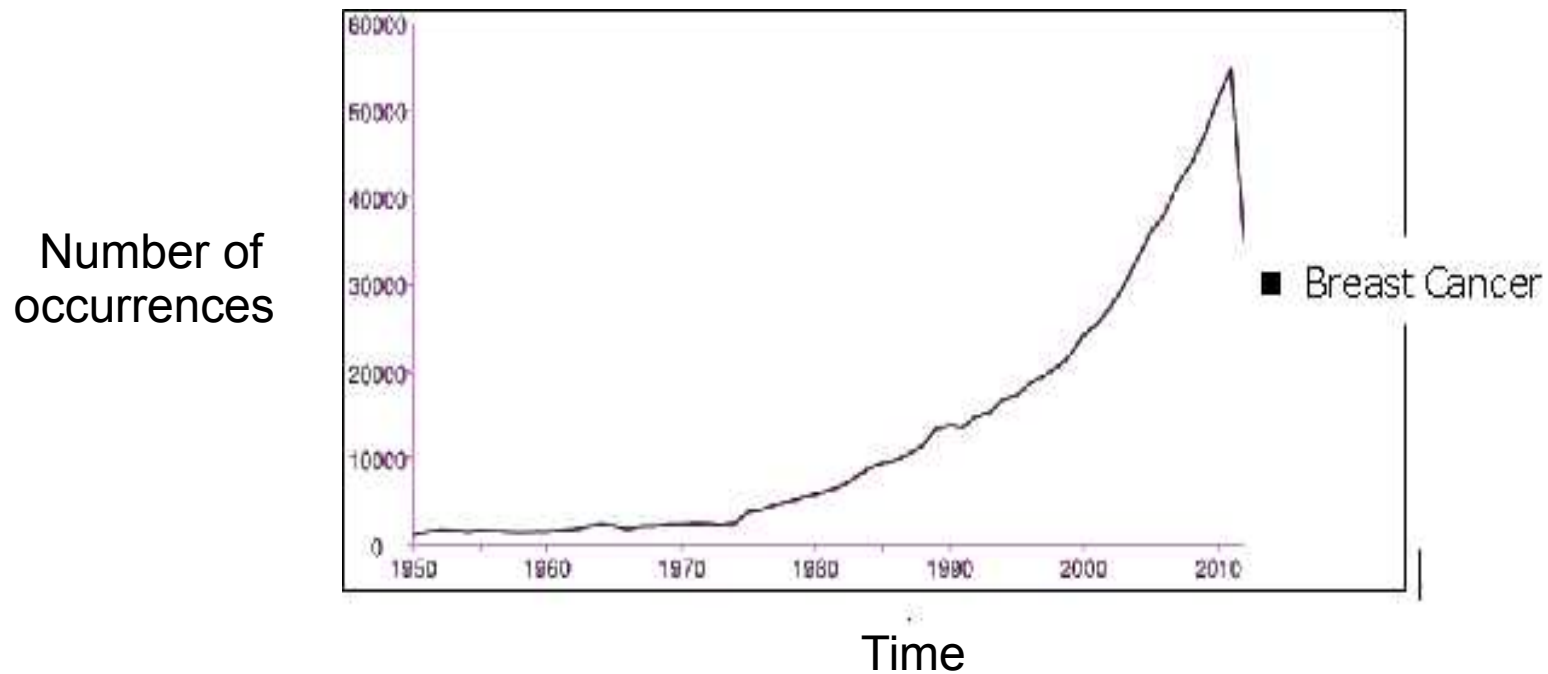


$(e \in G_2)$ and $(e \in G_1 \text{ or } e \in G_0)$

Methods

Time Relevance of a Biological Entity ?

The concept of **time relevance** has already been reported in [Palidwor GA, et al.] for graphing MEDLINE **keywords** over time using MLTrends.



When applied to recognized entities and events connecting them, this approach leads to much more informative functions

Methods

So what is a time relevant (t-relevant) biological entity ?

Set of abstract for query « Hepcidin 2005/05 [dp] »

Abstract +

J Biol Chem (2010) 285(18):4177-86. doi: 10.1073/jcb.100-029-0688-3. Epub 2010 Feb 19.

Natural and synthetic STAT3 inhibitors reduce hepcidin expression in differentiated mouse hepatocytes expressing the active phosphorylated STAT3 form.

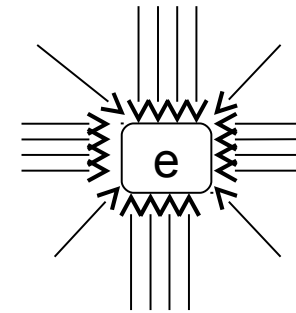
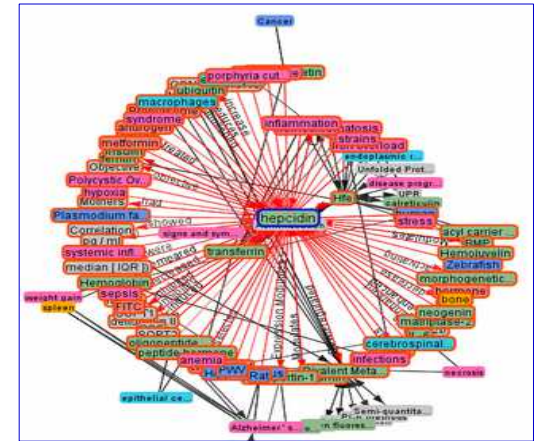
Fathallah¹, Cambresien E, Hamed M, Cofre A, Abouqoum E, Delvaux J, Laroye P, Brassat P, Liodet O.

Author information

Abstract

During the inflammatory process, [hepcidin](#) overexpression favours the development of anaemia of chronic diseases which represents the second most common form of anaemia worldwide. The identification of therapeutic agents decreasing hepcidin expression is therefore an important goal. The aim of this study was to target the STAT3 signaling involved in the development of increased hepcidin expression related to chronic inflammation. In a co-culture model associating mouse hepatocytes and rat liver epithelial cells, the mRNA levels of hepcidin1, albumin, aldolase B, Cyp3a4, Stat3, Smad4 and iron regulatory genes were measured by real-time PCR. STAT3 and phosphorylated [STAT3](#) proteins were analysed by Western blot. At variance of hepatocyte pure culture, co-culture provided high levels of hepcidin1 mRNA, reaching 400% of the freshly isolated hepatocyte values after 6 days of culture. Hepcidin expression was associated with the maintenance of hepatocyte phenotype. [STAT3](#) phosphorylation and functional BMP/SMAD pathway. Stat3 siRNAs inhibited the hepcidin1 mRNA expression. STAT3 inhibitors, including curcumin, AG490 and a peptide (PpYLKTK), reduced hepcidin1 mRNA expression even when cells were additionally exposed to IL-6. Hepcidin1 mRNA was expressed at high levels by hepatocytes in the co-culture model, and STAT3 pathway activation was controlled through STAT3 inhibitors. Such inhibitors could be useful to prevent anaemia related to hepcidin overexpression during chronic inflammation.

graph of events extracted (ali-baba)



Relevance should be revealed at **real time**, as entities may lose their relevance in future time points.

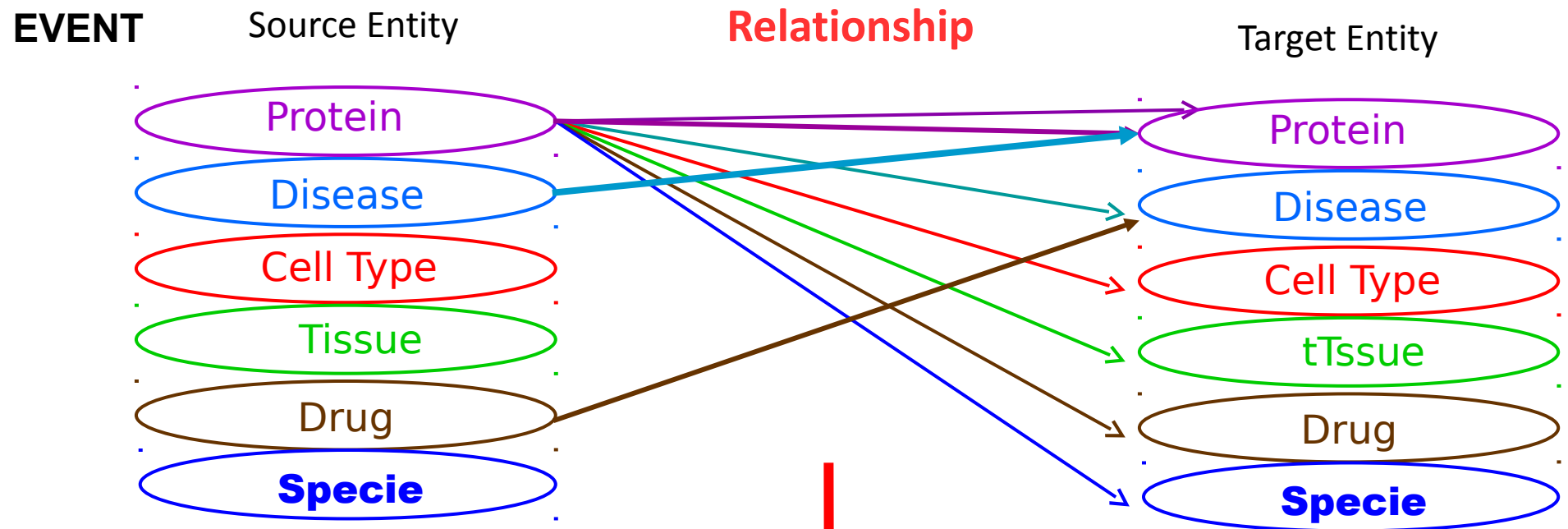
e is Highly Targeted by other bio-entities at time t

Methods

Definition 2

A recognized biological entity e is defined as being relevant at time t (or t -relevant) if it achieves a maximum of relationships at time t with other recognized biological entities.

Time relevance may be provided for different sorts of biological entities



Different valuable information

Our Use Case Study

10 years Heparidin publications

Large Scale Extraction of biological entities and events, using Ali-baba web service

do it for each period $p = 1$ month, from Dec 2000 to Dec 2011

Amount of background knowledge published during the Heparidin decade

Collected Events

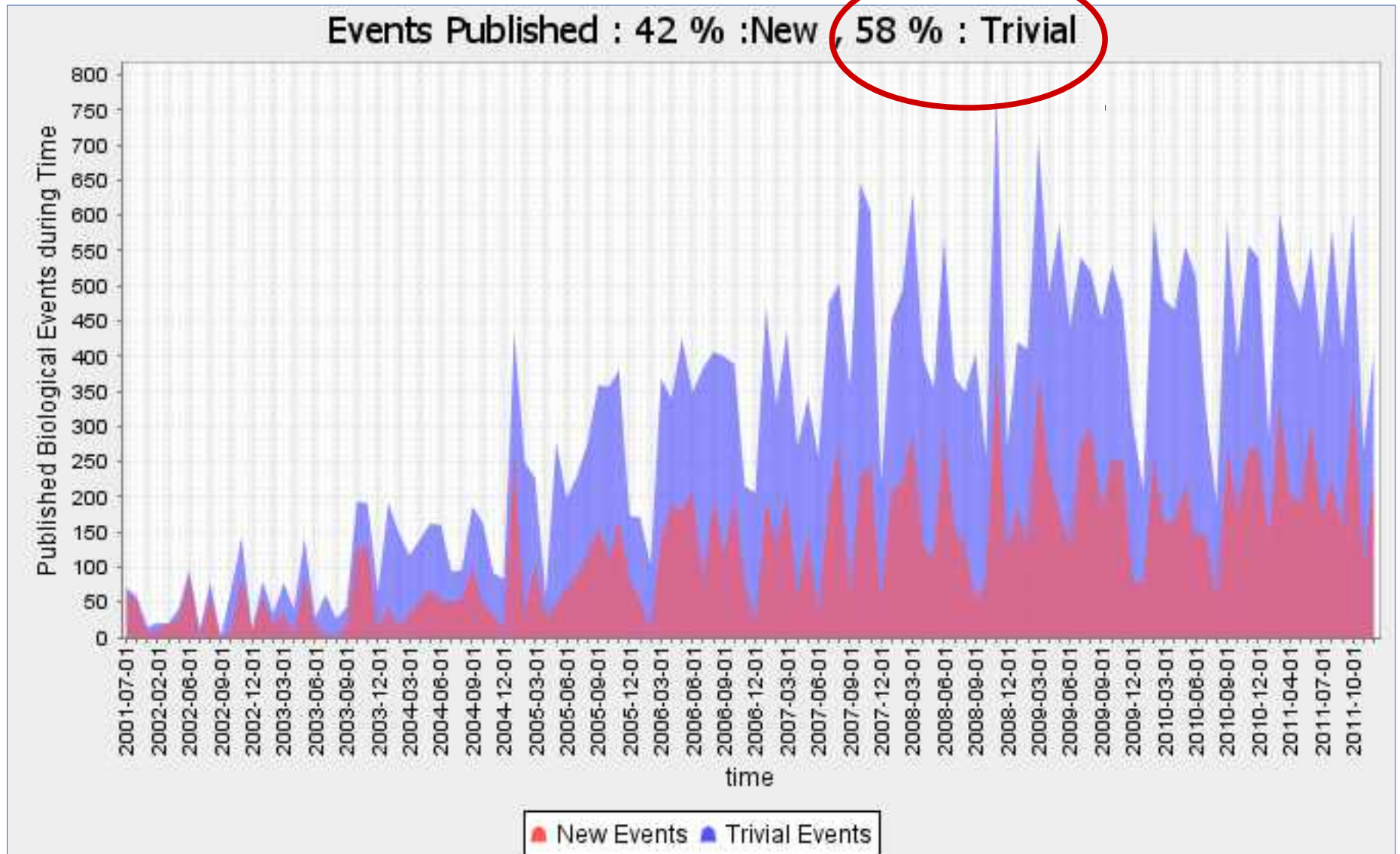


Highlight retrospectively

Time relevance :
Provide periodically highly targeted proteins diseases, drugs, cell types and tissues.

Results

Cumulative Quantification of Background Information Published

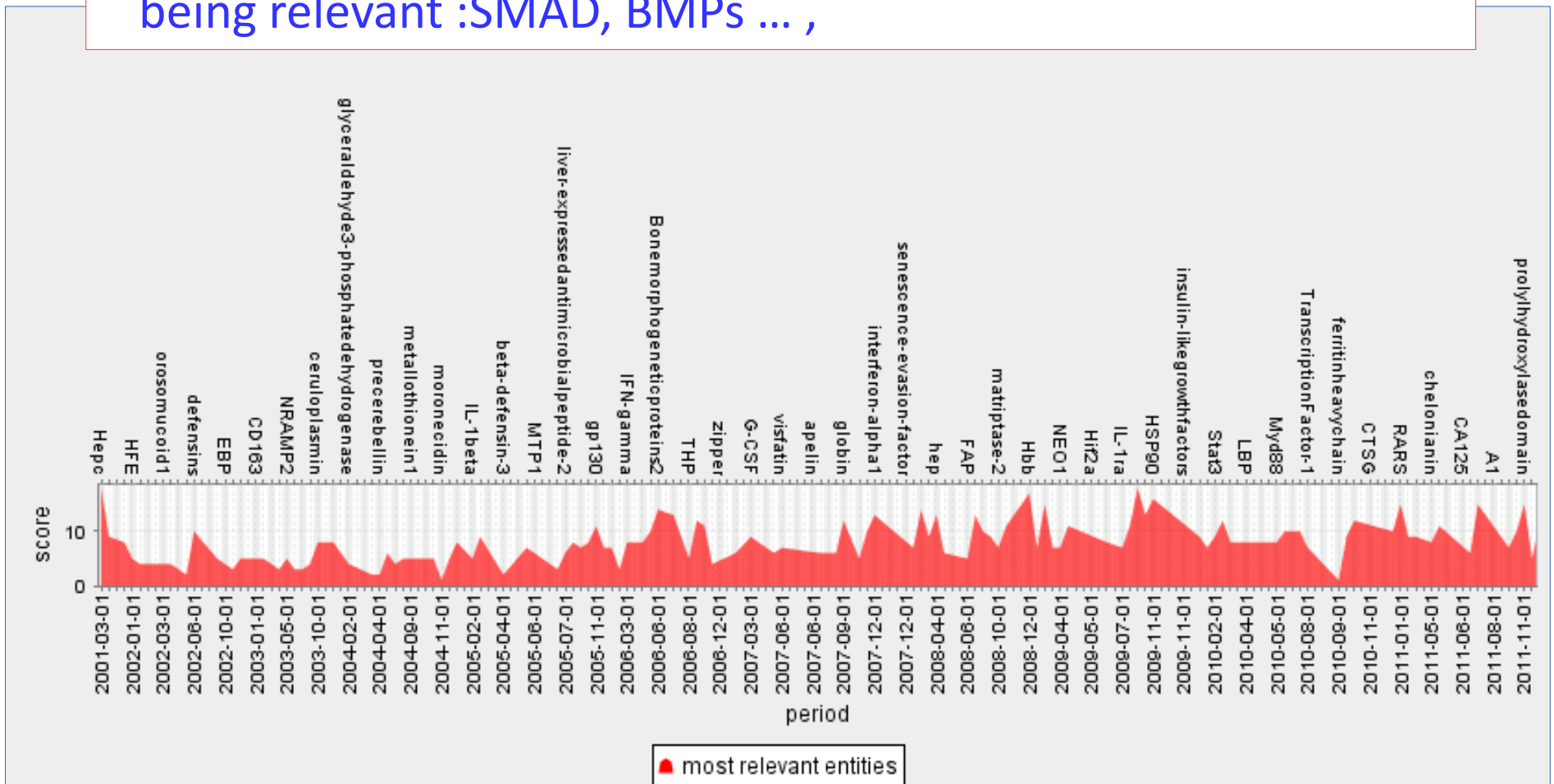


Results

Highly Targeted Proteins Over Time

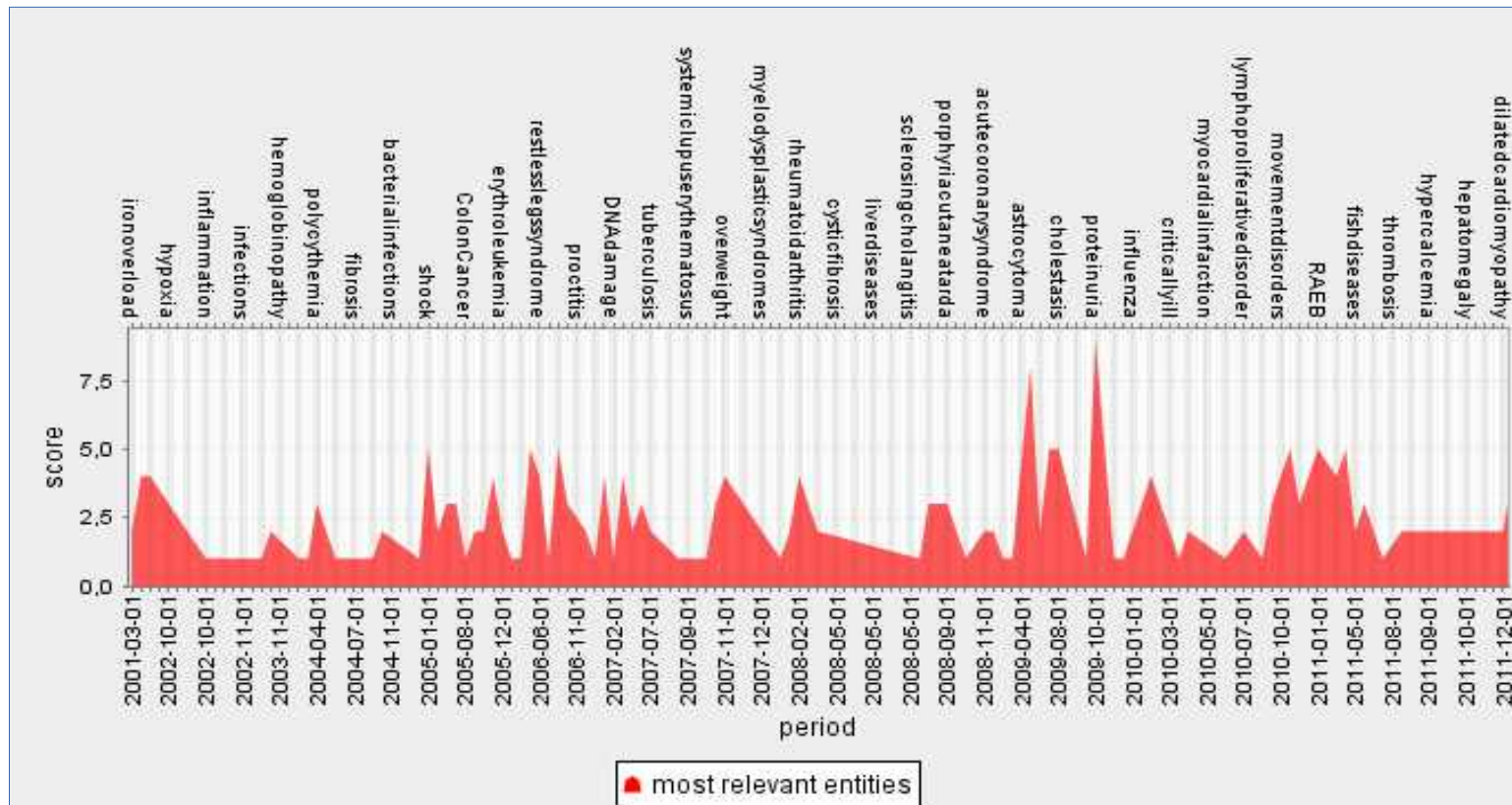
After filtering Background :

1. Drastic Fall of data (events)
2. From the expert side : non expected proteins emerge as being relevant :SMAD, BMPs ... ,



Results

Time Relevant Diseases

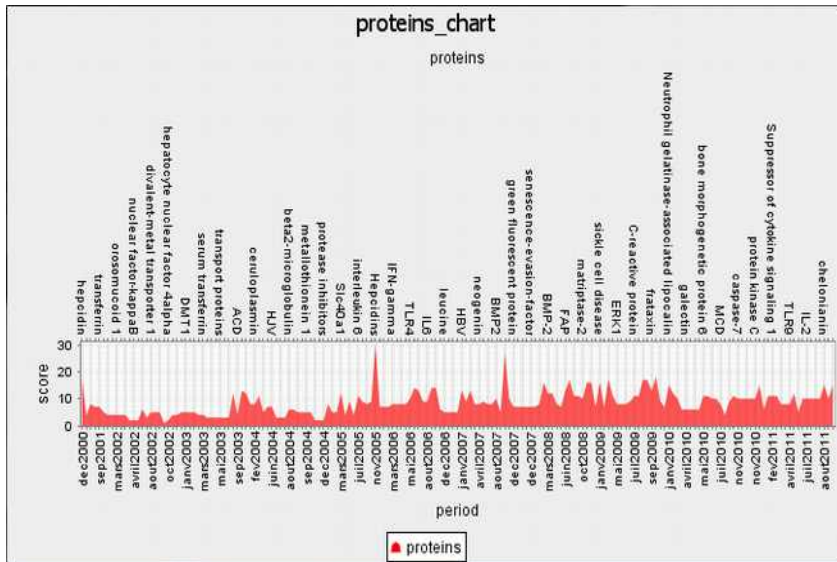


After Clearing Background

New diseases linked to Hepcidin and iron emerge as relevant like fish diseases and neurological diseases

Results

We may provide for biologists more annotations of the time relevant entities



Number	Date	HighTargetedEntity	Number of targets	#Abstracts	Swissprotid
28	mars2003	reverse transcriptase	4	1	http://www.uniprot.org/uniprot/...
29	mars2003	serum transferrin	4	3	http://www.uniprot.org/uniprot/...
30	avril2003	interleukin-6	3	1	http://www.uniprot.org/uniprot/...
31	avril2003	tumor necrosis factor alpha	3	1	http://www.uniprot.org/uniprot/...
32	mai2003	hepcidin 1	3	1	http://www.uniprot.org/uniprot/...
33	mai2003	transport proteins	3	1	http://www.uniprot.org/uniprot/...
34	juin2003	beta-2microglobulin	3	1	http://www.uniprot.org/uniprot/...
35	juin2003	TfR1	3	1	http://www.uniprot.org/uniprot/...
36	juil2003	IL6	12	8	http://www.uniprot.org/uniprot/...
37	sep2003	ACD	4	2	http://www.uniprot.org/uniprot/...
38	oct2003	ferroportin 1	13	6	http://www.uniprot.org/uniprot/...
39	nov2003	Nramp2	12	8	http://www.uniprot.org/uniprot/...
40	janv2004	reductase	8	5	http://www.uniprot.org/uniprot/...
41	fev2004	ceruloplasmin	8	1	http://www.uniprot.org/uniprot/...
42	mars2004	ferroportin-1	11	6	http://www.uniprot.org/uniprot/...
43	avril2004	hepc2	5	4	http://www.uniprot.org/uniprot/...
44	mai2004	IL-6	7	4	http://www.uniprot.org/uniprot/...
45	juin2004	HJV	7	4	http://www.uniprot.org/uniprot/...
46	juil2004	hepatocyte growth factor-lik...	3	1	http://www.uniprot.org/uniprot/...
47	juil2004				
48	juil2004				
49	aout2004				
50	aout2004				
51	sep2004				

Firefox

Serum hepcidin in clinical speci... x An Hfe-dependent pathway med... x +

www.ncbi.r

Br J Haematol. 2003 Sep;122(6):996-1000.

Serum hepcidin in clinical specimens.

Dallaglio G, Fleury T, Means RT.

Hematology Oncology Division, Department of Medicine, Ralph H. Johnson VA Medical Center a University of South Carolina, Charleston, SC 29425, USA. meansr@musc.edu

Abstract

The hepatic antimicrobial protein, hepcidin, is implicated in duodenal iron absor mobilization. Overexpression of the hepcidin gene is associated with a hypoferr iron refractory anemia. On the basis of these observations, it has been proposed

Firefox

Acyl-CoA dehydrogenase - Bacillus subt...

www.uniprot.org/uniprot

Names · Attributes · General annotation · Ontologies · Sequence annotation · Sequences · References · Cross-refs · Entry info · Documents · Customize order

Protein names

Recommended name:
Acyl-CoA dehydrogenase
EC=1.3.99.-

Gene names

Name: **acdA**
Synonyms: **acd**
Ordered Locus Names: **BSU37170**

Conclusion - Discussion

- ★ RetroMine approach is straightforward but is extremely helpful for providing **in-depth retrospective studies** to researchers on subjects of their interest.
- ★ Torrential data are extracted using advanced text mining tools and **mined in a second round** to draw unexpected patterns of biological entities and events behaviors over time.
- ★ RetroMine enhance comprehension of the extracted events :
 - by introducing **chronology** using TIME
 - by giving priority to non background biological entities, highly targeted over time
- ★ RetroMine revealed the considerable amount of **background information** published periodically in the biomedical literature
- ★ This work is still ongoing. Current developments :
 - Toward a generalization to any query on biological entities – substitute ali-baba
 - Mining Microbiota Litterature

Thanks to ...

Bertrand Ameline de Cadeville
Master **MSB** Université de Rennes1



Olivier Loréal
INSERM U991 **Iron and Liver** group
Supports from ANR IronReg



Ulf Leser
Astrid Rheinländer
Knowledge Management in Bioinformatics group
HU-Berlin



ESF
European Science Foundation

