The OntoGene system: an advanced information extraction application for biological literature

www.ontogene.org

Fabio Rinaldi

FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

NOVARTIS

University of Zurich[UZH]

Outline

- Motivation, brief history
- OntoGene approach
- Evaluation (shared tasks)
- SASEBio: from text mining to interactive curation
- Recent developments
  - PharmGKB
  - CTD
  - BioTermEvo (Gintare)

Motivations and History

- Motivation: prove that NLP technologies are mature enough for real world applications
- Target: biomedical text mining
  - Richness of terminological resources (grounding!)
  - Large text DBs - potential interest from bio comm.
- Goal: help organize the knowledge space of the biomedical sciences.
- Started in late 2004 with applications combining terminology structuring and dependency parsing.

OntoGene KB Query

OUTPUT FORMAT: XHTML CSV

CORPUS: ATCR GENIA

Relation: activate Agent: #protein Target: HIV

Submit

Use a verb or relational noun as the value for Relation (e.g. regulate, control, activation).
Use an entity name as the value for Agent and Target (e.g. cca1 for the ATCR corpus, hiv-1 for the GENIA corpus).
Alternatively, you can use a type restriction (which must start with the '#' character) as value for Agent or Target.

Some examples for GENIA:
Relation:*activate* Agent: Target:
Relation:*control* Agent: *#protein_molecule* Target:
Relation: Agent: *hiv-1* Target:

Some examples for ATCR:
Relation:*express* Agent: Target:
Relation:*control* Agent: *#GeneProtein* Target:
Relation: Agent: *cca1* Target:

Valid type restrictions for ATCR are: *#BiolProcess, #GeneProtein, #Compound*
Some possible type restrictions for GENIA are: *#protein, #virus, #organism, #nucleotide,...*
The full set of GENIA types is defined by the GENIA Ontology.
Notice that a type matches itself and all its subtypes, e.g. *#organism* matches *#virus*.

Temporary Notes: Currently only XHTML output is available.
A query can take up to 15 seconds, depending on its complexity.
WARNING: the more 'unrestricted' the query is (e.g. using very generic types or unspecified fields), the longer it might take.
In case of problems with this demo, please contact Fabio Rinaldi.

Last modified: Thu Jul 6 19:56:12 CEST 2006

OG-RM

sid	Sentence
m92013023-s1 SVG	Anti-CD2 receptor antibodies activate the HIV long terminal repeat in T lymphocytes .
m91355651-s5 SVG	We found that in both cell lines , both phorbol ester and TNF alpha were able to activate NF-kappa B .
m91355651-s5 SVG	We found that in both cell lines , both phorbol ester and TNF alpha were able to activate NF-kappa B .
m94148994-s9 SVG	These data suggest that interferon regulatory factor 1 not only triggers the activation of the interferon signal transduction pathway , but also may play a role in limiting the duration of this response by activating the transcription of IRF-2 .
m92107162-s5 SVG	The simian virus 40 early promoter is also synergistically activated by the Z/c-myb combination .
m91237803-s2 SVG	Human herpesvirus 6 (HHV-6) can activate the human immunodeficiency virus [HIV] promoter and accelerate cytopathic effects in HIV-infected human T cells .

GENIA

Netscape

ifi http://www.ifi.unizh....in/rinaldi/genia.cgi

MEDLINE:95221892

cDNA cloning of a NGFI-B/nur77-related transcription factor from an apoptotic human T cell line.

A human T lymphoid cell line, PEER, dies by apoptosis in the presence of PMA and calcium ionophore. A new gene, TINUR, was cloned from apoptotic PEER cells. The expression of the TINUR gene is induced within 1 h after the cross-linking of the T cell Ag receptor complex. TINUR belongs to the NGFI-B/nur77 family of the steroid receptor superfamily and is an orphan receptor. **TINUR binds to the same DNA sequence as NGFI-B/nur77.** We also propose that the NGFI-B/nur77 family can be classified into two subtypes.

MEDLINE:95363089

Multiple signals are required for function of the human granulocyte-macrophage colony-stimulating factor gene promoter in T cells.

The human granulocyte-macrophage CSF (GM-CSF) gene is expressed in T cells in response to TCR activation that can be mimicked by treatment of the cells with PMA and Ca2+ ionophore. The gene contains a proximal functional promoter region (-620 to +34), as well as a powerful enhancer located 3 kb upstream, both of which are involved in the response of the gene to TCR activation. The proximal promoter contains a region termed CLEO (-54 to -31) that consists of a purine-rich element abutting an activator protein-1 (AP-1)-like site, as well as an upstream nuclear factor-kappa B (NF-kappa B) site (-85 to -76) and a CK-1 element (-101 to -92). We show in this work that mutations in either the purine-rich region of the CLEO element or the NF-kappa B site result in reduced PMA/Ca2+ activation of a 620-bp human GM-CSF promoter-luciferase reporter construct in Jurkat T cells by 65% and 50%, respectively. **The major inducible protein complex that binds to the human CLEO (hCLEO) element is an AP-1-like complex that is inducible by PMA alone, but shows increased binding in response to PMA together with Ca2+ ionophore.** Although the binding of this complex is not cyclosporin-sensitive, promoter induction is inhibited by cyclosporin treatment. A second weak inducible complex resembling nuclear factor of activated T cells (NF-AT) was also observed binding to the hCLEO region. By using recombinant proteins, we confirmed that AP-1, NF-ATp, and a higher order NF-ATp/AP-1 complex could all form with the hCLEO element, and we have also defined the sequence requirements for binding of each of these complexes. We found that expression of a constitutively active form of calcineurin could substitute for Ca2+ ionophore and synergize with PMA to activate the GM-CSF promoter, and conversely that mutant-activated Ras could substitute for PMA and cooperate with Ca2+ ionophore. Co-expression of Ras and calcineurin, however, did not activate the GM-CSF promoter, but required the additional expression of NF-kappa B p65. These results imply that at least three signals are required to activate the GM-CSF proximal promoter, and that the signals impinge on distinct transcription factors that bind to the hCLEO and NF-kappa B regions of the promoter.

MEDLINE:95286632

The transcription factor, Nm23H2, binds to and activates the translocated c-myc allele in Burkitt's lymphoma.

We have identified an in vivo footprint over the PuF site on the translocated c-myc allele in Burkitt's lymphoma cells. The PuF site on the silent normal c-myc allele was unoccupied. We demonstrated by electrophoretic mobility shift assay, electrophoretic mobility shift assay with antibody, UV cross-linking followed by SDS-gel electrophoresis, and Western analysis that Nm23H2 in B cell nuclear extracts bound to the c-myc PuF site. Transfection experiments with c-myc promoter constructs in both DHL-9 and Raji cells revealed that the PuF site functioned as a positive regulatory element in B cells with a drop in activity with mutation of this site. Access to this site is blocked in the normal silent c-myc allele; these data suggest that the Nm23H2 protein is involved in deregulation of the translocated c-myc allele in Burkitt's lymphoma cells.

完了

References

- Fabio Rinaldi, Gerold Schneider, Kaarel Kaljurand, Michael Hess, Martin Romacker. An environment for relation mining over richly annotated corpora: the case of GENIA. BMC Bioinformatics 2006, 7(Suppl 3):S3. doi:10.1186/1471-2105-7-S3-S3

BC II (2006): approach

- Annotate entities using reference DBs as source
- Disambiguate proteins according to ORG distribution
- Give each ID a score according to freq and position
- Combine Ids in the same syntactic span
- Use manually constructed syn patterns to filter out unlikely pairs
- Use novel/background filter to identify sentences likely to convey the 'core' message
- Results: 3rd best

First SNF project

- "Detection of Biological Interactions from Biomedical Literature" (SNF 100014-118396/1)
- Funding: SNF and Novartis
- Duration: 18 months (April 2008 – October 2009)
- Main focus: IntAct database
  - Experimental methods (SMBM 2008)
  - Organisms (BioNLP 2009)
  - Entities (AIME 2009)
  - Interactions (CICLING 2009)

Experiment ? PSI-MI 1.0 PSI-MI 2.5	Name	Ac	Interaction detection	Participant identification	Host
	lucas-1999-1	EBI-1555830	anti bait coip	western blot	mouse-fdcp_1
Description	A novel spliced form of SH2-containing inositol phosphatase is expressed during myeloid development.				
Annotation	author-list	Lucas DM., Rohrschneider LR.			
	journal	Blood (0006-4971)			
	publication year	1999			
	dataset	Cancer - Interactions investigated in the context of cancer			
Xref	pubmed	10068665	-	Type: primary-reference	
	newt	10090	mouse	Type: target-species	
☐ Interaction ?	Name	Ac	Interaction type	Dissociation constant (Kd) M	
	grb2-inpp5d-1	EBI-1555843	physical association	-	
Description	Grb2 coimmunoprecipitates SHIP in FDC-P1 lysates				
Annotation	figure legend	7			
	agonist	M-CSF (macrophage colony-stimulating factor) - This interaction occurs when cells are treated with M-CSF, which causes tyrosine phosphorylation of SHIP.			

Interacting molecules	Name	Ac	Interactor type	Stoichiometry	Interactor description	Expression system	Identifier	Gene name	Role
	☐ Q9ES52-1	EBI-1452545	protein	-	Phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase 1	-	Q9ES52-1	-	prey
	☐ grb2_mouse	EBI-1688	protein	-	Growth factor receptor-bound protein 2	-	Q60631	-	bait

IntAct snippets

10369680 EBI-959516

Q8CCI5 RYBP_MOUSE RING1 and YY1-binding protein
O35730 RING1_MOUSE E3 ubiquitin-protein ligase RING1
MI:0096(pull down), pubmed:10369680, taxid:10090(mouse), taxid:10090(mouse), mouse-embryo:1

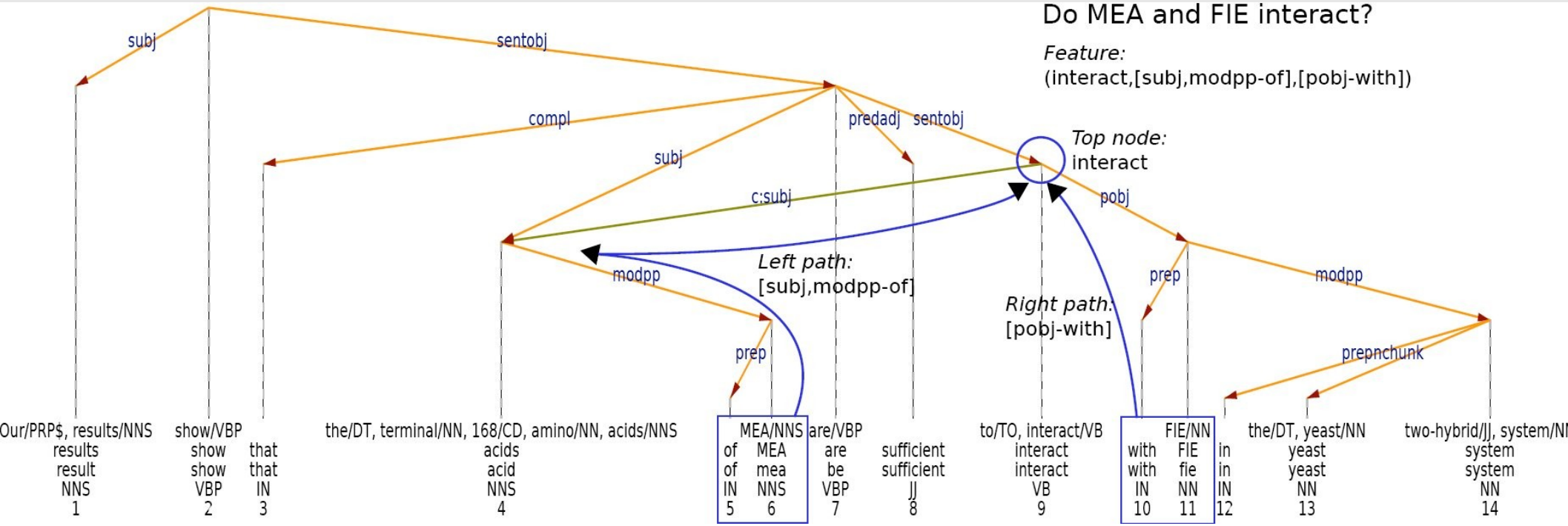
s61: **Ring1**[2] binds **RYBP**[2] and **M33** through the same **C** - terminal domain , whereas the **RYBP**[2] - **M33** interaction takes place through an **M33** domain not involved in **Ring1**[2] binding .

10369680 EBI-959540

P30658 CBX2_MOUSE Chromobox protein homolog 2
Q8CCI5 RYBP_MOUSE RING1 and YY1-binding protein
MI:0096(pull down), pubmed:10369680, taxid:10090(mouse), taxid:10090(mouse), human-293t:960

s62: **Ring1**[2] binds **RYBP**[2] and **M33** through the same **C** - terminal domain , whereas the **RYBP**[2] - **M33** interaction takes place through an **M33** domain not involved in **Ring1**[2] binding .

Syntactic Filters



PPI in BC II.5 (2009)

- All candidate pairs in a sentence are considered
  - Entity recognition and disamb. learnt from IntAct
  - One semi-automated submissions (ORG selection)
- Candidate pairs are scored, according to:
  - Pair salience; Zoning; Novelty score; Known interaction; Syntactic paths;
  - Syntax: now using learning to derive syn patterns from manually annotated corpus
- Results: best according to “raw” AUC iP/R

Annotated Abstract

The Cap - binding protein eIF4E promotes folding of a functional domain of yeast translation initiation factor eIF4G1 .

The association of eucaryotic tra IF4E_YEAST ation factor eIF4G with the cap - binding protein eIF4E establishes a critical link between the mRNA and the ribosome during translation initiation .

This association requires a conserved seven amino acid peptide within eIF4G that binds to eIF4E .

Here we report that a 98 - amino acid fragment of S . cerevisiae eIF4G1 that contains this eIF4E binding peptide undergoes an unfolded to folded transition upon binding to eIF4E .

The folding of the eIF4G1 domain was evidenced by the eIF4E - dependent changes in its protease sensitivity and (1) H - (15) N HSQC NMR spectrum .

Analysis of a series of charge - to - alanine mutations throughout the essential 55.4 - kDa core of yeast eIF4G1 also revealed substitutions within this 98 - amino acid region that led to reduced eIF4E binding in vivo and in vitro .

These data suggest that the association of yeast eIF4E with eIF4G1 leads to the formation of a structured domain within eIF4G1 that could serve as a specific site for interactions with other components of the translational apparatus .

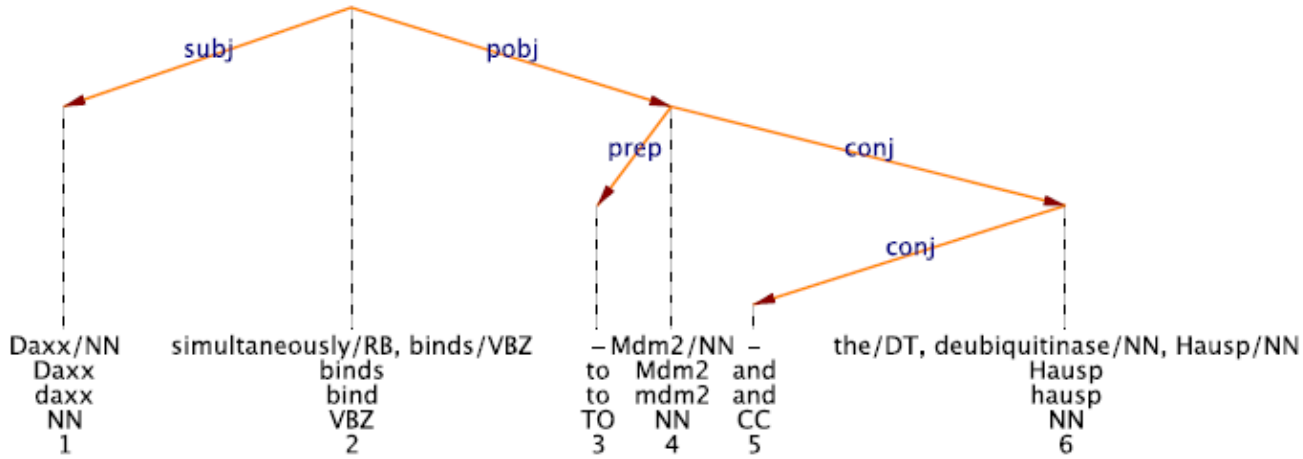
They also suggest that the stability of the native eIF4E - eIF4G complex is determined by amino acid residues outside of the conserved seven - residue consensus sequence .

Protein Interactions (IPS)

- Parse all positive sentences
- Apply lexico-syntactic patterns as filters
- Interactions which do not 'pass' a filter are discarded
- Results: P: 54.37%, R: 18.39%, F: 27.49%

The predicate dep(TYPE,HEAD,DEPENDENT) represents syntactic relations among the constituents of the sentence, the predicate prot(PROT) identifies a protein.

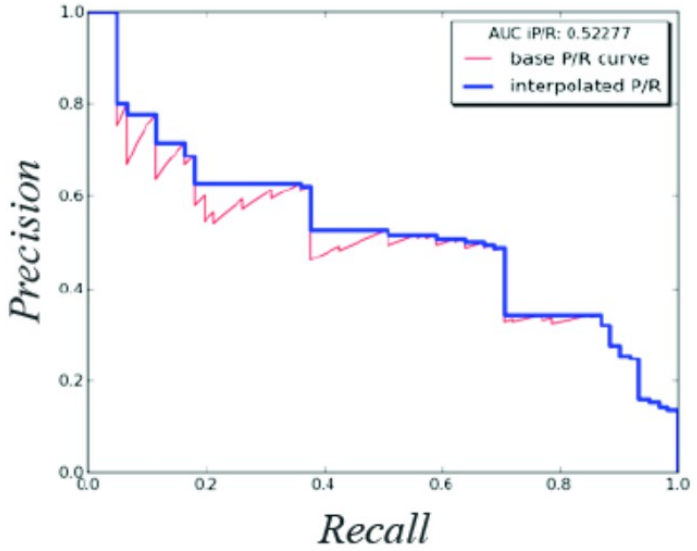
```
dep(subj,bind,Daxx),
dep(pobj,bind,Mdm2),
dep(conj,Mdm2,Hausp),
dep(prep,Mdm2,to),
dep(conj,Hausp,and),
prot(Daxx),
prot(Mdm2),
prot(Hausp).
```



Importance of ranking

Conf	Type 1	Concept 1	Name 1	Type 2	Concept 2	Name 2	N
1.00	Disease	PA446155	Precursor Cell Lymphobla...	Gene	PA245	MTHFR	
0.80	Disease	PA446155	Precursor Cell Lymphobla...	Gene	PA31236	MTHF...	
0.60	Drug	PA450428	methotrexate	Gene	PA245	MTHFR	
0.59	Drug	PA449692	folic acid	Gene	PA245	MTHFR	
0.58	Disease	PA445506	Recurrence	Gene	PA245	MTHFR	

- MRR
- MAP
- AUC iP/R
- TAP-k



SASEBio

- Semi-Automated Semantic Enrichment of the Biomedical Literature
- Funding by SNF (grant 105315_130558/1) and Novartis
- Duration: 3 years
- Positions: 2 post-docs, 1 PhD
- Goals:
  - Improve our text mining technologies
  - Make the tools relevant to potential users

SASEBio: activities so far

- CALBC: large scale entity extraction
- BC III (2010): successful participation to all tasks
- PharmGKB assisted curation experiment
- Terminology evolution studies
- BC 2012: best overall results in “triage” task for CTD

CALBC (2010)

- Large-scale entity extraction (900K abstracts)
- CALBC I: 3rd place for diseases (F:84%) and species (F:78%) against Silver Corpus I
- Best results for diseases and species against harmonized voting Silver Corpus II
- Challenges:
  - Processing large XML collections
  - Harmonize annotations
  - Efficiency of annotation process

BioCreative III (2010)

- Good results in all tasks
  - GN: Gene Normalization
    - Middle-rank results
  - PPI-ACT: binary classification of PPI papers
    - Top-rank results
  - PPI-IMT: find experimental methods in papers
    - Top-rank results
  - IAT: experimental interactive task
    - Positive comments from curators about usability

IAT: ODIN

File ▾ View ▾ Search ▾ Mode ▾ Help ▾

Inspectors

Term Inspector

Save | Cancel | Remove Term | Show

Term:

Gfi-1B

Term Type:

GEN

Concept Values:

14582_MOUSE 8328_HUMAN

Comment:

Search Databases

Search Terms

☑ EMBL ☑ UniProtKB

Document PMC 1226314

Show PubmedCentral | Show Pubmed Entry

GATA-1 mediates auto-regulation of Gfi-1B transcription in K562 cells

Abstract Gfi-1B (growth factor independence - 1B) gene is an erythroid-specific transcription factor , whose expression plays an essential role in erythropoiesis . Our laboratory has previously defined the human Gfi-1B promoter region and shown that GATA-1 mediates erythroid-specific Gfi-1B transcription . By further investigating the regulation of the Gfi-1B promoter , here we report that (i) Gfi - 1B transcription is negatively regulated by its own gene product , (ii) GATA-1 , instead of Gfi-1B , binds directly to the Gfi-1 - like sites in the Gfi-1B promoter and (iii) Gfi-1B suppresses GATA-1 - mediated stimulation of Gfi-1B promoter through their protein interaction . These results not only demonstrate that interaction of GATA-1 and Gfi-1B participates in a feedback regulatory pathway in controlling the expression of the Gfi-1B gene , but also provide the first evidence that Gfi-1B can exert its repression function by acting on GATA-1 - mediated transcription without direct binding to the Gfi-1 site of the target genes . Based on these data , we propose that this negative auto-regulatory feedback loop is important in restricting the expression level of Gfi-1B , thus optimizing its function in erythroid cells .

INTRODUCTION Gfi-1B (growth factor independence - 1B) is an erythroid-specific Gfi - family transcriptional factor , which was identified by low stringency hybridization screening with a partial Gfi-1 cDNA probe (1) . Both Gfi-1 and Gfi-1B contain a SNAG domain that mediates transcriptional repression and a zinc finger domain at its C - terminus for their DNA binding to the TAAATCAC (A/T) GCA recognition sequence (1 – 3) . Expression of Gfi-1B is confined in erythroid lineage cells and megakaryocytes in human (4,5) , whereas Gfi-1 is more abundant in the lymphopoietic thymus (6 – 8) . So far , p21 (cip1 / waf1) , Socs1 and Socs3 are known as the target genes of Gfi-1B - mediated transcriptional repression (1,9) . Since p21 is a cell cycle inhibitor and SOCS family members are known to suppress cytokine signaling , the functional role of Gfi-1B is considered to be important in controlling proliferation of erythrocyte/megakaryocyte-lineage cells . Its importance in erythropoiesis has been further highlighted by gene targeting experiment showing that Gfi-1B gene disruption results in embryonic lethality due to loss of red blood cell formation (10) . Enforced expression experiment in early erythroid progenitor cells has shown that Gfi-1B induces a drastic expansion of erythroblast independent of its SNAG repression domain with a parallel increase of GATA-2 expression , which is required for proliferation of erythroblasts (5) . On the other hand , a recent study has shown that Gfi-1B plays a critical role in terminal differentiation through its transcription repression function (11) . Likely , the function of Gfi-1B in erythropoiesis is highly dependent on cell stage and the sequence context of its targeted gene promoter . Despite the differential roles of Gfi-1B in different stages of differentiation , results of both studies indicate that elevation of Gfi-1B level alters the program of normal erythropoiesis (5,11) . However , it remains unclear how Gfi-1B expression is regulated in erythroid cells and whether there is a direct relationship between Gfi-1B and other transcription factor that is involved in erythropoiesis . The expression of many eukaryotic transcription factors has been shown to be auto-regulated positively and negatively (12 – 16) . In most auto-regulatory cases , a given factor binds to its own promoter and either activates or represses transcription . In this study , we observed negative auto-regulation of Gfi-1B in K562 cells . By analyzing the sequence of human Gfi-1B gene promoter region (17) , we found the presence of two tandem repeats of Gfi-1 - like sites located at − 59/− 56 and − 47/− 44 relative to its transcription start site . Very recently , a report has demonstrated that mouse Gfi-1B directly binds to the Gfi-1 binding sites near the mRNA transcription start site of the mouse Gfi-1B promoter and is able to auto-repress its own expression (18) . However , here we showed that mutations in these two Gfi-1 - like sites reduced the promoter activity of the human Gfi-1B promoter in K562 cells , indicating that these sites mediate transcriptional activation rather than silencing . By detailed DNA-binding analyses , we proved that GATA-1 , instead of Gfi-1B , is the main transcription factor preferentially binding to these non-typical GATA sites . Furthermore , we found that the Gfi-1B can form a complex with GATA-1 , by which GATA-1 - mediated transcription is repressed by Gfi-1B . Coincidentally , one recent report also showed that Gfi-1B forms a complex with GATA-1 and associates with the myc and myb promoters in mouse erythroleukemic (MEL) cells . Given the facts that overexpression of Gfi-1B in erythroid progenitors induces growth arrest and that expression of myc and myb is often associated with cell proliferation , they hypothesized that GATA-1 / Gfi-1B is a repressive complex that suppresses transcription of myc and myb genes (19) . Our results , on the other hand , present the first direct evidence that transcriptional repression function of Gfi-1B can work through its interaction with GATA-1 independent of its direct DNA binding to the gene promoter . Since our previous study has shown that GATA-1 is a necessary transcription factor for Gfi-1B expression , the auto-regulatory mechanism observed in this study reflects that the expression of Gfi-1B and the

Annotation

Genes/Proteins | Terms | Grouped Genes/Proteins

Refresh | Selected | Unselected | Export

i	Concept	Score	Freq	Type	Zone
☑	100041004_MOI	0	2	GEN	text
	Browse Entrez Gene entry 100041004 in separate window.				
☑	1026_HUMAN	0	2	GEN	text
	Browse Entrez Gene entry 1026 in separate window.				
☐	10661_HUMAN	0	1	GEN	text
☐	10993_HUMAN	1	4	GEN	text
☐	11262_HUMAN	0	1	GEN	text
☐	11770_MOUSE	0	2	GEN	text
☐	12575_MOUSE	0	2	GEN	text
☐	12700_MOUSE	0	1	GEN	text
☐	12702_MOUSE	0	2	GEN	text
☐	12703_MOUSE	0	2	GEN	text
☐	1387_HUMAN	0	1	GEN	text
☐	14247_MOUSE	0	1	GEN	text
☐	14248_MOUSE	0	1	GEN	text
☐	14460_MOUSE	0	128	GEN	title
☐	14461_MOUSE	0	1	GEN	text
☐	14581_MOUSE	0	25	GEN	abstract
☐	14582_MOUSE	0	230	GEN	title
☐	15452_MOUSE	0	1	GEN	text
☐	161882_HUMAN	0	1	GEN	text
☐	16909_MOUSE	0	1	GEN	text
☐	17274_MOUSE	0	2	GEN	text
☐	17863_MOUSE	0	5	GEN	text
☐	18045_MOUSE	0	1	GEN	text
☐	18521_MOUSE	0	1	GEN	text
☐	19165_MOUSE	0	2	GEN	text
☐	199699_HUMAN	0	2	GEN	text
☐	199_HUMAN	0	2	GEN	text
☐	20683_MOUSE	0	2	GEN	text
☐	21784_MOUSE	0	2	GEN	text
☐	22324_MOUSE	0	1	GEN	text

PharmGKB

- Provides manually annotated relationships between Drugs/Genes/Diseases (36557 as of Sep 30th, 2010)
- Annotation based on publications, pathways and RSIDs:
  - 26122 PMID
  - 5467 Pathway
  - 4968 RSID
- We consider only relationships derived from publications

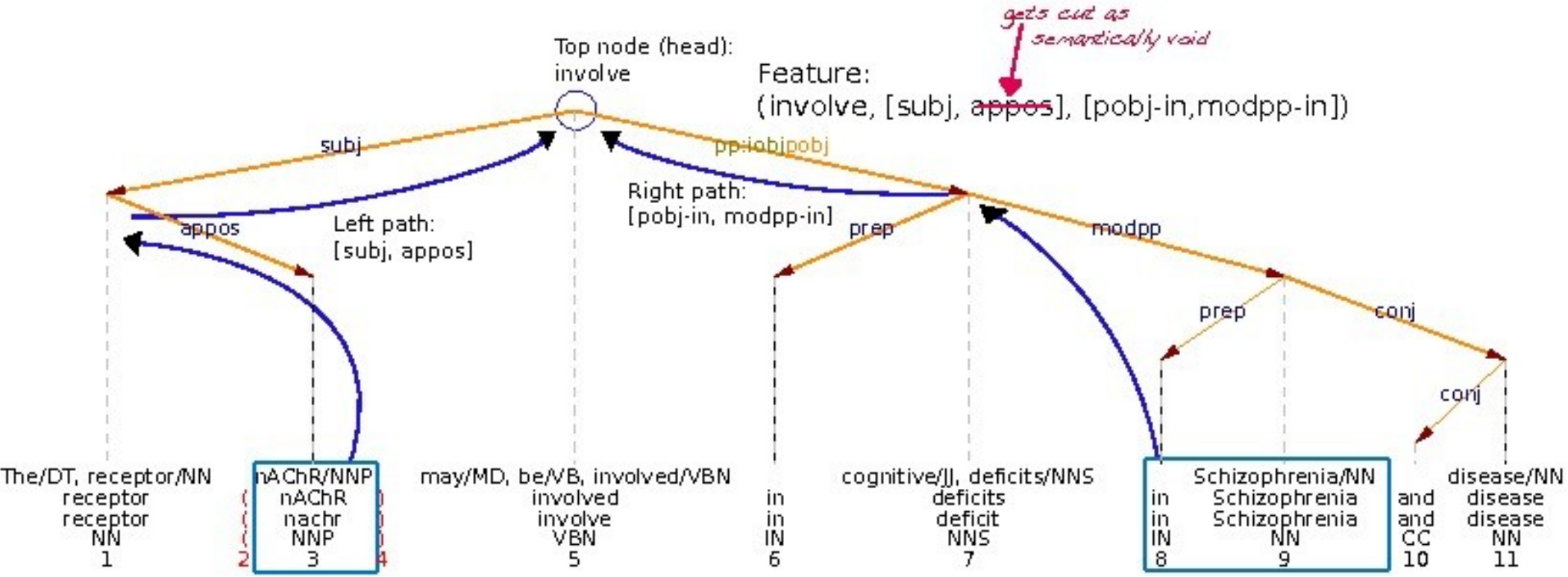
Approach

- Abstracts (5062) downloaded from PubMed
- Used the OG pipeline for entity annotation. Only terms derived from PharmGKB (Drugs: 30351 terms / 2986 ids, Diseases: 28633 terms / 3198 ids, Genes: 176366 terms / 28633 ids)
- Candidate interactions generated according to a set of different criterias (co-occurrence, syntax, ME)
- Comparison against “gold standard” using BioCreative II.5 PPI scorer

Creating a gold standard

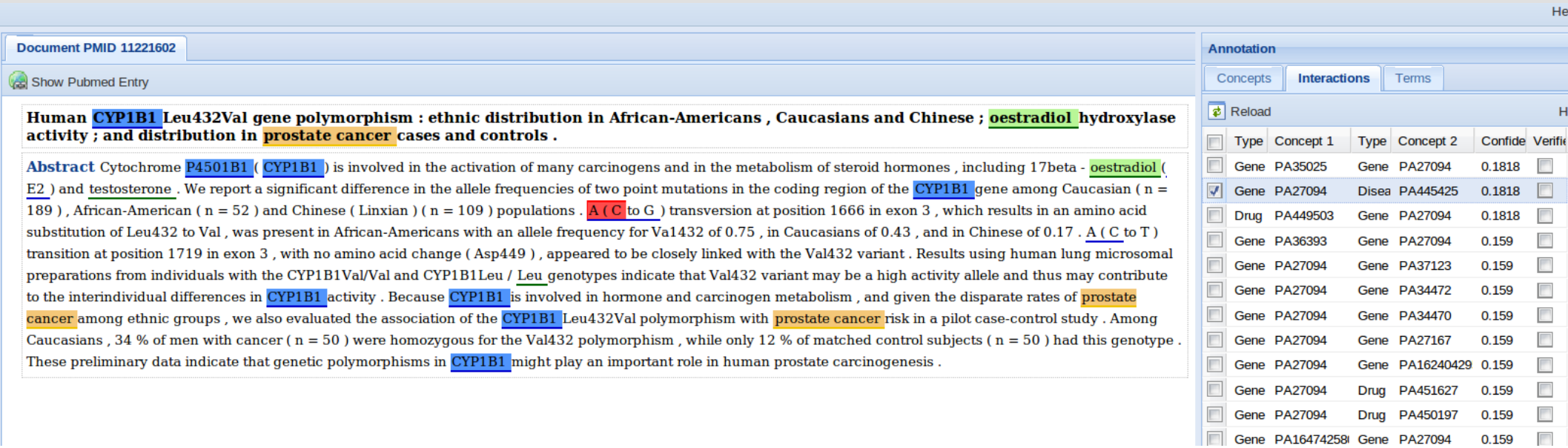
- The manually annotated interactions can be used to generate a gold standard
  - 10597 Gene/Drug
  - 9415 Gene/Disease
  - 4202 Drug/Disease
  - 928 Gene/Gene
  - 742 Drug/Drug
  - 238 Disease/Disease
- Total: 26122 interactions (24958 without duplicates)

Syntax-based approach

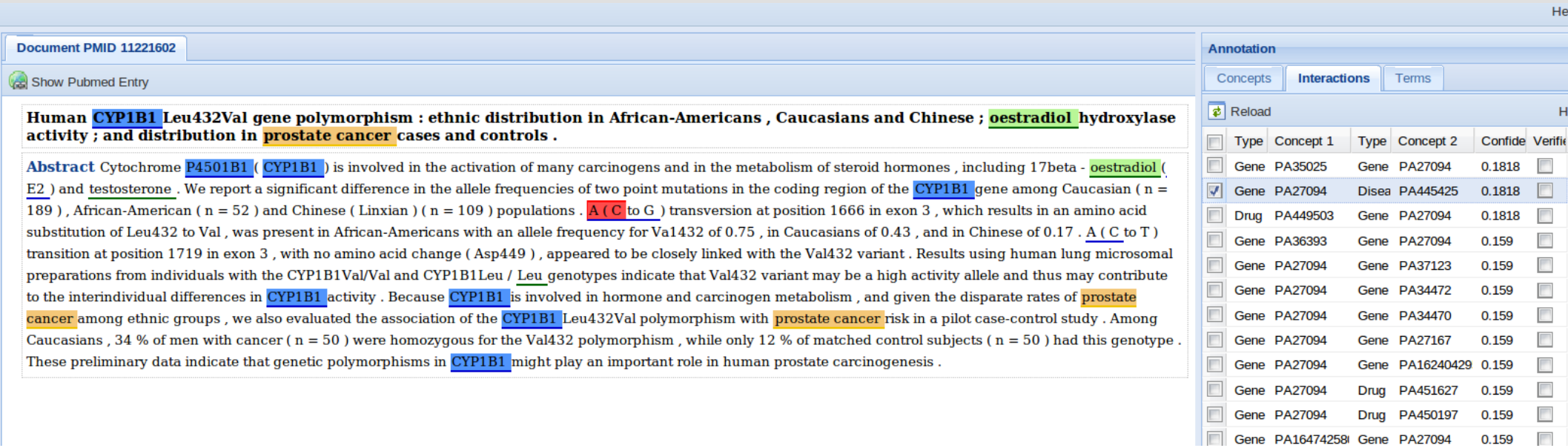


The neuronal nicotinic acetylcholine receptor alpha7 (nAChR alpha7) may be involved in cognitive deficits in Schizophrenia and Alzheimer's disease." [15695160]

Computed Interactions



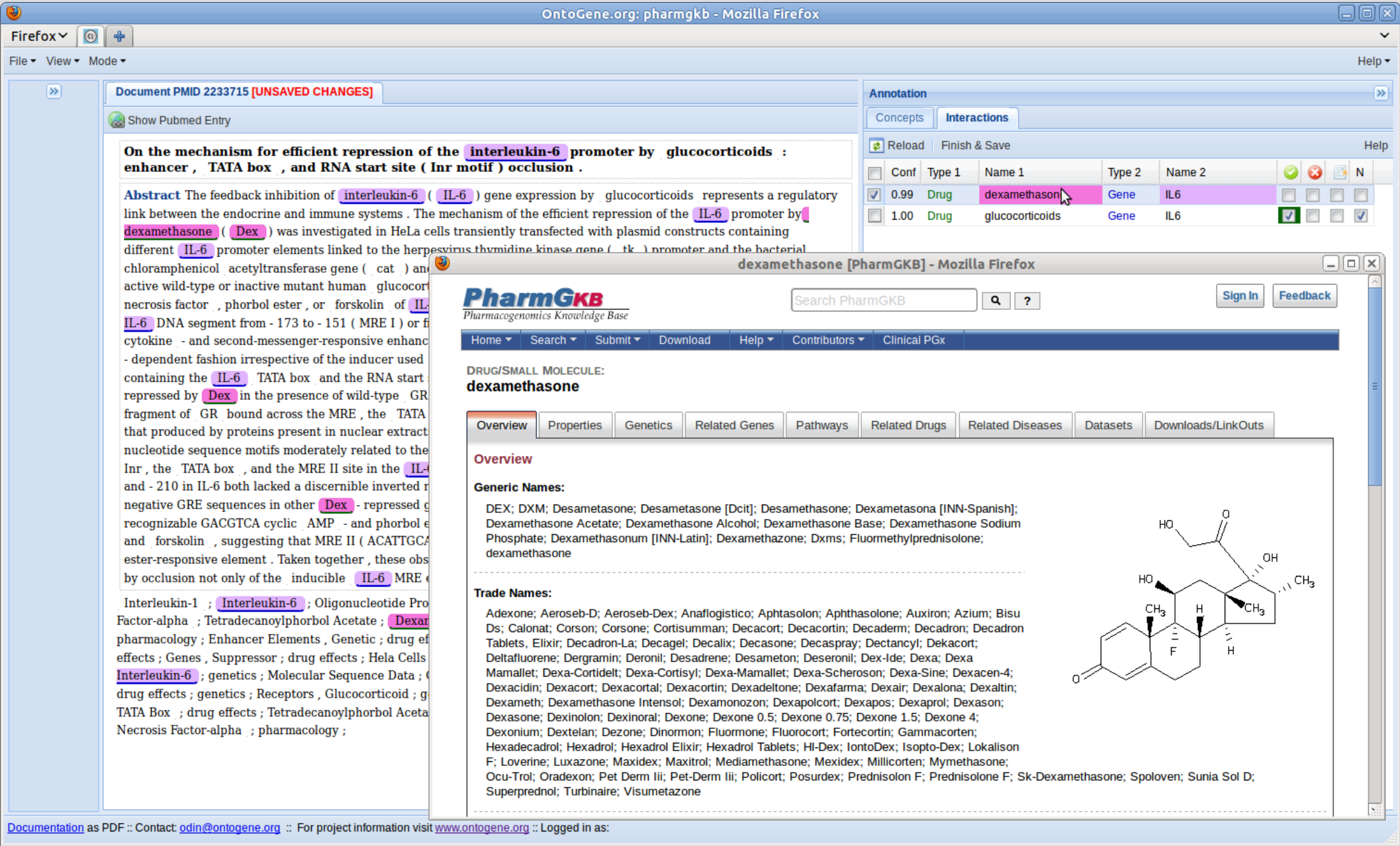
Computed Interactions

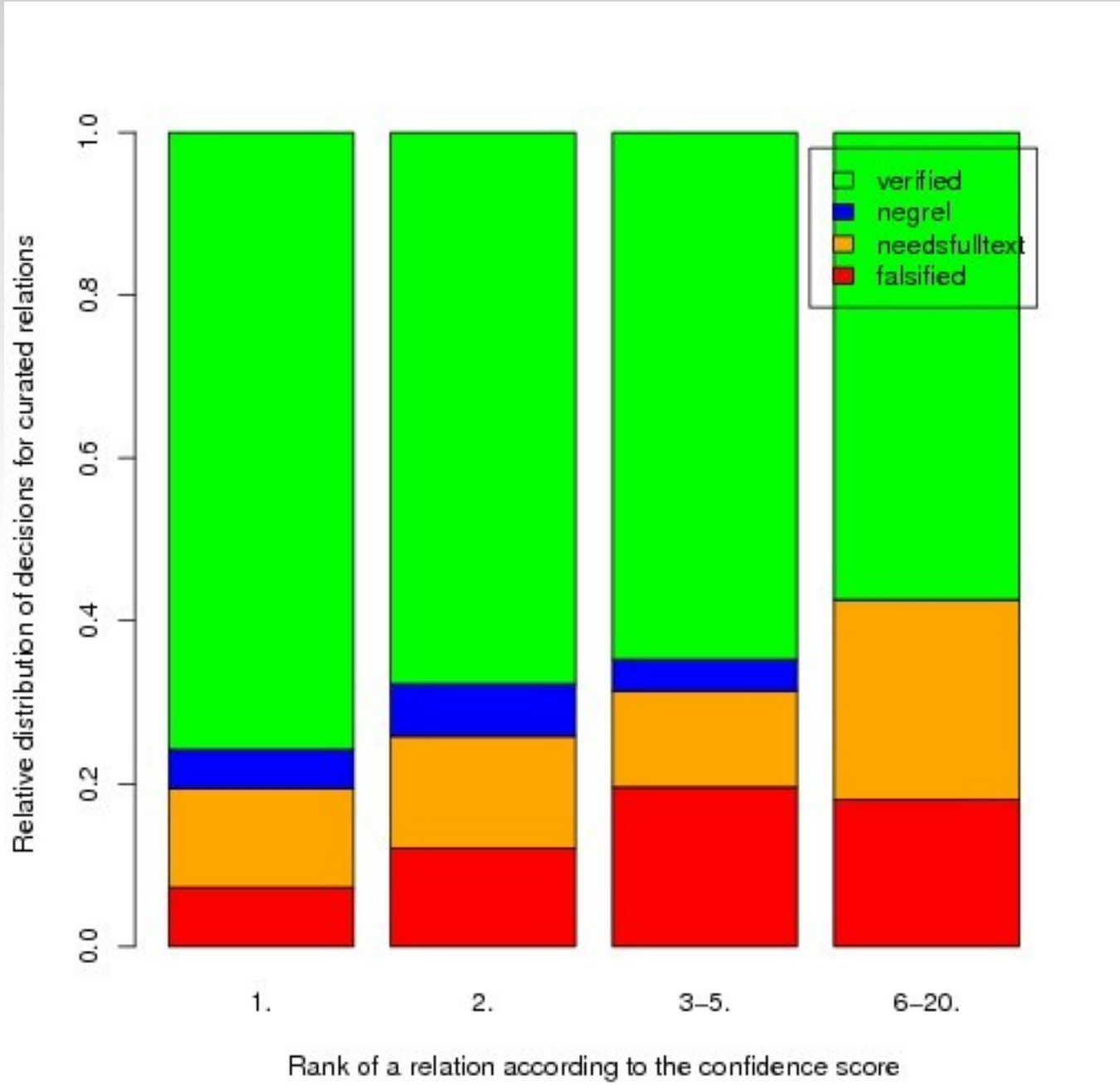


P = 30%, R = 28%, AUC = 22%

P = 7%, R = 66%, AUC = 28%

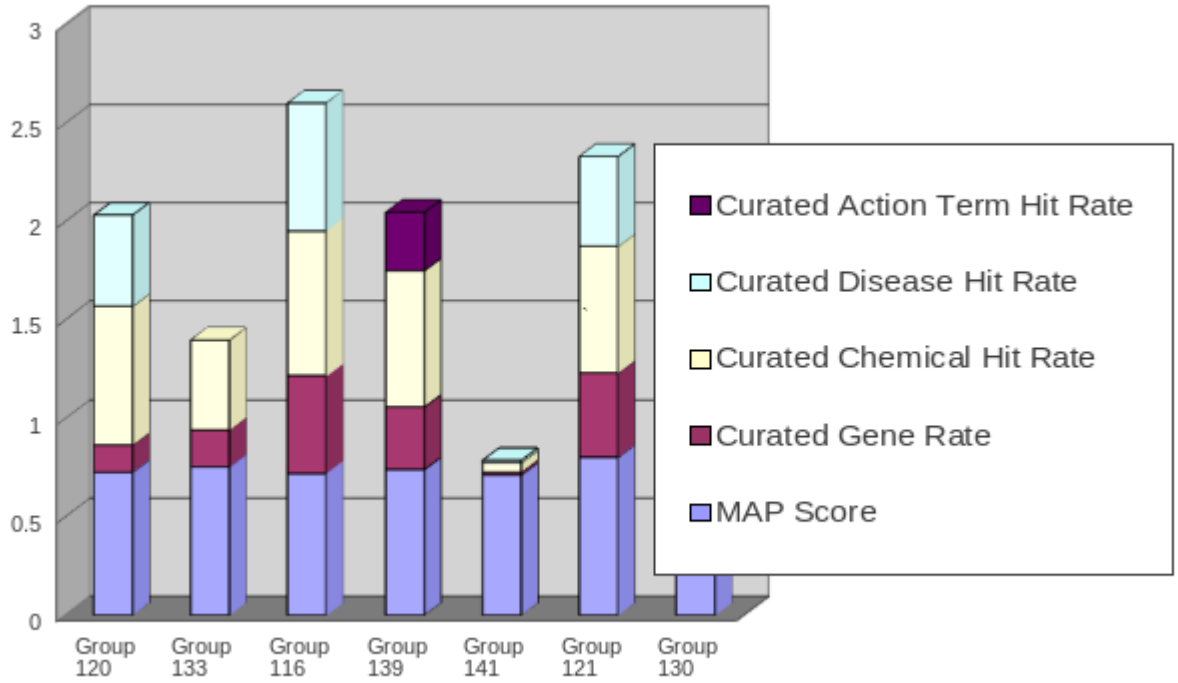
Interactive curation




Relative distribution of decisions for curated relations
0.0
0.2
0.4
0.6
0.8
1.0
verified
negrel
needsfulltext
falsified
1.
2.
3–5.
6–20.
Rank of a relation according to the confidence score

BioCreative 2012

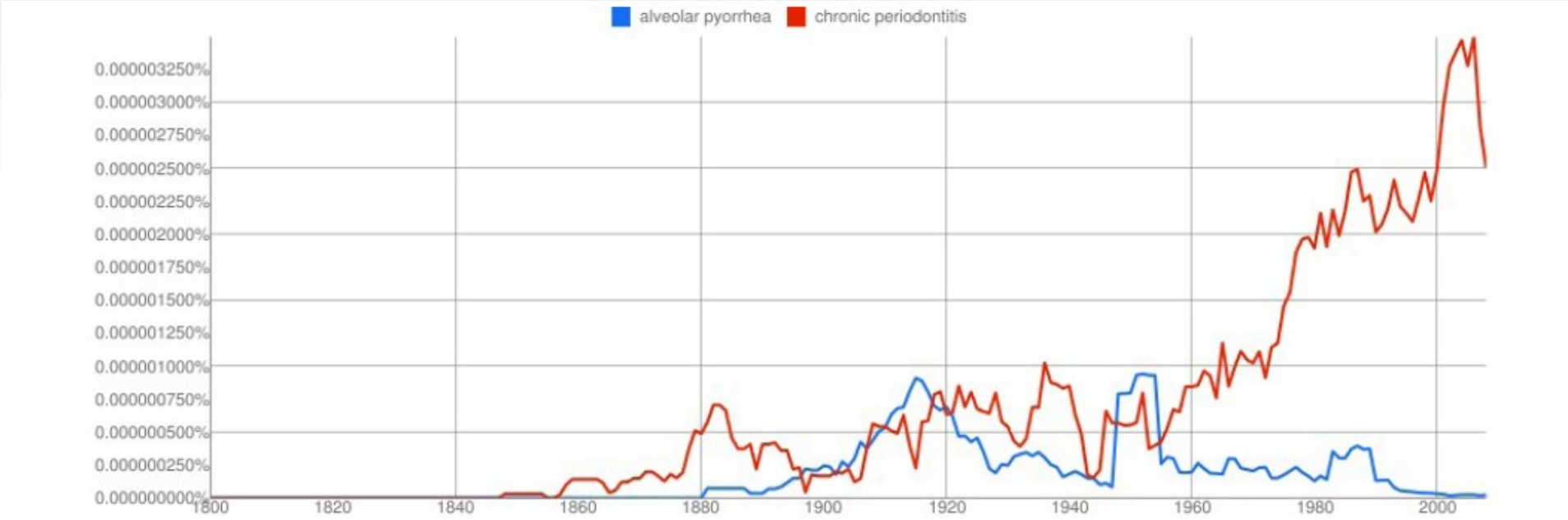
- Best overall results in Task 1 (triage for the Comparative Toxicogenomics Database)
- Best entity recognition for diseases and chemicals



Terminology evolution

- Goal: investigate appearance, disappearance and replacement of biomedical terminology over time
  - Quality terminology is essential for text mining
- Experiments with PharmGKB/CTD/UMLS as reference terminology (diseases)
- Using PubMed abstracts as reference collection

Term replacement?



Summary

- Goal: Develop innovative text mining technologies for the automatic extraction of information from the biomedical literature [application: assisted curation].
- OntoGene/SASEBio provide competitive text mining technologies (BC, CALBC prove quality)
- ODIN as a tool for text-mining supported interactive curation of the biomedical literature
- PharmGKB/CTD experiments provide case study
- Terminology studies

OntoGene highlights

- [2006] BioCreative II: PPI (3rd), IMT (best)
- [2009] BioCreative II.5 PPI (best results); BioNLP
- [2010] BioCreative III: ACT, IMT, IAT
- [2011] CALBC (large scale entity extraction), BioNLP
- [2012] PharmGKB/CTD assisted curation experiments
- 60 peer-reviewed publications, 17 journal papers

http://www.ontogene.org/

Acknowledgments

- Institute of Computational Linguistics UZH
  - Gerold Schneider (parsing, rel. extr., IMT, BioNLP)
  - Simon Clematide (ODIN, GN, ACT, CALBC)
  - Kaarel Kaljurand (pipeline, ODIN, BioNLP)
  - Gintare Grigonyte (Term evol.), Tilia Ellendorff
- NIBR-IT, Text Mining Services, Novartis
  - Therese Vachon, Martin Romacker
- Swiss National Science Foundation