DATABASE
The Journal of Biological Databases and Curation

Database, Vol. 2012, Article ID bas021, doi:10.1093/database/bas021

Original Article

Using ODIN for a PharmGKB revalidation experiment

Fabio Rinaldi[1,*], Simon Clematide[1], Yael Garten[2], Michelle Whirl-Carrillo[3], Li Gong[3], Joan M. Hebert[3], Katrin Sangkuhl[3], Caroline F. Thorn[3], Teri E. Klein[3] and Russ B. Altman[3,4,5]

[1]Institute of Computational Linguistics. Binzmuhlestrasse 171, 8050 Zurich, Switzerland, [2]Biomedical Informatics Program, [3]Department of Genetics, [4]Department of Bioengineering and [5]Department of Medicine, Stanford University, 1501 California Avenue, Palo Alto, CA 94304

***Corresponding author:** Tel: +41793006771; Fax: +41446356809; Email: fabio.rinaldi@uzh.ch

Submitted 17 October 2011; Revised 5 February 2012; Accepted 28 March 2012

The need for efficient text-mining tools that support curation of the biomedical literature is ever increasing. In this article, we describe an experiment aimed at verifying whether a text-mining tool capable of extracting meaningful relationships among domain entities can be successfully integrated into the curation workflow of a major biological database. We evaluate in particular (i) the usability of the system's interface, as perceived by users, and (ii) the correlation of the ranking of interactions, as provided by the text-mining system, with the choices of the curators.

Introduction

The increasing quantity of available biomedical data pose a challenge to life scientist wishing to explore a particular research problem. Although emerging bioinformatics services enable structured access to increasingly complex datasets, it is often the case that the primary data is only available in the published literature, and needs to be extracted and stored in a standardized format before it can be leveraged upon (1). This is the main motivation behind biomedical curation activities: 'to help the Life Sciences community to make sense of all the data that is accumulating' (2).

Although human curation offers the best guarantee of high-quality results, it suffers from severe bottlenecks that have long been recognized in the curation community. The most pressing problem is that of efficiency of the process: despite the fact that typically several databases attempt to focus on a particular type of biological data, and often collaborate at least sufficiently to prevent duplication of effort and ensure compatibility of resulting data formats, it is impossible for human curators to keep up with the growing pace of publication.

> Nobody will ever be able to manually annotate all the macromolecular biological entities that exist on this planet, and consequently automatization is the only solution. (2)

On the other hand automated text-mining tools cannot offer sufficient reliability to be employed indiscriminately without human supervision of the results that they deliver. Therefore, the ideal solution is to combine the best capabilities of automated systems with human supervision by highly qualified domain experts.

In this article, we describe recent experiments aimed at assessing the potential contribution of a specific curation tool (ODIN) to the curation process of a well-known database (PharmGKB). In the rest of this section we briefly describe both PharmGKB and ODIN.

The Pharmacogenomics Knowledge Base (PharmGKB) is a publicly available online worldwide resource (www.pharmgkb.org) (3,4). The mission of PharmGKB is to collect, encode and disseminate knowledge about the impact of human genetic variations on drug responses, contributing to the drive towards personalized medicine for better therapeutics. PharmGKB is an NIH-funded resource, which over the past 11 years has maintained a very high-quality manually curated knowledge base of pharmacogenomics facts, curated by a team of PhD and Masters level scientists.

One of the many tasks of the PharmGKB curators is to review past and current literature and add any relevant pharmacogenetic or genomic articles to the PharmGKB database. The curators identify relevant journal articles, largely selected from a set of about 20 journals that are followed, which include major pharmacogenomic journals and publications published by the PGRN (Pharmacogenomics Research Network). They then read the abstract or full text if necessary, and populate the knowledge base with information about the genes, drugs and phenotypes discussed. In the past, this information was gathered in Microsoft Excel spreadsheets and uploaded to the database. Today, the information is entered through a web-based graphical user interface (GUI) developed in-house to fit the curators' needs, and captures data that is far more structured than in the past, such as population characteristics of the study group described in an article, and *P*-values of associations found between genetic variants and drug response. Curators can manually enter drug and gene terms for each article using auto-complete fields that draw on PharmGKB's standardized vocabularies. Additionally, in the past, in each article curators captured the entities discussed in the form of a list of genes, drugs and phenotypes, which did not enable users to presume binary relationships between a single gene and a single drug in a PharmGKB Literature Annotation. Today, the relationships between entities are binary, such that for example a gene–drug relationship is explicitly captured, including some degree of specification regarding the type of interaction ('is associated with', 'inhibits', etc.). A detailed description of the types of annotations in PharmGKB has been published previously (5).

The current curator GUI assists the curators in their process using basic text mining by suggesting entities found in the article, but does not pre-populate fields or highlight any information found within the article text. The PharmGKB team is currently working on developing Natural Language Processing and machine learning methods to aid in the future in tasks such as document retrieval and information extraction.

The OntoGene group at the University of Zurich has developed advanced solutions for several text-mining tasks based upon advanced natural language processing technologies, which have been proven to be state-of-the-art by participation in several competitive evaluations (6–9). The OntoGene text-mining system is based on a standard NLP pipeline, composed of efficient modules for sentence splitting, tokenization, entity recognition, syntactic chunking and dependency parsing. Its entity recognition component has been shown in the recent CALBC shared evaluation (9) to be highly efficient and capable of delivering competitive results for several entity categories. Its relation mining component has been used in the BioCreative 2009 evaluation to deliver the best results for the identification of protein–protein interactions (7). [A more detailed description of the architecture of the OntoGene text-mining system is beyond the scope of this article, for further details the interested reader is invited to consult the following publications (6,7). Specific adaptations that were carried out for the PharmGKB task are described in separate forthcoming publications (10,11)].

The results of the OntoGene text-mining system are made accessible through a curation system called ODIN (OntoGene Document INspector; this tool is not connected in any way with the recently introduced commercial text analytics system called OdinText.) which allows a user to dynamically inspect the results of their text-mining pipeline. A previous version of ODIN was used for participation in the 'interactive curation' task (IAT) of the BioCreative III competition (12). This was an informal task without a quantitative evaluation of the participating systems. However, the curators who used the system commented extremely positively on its usability for a practical curation task.

More recently, the OntoGene group created a version of ODIN that allows inspection of abstracts automatically annotated with PharmGKB entities [the annotation is performed using the OntoGene pipeline (http://www.ontogene.org/pharmgkb/)]. Users can access either preprocessed documents, or enter any PubMed identifier and have the corresponding abstract processed 'on the fly'. For the documents already in PharmGKB it is also possible to inspect the gold standard and compare the results of the system against the gold standard. The curator can inspect all entities annotated by the system, and easily modify them if needed (removing false positives with a simple click, or adding missed terms if necessary). The modified documents can be sent back for reprocessing if desired, obtaining therefore modified candidate interactions. The user can also inspect the set of candidate interactions generated by the system, and act upon them just as on entities, i.e. confirm those that are correct, remove those that are incorrect. Candidate interactions are presented ordered according to the score that has been assigned to them by the text-mining system, therefore the curator can choose to work with only a small set of highly ranked candidates, ignoring all the rest.

ODIN, which is based on a client–server architecture, maintains a log of the interaction with the curator, which could be used for later revision by a supervisor or for reversing some specific annotation decisions. At the end of a session the modified document and its annotations are sent back to the server, together with the log, for permanent storage, and can be accessed again at the next session, which could take place on a different remote client. Additionally, the curator can choose to export the annotations to a local file in a simplified format (e.g. comma-separated values).

Related work

Automated tools have the potential to support the curation process in several phases. First of all, text-mining tools can provide a help in the initial triage stage in order to decide which papers should be inspected by the expert curators. Text classification tools are nowadays capable of reliably processing large sets of articles in order to score them and provide a ranked list of candidate papers, which can then be used to prevent inspection of less promising articles. This process is typically based on machine learning tools that can distinguish interesting and less interesting articles on the basis of similarities with previously classified articles.

During inspection of individual articles, it can be very helpful for curators to use a system capable of locating the entities of interest within the article, and disambiguate them as reliably as possible. This process is based on named entity recognition tools, which recently have made considerable progress and are now capable of recognizing several types of biomedical entities with great reliability. For example, recent results in the BioCreative competition (13) have shown that several systems are capable of recognizing and disambiguating gene names with *F*-scores above 80%. Databases that are entity-focused can immediately profit from such tools, as the curators will be able to manually filter the candidates suggested by the system at greater speed, compared with a manual extraction from the paper, which would involve (i) spotting the mentions in the paper, (ii) decide which database entities are actually intended.

The next major challenge for the introduction of text-mining systems within curation workflows is the automated detection of relations, which is relevant for several databases. Tools that can reliably detect entity interactions are in general much less efficient than named entity recognition tools due to the much greater complexity of the problem. In order to produce candidate interactions a tool needs first to identify the entities correctly. Given that some errors are inherent in this process, generation of candidate entity pairs will inevitably result in compounding that error, leading to lower performance. Contextual clues that can help to identify an interaction candidate are typically very sparse, making difficult to apply machine learning techniques.

Nonetheless, much progress has been achieved recently, as results in the BioCreative II (14) and II.5 (15) competitions show, and therefore it is now appropriate to start practical experimentation through collaborations between developers of text-mining solutions and database groups as potential users. Although immediate integration in the curation workflow might not be the goal, these joint experiment help both groups in deciding how to improve their activities.

Text-mining developers will receive feedback on the quality of their systems and gain an understanding of the specific needs of the curators [(12) stresses the 'importance of understanding the biocurator's curation workflow'], and curation groups will gain a better understanding of the current potential of technologies, which are now still experimental, but might soon become mainstream, and thus be able to choose the optimal point for integration in their workflows. Additionally, the feedback provided to text-mining developers will render future systems more usable in practical applications.

The need to pair developers with curators has been recognized by the organizers of the BioCreative competition. A new experimental task (IAT) dedicated to the evaluation of interactive curation environments was introduced in the last edition (12). Although the specific task chosen for the experimentation was an entity recognition task, several of the conclusions reached through analysis of participating systems are applicable to all types of interactive curation environments. Addressing usability of text-mining systems is a novel aspect of this task. 'Usability ... enables the users to find, interact with, share, compare and manipulate important information more effectively and efficiently' (12).

Although fully unsupervised extraction of information from the literature is, for some time at least, unrealistic, text-mining tools are already sufficiently reliable to be used to provide hints to the curators, in order to speed up their activities. Such a help is sorely needed, as it is already clear that manual curation cannot keep up with the rate of data generation (16). Curatorial work done with the assistance of a text-mining system has already been shown to be much more efficient than when done by human readers without support (17). The authors of this study state that: 'For biologists, an automated system with high recall and even moderate precision ... confers a great advantage over skimming text by eye.' Examples of well-known text-mining solutions are iHOP (18) and ChilliBot (19). Among the systems developed to support the curation process, one of the most interesting is (20). They use a manually annotated corpus (gold standard) to simulate an assisted curation environment, where the curators are given either gold standard data or the output of an (imperfect) NLP pipeline. (20,21) presents a system developed for the curators of FlyBase, a database for drosophila genetics and molecular biology. Although the document analysis is based on a conventional NLP pipeline, including the dependency parser RASP (23), the curator's interface has been developed in strict collaboration with the end-users. (24) discuss how well the performance of a text-mining system (in their case tailored to identify mentions of protein mutations), when evaluated with conventional techniques, translates into real utility of the system for a curation task. Textpresso is another well-known text-mining system that is characterized by the usage of ontological categories of biological concepts (17,25), as well as by processing full papers. A variant of Textpresso (Pharmspresso) has been used for automatic annotation of pharmacogenomic literature for PharmGKB, but was never integrated with the manual curation process (26).

Methods

Although the full OntoGene pipeline can deliver reliably a ranked list of candidate interactions, which can then be used by curators as prompts for annotation of novel articles, the experiment described in this article centered upon the validation of existing relations from PharmGKB. Revalidation of existing data is a common practice of several biological databases, for example (2) mentions several steps of re-annotation for Swiss-Prot, one of the most well-known and authoritative databases.

The main aim of the experiment was therefore to evaluate the usability of the interface, rather than the capabilities of the underlying text-mining tools. As (12) points out, an evaluation task must be chosen to be feasible in a given time frame, considering both the time needed by developers to adapt the existing text-mining system to the specific needs of the applications and the time available to curators for the verification of the results.

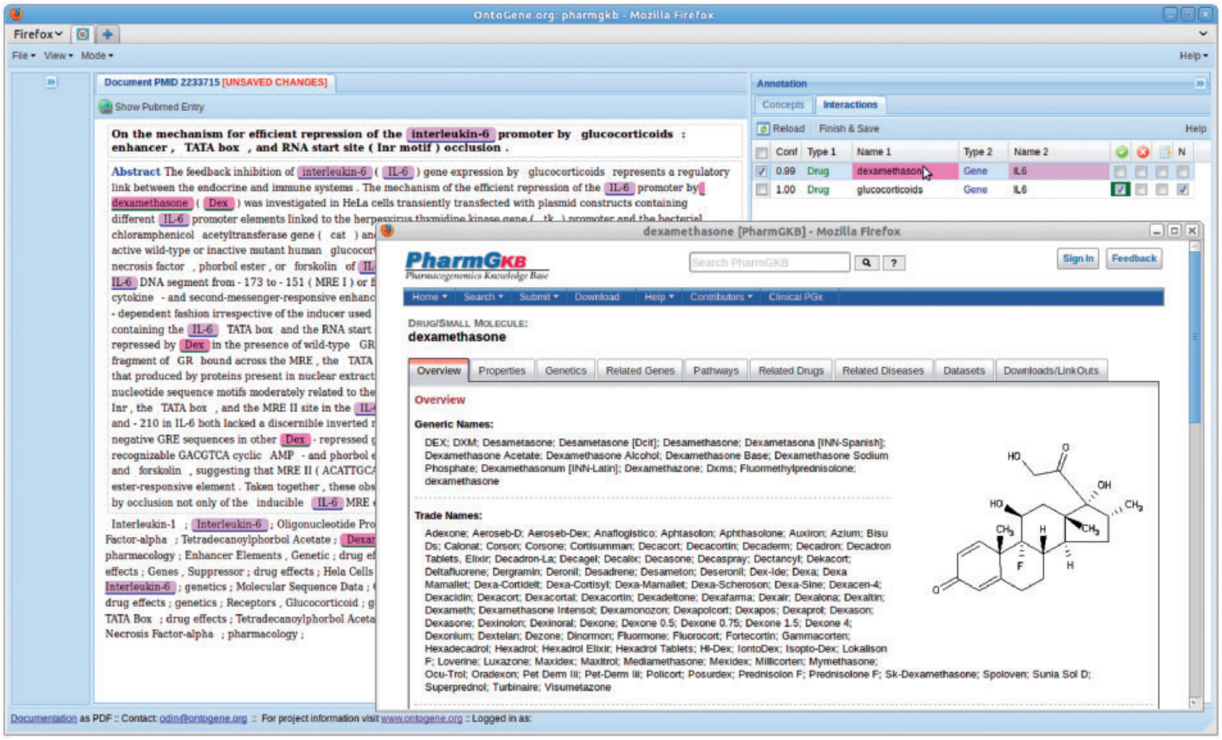
We started by considering the set of articles already curated by PharmGKB, processing them with the OntoGene relation extraction system. We then automatically compared for each article the results of the relation mining system with the manually extracted interactions, and computed the common subset. In general, we would expect the text-mining system to deliver a larger set of interactions than those manually curated (ideally covering all of them). In practice, since in this experiment only abstracts rather than full text were used, this was true only in 3059 articles out of 5378. All the remaining articles contain at least one 'false negative', i.e. a relation that was not detected by the text-mining system. (This could be due to several factors, for example an interaction that is mentioned only in the full text and not in the abstract will be obviously impossible for the system to detect with the current settings. Another possible source of false negatives is due to the way the PharmGKB data was created, i.e. for each paper, a list of genes and drugs discussed was kept, and in some cases interactions among those entities were simply hypothesized and not actually verified. Therefore, some of the interactions in the PharmGKB data are not expected to be true positives). If full articles had been processed, we would have expected the number of false negatives to be much lower. Full articles, however, are difficult to process for several reasons, most of which have little to do with text mining, such as widely different formats, or copyright restrictions, which in some cases explicitly prohibit text-mining applications.

We decided for this experiment to use only articles from the set where all interactions were found by the OntoGene pipeline. The aims of the experiment were the following: (i) evaluate the usability of the interface for revalidation of PharmGKB relationships, (ii) estimate whether the ranking of interactions provided by the text-mining system correlates well with decisions taken by the curators. The previously developed ODIN system was adapted to the needs of PharmGKB, on the basis of a close interaction with the curators. During development the following recommendations by (12) (adapted to the specific needs of our application) were taken into consideration:

(a) support for interactive disambiguation of domain entities;
(b) an editable list of candidate entities or interactions;
(c) a view of the document correlated with the candidate interactions (when an interaction is selected, the corresponding entities are highlighted);
(d) ability to sort the results according to different criteria;
(e) ability to collect event and timing information at the session level; and
(f) ability to export the results in a suitable format (e.g. CSV).

The ODIN system allows the user to verify and modify every single annotation provided by the system, at the entity level as well as at the interaction level (a). It provides a ranked list of candidate interactions (b), which additionally can be sorted by the user according to different criteria (d). The interface is structured around three panels: term editing panel, document panel and results panel. The term editing panel (not shown in the pictures in this article) can be used for (a). The document panel and results panel are actively connected, in that selection of items from the results panel will result in their visualization (highlighting) in the document panel. Every action of the user is stored in a log that is stored at the server level (e) with timing information (Figure 2). Finally, the results can be exported as a CSV file (f) and additional formats can easily be added upon request.

The close interaction among curators and system developers allowed the latter to implement a number of suggestions that made the usage of the system more effective. For example, in the initial demonstration the entities participating in an interaction were represented only by the PharmGKB identifier of the participating entities. The curators pointed out that it was not immediately obvious to them which entity was referred to by the identifier without consulting the database (that the ODIN system allows by simple click on the identifier, see Figure 1), so the 'reference name' of the entity was added. This, however, made the


OntoGene.org: pharmgkb - Mozilla Firefox
Firefox
File View Mode
Help
Document PMID 2233715 [UNSAVED CHANGES]
Show Pubmed Entry
Annotation
Concepts
Interactions
Reload
Finish & Save
Help
dexamethasone [PharmGKB] - Mozilla Firefox
PharmGKB
Pharmacogenomics Knowledge Base
Search PharmGKB
Sign In
Feedback
Home Search Submit Download Help Contributors Clinical PGx
Overview Properties Genetics Related Genes Pathways Related Drugs Related Diseases Datasets Downloads/LinkOuts
HO
O
OH
CH3
HO
CH3
H
CH3
F
H
O

On the mechanism for efficient repression of the interleukin-6 promoter by glucocorticoids : enhancer , TATA box , and RNA start site (Inr motif) occlusion .

Abstract The feedback inhibition of interleukin-6 (IL-6) gene expression by glucocorticoids represents a regulatory
link between the endocrine and immune systems . The mechanism of the efficient repression of the IL-6 promoter by
dexamethasone (Dex) was investigated in HeLa cells transiently transfected with plasmid constructs containing
different IL-6 promoter elements linked to the herpesvirus thymidine kinase gene (tk) promoter and the bacterial
chloramphenicol acetyltransferase gene (cat) an
active wild-type or inactive mutant human glucocor
necrosis factor , phorbol ester , or forskolin of IL
IL-6 DNA segment from - 173 to - 151 (MRE I) or f
cytokine - and second-messenger-responsive enhanc
- dependent fashion irrespective of the inducer used
containing the IL-6 TATA box and the RNA start
repressed by Dex in the presence of wild-type GR
fragment of GR bound across the MRE , the TATA
that produced by proteins present in nuclear extract
nucleotide sequence motifs moderately related to the
Inr , the TATA box , and the MRE II site in the IL-
and - 210 in IL-6 both lacked a discernible inverted r
negative GRE sequences in other Dex - repressed g
recognizable GACGTCA cyclic AMP - and phorbol e
and forskolin , suggesting that MRE II (ACATTGCA
ester-responsive element . Taken together , these obs
by occlusion not only of the inducible IL-6 MRE

Interleukin-1 ; Interleukin-6 ; Oligonucleotide Pro
Factor-alpha ; Tetradecanoylphorbol Acetate ; Dexar
pharmacology ; Enhancer Elements , Genetic ; drug ef
effects ; Genes , Suppressor ; drug effects ; Hela Cells
Interleukin-6 ; genetics ; Molecular Sequence Data ;
drug effects ; genetics ; Receptors , Glucocorticoid ; g
TATA Box ; drug effects ; Tetradecanoylphorbol Aceta
Necrosis Factor-alpha ; pharmacology ;

□	Conf	Type 1	Name 1	Type 2	Name 2	✓	✗		N
☑	0.99	Drug	dexamethason	Gene	IL6	□	□	□	□
□	1.00	Drug	glucocorticoids	Gene	IL6	☑	□	□	☑

Drug/Small Molecule:
dexamethasone

Overview

Generic Names:

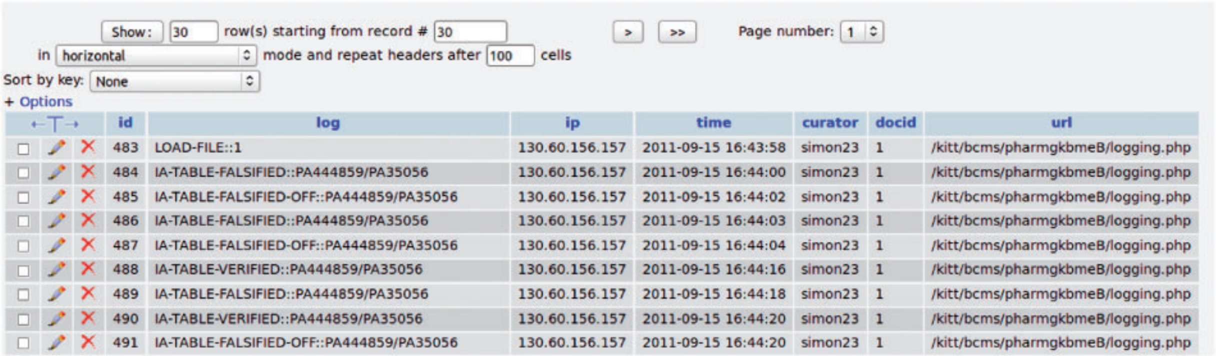
DEX; DXM; Desametasone; Desametasone [Dcit]; Desamethasone; Dexametasona [INN-Spanish]; Dexamethasone Acetate; Dexamethasone Alcohol; Dexamethasone Base; Dexamethasone Sodium Phosphate; Dexamethasonum [INN-Latin]; Dexamethazone; Dxms; Fluormethylprednisolone; dexamethasone

Trade Names:

Adexone; Aeroseb-D; Aeroseb-Dex; Anaflogistico; Aphtasolon; Aphthasolone; Auxiron; Azium; Bisu Ds; Calonat; Corson; Corsone; Cortisumman; Decacort; Decacortin; Decaderm; Decadron; Decadron Tablets, Elixir; Decadron-La; Decagel; Decalix; Decasone; Decaspray; Dectancyl; Dekacort; Deltafluorene; Dergramin; Deronil; Desadrene; Desameton; Deseronil; Dex-Ide; Dexa; Dexa Mamallet; Dexa-Cortidelt; Dexa-Cortisyl; Dexa-Mamallet; Dexa-Scheroson; Dexa-Sine; Dexacen-4; Dexacidin; Dexacort; Dexacortal; Dexacortin; Dexadeltone; Dexafarma; Dexair; Dexalona; Dexaltin; Dexameth; Dexamethasone Intensol; Dexamonozon; Dexapolcort; Dexapos; Dexaprol; Dexason; Dexasone; Dexinolon; Dexinoral; Dexone; Dexone 0.5; Dexone 0.75; Dexone 1.5; Dexone 4; Dexonium; Dextelan; Dezone; Dinormon; Fluormone; Fluorocort; Fortecortin; Gammacorten; Hexadecadrol; Hexadrol; Hexadrol Elixir; Hexadrol Tablets; Hl-Dex; IontoDex; Isopto-Dex; Lokalison F; Loverine; Luxazone; Maxidex; Maxitrol; Mediamethasone; Mexidex; Millicorten; Mymethasone; Ocu-Trol; Oradexon; Pet Derm Iii; Pet-Derm Iii; Policort; Posurdex; Prednisolon F; Prednisolone F; Sk-Dexamethasone; Spoloven; Sunia Sol D; Superprednol; Turbinaire; Visumetazone

Documentation as PDF :: Contact: odin@ontogene.org :: For project information visit www.ontogene.org :: Logged in as:

Figure 1. Inspection of PharmGKB entry associated with a given entity.



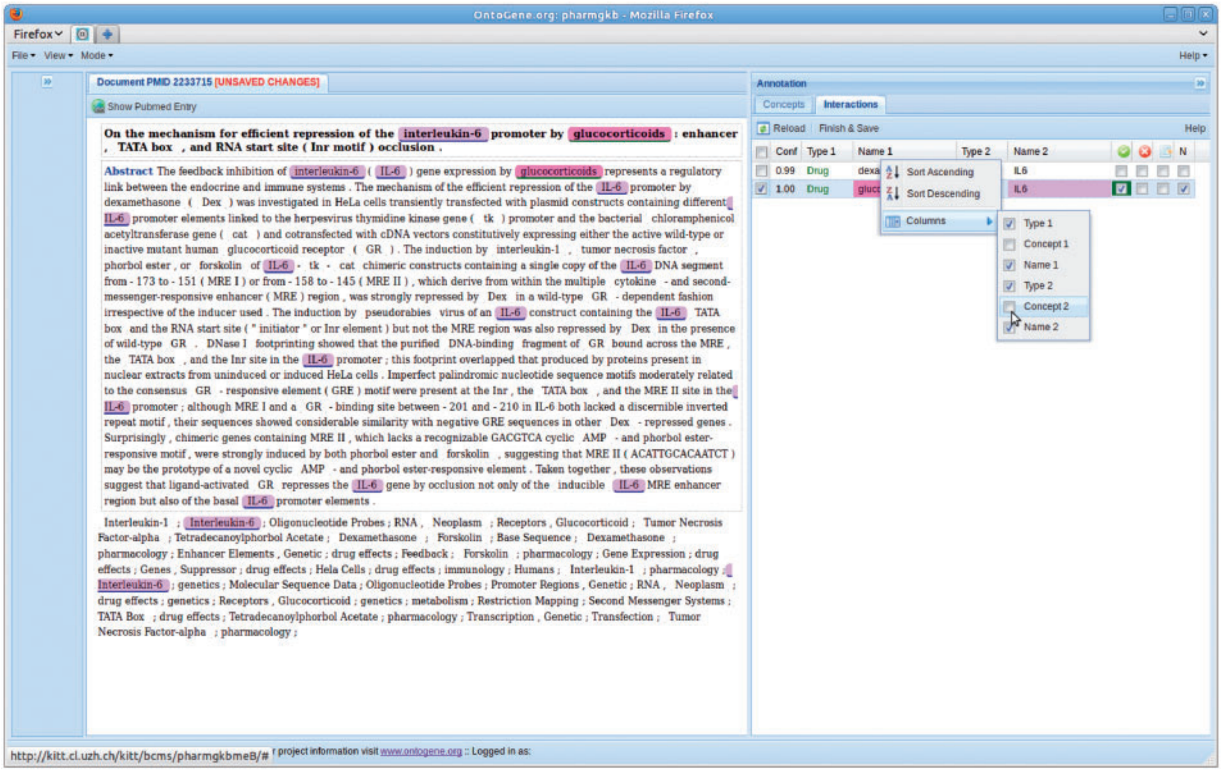
Show : 30 row(s) starting from record # 30 Page number: 1
in horizontal mode and repeat headers after 100 cells
Sort by key: None

←T→			id	log	ip	time	curator	docid	url
□	✎	✗	483	LOAD-FILE::1	130.60.156.157	2011-09-15 16:43:58	simon23	1	/kitt/bcms/pharmgkbmeB/logging.php
□	✎	✗	484	IA-TABLE-FALSIFIED::PA444859/PA35056	130.60.156.157	2011-09-15 16:44:00	simon23	1	/kitt/bcms/pharmgkbmeB/logging.php
□	✎	✗	485	IA-TABLE-FALSIFIED-OFF::PA444859/PA35056	130.60.156.157	2011-09-15 16:44:02	simon23	1	/kitt/bcms/pharmgkbmeB/logging.php
□	✎	✗	486	IA-TABLE-FALSIFIED::PA444859/PA35056	130.60.156.157	2011-09-15 16:44:03	simon23	1	/kitt/bcms/pharmgkbmeB/logging.php
□	✎	✗	487	IA-TABLE-FALSIFIED-OFF::PA444859/PA35056	130.60.156.157	2011-09-15 16:44:04	simon23	1	/kitt/bcms/pharmgkbmeB/logging.php
□	✎	✗	488	IA-TABLE-VERIFIED::PA444859/PA35056	130.60.156.157	2011-09-15 16:44:16	simon23	1	/kitt/bcms/pharmgkbmeB/logging.php
□	✎	✗	489	IA-TABLE-FALSIFIED::PA444859/PA35056	130.60.156.157	2011-09-15 16:44:18	simon23	1	/kitt/bcms/pharmgkbmeB/logging.php
□	✎	✗	490	IA-TABLE-VERIFIED::PA444859/PA35056	130.60.156.157	2011-09-15 16:44:20	simon23	1	/kitt/bcms/pharmgkbmeB/logging.php
□	✎	✗	491	IA-TABLE-FALSIFIED-OFF::PA444859/PA35056	130.60.156.157	2011-09-15 16:44:20	simon23	1	/kitt/bcms/pharmgkbmeB/logging.php

Figure 2. Log of user actions as stored on the OntoGene server.

table of interactions more cluttered, and the curators pointed to the fact that they might optionally want to remove some information from the table (such as the entity identifiers). The developers therefore modified the interface to allow precisely this type of modification directly by the user (i.e. selection of which fields they want to be displayed), see Figure 3. Some of the information hidden in this way could further be displayed as unobtrusive tooltip windows on mouseover by the user, another option that was added upon suggestion by the curators. (1) stresses the importance of being able 'to hide fields of negligible value to the curators thereby distracting their attention unnecessarily'.

Another example of the fruitfullness of the interaction between developers and curators is the addition of different types of confirmation boxes for an interaction. Instead of a simple confirm/reject choice, the maintainers of the database suggested the need for a more fine-grained choice. In particular, they wanted to be able to confirm 'negative' interactions, i.e. interactions that are stated in the paper as NOT to hold under the conditions investigated. Another wish was to be able to state that the



On the mechanism for efficient repression of the interleukin-6 promoter by glucocorticoids : enhancer , TATA box , and RNA start site (Inr motif) occlusion .

Abstract The feedback inhibition of interleukin-6 (IL-6) gene expression by glucocorticoids represents a regulatory link between the endocrine and immune systems . The mechanism of the efficient repression of the IL-6 promoter by dexamethasone (Dex) was investigated in HeLa cells transiently transfected with plasmid constructs containing different IL-6 promoter elements linked to the herpesvirus thymidine kinase gene (tk) promoter and the bacterial chloramphenicol acetyltransferase gene (cat) and cotransfected with cDNA vectors constitutively expressing either the active wild-type or inactive mutant human glucocorticoid receptor (GR) . The induction by interleukin-1 , tumor necrosis factor , phorbol ester , or forskolin of IL-6 - tk - cat chimeric constructs containing a single copy of the IL-6 DNA segment from - 173 to - 151 (MRE I) or from - 158 to - 145 (MRE II) , which derive from within the multiple cytokine - and second-messenger-responsive enhancer (MRE) region , was strongly repressed by Dex in a wild-type GR - dependent fashion irrespective of the inducer used . The induction by pseudorabies virus of an IL-6 construct containing the IL-6 TATA box and the RNA start site (" initiator " or Inr element) but not the MRE region was also repressed by Dex in the presence of wild-type GR . DNase I footprinting showed that the purified DNA-binding fragment of GR bound across the MRE , the TATA box , and the Inr site in the IL-6 promoter ; this footprint overlapped that produced by proteins present in nuclear extracts from uninduced or induced HeLa cells . Imperfect palindromic nucleotide sequence motifs moderately related to the consensus GR - responsive element (GRE) motif were present at the Inr , the TATA box , and the MRE II site in the IL-6 promoter ; although MRE I and a GR - binding site between - 201 and - 210 in IL-6 both lacked a discernible inverted repeat motif , their sequences showed considerable similarity with negative GRE sequences in other Dex - repressed genes . Surprisingly , chimeric genes containing MRE II , which lacks a recognizable GACGTCA cyclic AMP - and phorbol ester-responsive motif , were strongly induced by both phorbol ester and forskolin , suggesting that MRE II (ACATTGCACAATCT) may be the prototype of a novel cyclic AMP - and phorbol ester-responsive element . Taken together , these observations suggest that ligand-activated GR represses the IL-6 gene by occlusion not only of the inducible IL-6 MRE enhancer region but also of the basal IL-6 promoter elements .

Interleukin-1 ; Interleukin-6 ; Oligonucleotide Probes ; RNA , Neoplasm ; Receptors , Glucocorticoid ; Tumor Necrosis Factor-alpha ; Tetradecanoylphorbol Acetate ; Dexamethasone ; Forskolin ; Base Sequence ; Dexamethasone ; pharmacology ; Enhancer Elements , Genetic ; drug effects ; Feedback ; Forskolin ; pharmacology ; Gene Expression ; drug effects ; Genes , Suppressor ; drug effects ; Hela Cells ; drug effects ; immunology ; Humans ; Interleukin-1 ; pharmacology ; Interleukin-6 ; genetics ; Molecular Sequence Data ; Oligonucleotide Probes ; Promoter Regions , Genetic ; RNA , Neoplasm ; drug effects ; genetics ; Receptors , Glucocorticoid ; genetics ; metabolism ; Restriction Mapping ; Second Messenger Systems ; TATA Box ; drug effects ; Tetradecanoylphorbol Acetate ; pharmacology ; Transcription , Genetic ; Transfection ; Tumor Necrosis Factor-alpha ; pharmacology ;

☐	Conf	Type 1	Name 1	Type 2	Name 2	✓	✗		N
☐	0.99	Drug	dexa		IL6	☐	☐	☐	☐
☑	1.00	Drug	gluc		IL6	☑	☐	☐	☑

Figure 3. Modifying the presentation of the interactions.

abstract inspected for the experiment did not provide sufficient information to either confirm or reject the proposed relation. Figures 1, 3, and 4 show these options as four tick boxes in the top right corner of the picture, which correspond to 'confirm relation', 'reject', 'needs full text', 'negative relation confirmed'.

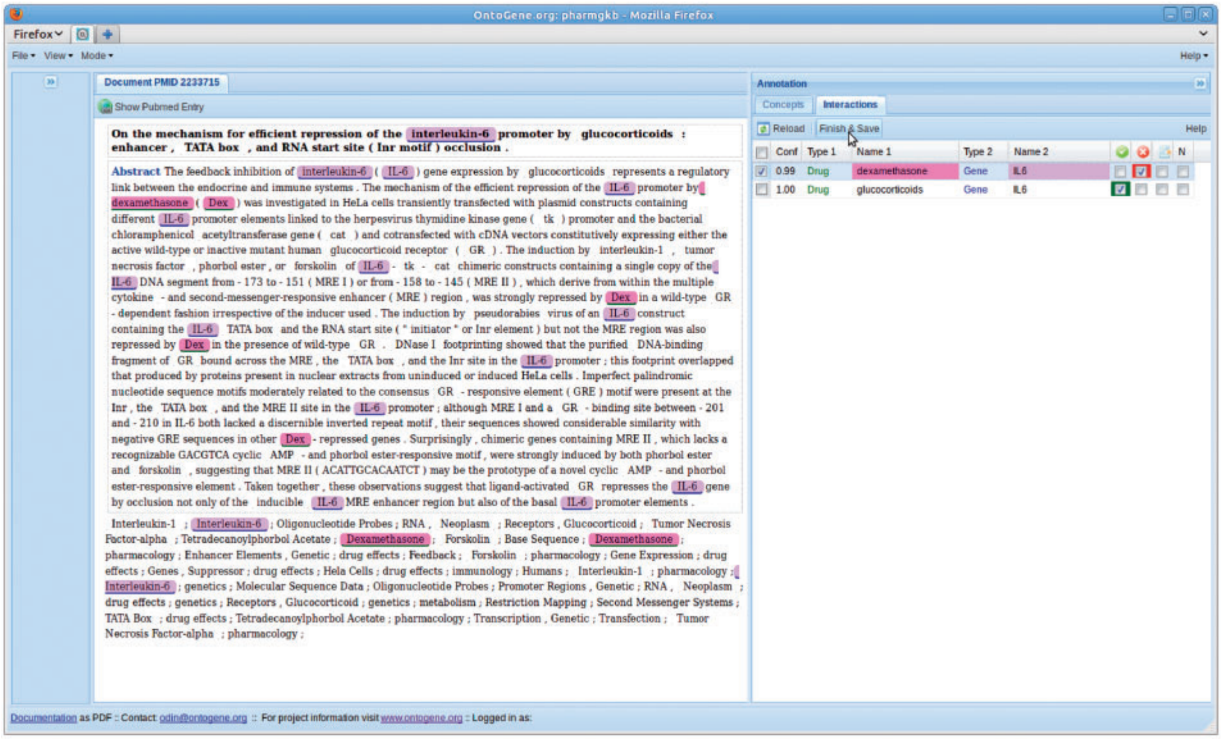
After a preliminary test phase on a few selected sample articles, which allowed the curators to gather some familiarity with ODIN, and the developers to fix the remaining issues, the validation experiment could start. A set of 125 articles was selected from the 3059 articles where the OntoGene pipeline could detect all of the relations originally annotated by PharmGKB. The selection was based on a randomized stratified sampling process, in order to generate a distribution of relations per article that would be roughly equivalent to the distribution in the whole set. This set was split into five sets of 25 articles each, which were then randomly assigned to PharmGKB curators. This sampling lead to the following distribution of articles per curator: 8 articles with 2 relations, 9 with 3 relations, 3 with 4 relations, 3 with 5 relations, 1 with 6–7 relations, 1 with 8–9 relations, 1 with 10–20 relations.

The data sets were at all stages identified only by a symbolic reference (A, B, C, D, E), which was randomly assigned to the curators (and known only to each of them), in order to ensure anonymity. This was done to avoid generating the impression that the result of the experiment could be used to evaluate individual performance. The full cooperation of the curators is of utmost importance to guarantee unskewed results, therefore we took care to prevent the possibility of identification. Curators were then asked to inspect the articles in the assigned set with the ODIN system and then use it to validate the interactions. During this process all of their actions were logged using the symbolic reference (that they had to enter into the system at the beginning of the process) as an identifier. The resulting validation decisions are automatically saved by the system and transferred to a server, together with detailed logs of the activity. This data set forms the basis of the evaluation presented in the next section.

At the end of the curation experiment we asked the curators to fill a questionnaire that was partly modeled on the questions used in the BCIII IAT Task (12). The feedback received through this survey is discussed at the end of the next section.

Evaluation

As explained in the previous section, the experiments described in this article were centered on the revalidation of relations already stored in PharmGKB. In order to evaluate the correlation of the rankings provided by the



On the mechanism for efficient repression of the interleukin-6 promoter by glucocorticoids : enhancer , TATA box , and RNA start site (Inr motif) occlusion .

Abstract The feedback inhibition of interleukin-6 (IL-6) gene expression by glucocorticoids represents a regulatory link between the endocrine and immune systems . The mechanism of the efficient repression of the IL-6 promoter by dexamethasone (Dex) was investigated in HeLa cells transiently transfected with plasmid constructs containing different IL-6 promoter elements linked to the herpesvirus thymidine kinase gene (tk) promoter and the bacterial chloramphenicol acetyltransferase gene (cat) and cotransfected with cDNA vectors constitutively expressing either the active wild-type or inactive mutant human glucocorticoid receptor (GR) . The induction by interleukin-1 , tumor necrosis factor , phorbol ester , or forskolin of IL-6 - tk - cat chimeric constructs containing a single copy of the IL-6 DNA segment from - 173 to - 151 (MRE I) or from - 158 to - 145 (MRE II) , which derive from within the multiple cytokine - and second-messenger-responsive enhancer (MRE) region , was strongly repressed by Dex in a wild-type GR - dependent fashion irrespective of the inducer used . The induction by pseudorabies virus of an IL-6 construct containing the IL-6 TATA box and the RNA start site (" initiator " or Inr element) but not the MRE region was also repressed by Dex in the presence of wild-type GR . DNase I footprinting showed that the purified DNA-binding fragment of GR bound across the MRE , the TATA box , and the Inr site in the IL-6 promoter ; this footprint overlapped that produced by proteins present in nuclear extracts from uninduced or induced HeLa cells . Imperfect palindromic nucleotide sequence motifs moderately related to the consensus GR - responsive element (GRE) motif were present at the Inr , the TATA box , and the MRE II site in the IL-6 promoter ; although MRE I and a GR - binding site between - 201 and - 210 in IL-6 both lacked a discernible inverted repeat motif , their sequences showed considerable similarity with negative GRE sequences in other Dex - repressed genes . Surprisingly , chimeric genes containing MRE II , which lacks a recognizable GACGTCA cyclic AMP - and phorbol ester-responsive motif , were strongly induced by both phorbol ester and forskolin , suggesting that MRE II (ACATTGCACAATCT) may be the prototype of a novel cyclic AMP - and phorbol ester-responsive element . Taken together , these observations suggest that ligand-activated GR represses the IL-6 gene by occlusion not only of the inducible IL-6 MRE enhancer region but also of the basal IL-6 promoter elements .

Interleukin-1 ; Interleukin-6 ; Oligonucleotide Probes ; RNA , Neoplasm ; Receptors , Glucocorticoid ; Tumor Necrosis Factor-alpha ; Tetradecanoylphorbol Acetate ; Dexamethasone ; Forskolin ; Base Sequence ; Dexamethasone ; pharmacology ; Enhancer Elements , Genetic ; drug effects ; Feedback ; Forskolin ; pharmacology ; Gene Expression ; drug effects ; Genes , Suppressor ; drug effects ; Hela Cells ; drug effects ; immunology ; Humans ; Interleukin-1 ; pharmacology ; Interleukin-6 ; genetics ; Molecular Sequence Data ; Oligonucleotide Probes ; Promoter Regions , Genetic ; RNA , Neoplasm ; drug effects ; genetics ; Receptors , Glucocorticoid ; genetics ; metabolism ; Restriction Mapping ; Second Messenger Systems ; TATA Box ; drug effects ; Tetradecanoylphorbol Acetate ; pharmacology ; Transcription , Genetic ; Transfection ; Tumor Necrosis Factor-alpha ; pharmacology ;

□	Conf	Type 1	Name 1	Type 2	Name 2	✓	✗	📄	N
☑	0.99	Drug	dexamethasone	Gene	IL6	□	☑	□	□
□	1.00	Drug	glucocorticoids	Gene	IL6	☑	□	□	□

Documentation as PDF :: Contact odin@ontogene.org :: For project information visit www.ontogene.org :: Logged in as:

Figure 4. Entities which participate in the selected interaction are highlighted in the document panel.

text-mining system with curator's validation decisions, we were limited to use only articles for which all PharmGKB interactions could be detected by the text-mining system. Since only abstracts were used for automatic processing, only about 56% of PharmGKB curated articles (3059) could be taken into consideration.

We also observed that articles that contained a single curated relation would not be particularly interesting for this experiment, since it can be presumed that the vast majority of these cases are correct, and in any case there is no ranking to evaluate. Additionally, articles containing more than 20 interactions were also excluded, because there was a very limited number of them, and they would require too much time for revalidation. Excluding these cases, we were left with a set of 1407 articles. Out of this set, we selected by stratified random sampling five sets of 25 articles each, as described in the previous section.

In the rest of this section, we describe in detail the results of our experiments through descriptive statistics computed from the logs of the interactions. This is followed by a qualitative analysis of the final survey.

The curators could take for each relation one of four decisions: 'confirm' (the abstract supports the interactions), *reject* (there is no support in the abstract for the interaction), 'negative' (the abstract states that the mentioned entities DO NOT interact), 'needs full text' (there is no sufficient information in the abstract to decide either way). The pie chart on the left of Figure 5 shows the total distribution of these decision across all articles. Nearly 3/4 of the relations were confirmed as positive. However, this distribution appears to be strongly dependent on the type of the entities participating in a relationship. The distribution of such decision by relation type is shown on the left of Figure 6. The relationship Drug/Gene has been chosen to be the main focus of future revalidation work and the results show that this type of interaction has a relatively low rejection rate. The distribution by curator is shown on the right of the same figure.

One of the aims of the experiment was to verify how the ranking of interactions produced by the text-mining system correlates with validation decisions by the curators. The bar chart on the right of Figure 5 shows a clear positive correlation at least for the best ranked cases (ranks 1–5). The proportion of interactions that the curators confirm as positive is greater at rank 1 and gradually decreases. At higher ranks there is no visible correlation, but this is partially due to the sparsity of data (in general there are fewer articles that have 5 interactions, and very few that have more than 20) (Among the articles selected for the experiment, 30.06% have 2 interactions, 37.81% 3 interactions, 6.39% 4 interactions, 11.02% 5, 5.90% 6 or 7, 4.33% 8 or 9 and 4.48% 10–20.).

After validating all relations in each document, the curators were asked to express their opinion about the quality

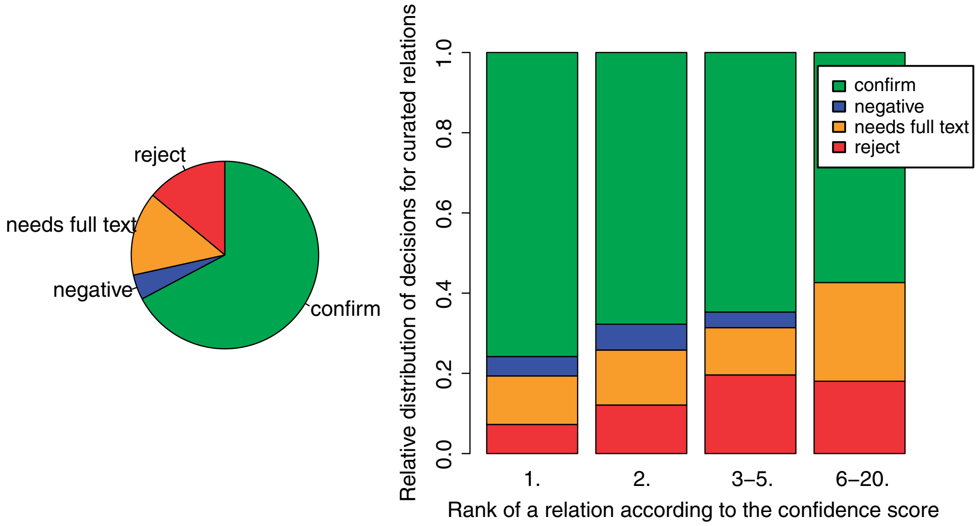


Figure 5. Distribution of validation decisions taken by curators.

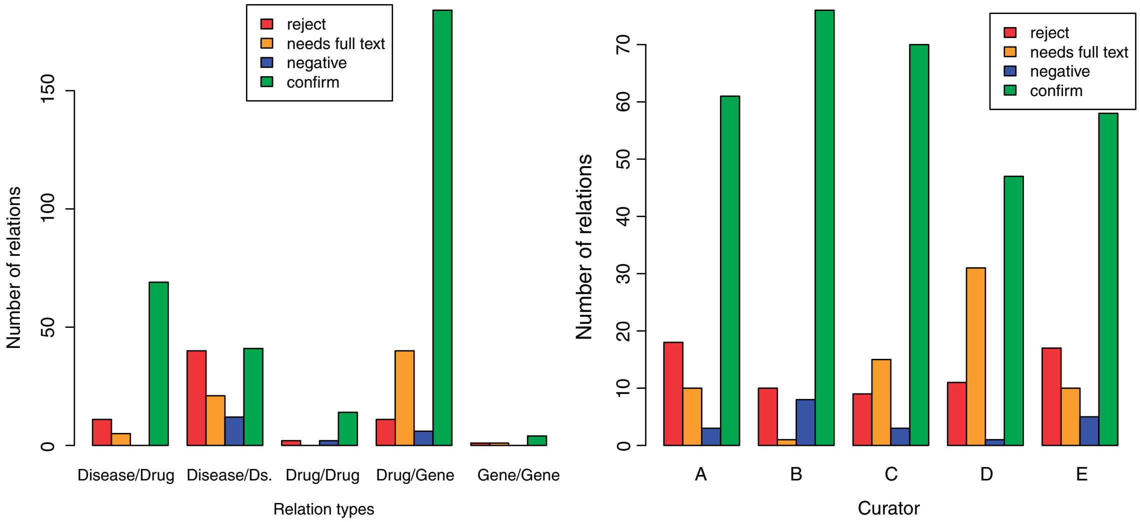


Figure 6. Validation decisions by category and by curator.

of concept identification (provided by the text-mining system) for that particular document. They could only chose among three values (bad, ok, good). However, since this comment was not mandatory, in about 1/4 of articles such judgments are missing. Figure 7 on the left shows the totals, and on the right distributed per curator. These values represent the perceived quality of concept identifications, i.e. the subjective judgment of the curators about the correctness of the entities suggested by the system.

In our experiment we also measured (through the logs) the exact time span between opening of a document and saving it after completing the validation of its interactions. This time was then divided by the number of interactions that had to be validated in each specific document. The average time needed for the validation of each interaction varies strongly among different curators, from 15 s up to 122 s. The box-and-whisker plots in Figure 8 illustrate the distribution of the mean time (in seconds) used for the curation of all relations of an article. The graph on the left shows these timings in relation to the curator's subjective judgment about the quality of concept recognition in the article, the graph on the right shows the timings per curator. The bottom whisker gives the minimum mean time, the top whisker gives the maximum mean time. Whiskers are shortened as usual to a length of 1.5× the box length and possible outliers are plotted separately with points. The bottom line of box is the first quantile, the upper line of box is the third quantile. The median is marked by the strong black line in the box.

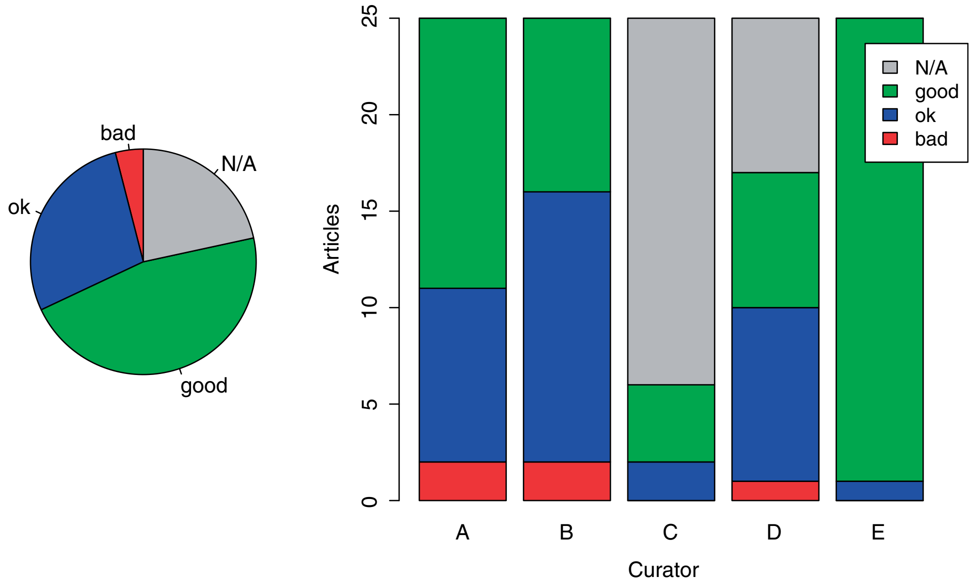


Figure 7. Distribution of concept identification judgments.

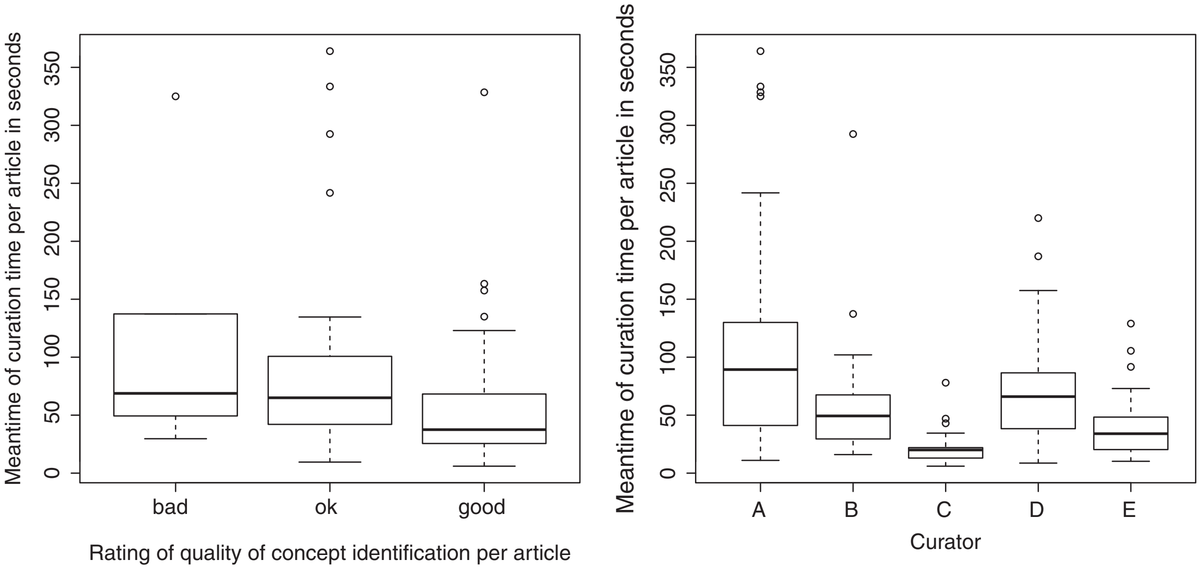


Figure 8. Box-and-whisker plots illustrating curation time (on the left according to the decision taken, on the right per curator).

An interesting research question is to evaluate whether the quality of a text-mining solution has a correlation with the time necessary for the curation of a specific article. The left graph in Figure 8 clearly shows such a correlation for our experiment: articles where concept identification was regarded by curators as 'good' required a much lower curation time than for the other two categories. It should be noticed, however, that the category 'bad' was selected in a too small number of cases, and therefore results in this category might not be very informative.

As mentioned in the previous section, at the end of the experiment the curators were asked to fill in a brief questionnaire in order to collect subjective feedback about their experience with the curation environment. One of the questionnaires was not returned. Below we list the questions that were asked and the feedback received.

Q1 Do you consider the system easy and intuitive to use?

(1) not intuitive at all
(2) partly intuitive (25%)
(3) mostly intuitive (75%)
(4) very intuitive

Q2 Do you consider the organization of the panels to be practical?

(1) not practical at all
(2) partly practical
(3) mostly practical (100%)
(4) very practical

Q3 What aspects of the system are most appealing to you?

Having the abstract to the left of the terms makes sense to me. It did help in that the entities in a proposed relationship were listed already so matched a PharmGKB term and were spelled correctly. That saves a bit of time.
Speeds things up because relationships are already entered just need to verify them
Highlighting the genes and drugs in the abstract, especially matching gene synonyms to gene symbol.

Q4 What problems/limitations of the system did you notice?

I would like to see the mapping of terms that are different in the abstract versus displayed gene names or diseases in the panel. I would like if just the gene and drug would be highlighted in the sentences used by the program to define the relationship not every mention of the gene and drug in the abstract. At the moment too many objects are highlighted.
Whenever I checked the green checkbox, it would turn on the highlighting for that row(I did not have that experience with the other choices).
I think it missed some relationships.
Sometimes needs to resort to full text. Some gene/drug relationships may not be identified through the system but are true relationships. Would be more valuable to extract the types of relationships between the concepts (eg. metabolize, transport, inhibit, induce etc)

Q5 Please mention any aspect of the system that did not appeal to you, or suggestions for changes.

I would have liked an option to turn off the underlining in the abstract. I find highlighting and underlining to be distracting and to clutter up the thing Im trying to read; I do not find them helpful. Maybe there was an option to do that and I simply did not notice it.
It would be helpful to have a place to add free text note to each relationship
Would be more valuable to extract the types of relationships between the concepts (eg. metabolize, transport, inhibit, induce, treat etc), also would be nice to extract genetic variations as another type of concept.

Q6 Was the system helpful in performing the validation task?

(1) not helpful at all
(2) partly helpful
(3) mostly helpful (75%)
(4) very helpful (25%)

Q7 Would you consider using a similar system for your regular curation task?

(1) no
(2) probably not
(3) probably yes (75%)
(4) yes (25%)

Q8 Do you agree that a similar system could increase the efficiency of the manual curation process?

(1) no
(2) probably not
(3) probably yes (75%)
(4) yes (25%)

Discussion

Usability issues are crucial for the acceptance of any specific IT tool by the end users. In the case of biomedical curation, it is essential that text-mining results are delivered to the curator in a transparent fashion, without the need of dealing with system technicalities, in order to prevent cognitive overload. The tool should not disrupt the 'rhythm, flow of thinking and mental modeling process' of the users, otherwise even minor problems with the interface could turn into major disruptions (1).

Ideally the user should be put in a situation where he/she can make a quick but well-motivated decision based on the information provided by the system. According to (1) the output of a text-mining system should respect the following five criteria in order to be really helpful in the curation process:

- Relevance: a connection to the disease of interest or to synonyms, homologs of interest.
- Valid: not likely to be a false positive (has some statistic of significance associated with it).
- Credible: trustworthy methods generated the evidence plus numerous lines of evidence, number of research publications, convincing public metadata.
- Plausible: function, location/structure, interaction type, biological process, and cellular component suggest an explanatory story.

- Manageable: enough interactions for promising insights but not so many as to be overwhelming (roughly between 10 and 50).

Another point mentioned by the same author is that it would be very helpful if the tool could present an interaction type: 'Interaction type was crucial for scientists' judgments about whether results might help construct a plausible explanatory story.' This is a wish that has also been expressed by the curators in the experiment described in this article.

On the basis of the final survey, it appears that the users appreciate the comfort and support that ODIN gives them. They consider it as helpful in several ways. According to the qualitative feedback provided in the survey, visual highlighting depends on personal preferences, and therefore users should be given the possibility to customize some additional aspects of the interface. Several curators mentioned that they would like to be able to add more specific information at the level of individual relationships, for example by means of a free text comment. Particularly useful would be to add to each relation an indication of its type as mentioned in the document (inhibition, activation, etc.). [The PharmGKB group is separately researching similar issues (27,28)]. If the system could provide hints in this direction, this would be a very helpful feature. The OntoGene text-mining system is capable of extracting this kind of interaction type indicators, however this feature was not used in the experiment as the developers assumed it would not be needed. This is another example that shows the importance of close collaboration between system developers and database curators.

Conclusion and future work

The experiment described in this article aims primarily at verifying the usability of the ODIN system in the context of curation of the PharmGKB database. The initial assumption was that we could separate an evaluation of the interface from an evaluation of the underlying text-mining system by asking curators to perform a revalidation task rather than a novel extraction task. The revalidation task consists in using the ODIN functionalities to quickly check the correctness of interactions already stored in PharmGKB.

The results of the experiment confirm the initial assumptions: (i) the ODIN system offers a comfortable environment for relation validation that can considerably speed up this particular curation task (ii) the positive validation decisions are strongly correlated with the rankings provided by the text-mining system.

As a next step, we intend to verify the quality of the interaction mining component on novel unseen articles but using the same interface. Since the curators at this stage are already familiar with the interface, we will be able to verify how the results delivered by the text-mining system can actually improve their effectiveness in annotating interactions, without being impaired by an unfamiliar interface.

Funding

The OntoGene group (F.R. and S.C.) is supported by the Swiss National Science Foundation (grant 100014-118396/1) and Novartis Pharma AG, NIBR-IT, Text Mining Services, CH-4002, Basel, Switzerland. The PharmGKB group (all other authors) is supported by National Institute of Health/National Institute of General Medical Sciences (R24 GM61374) and National Library of Medicine (contract HHSN276201000025C).

Conflict of interest. None declared.

References

1. Mirel,B. (2007) Usability and usefulness in bioinformatics: evaluating a tool for querying and analyzing protein interactions based on scientists' actual research questions. In: *Professional Communication Conference, IEEE International*. The Crowne Plaza Seattle Hotel, Seattle, Washington, USA, 1–8.
2. Bairoch,A. (2009) The future of annotation/biocuration. *Nature Precedings.* http://precedings.nature.com/documents/3092/version/1 (13 April 2012, date last accessed).
3. Klein,T.E., Chang,J.T., Cho,M.K. *et al*. (2001) Integrating genotype and phenotype information: an overview of the PharmGKB project. *Pharmacogenomics J.*, **1**, 167–170.
4. Sangkuhl,K., Berlin,D.S., Altman,R.B. *et al*. (2008) PharmGKB: understanding the effects of individual genetic variants. *Drug Metabolism Rev.*, **40**, 539–551.
5. McDonagh,E.M., Whirl-Carrillo,M., Garten,Y. *et al*. (2011) From pharmacogenomic knowledge acquisition to clinical applications: the PharmGKB as a clinical pharmacogenomic biomarker resource. *Biomarkers Med*, **5**, 795–806.
6. Rinaldi,F., Kappeler,T., Kaljurand,K. *et al*. (2008) OntoGene in BioCreative II. *Genome Biol.*, **9**(Suppl 2), S13.
7. Rinaldi,F., Schneider,G., Kaljurand,K. *et al*. (2010) OntoGene in BioCreative II.5. *IEEE/ACM Trans. Computa. Biol. Bioinfor.*, **7**, 472–480.
8. Schneider,G., Clematide,S. and Rinaldi,F. (2011) Detection of interaction articles and experimental methods in biomedical literature. *BMC Bioinformatics*, **12**(Suppl 8), S13.
9. Rebholz-Schuhmann,D., Yepes,A., Li,C. *et al*. (2011) Assessment of ner solutions against the first and second calbc silver standard corpus. *J. Biomed. Semantics*, **2**(Suppl 5), S11.
10. Rinaldi,F., Schneider,G. and Clematide,S. (2012) Relation mining experiments in the pharmacogenomics domain. *J. Biomed. Inform*, in press.
11. Clematide,S. and Rinaldi,F. (2012) Ranking interactions for a curation task. *Journal of Biomedical Semantics*, in press.
12. Arighi,C., Roberts,P., Agarwal,S. *et al*. (2011) Biocreative III interactive task: an overview. *BMC Bioinformatics*, **12**(Suppl 8), S4.
13. Lu,Z., Kao,H.-Y., Wei,C.-H. *et al*. (2011) The gene normalization task in biocreative III. *BMC Bioinformatics*, **12**(Suppl 8), S2.

14. Krallinger,M., Leitner,F., Rodriguez-Penagos,C. *et al*. (2008) Overview of the protein-protein interaction annotation extraction task of BioCreative II. *Genome Biol.*, **9**(Suppl 2), S4.

15. Leitner,F., Mardis,S.A., Krallinger,M. *et al*. (2010) An overview of Biocreative II.5. *IEEE/ACM T. Computat. Biol. Bioinform.*, **7**, 385–399.

16. Baumgartner,W.A., Cohen,K.B., Fox,L.M. *et al*. (2007) Manual curation is not sufficient for annotation of genomic databases. *Bioinformatics*, **23**, i41–i48.

17. Mller,H.-M., Kenny,E.E. and Sternberg,P.W. (2004) Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS Biol.*, **2**, e309,09.

18. Hoffmann,R. and Valencia,A. (2004) A gene network for navigating the literature. *Nat. Genet.*, **36**, 664.

19. Chen,H. and Sharp,B.M. (2004) Content-rich biological network constructed by mining pubmed abstracts. *BMC Bioinformatics*, **5**, 147.

20. Alex,B., Grover,C., Haddow,B. *et al*. (2008) Assisted curation: does text mining really help. In: Altman,R.B., Dunker,A.K., Hunter,L. *et al*. (eds), *BIOCOMPUTING 2008. Proceedings of the Pacific Symposium on Biocomputing*. Kohala Coast, Hawaii, USA.

21. Karamanis,N., Seal,R., Lewin,I. *et al*. (2008) Natural language processing in aid of flybase curators. *BMC Bioinformatics*, **9**, 193.

22. Karamanis,N., Lewin,I., Seal,R. *et al*. (2007) Integrating natural language processing with flybase curation. Grand Wailea, Maui, Hawaii. In: *Pacific Symposium on Biocomputing*, 245–256.

23. Briscoe,T., Carroll,J. and Watson,R. (2006) The second release of the RASP system. In: *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*July. Association for Computational Linguistics, Sydney, Australia, pp. 77–80.

24. Caporaso,J.G., Deshpande,N., Fink,J.L. *et al*. (2008) Intrinsic evaluation of text mining tools may not predict performance on realistic tasks. Fairmont Orchid, Big Island of Hawaii. *Pacific Symposium on Biocomputing*, **13**, 640–651.

25. Mller,H.M., Rangarajan,A., Teal,T.K. *et al*. (2008) Textpresso for neuroscience: searching the full text of thousands of neuroscience research papers. *Neuroinformatics*, **6**, 195–20.

26. Garten,Y. and Altman,R. (2009) Pharmspresso: a text mining tool for extraction of pharmacogenomic concepts and relationships from full text. *BMC Bioinformatics*, **10**(Suppl 2), S6.

27. Garten,Y., Coulet,A. and Altman,R.B. (2010) Recent progress in automatically extracting information from the pharmacogenomic literature. *Pharmacogenomics*, **11**, 1467–1489.

28. Coulet,A., Garten,Y., Dumontier,M. *et al*. (2011) Integration and publication of heterogeneous text-mined relationships on the semantic web. *J. Biomed. Semantics*, **2**(Suppl 2), S10.