

The **BioMeXT** group specializes in Information Extraction (IE), from the biomedical literature, as well as from other textual sources. Information extraction consists in the task of automatically extracting structured information from textual documents, and it is an important component of Text Mining systems. We focus in particular on the extraction of domain-specific entities (such as genes, proteins, drugs, diseases), and their semantic relations (e.g. protein-protein interactions, gene-disease associations). Our tools are often evaluated through participation in community-run evaluation challenges (e.g. **BioCreative**: Critical Assessment of Information Extraction systems in Biology). Additionally, we provide an environment for Assisted Curation (ODIN), which is currently being used in the curation pipeline of the RegulonDB database in a project funded by the US National Institutes of Health.

### Bio Term Hub

**Resource Selection**

Please select the resources to be included:

select all

- Cell Ontology (→ source) Up-to-date. [Update]
- Cellosaurus (→ source) Up-to-date. [Update]
- ChEBI (→ source) Up-to-date. [Update]
- CTD chemicals (→ source) Up-to-date. [Update]
- CTD diseases (→ source) Up-to-date. [Update]
- EntrezGene (→ source) Update available. [Update]
- Gene Ontology (→ source) Update available. [Update]
- MeSH (→ source) Update available. [Update]
- NCBI Taxonomy (→ source) Update available. [Update]
- Protein Ontology (→ source) Up-to-date. [Update]
- Sequence Ontology (→ source) Update available. [Update]
- Swiss-Prot (→ source) Up-to-date. [Update]

The Bio Term Hub provides access to biomedical terminology resources in a unified format. We envision its main use as a basis for text mining systems.

Select any of the external resources on the left to obtain a custom terminology list. The list is a tab-separated table including terms (names, synonyms), preferred name, concept identifier, entity type, and original resource. The contents of the last two fields can be modified using replacement patterns (uncover the **Renaming** section below using the triangle button). Submit your request with the **Create resource** button at the bottom of the page.

Aggregating, filtering, reformatting, and optionally renaming all this information takes time. Creation time mainly depends on the size of each resource. Compiling a list of all terminologies except for EntrezGene takes approximately 90 seconds; however, EntrezGene is very large, and processing it requires 3 to 4 minutes.

The requested terminology list is compiled on the fly, based on local copies of the external terminology resources. Whenever one of the local copies becomes out-of-date with respect to its remote source, an **Update** button is shown next to the corresponding resource name. This button triggers downloading the latest version to the corresponding resource name. Please note that, due to data volume and bandwidth restrictions, updating may take several minutes for some of the resource.

The Bio Term Hub (BTH) is a combined terminological resource created by dynamically sourcing entity names and their identifiers from reference databases. A web interface allows a user to access selected resources and download them in an uniform format.

<http://www.ontogene.org/resources/termdb>

### OntoGene Entity Recognition (OGER)

ChEBI, MeSH, NCB, Expasy, SIB

seconds/document	1.06943 s	1st
seconds/Byte	0.00086 s	1st
seconds/annotation	0.07227 s	
annotations/document	14.7923	
time between failures	[no failure]	1st (shared)
time to repair	[no failure]	1st (shared)

The OntoGene's Biomedical Entity Recogniser (OGER) is a RESTful web service implemented on top of the BTH which allows a remote user to batch annotate a collection of documents. Recently, we have participated in a community-organized evaluation of Bio Text Mining services (BioCreative/TIPS), in which our system obtained the best results according to several of the evaluation metrics.

<http://www.ontogene.org/resources/oger>

**Methotrexate enhances the anti-inflammatory effect of CF101 via up-regulation of the A3 adenosine receptor expression.**

**Abstract** Methotrexate (MTX) exerts an anti-inflammatory effect via its metabolite adenosine, which activates adenosine receptors. The A3 adenosine receptor (A3AR) was found to be highly expressed in inflammatory tissues and peripheral blood mononuclear cells (PBMCs) of rats with adjuvant-induced arthritis (AIA). CF101 (IB-MECA), an A3AR agonist, was previously found to inhibit the clinical and pathological manifestations of AIA. The aim of the present study was to examine the effect of MTX on A3AR expression level and the efficacy of combined treatment with CF101 and MTX in AIA rats. AIA rats were treated with MTX, CF101, or both agents combined. A3AR mRNA, protein expression and exhibition were tested in paw and PBMC extracts from AIA rats utilizing immunohistochemistry staining, RT-PCR and Western blot analysis. A3AR level was tested in PBMC extracts from patients chronically treated with MTX and healthy individuals. The effect of CF101, MTX and combined treatment on A3AR expression level was also tested in PHA-stimulated PBMCs from healthy individuals and from MTX-treated patients with rheumatoid arthritis (RA). Combined treatment with CF101 and MTX resulted in an additive anti-inflammatory effect in AIA rats. MTX induced A3AR and A3AR over-expression in paw cells from treated animals. Moreover, increased A3AR expression level was detected in PBMCs from MTX-treated RA patients compared with cells from healthy individuals. MTX also increased the protein expression level of PHA-stimulated PBMCs from healthy individuals. The increase in A3AR level was counteracted in vitro by

**Disambiguation challenges**

Human prostate cancer metastases target the hematopoietic stem cell niche to establish footholds in mouse bone marrow. HSC homing, quiescence, and self-renewal depend on the bone marrow HSC niche. A large proportion of solid tumor metastases are bone metastases, known to usurp HSC homing pathways to establish footholds in the bone marrow. However, it is not clear whether tumors target the HSC niche during metastasis.

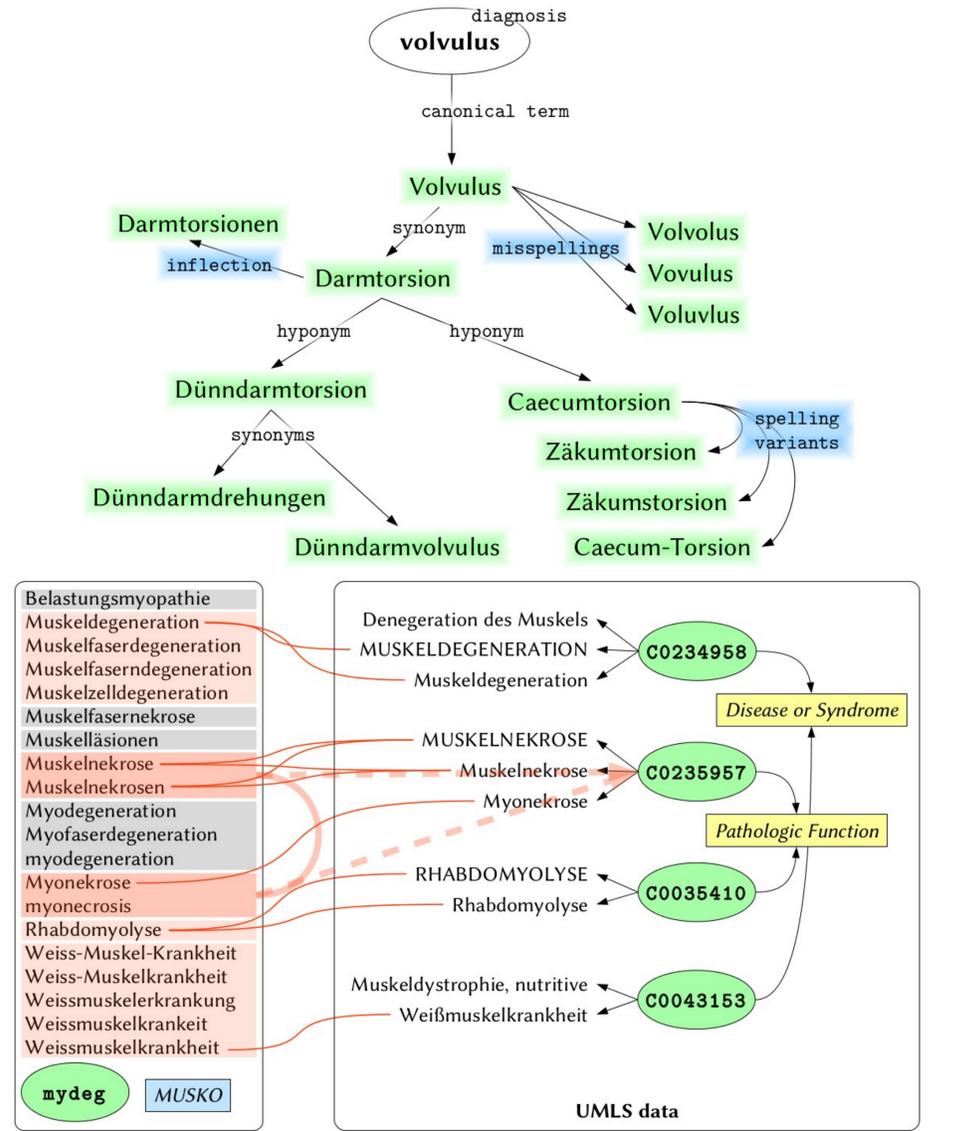
**Legend**

- disease
- chemical
- sequence
- gene/protein
- biological process
- organism
- cell

gene/protein	galactoside 2-alpha-L-fucosyltransferase 1	PR:00007702	Protein Ontology
cellular component	Hedgehog signaling complex	GO:0035301	Gene Ontology
gene/protein	chaperone protein HscA	PR:00022925	Protein Ontology
chemical	N-(3-carboxypropanoyl)-N-hydroxycadaverine	CHEBI:50443	CHEBI

The biggest problem in dealing with biomedical text entities is the large degree of ambiguity. The same name can refer to several different entities, and besides some common language names can be used occasionally as domain terms. Therefore disambiguation is a major challenge. In some types of text (e.g. clinical reports), other problems might emerge, such as misspellings and the usage of variants. The usage of reference Ontologies can help to deal with this problem.

### Terminologies and Ontologies



### Assisted Curation

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

In an NIH-funded project in collaboration with the RegulonDB group (CCG, UNAM), we are developing an interface for assisted curation (ODIN) which offers efficient manipulation of text and annotations. This system can use the predefined entities provided by the Bio Term Hub, or allow for customized vocabularies. The introduction of text mining technologies in an assisted curation pipeline is expected to increase the throughput of the curation process without compromising the quality of the results.

### Funding

The OntoGene/BioMeXT group is partially supported by the Swiss National Science Foundation (SNF): grants 100014-118396/1, 105315-130558/1, CR3011\_162758 (F. Rinaldi). We are also involved in an SNF-NRP74 project, and in a large project funded by the US-NIH. Additionally we received funding from the COGITO foundation and the Federal Food Safety and Veterinary Office (FSVO). Recently a new project proposal has been approved by the Commission for Technology and Innovation (CTI).