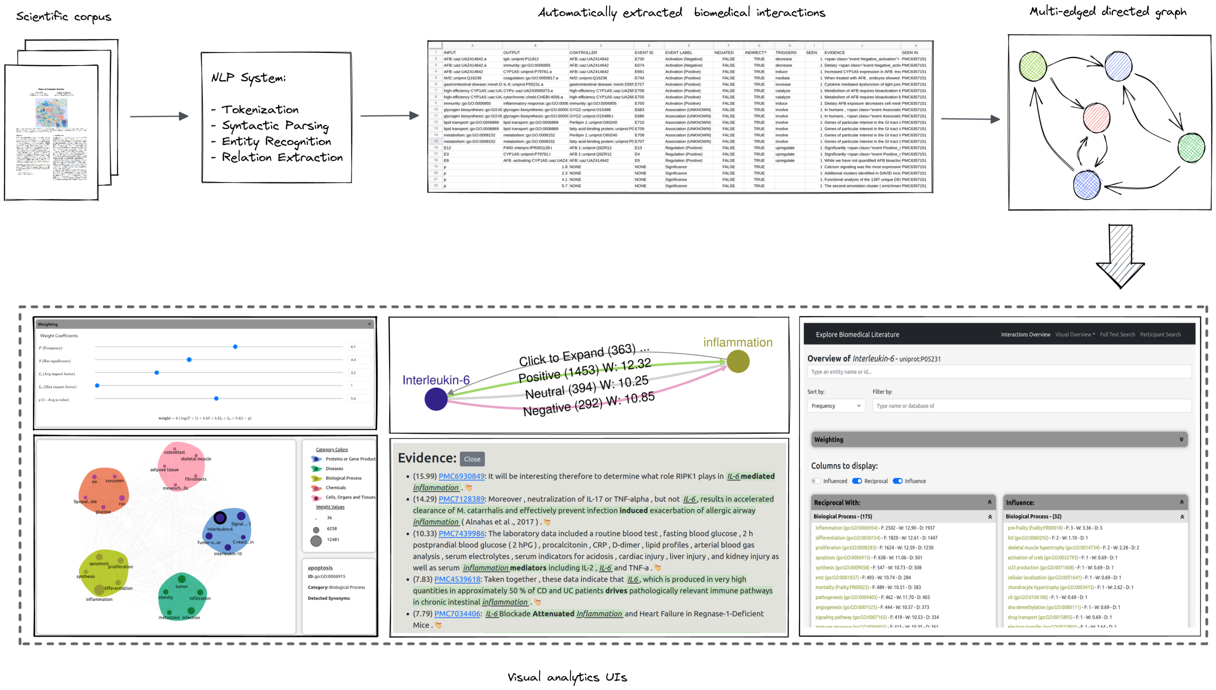
Visualizing Interaction Networks and Evidence in Biomedical Corpora

Enrique Noriega-Atala* Md Rahat-Uz-Zaman† Ruchika Bhat‡ Mladen Jergović§
Stephen G. Kobourov¶ Janko Nikolich-Žugich‖

The University of Arizona



- (15.99) PMC6930849: It will be interesting therefore to determine what role RIPK1 plays in *IL-6* **mediated** *inflammation* .
- (14.29) PMC7128389: Moreover , neutralization of IL-17 or TNF-alpha , but not *IL-6* , results in accelerated clearance of M. catarrhalis and effectively prevent infection **induced** exacerbation of allergic airway *inflammation* (Alnahas et al ., 2017) .
- (10.33) PMC7439986: The laboratory data included a routine blood test , fasting blood glucose , 2 h postprandial blood glucose (2 hPG) , procalcitonin , CRP , D-dimer , lipid profiles , arterial blood gas analysis , serum electrolytes , serum indicators for acidosis , cardiac injury , liver injury , and kidney injury as well as serum *inflammation* **mediators** including IL-2 , *IL-6* and TNF-a .
- (7.83) PMC4539618: Taken together , these data indicate that *IL6* , which is produced in very high quantities in approximately 50 % of CD and UC patients **drives** pathologically relevant immune pathways in chronic intestinal *inflammation* .
- (7.79) PMC7034406: *IL-6* Blockade **Attenuated** *Inflammation* and Heart Failure in Regnase-1-Deficient Mice .

Reciprocal With:

Biological Process - (175)

inflammation (go:GO:0006954) - F: 2502 - W: 12.90 - D: 1937
differentiation (go:GO:0030154) - F: 1820 - W: 12.61 - D: 1447
proliferation (go:GO:0008283) - F: 1624 - W: 12.59 - D: 1230
apoptosis (go:GO:0006915) - F: 638 - W: 11.06 - D: 501
synthesis (go:GO:0009058) - F: 547 - W: 10.73 - D: 508
emt (go:GO:0001837) - F: 493 - W: 10.74 - D: 284
mortality (frailty:FR00023) - F: 489 - W: 10.51 - D: 383
pathogenesis (go:GO:0009405) - F: 462 - W: 11.70 - D: 403
angiogenesis (go:GO:0001525) - F: 444 - W: 10.57 - D: 373
signaling pathway (go:GO:0007165) - F: 419 - W: 10.53 - D: 334

Influence:

Biological Process - (32)

pre-frailty (frailty:FR00018) - F: 3 - W: 3.36 - D: 3
ltd (go:GO:0060292) - F: 2 - W: 1.10 - D: 1
skeletal muscle hypertrophy (go:GO:0014734) - F: 2 - W: 2.28 - D: 2
activation of creb (go:GO:0032793) - F: 1 - W: 0.69 - D: 1
ccl3 production (go:GO:0071608) - F: 1 - W: 0.69 - D: 1
cellular localization (go:GO:0051641) - F: 1 - W: 0.69 - D: 1
chondrocyte hypertrophy (go:GO:0003415) - F: 1 - W: 2.62 - D: 1
cit (go:GO:0106106) - F: 1 - W: 0.69 - D: 1
dna demethylation (go:GO:0080111) - F: 1 - W: 0.69 - D: 1
drug transport (go:GO:0015893) - F: 1 - W: 0.69 - D: 1

Figure 1: An overview of the proposed system designed to help biomedical researchers to search, visualize and interact with biological networks derived via information extraction tools. The input is texts from scientific papers. An information extraction system is used to detect and assemble structured relations between biological concepts (e.g., chemicals, proteins, diseases). The structured data is then used to create a network, connecting all the individual pieces together. Finally, the domain expert interacts with the network and text via multiple user interfaces, to find relevant interactions and backing evidences.

Abstract

The abundance of scientific articles published and indexed in publicly accessible repositories has spurred the research and development of automated information extraction systems. The output of such systems can be used to assemble large networks capturing the understanding of mechanistic pathways and their interactions as represented in the underlying body of research.

We describe a system designed to help researchers search, visualize and interact with biological networks derived via information extraction tools. As input, the system takes a dataset of biological and biochemical interactions automatically generated by an information extraction system and provides an interface designed to search, visualize and interact with the data. The usage paradigm consists of identifying a starting point for a search, then using the data's network structure by incrementally exploring the immediate neighborhood of the elements displayed by the system.

Our system differs from prior work as it leverages both the network structure in the data and the natural language text backing those connections: every connection displayed is traceable back to the documents and phrases in the corpus that support that specific piece of information. We also present two case studies with immunobiology researchers using the system to find previously unknown relationships between biological entities. While the evidence suggesting these relationships already existed, it was scattered across the literature, and existing specialized web databases and domain-search engines could not find it. The system is open-source, with the code publicly available on GitHub.

Index Terms: Information systems—Search interfaces; Computing methodologies—Visualization—Information extraction

*e-mail: enoriega@arizona.edu
†e-mail: rahatzamancse@arizona.edu
‡e-mail: ruchikabhat@arizona.edu
§e-mail: mjergovic@arizona.edu
¶e-mail: kobourov@cs.arizona.edu
‖e-mail: nikolich@arizona.edu

This research was, in part, funded by the U.S. Government. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government.

1 Introduction

For more than a decade, the research productivity in several life sciences has seen an astonishing growth rate. Between the years 2004-2013, an average of 730,000 research publications per year were added in PubMed [41]. Since 2013, this average has grown substantially to 1,970,000 per year [1].

Certainly, the abundance of available information that results from the aforementioned growth is a positive development. It is, however, becoming difficult to navigate through a vast body of research to search for, identify and assimilate all the relevant related work without accidentally missing key information. Just by size alone, the task of searching for and assimilating knowledge contained within vast scientific repositories has become infeasible for individuals and even small groups of people.

In the biomedical domain, this situation has fostered the research and development of specialized machine reading systems [26, 39]. Such tools are aimed at helping information seekers increase recall when conducting a literature search and automatically extract structured information that describes the mechanistic building blocks of biological processes and biochemical interactions.

Machine reading tools ameliorate the issue of information overload, but do not eliminate it. When machine reading tools are deployed at scale, the volume of extracted information tends to be high, and when assembled, it often exhibits a complex network structure. As a result, the visualization of those structures also comes with its own set of challenges.

Visualization tools for the biomedical domain need to display large amounts of structured data while highlighting the biological semantics encoded in the data [33]. Machine reading and visualization tools complement each other but are not necessarily well integrated with each other.

Our main contribution is an integrated web-based system that (1) combines a Natural Language Processing (NLP) pipeline tailored to biomedical researchers, (2) provides a user interface to efficiently search for and locate mechanistic interactions, the underlying textual evidence, and a pointer to the source of the information; and (3) facilitates the exploration of the underlying network structure exhibited by the data. The design goals of the system are:

- Index large collections of biomedical research publications by their natural language text, augmented with extracted mechanistic information that encodes the interactions between different biomedical entities (e.g., proteins, biological processes)
- Provide a user interface to efficiently search for and locate mechanistic interactions, the underlying textual evidence, and a pointer to the source of the information.
- Facilitate exploration of the network structure by inspection of the neighborhood of a given interaction or any of its participant elements.
- Provide a flexible and extensible domain agnostic configurable system, with a simple ingestion data format that decouples the system from any specific information extraction tool.

A live version of the system is publicly available[2] and the source code and data are open source[3]. In addition, we provide a walkthrough video that describes the main features of the system in the supplementary materials.

2 Related Work

PubMed Central (PMC) [31] is a biomedical and life sciences archive at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM). PMC also includes NLM's extensive print and licensed electronic journal holdings and supports biomedical research. PMC is the premier repository for biomedical and life sciences research, publicly available since 2000 with a web interface [43] to explore the literature using keyword and phrase queries.

Information Retrieval (IR) and Information Extraction (IE) technology, also referred to in tandem as *Machine Reading*, aid the search and processing of documents present in large corpora, such as PMC. IR systems [17, 24] allow to efficiently query, retrieve and rank documents, and IE systems [8, 20, 40] process text to automatically detect and extract entities and their relations, inducing structure to the information present in natural language, as well as inferring the context of those structures [27, 28]. While machine reading is a domain-agnostic field, a large body of research is devoted to specialize these technologies for the biomedical and life sciences domains [12, 25, 39]. In this work, we propose a system that complements the features of keyword and phrase searching with the finer-grained detail achieved by extracting structured information with machine reading and natural language processing tools.

Biological processes studied in medical and life sciences research often exhibit complex network structures. Naturally, there is a great deal of work on networks that considers different algorithms and visualization metaphors, among others [1, 9, 29, 34]. Domain-specific visualization tools, most notably Cytoscape [33], help integrate, display and navigate large databases of biomolecular interaction networks while being agnostic of the difference in biological semantics of the different biomedical disciplines. Our proposed system differs from such work by providing different views that combine the visual representation of the underlying data and the text representation (evidence) for the data, as well as providing custom ranking features for the biological interactions to help narrow down the search space.

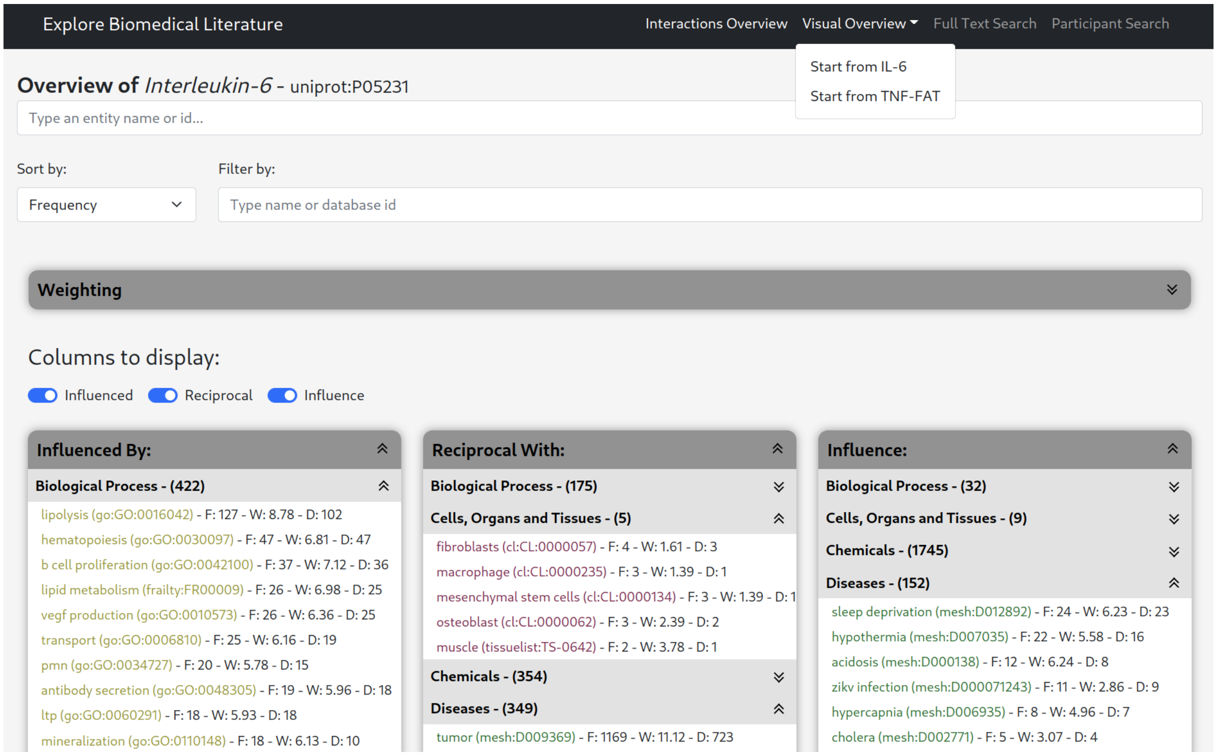
Many approaches have been proposed for extracting and visualizing data from scientific publications. Isenberg et al. [14] assembled a dataset composed of curated information from all papers appearing in visualization conferences. Continuing this line of research, Isenberg et al. [15] propose an approach to build a taxonomy of visualization papers based on the most prominent keywords, designed to help domain experts narrow down the search space to relevant articles of their interest. Some visual analytics tools operate directly on structured tabular data present in scientific articles. The approach presented in [42] proposes a time navigation technique to update tables by directly manipulating its values. The technique allows one to drag a table's cells to change the time period, while a line chart overlays on top of the table to provide an overview of the changes. Yalçın et al. [44] introduce a system for tabular data exploration, that aggregates records and visualizes aggregation characteristics. Literature search systems such as [30] show yearly publication trends, keywords, and authors. VAiRoma [5] is a filterable Roman history discovery system that uses Wikipedia text and metadata such as location and time to find relevant information about specific locations and times. While these approaches help to find relevant documents, topics and trends, they do not get to the level of detail needed to visualize the relevant phrases and sentences that contain interactions between different entities in the underlying text.

LitSense [35], a graph visualization system based on the citation network of documents with similar topics, relies on seeding with a small number of documents relevant to a topic and then expanding the search via the citation network. Visualization systems for relevant document discovery based on context [2] and keywords [6] have also been proposed. GRAM visualizes worldwide scholarship activity and supports map-based interactive features, including semantic zooming, panning, and searching [3]. ReMatch [13] analyzes research papers and collaborations to identify experts in a given field.

[1] https://www.ncbi.nlm.nih.gov/pmc/about/intro/

[2] http://vpe.cs.arizona.edu/

[3] https://github.com/clulab/bioliterature_network_visualizer



Overview of *Interleukin-6* - uniprot:P05231

Influenced By:

Biological Process - (422)

lipolysis (go:GO:0016042) - F: 127 - W: 8.78 - D: 102
hematopoiesis (go:GO:0030097) - F: 47 - W: 6.81 - D: 47
b cell proliferation (go:GO:0042100) - F: 37 - W: 7.12 - D: 36
lipid metabolism (frailty:FR00009) - F: 26 - W: 6.98 - D: 25
vegf production (go:GO:0010573) - F: 26 - W: 6.36 - D: 25
transport (go:GO:0006810) - F: 25 - W: 6.16 - D: 19
pmn (go:GO:0034727) - F: 20 - W: 5.78 - D: 15
antibody secretion (go:GO:0048305) - F: 19 - W: 5.96 - D: 18
ltp (go:GO:0060291) - F: 18 - W: 5.93 - D: 18
mineralization (go:GO:0110148) - F: 18 - W: 6.13 - D: 10

Reciprocal With:

Biological Process - (175)

Cells, Organs and Tissues - (5)

fibroblasts (cl:CL:0000057) - F: 4 - W: 1.61 - D: 3
macrophage (cl:CL:0000235) - F: 3 - W: 1.39 - D: 1
mesenchymal stem cells (cl:CL:0000134) - F: 3 - W: 1.39 - D: 1
osteoblast (cl:CL:0000062) - F: 3 - W: 2.39 - D: 2
muscle (tissuelist:TS-0642) - F: 2 - W: 3.78 - D: 1

Chemicals - (354)

Diseases - (349)

tumor (mesh:D009369) - F: 1169 - W: 11.12 - D: 723

Influence:

Biological Process - (32)

Cells, Organs and Tissues - (9)

Chemicals - (1745)

Diseases - (152)

sleep deprivation (mesh:D012892) - F: 24 - W: 6.23 - D: 23
hypothermia (mesh:D007035) - F: 22 - W: 5.58 - D: 16
acidosis (mesh:D000138) - F: 12 - W: 6.24 - D: 8
zikv infection (mesh:D000071243) - F: 11 - W: 2.86 - D: 9
hypercapnia (mesh:D006935) - F: 8 - W: 4.96 - D: 7
cholera (mesh:D002771) - F: 5 - W: 3.07 - D: 4

Figure 2: Entity overview interface. The display anchored entity is IL-6. The anchor entity can be changed to search by name or database identifier in the entity search box. The elements in the entity columns can be sorted by multiple criteria and filtered by name or database identifier. The entity columns can be hidden by toggling their display indicator.

Visual analytics tools can also be deployed to identify and verify factual claims from large transcripts of conversations [32].

Our proposed system contains a feature set that partially overlaps and complements related work in this space: First, it leverages machine reading technology by indexing natural language text and structured biochemical and biological interactions present in the literature. Second, it provides visualization interfaces to interactively explore the network structure, while pointing to the specific underlying evidence in the literature from which it was extracted.

The closest related work is a visual analytics system by Peng-Hsuan et al. [21] to explore interactions between biomedical entities present in PubMed abstracts, automatically extracted by natural language processing tools, and creating an interactive semantic graph. The goal is to go beyond the basic search capabilities presented by PubMed, by providing additional knowledge insights in the presented results. Our system, while having similar design goals, considers the full text rather than abstracts and is not specific to PubMed, but can be used for any scientific corpus.

3 Machine Reading and Data Format

The primary source of information handled by our information system is *biomedical research papers*. In contrast to related work (§2), we seek to provide a search engine and interactive visualization over a dataset composed of a large collection of entity interactions and their evidence, rather than the full text of the publications.

For the purpose of this paper, an *entity interaction* is any relation that associates two elements and has biological semantics. Examples of such interactions include the regulation of a disease by a chemical or the inhibition of a chemical reaction in a protein by another protein. These interactions can be complex nested structures; e.g., `RAF kinase, which regulates the phosphorylation of MEK`, includes one entity that regulates the phosphorylation of another entity. When we encounter these types of interactions, we "binarize" them into directed interactions, captured by directed edges in a graph (instead of necessitating the use of hypergraphs).

While different types of interactions and their semantics vary in the medical an biological sub-fields, they share a common underlying structure: there are two participating entities and a named relation between them. The relation may also have *polarity*, either positive or negative. In a *directed* interaction the participants take the roles of *controller* and *controlled* entities, with the former influencing the later.

We adapted the REACH biomedical information extraction tool [39], to process biomedical research papers, and extract the entity interactions indexed and visualized in our system. REACH takes as input the text of a scientific article (either as an unformatted text file or as an XML file following the MEDLINE/PubMed data type definition[4]). The input text is processed by a pipeline of natural language processing (NLP) components, that tokenize it, split it into sentences, annotate it with syntax information, etc. These annotations are then used by a set of rules, or a *grammar*, to recognize syntactic and lexical patterns that describe interactions between different types of biological concepts, or *entities*. Once a sentence in the input text matches a rule, REACH assembles a data structure that represents the interaction between a pair of entities. Finally, these interactions are used to build an output text file, in tabular format, containing the extracted information. The upper row of Fig. 1 illustrates the workflow of REACH and table 1 shows the specific types of interactions and participant entities handled by the system.

REACH handles the normalization of entities and interactions with robust detection algorithms. Entities can have synonyms, different spellings or abbreviations. REACH can handle such variations

[4] `https://www.nlm.nih.gov/bsd/licensee/data_elements_doc.html`

Four interaction types: Promotions, inhibitions, associations, regulations.
Five entity types: Protein & gene products, diseases, biological processes, chemicals, cells/organs/tissues.

Table 1: Entity types and interactions types recognized by the NLP system.

and normalizes participant entities to database identifiers, such as UniProt [38] and MeSH [22]. For example, *Interleukin 6*, a cytokine frequently associated with inflammatory processes, is commonly referred to as *IL-6* or *IL 6*. REACH assigns the UniProt identifier `P05231` to all its variants and ensure consistency on all the interactions that have it as a participant. Interaction names are handled similarly, using a list of terms that refer to each interaction type. Every interaction extracted, in addition to its structural elements, also contains its *evidence*: the natural language phrase from which the interaction was obtained.

The REACH output is stored as tabular data files, which are later ingested to generate a database that indexes the previously extracted entities, regulations, interactions and evidence sentences. Table 2 contains a summary of the primary fields of the tabular structure of the output files.

Field	*Description*
Controller	Identity of the controller entity
Controlled	Identity of the controlled entity
Interaction	Name and polarity of the interaction between the entities
Frequency	# of detections of the interaction in the document
Evidence	Phrases describing the interaction in the document

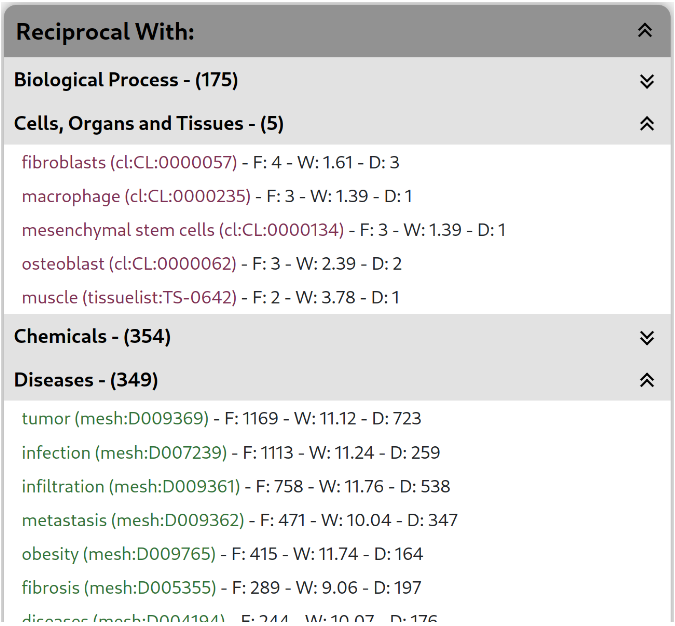
Table 2: Input file format fields.

In addition to entity interactions, we extend REACH to detect and extract the evidence of statistical quantities associated with experimental results, e.g., statistical significance numbers, correlation coefficients and confidence intervals. While statistical evidence extractions are not part of the primary database and are not part of the network structure induced by the interactions, they are used as component parts to rank the extractions, described in §4.1.

The original implementation of REACH was independently evaluated by a team of domain experts [39]. REACH was used to extract interactions from one thousand papers related to the *Ras signaling pathway*. The evaluation looked at precision, defined as the proportion of extractions considered to be "largely correct," and throughput, the number of interactions extracted by the system per day. REACH had the highest throughput and second highest precision among the participant systems in the evaluation.

We note that wile these adaptations of REACH are needed to meet the design goals of our system, the system is not tied to REACH. Any information extraction tool that can be adapted to produce the tabular data format or whose output can be transformed into the tabular data format can be used as part of our system. A sample input file following the data format specified in this section is shown in Fig. 11, in the supporting material. The source code of our fork of REACH along with the adaptations mentioned in this section, as well as the parameters used to generate the data, are open source and available online[5].

[5] https://github.com/clulab/reach/releases/tag/pacificvis2023



Reciprocal With:

Biological Process - (175)

Cells, Organs and Tissues - (5)

fibroblasts (cl:CL:0000057) - F: 4 - W: 1.61 - D: 3
macrophage (cl:CL:0000235) - F: 3 - W: 1.39 - D: 1
mesenchymal stem cells (cl:CL:0000134) - F: 3 - W: 1.39 - D: 1
osteoblast (cl:CL:0000062) - F: 3 - W: 2.39 - D: 2
muscle (tissuelist:TS-0642) - F: 2 - W: 3.78 - D: 1

Chemicals - (354)

Diseases - (349)

tumor (mesh:D009369) - F: 1169 - W: 11.12 - D: 723
infection (mesh:D007239) - F: 1113 - W: 11.24 - D: 259
infiltration (mesh:D009361) - F: 758 - W: 11.76 - D: 538
metastasis (mesh:D009362) - F: 471 - W: 10.04 - D: 347
obesity (mesh:D009765) - F: 415 - W: 11.74 - D: 164
fibrosis (mesh:D005355) - F: 289 - W: 9.06 - D: 197
diseases (mesh:D004194) - F: 244 - W: 10.07 - D: 176

Figure 3: Entity column widget. Groups entities by type and for each entry, it displays the name, database identifier, frequency of the interaction in the database (F), custom weight value (W) and number of documents that contain the interaction (D).

4 VISUALIZATION SYSTEM

Here we described the web-based user interface for exploring the database of entity interactions by querying the structure of the network and the text of the underlying evidence.

Node-link diagram representations of networks are intuitive for small instances, but start to look like hairballs as the underlying data gets larger as explained by Gibson et al. [10]. Instead, we present several user interfaces to enable the domain experts to narrow down the search space, find a starting point, and then gradually and on demand, explore the local neighborhood of the currently displayed elements.

The interfaces vary by the degree of detail that they present and by the type of query that they support. They can be grouped into two high-level categories: *structural search* (§4.1, §4.2, §4.4) and *textual search* (§4.5, §4.6)

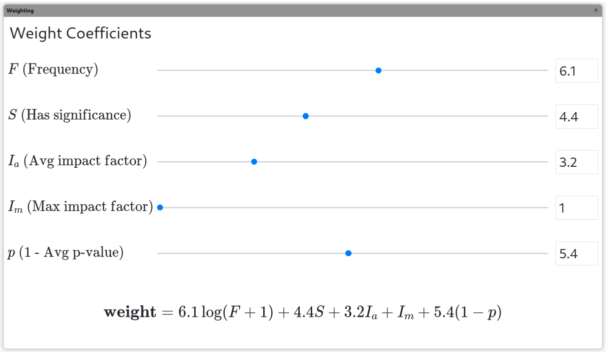
4.1 Entity Overview

The *entity overview* page is the entry point to the web interface. It is designed to explore the direct interactions with respect to an entity of interest, the *anchor entity*, which can be chosen and changed at any time via a search box widget, with auto-complete functionality. Fig. 2 depicts the entity overview interface with *IL-6* selected as the anchor.

The main components of this view are three columns that display the entities that are participants in a relation alongside the anchor entity. Every column lists the entities stratified by their type (listed in table 1) and displays their information and corpus statistics. Clicking on an entity will redirect the interface to the *Node-Link* view (§4.2). Fig. 3 shows a close-up view of a part of an entity column.

The three entity columns are functionally equivalent. The difference between them lies on the direction of the interactions. Columns `Influence`, `Influenced by` and `Reciprocal` contain the entities for which the anchor entity is the *controller*, *controlled* or for which there exist interactions on both directions, respectively.

To make the inspection of the overview simpler, the elements in the columns can be filtered and sorted by name, database identifier,



Weighting

Weight Coefficients

F (Frequency) 6.1

S (Has significance) 4.4

I_a (Avg impact factor) 3.2

I_m (Max impact factor) 1

p (1 - Avg p-value) 5.4

$$\textbf{weight} = 6.1\log(F+1) + 4.4S + 3.2I_a + I_m + 5.4(1-p)$$

Figure 4: Custom interactions weight widget. Coefficients for each variable can be tuned either by adjusting the slider controls or by directly typing values in. The value returned by this formula is used for all interactions displayed in the current view and for ranking the results. An explanation of each variable is found in §4.1.

or corpus frequency.

Additionally, we provide a widget (depicted in Fig. 4) to tune a *custom weight function* evaluated over each interaction. The function is a linear combination of multiple metadata variables, allowing one to adjust the importance given to each according to his or her needs. The different variables detected by the system are:

- *Interaction Frequency*: Number of times an interaction of between both entities was detected in the literature by the information extraction system.
- *Significance*: Number of documents where the interaction was detected and reported statistical significance. The information extraction system has rules to detect and extract p-values and confidence intervals.
- *Average impact factor*: Computed by considering the impact factors of the journal/conference that published the documents where the interaction was detected (the value defaults to zero if the venue lacks an impact factor).
- *Max impact factor*: Similarly, the maximum impact factor of the journals or venues that published the documents where the interaction was detected.
- *1 - $\bar{p}$*: Detected p-values are averaged and used as a proxy of the statistical power of the experiments reported by the documents. For the sake of consistency we use 1 - $\bar{p}$, so that higher values represent stronger contributions to the custom weight function.

The set of 5 variables was obtained after several rounds of interactions with domain experts and the weight function helps combine them in order to rank the interactions by the relative importance given to each of them. For the study described in §5.1, the domain experts converged to weights that produced the results they expected to see for input instances they were familiar with and our system can default to these values.

4.2 Node-Link View

If the Entity Overview serves as a starting point to conduct a search, the *Node-Link* view is the interface designed to “zoom into” the mechanistic details of individual interactions or small sub-networks.

This interface displays interactions as a node-link diagram where each edge represents a specific relation between the entities/nodes. Nodes are labeled with human-friendly entity names. A domain expert may be used to refer to the same entity by other names (e.g., Interleukin 6, IL-6, IL6) and in order to reduce the ambiguity introduced by potential variations, we also include the entity’s database identifier. The sub-figure in top-left of Fig. 5 depicts an expanded view of the interactions between `IL-6` and `Inflammation`.

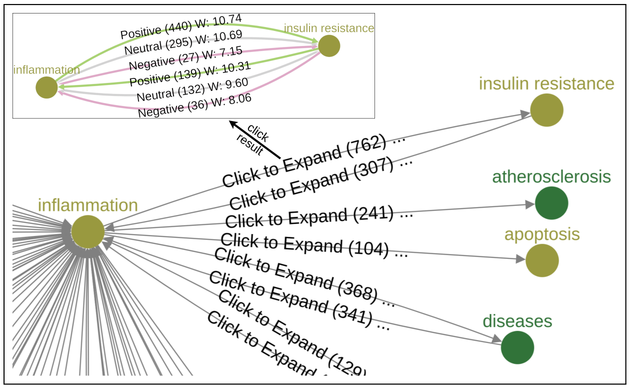


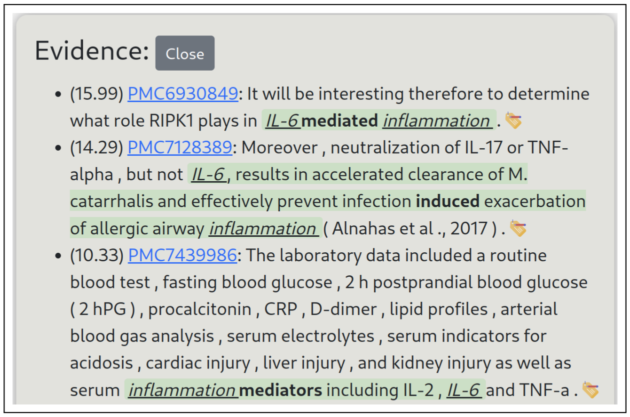
Figure 5: The main figure shows the Node-Link view after double-clicking on `Inflammation` which fetches additional neighbors. The sub-figure (top-left) shows the Node-Link view of `Inflammation` and `insulin resistance`. Polarity is represented by both color and label. Interaction direction is represented by arrows. The numbers correspond to interaction frequency in the database and the value of W is the customized interaction weight (see Fig. 4).

A pair of entities can be related by different types of interactions. The interactions is grouped by direction and polarity and each group is represented by one directed edge in the visualization. The edge color encodes the polarity of the relation and the label displays the frequency of the interaction in the corpus. Note that the frequencies can be normalized (e.g., based on usage in the underlying corpus) but the domain experts opted for raw frequencies. Edges also display weights, representing the importance of the interactions, based on the custom weights described in §4.1. Clicking on an edge fetches all the evidence in the database backing the particular interaction and displays it in an evidence panel (§4.3).

This interface allows for sorting the elements in the network of interactions by inspecting the structural properties of the edges (i.e., direction and polarity) and retrieving the evidence backing findings. By design, the scope of the information shown by the Node-Link view is narrow, to reduce clutter and help focus on the details of a particular interaction. Nonetheless, sometimes it is necessary to inspect one step further. Double-clicking a node retrieves all the adjacent entities in the database, along with their interactions. This feature allows for an incremental analysis by following a “trail” of edges. Fetching additional entities and interactions increases the amount of data shown and the interface allows for zooming and panning, as well as grouping of edges. When a grouped edge is clicked, it expands into multiple edges, each representing a different polarity. The main portion of Fig. 5 shows an example of a node-link visualization after retrieving the neighbors of one of the participants.

4.3 Evidence Panel

The evidence panel is a widget that displays the phrases and sentences in the literature containing an interaction. First mentioned in §4.2, the evidence panel widget is reused anywhere in the system where the evidence is displayed, bringing along its complete feature set. Fig. 6 shows an evidence panel with phrases that contain positive interactions between `IL-6` and `inflammation`. The widget features visual aids to quickly identify the elements of the interaction within a phrase. The text span of the interaction is highlighted and color-coded to indicate its polarity. The participant entities are underlined, the interaction keyword is bold, and each phrase has a link to its source document with the SJR journal rank indicator [11].



Evidence:

- (15.99) PMC6930849: It will be interesting therefore to determine what role RIPK1 plays in *IL-6* **mediated** *inflammation* .
- (14.29) PMC7128389: Moreover , neutralization of IL-17 or TNF-alpha , but not *IL-6* , results in accelerated clearance of M. catarrhalis and effectively prevent infection **induced** exacerbation of allergic airway *inflammation* (Alnahas et al ., 2017) .
- (10.33) PMC7439986: The laboratory data included a routine blood test , fasting blood glucose , 2 h postprandial blood glucose (2 hPG) , procalcitonin , CRP , D-dimer , lipid profiles , arterial blood gas analysis , serum electrolytes , serum indicators for acidosis , cardiac injury , liver injury , and kidney injury as well as serum *inflammation* **mediators** including IL-2 , *IL-6* and TNF-a .

Figure 6: Evidence panel widget displaying phrases that contain evidence of interactions between `IL-6` and `Inflammation`. The sentences display visual aids to easily identify the elements of the interaction, a link to the source article and the impact factor of the journal containing the evidence phrase. Also depicted is the tagger widget to assign labels to each piece of evidence if necessary.

The goal of this widget is to help domain experts make a final determination of the veracity and utility of the interactions in the dataset. We include a *tagging* feature that allows to label on-the-fly individual phrases with any tag defined by the domain expert, e.g. *incorrect detection*, *relevant interaction*, *duplicate*, etc. The labels can be used to design and implement potentially customized features, such as custom filters, or text similarity models.

4.4 Visual Summary Overview

The visual summary overview interface is designed to be a "bird's eye" view of the most relevant entities in the network with respect to one or more *anchor entities*, similar to the entity of interest in §4.1. The system displays a node-link diagram with the *top n* entities by entity type, as described in Table 1. By default, the system computes a layout for the top 5 elements in each of the 5 types. The number of elements of each type can be changed on the fly and the layout is dynamically adjusted to accommodate the change. Fig. 7 shows an example of the visual summary overview with IL-6 as the anchor entity.

In this view of the data, the importance assigned to each entity/node is defined by the frequency with which it participates in an interaction with the anchors. This definition contextualizes the salience of each entity as a function of co-occurrence with the predefined set of anchors.

After every entity in the database is scored with respect to the anchor set, the top n entities per category are displayed in groups. The radius of a node is a function of the degree of entity in the network and the layout is computed to represent their pair-wise relations. This layout is governed by three main *force vectors*. The final force on a single node is the vector addition of all the forces.

1. **Category-centric force:** This force helps to keep nodes of the same category together in a pre-calculated location, determined by equation 1. (x_i, y_i) represents center of each category $C_i \in C$ where $0 \leq i \leq |C| - 1$, C being the set of categories. The width and height of the canvas are w and h respectively. Note that the scalar value of category-centric force for a node n_a is $|F_{ce}(n_a)| = C_{ce}$ where C_{ce} is constant.

$$\begin{aligned} x_i &= \frac{w}{2} + r\cos\frac{2\pi i}{|C|} \\ y_i &= \frac{h}{2} + r\sin\frac{2\pi i}{|C|} \end{aligned} \tag{1}$$

2. **Link force:** This force ensures that more tightly connected entities are closer in the layout, by assigning edge values proportional to the frequency of evidence between the endpoints. This can be presented as $F_l(n_a, n_b) = k_l E(n_a, n_b)$ where E gives us the frequency between n_a and n_b and k_l being a constant.

3. **Node-charge and collision forces:** These two repulsion forces help improve readability and prevent overlaps between nodes. Node-charge helps spread nodes in the canvas, while collision repulses nodes within the radius of a node. These two forces are defined in equation 2 where C_{co} and k_{ch} are constants, F_{co} and F_{ch} are the force vectors between node n_a and n_b at r distance for all nodes in the view.

$$|F_{co}(n_a, n_b)| = C_{co} \tag{2}$$

$$|F_{ch}(n_a, n_b)| = k_{ch}/r^2 \tag{3}$$

A smooth convex hull is drawn around all entities that belong to the same category to separate and highlight each group. We use a consistent colorblind-safe color scheme across all the webpage, interface and visualization components. This required selecting 13 colorblind-safe colors: 5 qualitative colors for the convex hulls, 5 (matching) qualitative colors for the nodes in the hulls, and 3 diverging colors to encode the relation edges in the detailed relation view page. The color assignment is shown in the legend (to the right of the main visualization window). We used a careful combination of the *Light qualitative color scheme* and *Muted qualitative color scheme* from [37]. Finally, we used an online tool, Adobe Color [6], to verify that our choices are colorblind safe.

Clicking on an edge (or a pair of nodes), modifies the view to provide details about the underlying interactions as shown in Fig. 8. In this modality, clicking on one of the 6 possible interactions fetches the evidence phrases that back the specific interaction and shows it in an evidence panel, along with its complete feature set.

At the right of the interface, a legend explains the color codes used for each entity category, a scale indicating the relation between the nodes radii and their frequency in the dataset and a panel showing metadata for the currently selected node, such as it is human friendly name, database id, and other synonyms assigned to the same concept that the node represents.

In addition to the core functionality in the visualization, this view has a control panel to the left side of the network visualization to change the configuration parameters to dynamically change the anchor entity or entities, the number of most relevant entities displayed per group, the opacity of the edges and the scaling of the node radii. This panel also has the controls to change the opacity of individual components like text labels, inter-group and intra-group edges, and some graph layout parameters (e.g. force constants discussed in §4.4). This helps domain experts share different findings with screenshots to their liking.

4.5 Full-Text Search Interface

So far, the user interfaces presented make an assumption about a previously known starting point to initiate a search. However, this is not always the case. We provide a full-text search interface for the cases where there is no clear definition of a starting point, such as any particular entity or entities of interest. In this interface, one can type a search query and the system returns the sentences that contain an extraction (evidence phrases) displayed inside an evidence panel (§4.3). Fig. 9 displays the query widget in the full-text search view.

A query can be any word or phrase that resembles the kind of interactions that the domain expert is looking for. The query is

[6] `https://color.adobe.com/create/color-accessibility`

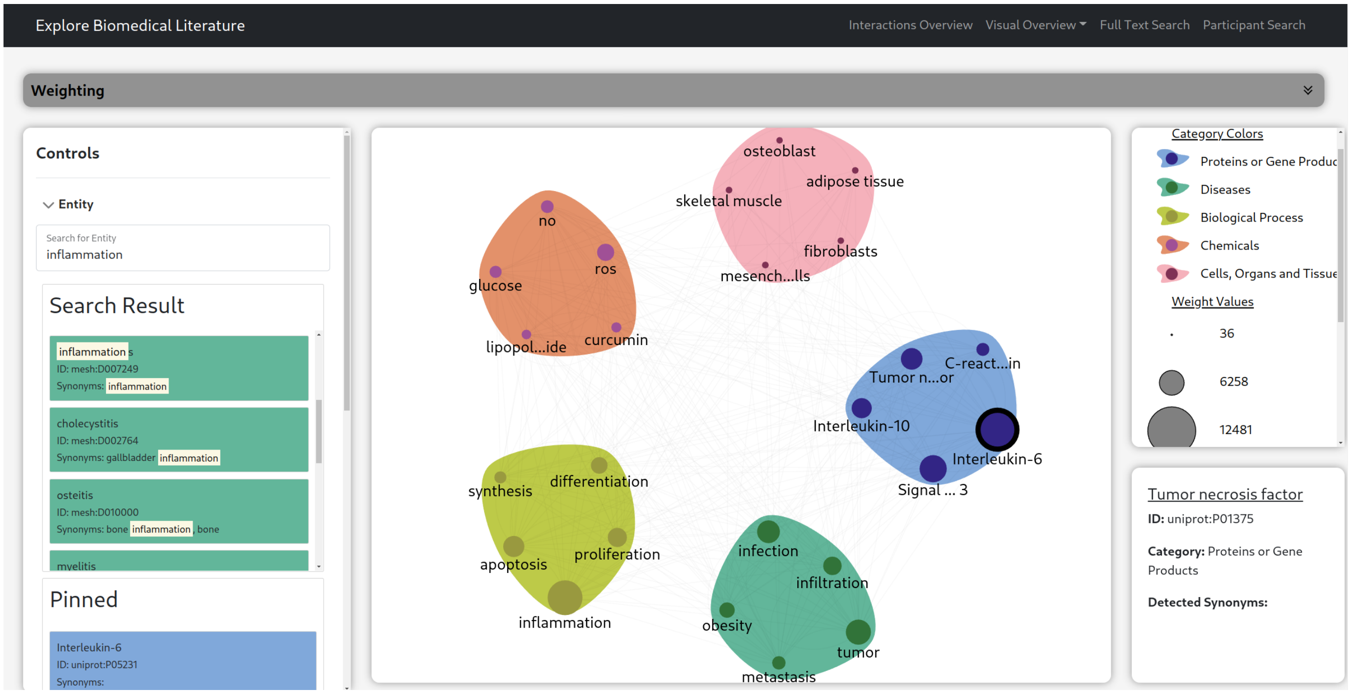


Figure 7: Visual summary overview page. The center canvas contains the top five entities grouped by type, with respect to IL-6 (highlighted by a black ring). The left panel contains controls to adjust the visualization, search and pin entities. The right panel contains legends. The bottom right section shows additional information about a hovered entity.

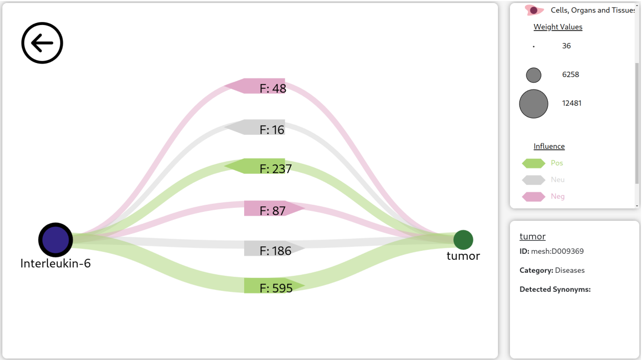
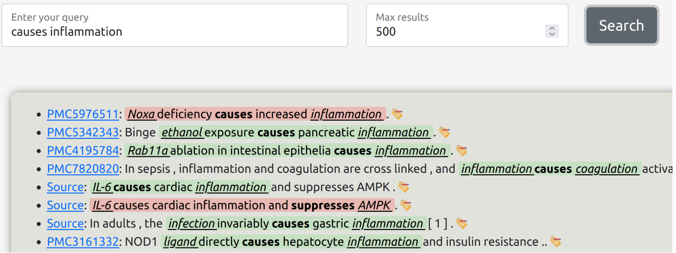


Figure 8: Transition to detailed relation view page. After clicking on 2 nodes (IL-6 and tumor) or an edge between them in the center panel of Fig. 7, all nodes move away from center and the 2 selected nodes create a new visualization from their relations. The arrows indicate the direction of influence, and color denotes type of influence (positive by green, negative by red and neutral by grey). Clicking on these relation edges will show that number of evidences at the bottom panel. The back button on top left corner will transition to Fig. 7.

evaluated by an information retrieval (IR) inverted index built on top of all the evidence phrases in the database to fetch and rank the matches that will be shown in the evidence panel. Since this is an IR search, the results are not necessarily exact string matches. The phrases returned by the index are more likely to be relevant because the data stored in the index is built exclusively from evidence phrases, although they are also ranked by a BM25 retrieval model [18]. If an interaction described in an evidence sentence is interesting, one can click on its text and the system transitions to the Node-Link view (§4.2) of that specific interaction.


Enter your query
causes inflammation

Max results
500

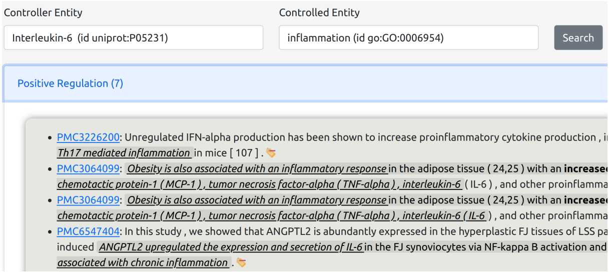
- PMC5976511: *Noxa* deficiency **causes** increased *inflammation* .
- PMC5342343: Binge *ethanol* exposure **causes** pancreatic *inflammation* .
- PMC4195784: *Rab11a* ablation in intestinal epithelia **causes** *inflammation* .
- PMC7820820: In sepsis , inflammation and coagulation are cross linked , and *inflammation* **causes** *coagulation* activa
- Source: *IL-6* **causes** cardiac *inflammation* and suppresses AMPK .
- Source: *IL-6* causes cardiac inflammation and **suppresses** *AMPK* .
- Source: In adults , the *infection* invariably **causes** gastric *inflammation* [1] .
- PMC3161332: NOD1 *ligand* directly **causes** hepatocyte *inflammation* and insulin resistance ..

Figure 9: Segment of the full-text search interface. At the top, the search box takes a query and a limit on the number of results. At the bottom, an evidence panel displays the results of the search. Clicking on the text of an interaction will open a page with a Node-Link view of the interaction.

4.6 Participant Search Interface

The participant search interface is designed for the cases in which maximum flexibility is needed while searching for interactions. In this interface, the domain expert specifies two entities of interest by filling the name or database identifier of each in the search fields presented by the interface. For convenience, the fields have auto-complete functionality for entity names and identifiers.

When either search field is specified, the choices available to fill the opposite one are narrowed down to contain only valid elements present in the database. After the search fields are specified, the system fetches all the evidence of interactions containing the chosen participants. The evidence phrases are grouped by interaction type and each interaction type is shown in an evidence panel. Fig. 10 depicts an instance of the participant search view for `IL-6` and `Inflammation`.



Controller Entity

Interleukin-6 (id uniprot:P05231)

Controlled Entity

inflammation (id go:GO:0006954)

Positive Regulation (7)

- PMC3226200: Unregulated IFN-alpha production has been shown to increase proinflammatory cytokine production , i
  Th17 mediated inflammation in mice [107] .
- PMC3064099: *Obesity is also associated with an inflammatory response* in the adipose tissue (24,25) with an **increase**
  chemotactic protein-1 (MCP-1) , tumor necrosis factor-alpha (TNF-alpha) , interleukin-6 (IL-6) , and other proinflamma
- PMC3064099: *Obesity is also associated with an inflammatory response* in the adipose tissue (24,25) with an **increase**
  chemotactic protein-1 (MCP-1) , tumor necrosis factor-alpha (TNF-alpha) , interleukin-6 (IL-6) , and other proinflamma
- PMC6547404: In this study , we showed that ANGPTL2 is abundantly expressed in the hyperplastic FJ tissues of LSS pa
  induced *ANGPTL2 upregulated the expression and secretion of IL-6* in the FJ synoviocytes via NF-kappa B activation and
  associated with chronic inflammation .

Figure 10: Participant search interface. At the top are the participant search boxes. At the bottom is an “accordion” widget displaying the interactions between both participants, grouped by interaction type.

5 Two Case Studies

We describe a use case with immunobiology domain experts who used our system to search, visualize and discover knowledge by looking for evidence in the literature.

5.1 Uncovering a Link Between IL-6 and Lipolysis

In the first study the immunobiologists found a new and relevant interaction between `IL-6` and `lipolysis`. While the link between both biological entities was already hypothesized, there was no clear and direct evidence of a non-immunological metabolic link between them in existing visual analytic systems and domain-specific online databases. To understand if our tool is useful, we compare the findings with the information returned by other online tools and databases to highlight the value of our methods.

We created a database by retrieving full-text articles from PMC and xDD[7] that mention either *IL-6*, *Tumor Necrosis Factor*, *fat tissue* or *adipocytes* by their name or any of their synonyms or variations. The result set contains 948,759 documents. After processing the documents with REACH, the database contains 42,619 different entities and 2,103,351 different unique interactions.

Starting from the interactions overview page, the domain expert selected `IL-6` as anchor entity and focused on the entities *influenced* by it. Under biological processes group, the top result is *lipolysis* with frequency of 127 followed by *hematopoiesis* (freq. 47) and *b-cell proliferation* (freq. 37); see Fig. 12(left subfigure), which along the other pictures supporting this section, is included in the supporting material. Lipolysis is the biochemical pathway through which catabolism of triacylglycerols stored in cellular lipid droplets occurs. This process is used to mobilize stored energy during fasting or exercise and occurs in different tissues and cell types but predominantly fat adipocytes in white and brown adipose tissue.

The visual summary overview anchored on IL-6, displays abundant connections of IL-6 to both adipose tissue and adipocytes, both of them related to lipolysis, illustrated in right subfigure of Fig. 12. Clicking on the edge connecting lipolysis to IL-6 to examine in detail the interactions between both entities reveals that the evidence points to a positive relationship with 120 instances of IL-6 positively influencing lipolysis compared to only 5 claiming negative interaction (Fig. 13). A detailed examination of the evidence backing the positive interactions shows multiple studies published in non-immunological journals (the top result comes from the *Sports Medicine* journal) demonstrating IL-6 directly increases lipolysis, glycogenolysis and fatty acid oxidation (Fig. 14).

The visual analytics system was very useful to the domain experts by virtue of pointing to important non immunological, metabolic roles of IL-6. The research group recently generated and characterized a mouse model with inducible IL-6 expression [16]. In this work they showed that overexpression of IL-6, within the range of that seen in old frail mice, induces early onset of frailty and muscle frailty. Also, very high induction of IL-6 was followed by very pronounced weight loss in IL-6 overexpressing mice. With the aid of the information found using our information system, the research group has further characterized global adipose tissue content in IL-6 overexpressing mice by magnetic resonance imaging (MRI). Fig. 15 shows images of fat (yellow)/water(black) contrast of the mouse abdomen showed that IL-6 overexpressing mice lost most of their body fat after 2 weeks of high IL-6 induction. Therefore, findings uncovered with help from our proposed visual analytics tool provided perhaps the most direct biological evidence to date of the fascinating rate at which IL-6 can increase lipolysis.

The description of this use case is useful to understand the utility of using visual analytics to navigate the information generated by a biomedical information extraction systems. But it is also important to understand how our system compares to and complements the other systems in the space.

To contrast how much of this information was not as readily available using other similar search tools, the domain experts looked for interactions contained in different pathways involving IL-6 with lipolysis. Querying the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database [19] with keyword as “IL6” returned 53 pathways associated with IL-6, and none showed any direct or indirect connection with lipolysis; see Fig. 16. Searching for pathways in STRING [36] returns multiple pathways categorized into different groups, however there was no mention of lipolysis or adipocytes associated with IL-6 in any of the pathway results; see Fig. 17. Searching “IL-6” in Pathway Commons [4] against “any data source” available yielded a total of 59 pathways associated with IL-6; see Fig. 18. All these pathways were mainly associated with the immunological activity and response of IL-6.

According to the domain experts, our system was very useful to mine pleiotropic impacts of a molecule of interest and especially examining results published in journals outside their field. In comparison, a search of ‘interleukin-6 biological processes’ on the PubMed search engine yielded 8,886 results where the first mention of adipose or lipolysis did not occur until result 37 on page 4, requiring significantly more engagement with the website before finding possibly useful information for this research question.

These results show that deploying the machine reading, search and visualization features present in our system over a large corpus of research publications worked better than the standard options. Specifically, our system helped find the correct association between lipolysis with IL-6, something that other databases and tools either failed to do or did not do effectively.

5.2 MPC-1: unexpected supressor of tumor progression

A immunobiology researcher, also interested in the effects of IL-6, analyzed RNA sequencing data of the *gastrocnemius muscle* of mice and found that Mitochondrial Pyruvate Carrier 1 (MPC-1) is one of the most up-regulated genes in his experiments. The up-regulation of MPC-1 was an unexpected finding which motivated further inspection by the immunobiologist, who used the system to see what kind of interactions were present in our database that had MPC-1 as a participant.

Using the visual overview, the researcher quickly found that MPC-1 has a reciprocal inetraction with *tumor progression* and, particularly, one of the evidence phrases states that MPC-1 supresses tumor progression through interaction with mitochondrial STAT-3 [8], a mechanistic relation that is biologically relevant to the study (Figs. 19, 20).

The source article of the relevant interaction mentions multiple times how IL-6 is an agonist of the phosphorylation of STAT-3 and that MPC-1 binds to STAT-3, which suggests that if we observe an up-regulation of MPC-1, then STAT-3 is an interesting subject for

[7] `https://xdd.wisc.edu`

[8] `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6377639/`

further investigation. This piece of information was an incidental discovery that became evident fairly quickly with the search mechanisms of the visual analytics system, because its search space is the network structure created out of the extractions from the literature.

Searching PubMed simultaneously for both IL-6 and MPC-1[9], does not return the aforementioned article (Fig. 21), even though it mentions IL-6 least twelve times and MPC-1 at least thirty times. However, the article can be found by explicitly searching for it by name. This is an example of how the proposed system complements the other tools in this space, such as PubMed search, because leveraging the information of the interactions network built upon the automated extractions can help find relevant literature quicker.

Upon this finding, the researcher used the full-text search feature of the system to look for otehr biologically relevant interactions. He used the same query string that was used in PubMed search and was able to find another relevant article[10] that did not appear in the results of PubMed (Fig. 22).

While these observations are motivating, our system is not perfect. By virtue of being a domain expert, the researcher conducting this case study is aware of another extremely relevant article which does not appear in our results.[11] There are two reasons for this omission: First, the dataset that is currently deployed with the system is composed of open access full texts. The aforementioned extremely relevant article is not part of the open access subset of PMC. Second, while we have some coverage of non-open publications, the processing and indexing happened before the relevant article was published by the journal. This is a limitation of the system in its current state, that we discuss in Section 6.

6 Discussion, Limitations and Future Work

We introduced a visual analytics system designed to aid biomedical and life-sciences researchers with literature search and knowledge visualization and exploration. Our system leverages machine reading and visualization techniques to make it possible to search the vast underlying research literature and interact with the results. The source code, data, and a live version are publicly available. A walk-through video describing the system is included as part of the supplementary material.

The design of the system is the result of multiple iterations between Immunology domain experts, NLP experts, and Visualization researchers. The main design goal was to integrate an NLP pipeline processing hundreds of thousands of research papers, with a visualization interface to search for and explore biological interactions, while also providing the underlying evidence. Facilitating the exploration is the underlying network structure, which is also extensively used in the visualization, search and navigation. The choices made reflect the incremental refinements needed until the system became easy to use by the domain experts. We aimed to keep the system as general as possible, so that it could be used in other domains, but we recognize that this was not always possible. For example, we used a color palette intuitive for interactions between biological entities, but that may not work in another, unrelated domain. Similarly, the metaphor of multiple directed edges between pairs of nodes (Fig. 8) makes sense in the context of biological interactions to help visualize what role each element plays in the relation (i.e. the controller entity, or controlled process), but may not be suitable in other domains.

The case studies that we presented demonstrate how the elements and interfaces of the system helped researchers with knowledge discovery. Several of the relevant findings where made possible due to the network structure, constructed using information extraction techniques, and highlighted in different views.

The domain experts who conducted their research with the system provided valuable feedback, ranging from simple "quality of life" (QoL) issues that caused minor grievances while using the interface, to fundamental improvements necessary to realize the potential of the system. Some of the QoL issues include

- The zoom gesture in the node-link diagram is too sensitive, making it easy to "loose" the diagram altogether by excessively shrinking it, needing to refresh the page to restore it.
- The visual overview can be slow (especially on older machines) due to its heavy CPU needs.
- Unlike other tools in the space, there is no user manual that explains in detail all the elements of the system.

QoL issues are relatively simple to fix, but there are other areas of opportunity that should be addressed in order to make the system truly useful beyond a narrow biomedical domain.

A significant one, which arose in section 5.2, is to increase the coverage of the literature indexed by the system to include all of the open-access subset of PMC. This would require processing and indexing an order of magnitude more data. Including PubMed abstracts for non-open-access articles, would further increase the number of currently indexed articles in the system. Increasing the size of the database presents an interesting engineering challenge in order to keep the same speed and responsiveness shown by the system in its current state.

More data also means more noise. We added a labeling widget that allows the experts to provide feedback about individual extractions, which can be used in the future to train supervised learning classifiers to filter out noisy extractions, or to rank the extractions based on one's preferences. However, we have not done this and it remains open for future research.

We recognize that the visualizations we currently reply on can be strained when one expands the underlying network (currently visualized with a standard node-link diagram). Larger networks than those shown in the visual summary overview and the node-link view figures can quickly overload the limited display real estate. We leave for future work the exploration and integration of approaches such as semantic zooming [23] and clustering techniques to dynamically summarize the network, such as motif simplification [7].

Acknowledgments

We thank Skye Rounseville, Clayton T. Morrison and Mihai Surdeanu for feedback, discussions and suggestions for improvement. This work was supported by the Strategic Initiative in Personalized Defense from the University of Arizona and the NIH grant AG020719 and in part by the DARPA project SKEMA: Scientific Knowledge Extraction and Model Analysis, federal contract no: HR00112290092.

References

[1] V. Batagelj and A. Mrvar. Pajek-program for large network analysis. *Connections*, 21(2):47–57, 1998.

[2] A. Benito-Santos and R. Theron. GlassViz: Visualizing Automatically-Extracted Entry Points for Exploring Scientific Corpora in Problem-Driven Visualization Research. In *Proceedings - 2020 IEEE Visualization Conference, VIS 2020*, pp. 226–230, 10 2020.

[3] R. Burd, K. A. Espy, M. I. Hossain, S. G. Kobourov, N. C. Merchant, and H. C. Purchase. GRAM: global research activity map. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces (AVI)*, pp. 31:1–31:9. ACM, 2018.

[4] E. G. Cerami, B. E. Gross, E. Demir, I. Rodchenkov, Ö. Babur, N. Anwar, N. Schultz, G. D. Bader, and C. Sander. Pathway Commons, a web resource for biological pathway data. *Nucleic acids research*, 39(suppl_1):D685–D690, 2010.

[5] I. Cho, W. Dou, D. X. Wang, E. Sauda, and W. Ribarsky. VAiRoma: A Visual Analytics System for Making Sense of Places, Times, and Events in Roman History. *IEEE Trans. Vis. Comput. Graph.*, 22(1):210–219, 2016.

[9] https://pubmed.ncbi.nlm.nih.gov/?term=IL-6%2C%20MPC-1

[10] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4486456/

[11] https://pubmed.ncbi.nlm.nih.gov/33242651/

[6] K. Choe, S. Jung, S. Park, H. Hong, and J. Seo. Papers101: Supporting the Discovery Process in the Literature Review Workflow for Novice Researchers. In *IEEE Pacific Visualization Symposium*, vol. 2021-April, pp. 176–180, 4 2021. doi: 10.1109/PacificVis52677.2021.00037

[7] C. Dunne and B. Shneiderman. Motif Simplification: Improving Network Visualization Readability with Fan, Connector, and Clique Glyphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, p. 3247–3256. Association for Computing Machinery, New York, NY, USA, 2013.

[8] O. Etzioni, M. Banko, S. Soderland, and D. S. Weld. Open information extraction from the web. *Communications of the ACM*, 51(12):68–74, 2008.

[9] E. R. Gansner, Y. Hu, and S. G. Kobourov. GMap: Visualizing graphs and clusters as maps. In *IEEE Pacific Visualization Symposium PacificVis 2010, Taipei, Taiwan, March 2-5, 2010*, pp. 201–208. IEEE Computer Society, 2010. doi: 10.1109/PACIFICVIS.2010.5429590

[10] H. Gibson, J. Faith, and P. Vickers. A survey of two-dimensional graph layout techniques for information visualisation. *Information visualization*, 12(3-4):324–357, 2013.

[11] B. González-Pereira, V. P. Guerrero-Bote, and F. Moya-Anegón. A new approach to the metric of journals' scientific prestige: The SJR indicator. *Journal of informetrics*, 4(3):379–391, 2010.

[12] Y. Gu, R. Tinn, H. Cheng, M. Lucas, N. Usuyama, X. Liu, T. Naumann, J. Gao, and H. Poon. Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. *ACM Trans. Comput. Healthcare*, 3(1), oct 2021. doi: 10.1145/3458754

[13] M. I. Hossain, S. G. Kobourov, H. C. Purchase, and M. Surdeanu. REMatch: Research Expert MAtching System. In *2018 International Symposium on Big Data Visual and Immersive Analytics (BDVA)*, pp. 1–10. IEEE, 2018. doi: 10.1109/BDVA.2018.8534021

[14] P. Isenberg, F. Heimerl, S. Koch, T. Isenberg, P. Xu, C. D. Stolper, M. Sedlmair, J. Chen, T. Moller, and J. Stasko. Vispubdata.org: A Metadata Collection about IEEE Visualization (VIS) Publications. *IEEE Transactions on Visualization and Computer Graphics*, 23:2199–2206, 9 2017. doi: 10.1109/TVCG.2016.2615308

[15] P. Isenberg, T. Isenberg, M. Sedlmair, J. Chen, and T. Möller. Visualization as Seen through its Research Paper Keywords. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):771–780, 2017.

[16] M. Jergović, H. L. Thompson, C. M. Bradshaw, S. A. Sonar, A. Ashgar, N. Mohty, B. Joseph, M. J. Fain, K. Cleveland, R. G. Schnellman, and J. Nikolich-Žugich. IL-6 can singlehandedly drive many features of frailty in mice. *GeroScience*, 43(2):539–549, Apr. 2021.

[17] J. Johnson, M. Douze, and H. Jégou. Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, 7(3):535–547, 2019.

[18] K. S. Jones, S. Walker, and S. E. Robertson. A probabilistic model of information retrieval: development and comparative experiments: Part 2. *Information processing & management*, 36(6):809–840, 2000.

[19] M. Kanehisa and S. Goto. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.

[20] J.-D. Kim, T. Ohta, S. Pyysalo, Y. Kano, and J. Tsujii. Overview of BioNLP'09 shared task on event extraction. In *Proceedings of the BioNLP 2009 workshop companion volume for shared task*, pp. 1–9, 2009.

[21] P.-H. Li, T.-F. Chen, J.-Y. Yu, S.-H. Shih, C.-H. Su, Y.-H. Lin, H.-K. Tsai, H.-F. Juan, C.-Y. Chen, and J.-H. Huang. pubmedKB: an interactive web server for exploring biomedical entity relations in the biomedical literature. *Nucleic Acids Research*, 50(W1):W616–W622, 05 2022. doi: 10.1093/nar/gkac310

[22] C. E. Lipscomb. Medical subject headings (MeSH). *Bulletin of the Medical Library Association*, 88(3):265, 2000.

[23] F. D. Luca, M. I. Hossain, S. G. Kobourov, and K. Börner. Multi-level tree based approach for interactive graph visualization with semantic zoom. *CoRR*, abs/1906.05996, 2019.

[24] C. Manning, P. Raghavan, and H. Schütze. Introduction to information retrieval. *Natural Language Engineering*, 16(1):100–103, 2010.

[25] C. Nédellec, R. Bossy, J.-D. Kim, J.-J. Kim, T. Ohta, S. Pyysalo, and P. Zweigenbaum. Overview of BioNLP shared task 2013. In *Proceedings of the BioNLP shared task 2013 workshop*, pp. 1–7, 2013.

[26] M. Neumann, D. King, I. Beltagy, and W. Ammar. ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing. In *Proceedings of the 18th BioNLP Workshop and Shared Task*, pp. 319–327. Association for Computational Linguistics, Florence, Italy, Aug. 2019.

[27] E. Noriega-Atala, P. D. Hein, S. S. Thumsi, Z. Wong, X. Wang, and C. T. Morrison. Inter-Sentence Relation Extraction for Associating Biological Context with Events in Biomedical Texts. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 722–731, 2018. doi: 10.1109/ICDMW.2018.00110

[28] E. Noriega-Atala, P. M. Lovett, C. T. Morrison, and M. Surdeanu. Neural Architectures for Biological Inter-Sentence Relation Extraction. In A. Pouran Ben Veyseh, F. Dernoncourt, T. Huu Nguyen, W. Chang, and V. Dack Lai, eds., *2^{nd} Workshop on Scientific Document Understanding*, number 3164 in CEUR Workshop Proceedings, 2022.

[29] S. I. O'Donoghue, B. F. Baldi, S. J. Clark, A. E. Darling, J. M. Hogan, S. Kaur, L. Maier-Hein, D. J. Mccarthy, W. J. Moore, E. Stenau, J. R. Swedlow, J. Vuong, and J. B. Procter. Visualization of Biomedical Data. *Annual Review of Biomedical Data Science*, 1(1):275–304, July 2018. doi: 10.1146/annurev-biodatasci-080917-013424

[30] A. Rind, A. Haberson, K. Blumenstein, C. Niederer, M. Wagner, and W. Aigner. PubViz: Lightweight Visual Presentation of Publication Data. In B. Kozlikova, T. Schreck, and T. Wischgoll, eds., *EuroVis 2017 - Short Papers*. The Eurographics Association, 2017.

[31] R. J. Roberts. PubMed Central: The GenBank of the published literature. In *Proceedings of the National Academy of Sciences*, vol. 98, pp. 381–382. National Acad Sciences, 2001.

[32] M. M. U. Rony, E. Hoque, and N. Hassan. ClaimViz: Visual Analytics for Identifying and Verifying Factual Claims. In *Proceedings - 2020 IEEE Visualization Conference, VIS 2020*, pp. 246–250. Institute of Electrical and Electronics Engineers Inc., 10 2020.

[33] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11):2498–2504, 2003.

[34] C. Su, J. Tong, Y. Zhu, P. Cui, and F. Wang. Network embedding in biomedical data science. *Briefings in Bioinformatics*, 21(1):182–197, 12 2018. doi: 10.1093/bib/bby117

[35] N. Sultanum, C. Murad, and D. Wigdor. Understanding and Supporting Academic Literature Review Workflows with LitSense. In *ACM International Conference Proceeding Series*. Association for Computing Machinery, 9 2020. doi: 10.1145/3399715.3399830

[36] D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*, 47(D1):D607–D613, 2019.

[37] P. Tol. Colour schemes. *SRON Technical Note*, (2.2):SRON–EPS, 2012.

[38] UniProt Consortium. UniProt: a hub for protein information. *Nucleic acids research*, 43(D1):D204–D212, 2015.

[39] M. A. Valenzuela-Escárcega, Ö. Babur, G. Hahn-Powell, D. Bell, T. Hicks, E. Noriega-Atala, X. Wang, M. Surdeanu, E. Demir, and C. T. Morrison. Large-scale automated machine reading discovers new cancer-driving mechanisms. *Database*, 2018, 09 2018. bay098.

[40] M. A. Valenzuela-Escárcega, G. Hahn-Powell, and M. Surdeanu. Odin's runes: A rule language for information extraction. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pp. 322–329, 2016.

[41] K. Z. Vardakas, G. Tsopanakis, A. Poulopoulou, and M. E. Falagas. An analysis of factors contributing to PubMed's growth. *Journal of Informetrics*, 9(3):592–617, 2015.

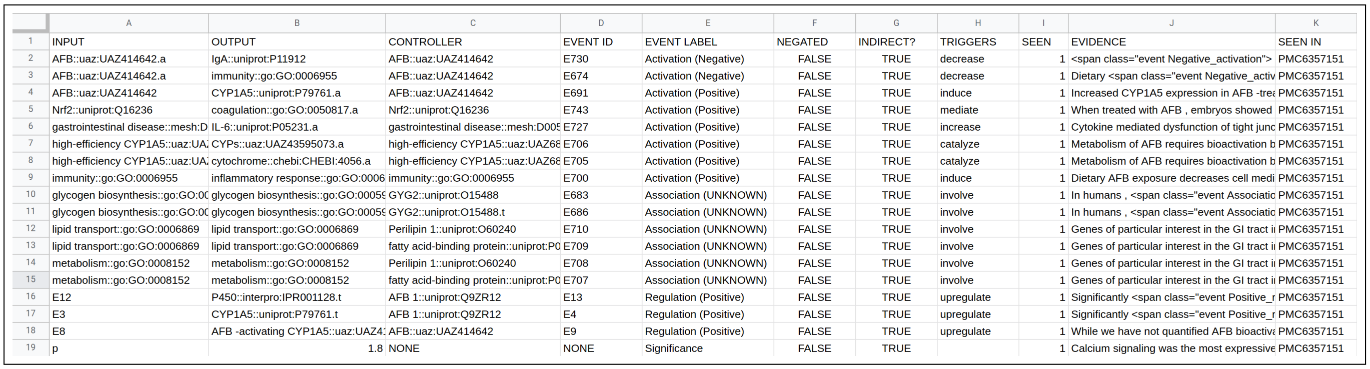
[42] R. Vuillemot and C. Perin. Investigating the Direct Manipulation of Ranking Tables for Time Navigation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, p. 2703–2706. Association for Computing Machinery, New York, NY, USA, 2015.

[43] J. White. PubMed 2.0. *Medical Reference Services Quarterly*, 39(4):382–387, 2020. PMID: 33085945.

[44] M. A. Yalçın, N. Elmqvist, and B. B. Bederson. Keshif: Rapid and Expressive Tabular Data Exploration for Novices. *IEEE Transactions on Visualization and Computer Graphics*, 24(8):2339–2352, 2018.

Supplementary Material

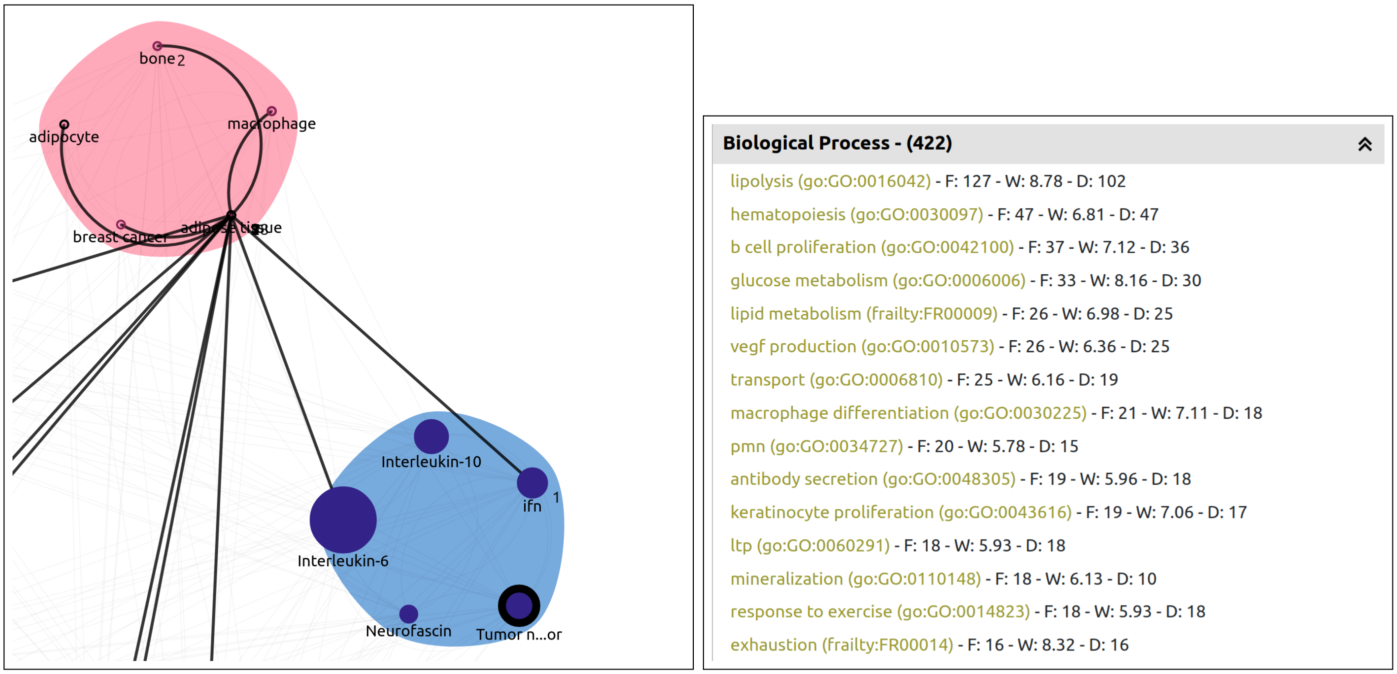
A Example Tabular Data Format



	A	B	C	D	E	F	G	H	I	J	K
1	INPUT	OUTPUT	CONTROLLER	EVENT ID	EVENT LABEL	NEGATED	INDIRECT?	TRIGGERS	SEEN	EVIDENCE	SEEN IN
2	AFB::uaz:UAZ414642.a	IgA::uniprot:P11912	AFB::uaz:UAZ414642	E730	Activation (Negative)	FALSE	TRUE	decrease	1	<span class="event Negative_activation">	PMC6357151
3	AFB::uaz:UAZ414642.a	immunity::go:GO:0006955	AFB::uaz:UAZ414642	E674	Activation (Negative)	FALSE	TRUE	decrease	1	Dietary <span class="event Negative_activ	PMC6357151
4	AFB::uaz:UAZ414642	CYP1A5::uniprot:P79761.a	AFB::uaz:UAZ414642	E691	Activation (Positive)	FALSE	TRUE	induce	1	Increased CYP1A5 expression in AFB -trea	PMC6357151
5	Nrf2::uniprot:Q16236	coagulation::go:GO:0050817.a	Nrf2::uniprot:Q16236	E743	Activation (Positive)	FALSE	TRUE	mediate	1	When treated with AFB , embryos showed	PMC6357151
6	gastrointestinal disease::mesh:D	IL-6::uniprot:P05231.a	gastrointestinal disease::mesh:D005	E727	Activation (Positive)	FALSE	TRUE	increase	1	Cytokine mediated dysfunction of tight junc	PMC6357151
7	high-efficiency CYP1A5::uaz:UA	CYPs::uaz:UAZ43595073.a	high-efficiency CYP1A5::uaz:UAZ68	E706	Activation (Positive)	FALSE	TRUE	catalyze	1	Metabolism of AFB requires bioactivation b	PMC6357151
8	high-efficiency CYP1A5::uaz:UA	cytochrome::chebi:CHEBI:4056.a	high-efficiency CYP1A5::uaz:UAZ68	E705	Activation (Positive)	FALSE	TRUE	catalyze	1	Metabolism of AFB requires bioactivation b	PMC6357151
9	immunity::go:GO:0006955	inflammatory response::go:GO:0006	immunity::go:GO:0006955	E700	Activation (Positive)	FALSE	TRUE	induce	1	Dietary AFB exposure decreases cell medi	PMC6357151
10	glycogen biosynthesis::go:GO:00	glycogen biosynthesis::go:GO:00059	GYG2::uniprot:O15488	E683	Association (UNKNOWN)	FALSE	TRUE	involve	1	In humans , <span class="event Associatio	PMC6357151
11	glycogen biosynthesis::go:GO:00	glycogen biosynthesis::go:GO:00059	GYG2::uniprot:O15488.t	E686	Association (UNKNOWN)	FALSE	TRUE	involve	1	In humans , <span class="event Associatio	PMC6357151
12	lipid transport::go:GO:0006869	lipid transport::go:GO:0006869	Perilipin 1::uniprot:O60240	E710	Association (UNKNOWN)	FALSE	TRUE	involve	1	Genes of particular interest in the GI tract i	PMC6357151
13	lipid transport::go:GO:0006869	lipid transport::go:GO:0006869	fatty acid-binding protein::uniprot:P0	E709	Association (UNKNOWN)	FALSE	TRUE	involve	1	Genes of particular interest in the GI tract i	PMC6357151
14	metabolism::go:GO:0008152	metabolism::go:GO:0008152	Perilipin 1::uniprot:O60240	E708	Association (UNKNOWN)	FALSE	TRUE	involve	1	Genes of particular interest in the GI tract i	PMC6357151
15	metabolism::go:GO:0008152	metabolism::go:GO:0008152	fatty acid-binding protein::uniprot:P0	E707	Association (UNKNOWN)	FALSE	TRUE	involve	1	Genes of particular interest in the GI tract i	PMC6357151
16	E12	P450::interpro:IPR001128.t	AFB 1::uniprot:Q9ZR12	E13	Regulation (Positive)	FALSE	TRUE	upregulate	1	Significantly <span class="event Positive_r	PMC6357151
17	E3	CYP1A5::uniprot:P79761.t	AFB 1::uniprot:Q9ZR12	E4	Regulation (Positive)	FALSE	TRUE	upregulate	1	Significantly <span class="event Positive_r	PMC6357151
18	E8	AFB -activating CYP1A5::uaz:UAZ4	AFB::uaz:UAZ414642	E9	Regulation (Positive)	FALSE	TRUE	upregulate	1	While we have not quantified AFB bioactiva	PMC6357151
19	p	1.8	NONE	NONE	Significance	FALSE	TRUE		1	Calcium signaling was the most expressive	PMC6357151

Figure 11: Data file example. This the segment of a file outputted by an information extraction system and digested to be assembled and visualized as described in §3 of the main document.

B Use Cases Supporting Material



Biological Process - (422)

- lipolysis (go:GO:0016042) - F: 127 - W: 8.78 - D: 102
- hematopoiesis (go:GO:0030097) - F: 47 - W: 6.81 - D: 47
- b cell proliferation (go:GO:0042100) - F: 37 - W: 7.12 - D: 36
- glucose metabolism (go:GO:0006006) - F: 33 - W: 8.16 - D: 30
- lipid metabolism (frailty:FR00009) - F: 26 - W: 6.98 - D: 25
- vegf production (go:GO:0010573) - F: 26 - W: 6.36 - D: 25
- transport (go:GO:0006810) - F: 25 - W: 6.16 - D: 19
- macrophage differentiation (go:GO:0030225) - F: 21 - W: 7.11 - D: 18
- pmn (go:GO:0034727) - F: 20 - W: 5.78 - D: 15
- antibody secretion (go:GO:0048305) - F: 19 - W: 5.96 - D: 18
- keratinocyte proliferation (go:GO:0043616) - F: 19 - W: 7.06 - D: 17
- ltp (go:GO:0060291) - F: 18 - W: 5.93 - D: 18
- mineralization (go:GO:0110148) - F: 18 - W: 6.13 - D: 10
- response to exercise (go:GO:0014823) - F: 18 - W: 5.93 - D: 18
- exhaustion (frailty:FR00014) - F: 16 - W: 8.32 - D: 16

Figure 12: Left: Visual overview clip anchored on *Tumor necrosis factor α*. Observe how IL-6 is one of the most interconnected entities in the data set and how there is a path from it to *adipocyte* going through *adipose tissue*.
Right: Biological processes influenced by IL-6 as extracted from the use case corpus of scientific papers.

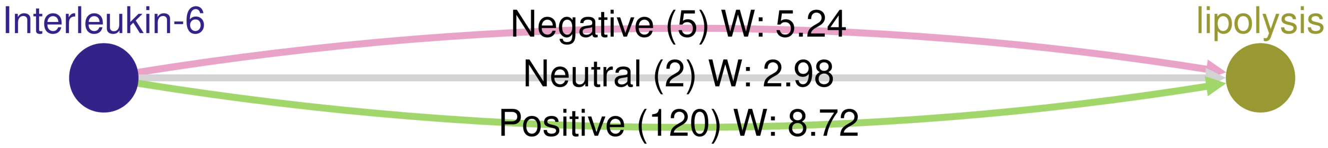
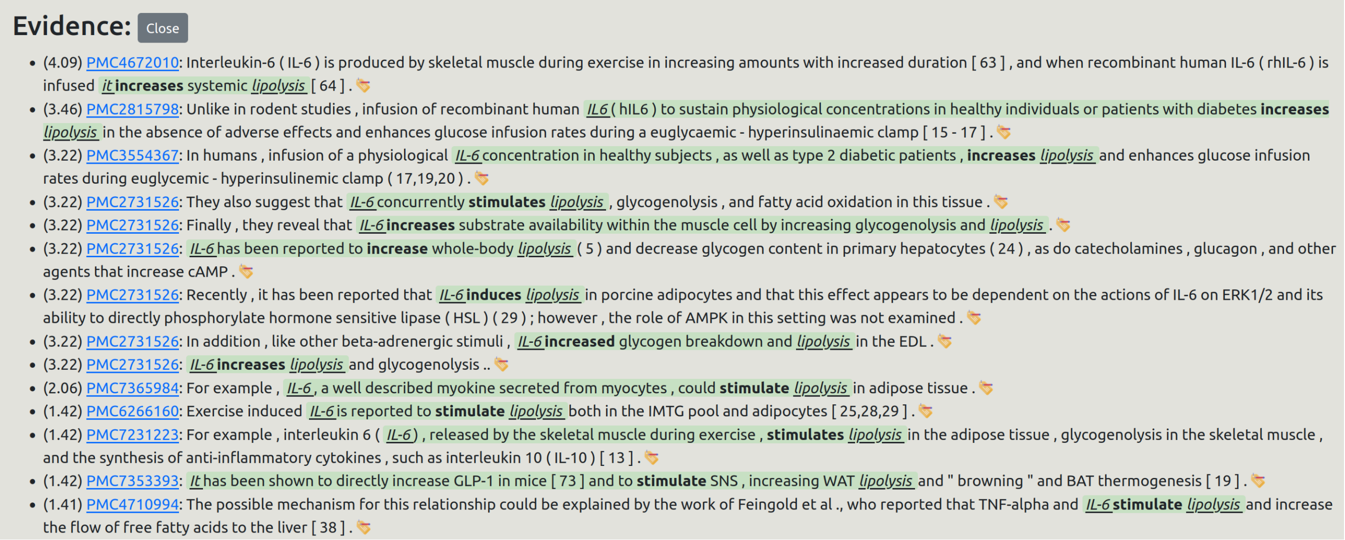


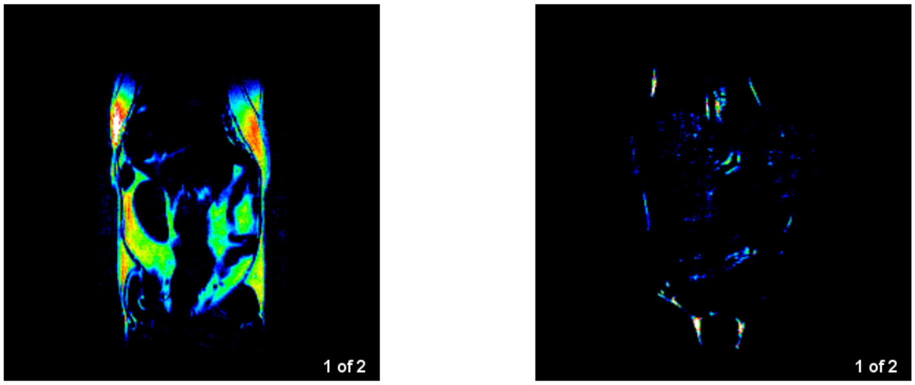
Figure 13: Node-link view showing the breakout of the interactions between IL-6 and lipolysis by polarity. Observe how the evidence only points to interactions where IL-6 is the controller in the interactions.



Evidence: Close

- (4.09) PMC4672010: Interleukin-6 (IL-6) is produced by skeletal muscle during exercise in increasing amounts with increased duration [63] , and when recombinant human IL-6 (rhIL-6) is infused *it* **increases** systemic *lipolysis* [64] .
- (3.46) PMC2815798: Unlike in rodent studies , infusion of recombinant human *IL6* (hIL6) to sustain physiological concentrations in healthy individuals or patients with diabetes **increases** *lipolysis* in the absence of adverse effects and enhances glucose infusion rates during a euglycaemic - hyperinsulinaemic clamp [15 - 17] .
- (3.22) PMC3554367: In humans , infusion of a physiological *IL-6* concentration in healthy subjects , as well as type 2 diabetic patients , **increases** *lipolysis* and enhances glucose infusion rates during euglycemic - hyperinsulinemic clamp (17,19,20) .
- (3.22) PMC2731526: They also suggest that *IL-6* concurrently **stimulates** *lipolysis* , glycogenolysis , and fatty acid oxidation in this tissue .
- (3.22) PMC2731526: Finally , they reveal that *IL-6* **increases** substrate availability within the muscle cell by increasing glycogenolysis and *lipolysis* .
- (3.22) PMC2731526: *IL-6* has been reported to **increase** whole-body *lipolysis* (5) and decrease glycogen content in primary hepatocytes (24) , as do catecholamines , glucagon , and other agents that increase cAMP .
- (3.22) PMC2731526: Recently , it has been reported that *IL-6* **induces** *lipolysis* in porcine adipocytes and that this effect appears to be dependent on the actions of IL-6 on ERK1/2 and its ability to directly phosphorylate hormone sensitive lipase (HSL) (29) ; however , the role of AMPK in this setting was not examined .
- (3.22) PMC2731526: In addition , like other beta-adrenergic stimuli , *IL-6* **increased** glycogen breakdown and *lipolysis* in the EDL .
- (3.22) PMC2731526: *IL-6* **increases** *lipolysis* and glycogenolysis ..
- (2.06) PMC7365984: For example , *IL-6* , a well described myokine secreted from myocytes , could **stimulate** *lipolysis* in adipose tissue .
- (1.42) PMC6266160: Exercise induced *IL-6* is reported to **stimulate** *lipolysis* both in the IMTG pool and adipocytes [25,28,29] .
- (1.42) PMC7231223: For example , interleukin 6 (*IL-6*) , released by the skeletal muscle during exercise , **stimulates** *lipolysis* in the adipose tissue , glycogenolysis in the skeletal muscle , and the synthesis of anti-inflammatory cytokines , such as interleukin 10 (IL-10) [13] .
- (1.42) PMC7353393: *It* has been shown to directly increase GLP-1 in mice [73] and to **stimulate** SNS , increasing WAT *lipolysis* and " browning " and BAT thermogenesis [19] .
- (1.41) PMC4710994: The possible mechanism for this relationship could be explained by the work of Feingold et al ., who reported that TNF-alpha and *IL-6* **stimulate** *lipolysis* and increase the flow of free fatty acids to the liver [38] .

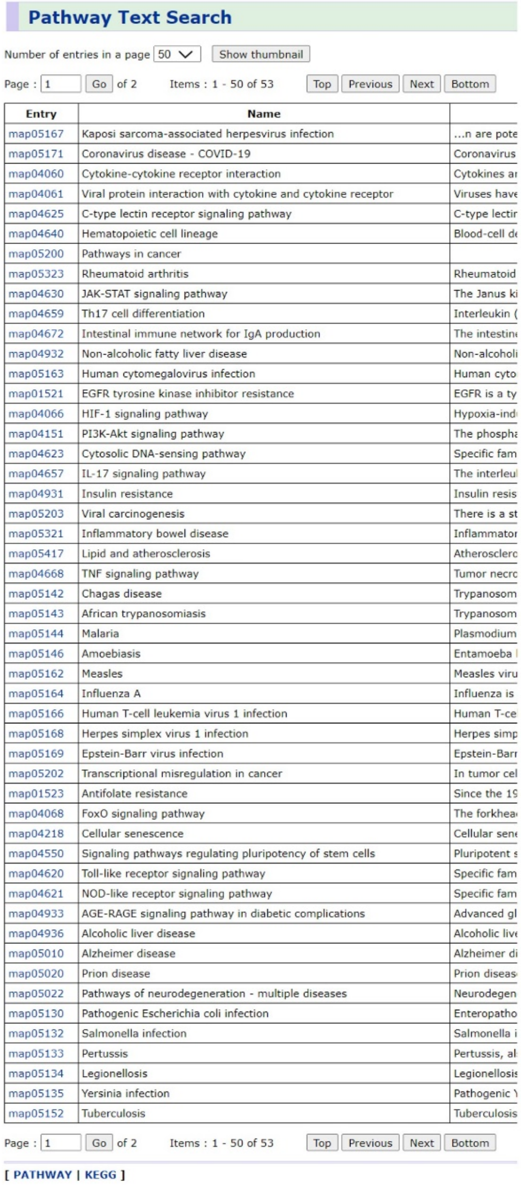
Figure 14: Evidence panel displaying a few sentences, found in the literature, containing a positive interaction where IL-6 controls lipolysis.



Normal IL-6 level mouse

IL-6 overexpressed mouse

Figure 15: Scan images of fat (yellow)/water(black) contrast of fat tissue in a mouse's abdomen. Left shows abdomen fat before IL-6 induction. Right shows abdomen fat two weeks after inducing IL-6 over-expression.



Pathway Text Search

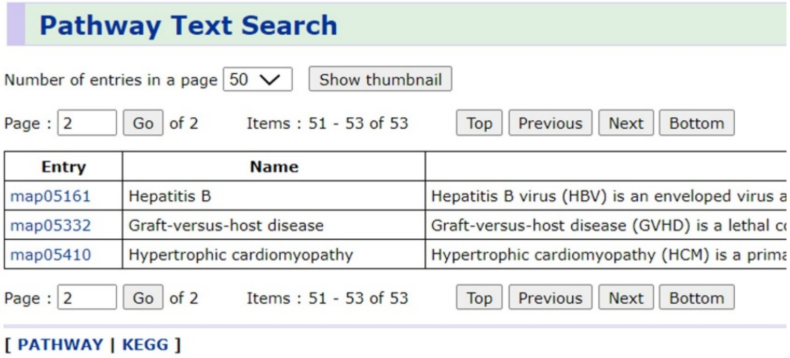
Number of entries in a page 50 Show thumbnail

Page : 1 Go of 2 Items : 1 - 50 of 53 Top Previous Next Bottom

Entry	Name	
map05167	Kaposi sarcoma-associated herpesvirus infection	...n are pote
map05171	Coronavirus disease - COVID-19	Coronavirus
map04060	Cytokine-cytokine receptor interaction	Cytokines ar
map04061	Viral protein interaction with cytokine and cytokine receptor	Viruses have
map04625	C-type lectin receptor signaling pathway	C-type lectir
map04640	Hematopoietic cell lineage	Blood-cell de
map05200	Pathways in cancer	
map05323	Rheumatoid arthritis	Rheumatoid
map04630	JAK-STAT signaling pathway	The Janus ki
map04659	Th17 cell differentiation	Interleukin (
map04672	Intestinal immune network for IgA production	The intestine
map04932	Non-alcoholic fatty liver disease	Non-alcoholi
map05163	Human cytomegalovirus infection	Human cyto
map01521	EGFR tyrosine kinase inhibitor resistance	EGFR is a ty
map04066	HIF-1 signaling pathway	Hypoxia-ind
map04151	PI3K-Akt signaling pathway	The phospha
map04623	Cytosolic DNA-sensing pathway	Specific fam
map04657	IL-17 signaling pathway	The interleul
map04931	Insulin resistance	Insulin resis
map05203	Viral carcinogenesis	There is a st
map05321	Inflammatory bowel disease	Inflammator
map05417	Lipid and atherosclerosis	Atherosclerc
map04668	TNF signaling pathway	Tumor necrc
map05142	Chagas disease	Trypanosom
map05143	African trypanosomiasis	Trypanosom
map05144	Malaria	Plasmodium
map05146	Amoebiasis	Entamoeba
map05162	Measles	Measles viru
map05164	Influenza A	Influenza is
map05166	Human T-cell leukemia virus 1 infection	Human T-ce
map05168	Herpes simplex virus 1 infection	Herpes simp
map05169	Epstein-Barr virus infection	Epstein-Barr
map05202	Transcriptional misregulation in cancer	In tumor cel
map01523	Antifolate resistance	Since the 19
map04068	FoxO signaling pathway	The forkhea
map04218	Cellular senescence	Cellular sene
map04550	Signaling pathways regulating pluripotency of stem cells	Pluripotent s
map04620	Toll-like receptor signaling pathway	Specific fam
map04621	NOD-like receptor signaling pathway	Specific fam
map04933	AGE-RAGE signaling pathway in diabetic complications	Advanced gl
map04936	Alcoholic liver disease	Alcoholic live
map05010	Alzheimer disease	Alzheimer di
map05020	Prion disease	Prion diseas
map05022	Pathways of neurodegeneration - multiple diseases	Neurodegen
map05130	Pathogenic Escherichia coli infection	Enteropatho
map05132	Salmonella infection	Salmonella i
map05133	Pertussis	Pertussis, al
map05134	Legionellosis	Legionellosis
map05135	Yersinia infection	Pathogenic Y
map05152	Tuberculosis	Tuberculosis

Page : 1 Go of 2 Items : 1 - 50 of 53 Top Previous Next Bottom

[PATHWAY | KEGG]



Pathway Text Search

Number of entries in a page 50 Show thumbnail

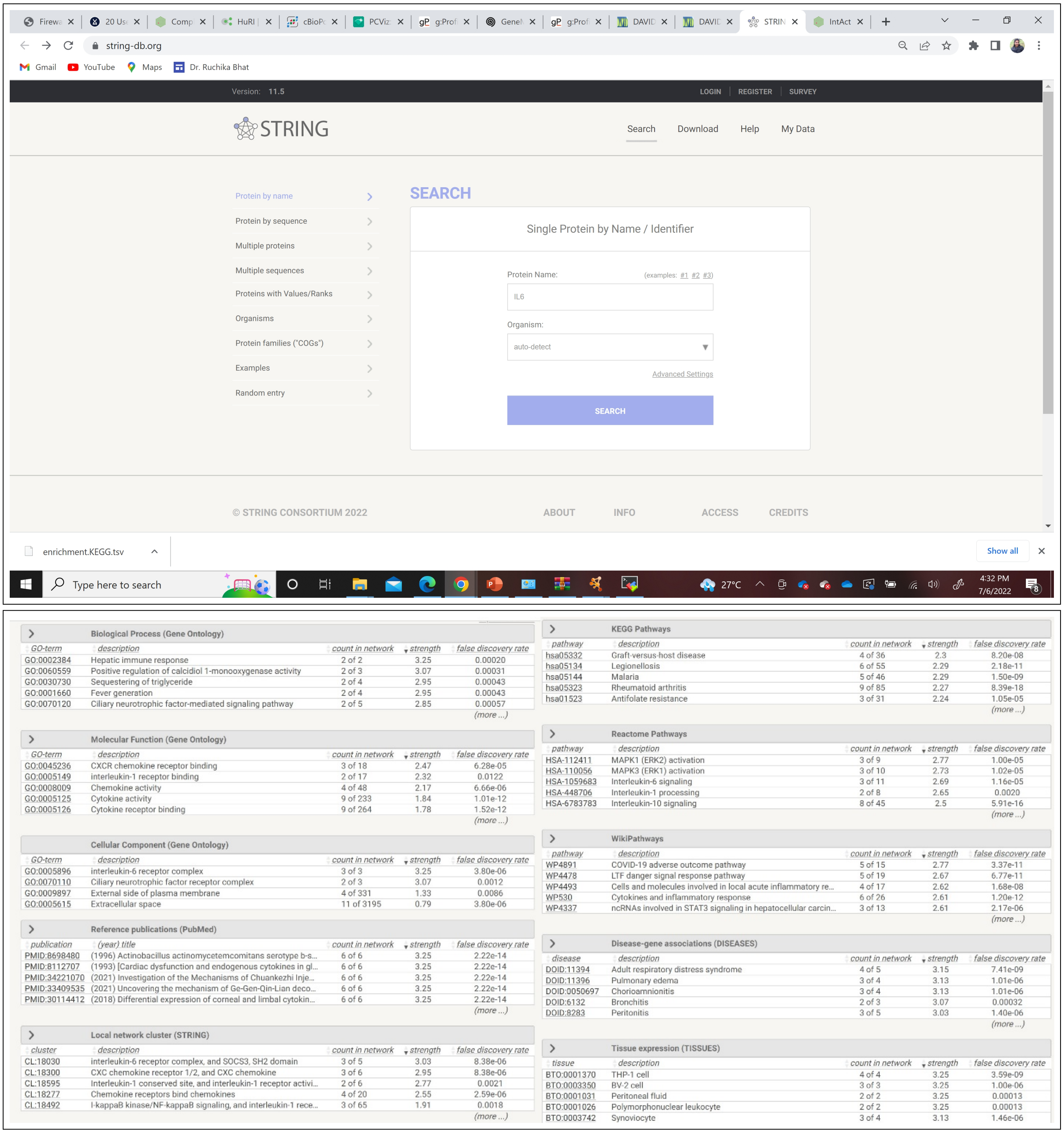
Page : 2 Go of 2 Items : 51 - 53 of 53 Top Previous Next Bottom

Entry	Name	
map05161	Hepatitis B	Hepatitis B virus (HBV) is an enveloped virus a
map05332	Graft-versus-host disease	Graft-versus-host disease (GVHD) is a lethal co
map05410	Hypertrophic cardiomyopathy	Hypertrophic cardiomyopathy (HCM) is a prima

Page : 2 Go of 2 Items : 51 - 53 of 53 Top Previous Next Bottom

[PATHWAY | KEGG]

Figure 16: Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database results after searching for term “IL-6”.



Version: 11.5

LOGIN | REGISTER | SURVEY

STRING

Search Download Help My Data

Protein by name
Protein by sequence
Multiple proteins
Multiple sequences
Proteins with Values/Ranks
Organisms
Protein families ("COGs")
Examples
Random entry

SEARCH

Single Protein by Name / Identifier

Protein Name: (examples: #1 #2 #3)

IL6

Organism:

auto-detect

Advanced Settings

SEARCH

© STRING CONSORTIUM 2022

ABOUT INFO ACCESS CREDITS

Biological Process (Gene Ontology)

GO-term	description	count in network	strength	false discovery rate
GO:0002384	Hepatic immune response	2 of 2	3.25	0.00020
GO:0060559	Positive regulation of calcidiol 1-monooxygenase activity	2 of 3	3.07	0.00031
GO:0030730	Sequestering of triglyceride	2 of 4	2.95	0.00043
GO:0001660	Fever generation	2 of 4	2.95	0.00043
GO:0070120	Ciliary neurotrophic factor-mediated signaling pathway	2 of 5	2.85	0.00057

(more ...)

Molecular Function (Gene Ontology)

GO-term	description	count in network	strength	false discovery rate
GO:0045236	CXCR chemokine receptor binding	3 of 18	2.47	6.28e-05
GO:0005149	interleukin-1 receptor binding	2 of 17	2.32	0.0122
GO:0008009	Chemokine activity	4 of 48	2.17	6.66e-06
GO:0005125	Cytokine activity	9 of 233	1.84	1.01e-12
GO:0005126	Cytokine receptor binding	9 of 264	1.78	1.52e-12

(more ...)

Cellular Component (Gene Ontology)

GO-term	description	count in network	strength	false discovery rate
GO:0005896	interleukin-6 receptor complex	3 of 3	3.25	3.80e-06
GO:0070110	Ciliary neurotrophic factor receptor complex	2 of 3	3.07	0.0012
GO:0009897	External side of plasma membrane	4 of 331	1.33	0.0086
GO:0005615	Extracellular space	11 of 3195	0.79	3.80e-06

Reference publications (PubMed)

publication	(year) title	count in network	strength	false discovery rate
PMID:8698480	(1996) Actinobacillus actinomycetemcomitans serotype b-s...	6 of 6	3.25	2.22e-14
PMID:8112707	(1993) [Cardiac dysfunction and endogenous cytokines in gl...	6 of 6	3.25	2.22e-14
PMID:34221070	(2021) Investigation of the Mechanisms of Chuankezhi Inje...	6 of 6	3.25	2.22e-14
PMID:33409535	(2021) Uncovering the mechanism of Ge-Gen-Qin-Lian deco...	6 of 6	3.25	2.22e-14
PMID:30114412	(2018) Differential expression of corneal and limbal cytokin...	6 of 6	3.25	2.22e-14

(more ...)

Local network cluster (STRING)

cluster	description	count in network	strength	false discovery rate
CL:18030	interleukin-6 receptor complex, and SOCS3, SH2 domain	3 of 5	3.03	8.38e-06
CL:18300	CXC chemokine receptor 1/2, and CXC chemokine	3 of 6	2.95	8.38e-06
CL:18595	Interleukin-1 conserved site, and interleukin-1 receptor activi...	2 of 6	2.77	0.0021
CL:18277	Chemokine receptors bind chemokines	4 of 20	2.55	2.59e-06
CL:18492	I-kappaB kinase/NF-kappaB signaling, and interleukin-1 rece...	3 of 65	1.91	0.0018

(more ...)

KEGG Pathways

pathway	description	count in network	strength	false discovery rate
hsa05332	Graft-versus-host disease	4 of 36	2.3	8.20e-08
hsa05134	Legionellosis	6 of 55	2.29	2.18e-11
hsa05144	Malaria	5 of 46	2.29	1.50e-09
hsa05323	Rheumatoid arthritis	9 of 85	2.27	8.39e-18
hsa01523	Antifolate resistance	3 of 31	2.24	1.05e-05

(more ...)

Reactome Pathways

pathway	description	count in network	strength	false discovery rate
HSA-112411	MAPK1 (ERK2) activation	3 of 9	2.77	1.00e-05
HSA-110056	MAPK3 (ERK1) activation	3 of 10	2.73	1.02e-05
HSA-1059683	Interleukin-6 signaling	3 of 11	2.69	1.16e-05
HSA-448706	Interleukin-1 processing	2 of 8	2.65	0.0020
HSA-6783783	Interleukin-10 signaling	8 of 45	2.5	5.91e-16

(more ...)

WikiPathways

pathway	description	count in network	strength	false discovery rate
WP4891	COVID-19 adverse outcome pathway	5 of 15	2.77	3.37e-11
WP4478	LTF danger signal response pathway	5 of 19	2.67	6.77e-11
WP4493	Cells and molecules involved in local acute inflammatory re...	4 of 17	2.62	1.68e-08
WP530	Cytokines and inflammatory response	6 of 26	2.61	1.20e-12
WP4337	ncRNAs involved in STAT3 signaling in hepatocellular carcin...	3 of 13	2.61	2.17e-06

(more ...)

Disease-gene associations (DISEASES)

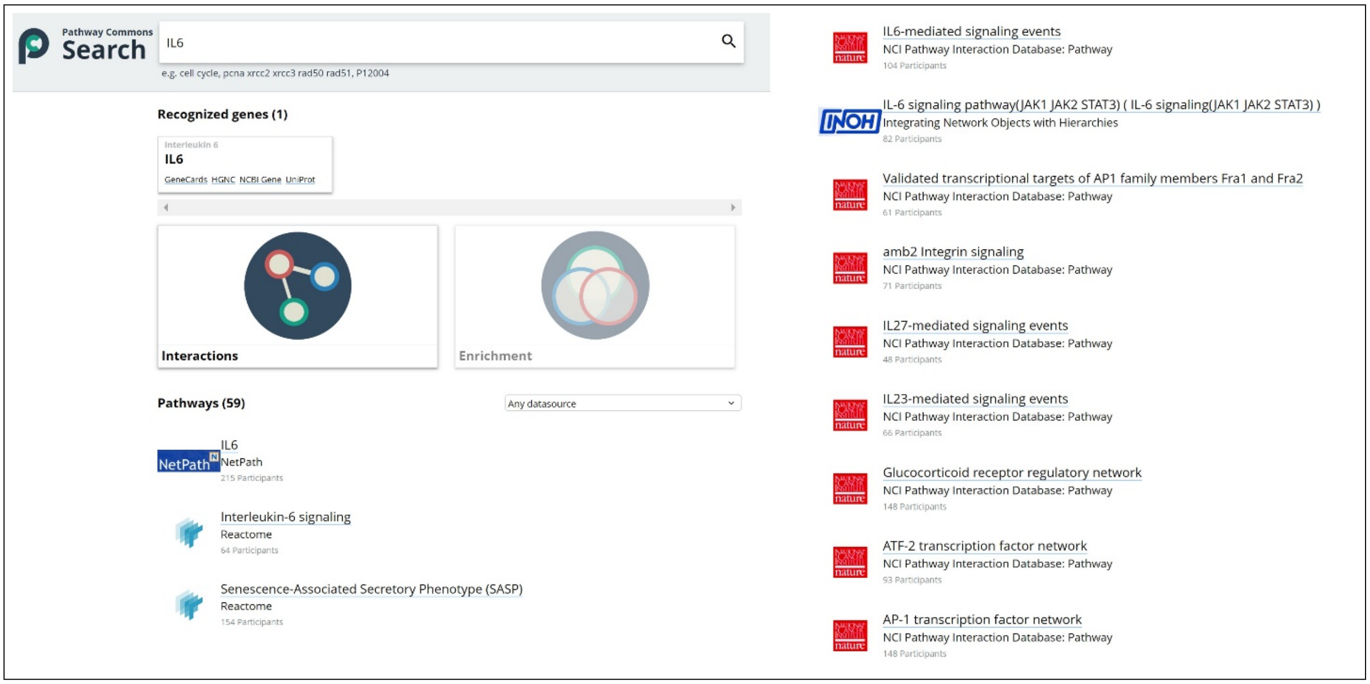
disease	description	count in network	strength	false discovery rate
DOID:11394	Adult respiratory distress syndrome	4 of 5	3.15	7.41e-09
DOID:11396	Pulmonary edema	3 of 4	3.13	1.01e-06
DOID:0050697	Chorioamnionitis	3 of 4	3.13	1.01e-06
DOID:6132	Bronchitis	2 of 3	3.07	0.00032
DOID:8283	Peritonitis	3 of 5	3.03	1.40e-06

(more ...)

Tissue expression (TISSUES)

tissue	description	count in network	strength	false discovery rate
BTO:0001370	THP-1 cell	4 of 4	3.25	3.59e-09
BTO:0003350	BV-2 cell	3 of 3	3.25	1.00e-06
BTO:0001031	Peritoneal fluid	2 of 2	3.25	0.00013
BTO:0001026	Polymorphonuclear leukocyte	2 of 2	3.25	0.00013
BTO:0003742	Synoviocyte	3 of 4	3.13	1.46e-06

Figure 17: STRING Database results after searching for term “IL-6”.



Pathway Commons Search

IL6

e.g. cell cycle, pcna xrcc2 xrcc3 rad50 rad51, P12004

Recognized genes (1)

Interleukin 6
IL6
GeneCards HGNC NCBI Gene UniProt

Interactions

Enrichment

Pathways (59)

Any datasource

IL6
NetPath
215 Participants

Interleukin-6 signaling
Reactome
64 Participants

Senescence-Associated Secretory Phenotype (SASP)
Reactome
154 Participants

IL6-mediated signaling events
NCI Pathway Interaction Database: Pathway
104 Participants

IL-6 signaling pathway(JAK1 JAK2 STAT3) (IL-6 signaling(JAK1 JAK2 STAT3))
Integrating Network Objects with Hierarchies
82 Participants

Validated transcriptional targets of AP1 family members Fra1 and Fra2
NCI Pathway Interaction Database: Pathway
61 Participants

amb2 Integrin signaling
NCI Pathway Interaction Database: Pathway
71 Participants

IL27-mediated signaling events
NCI Pathway Interaction Database: Pathway
48 Participants

IL23-mediated signaling events
NCI Pathway Interaction Database: Pathway
66 Participants

Glucocorticoid receptor regulatory network
NCI Pathway Interaction Database: Pathway
148 Participants

ATF-2 transcription factor network
NCI Pathway Interaction Database: Pathway
93 Participants

AP-1 transcription factor network
NCI Pathway Interaction Database: Pathway
148 Participants

Figure 18: Pathway Commons results after searching for term “IL-6”.

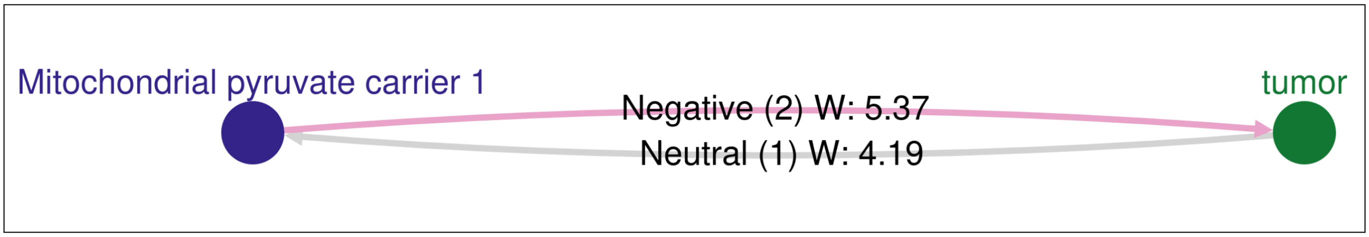
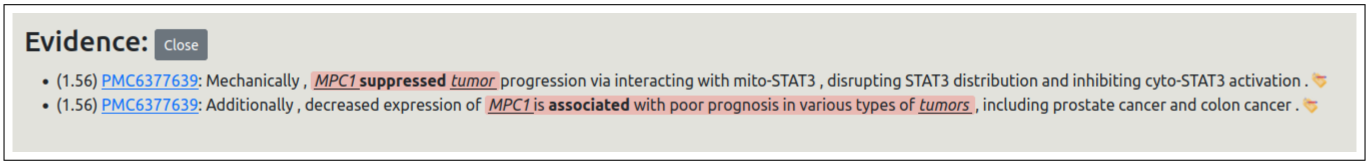


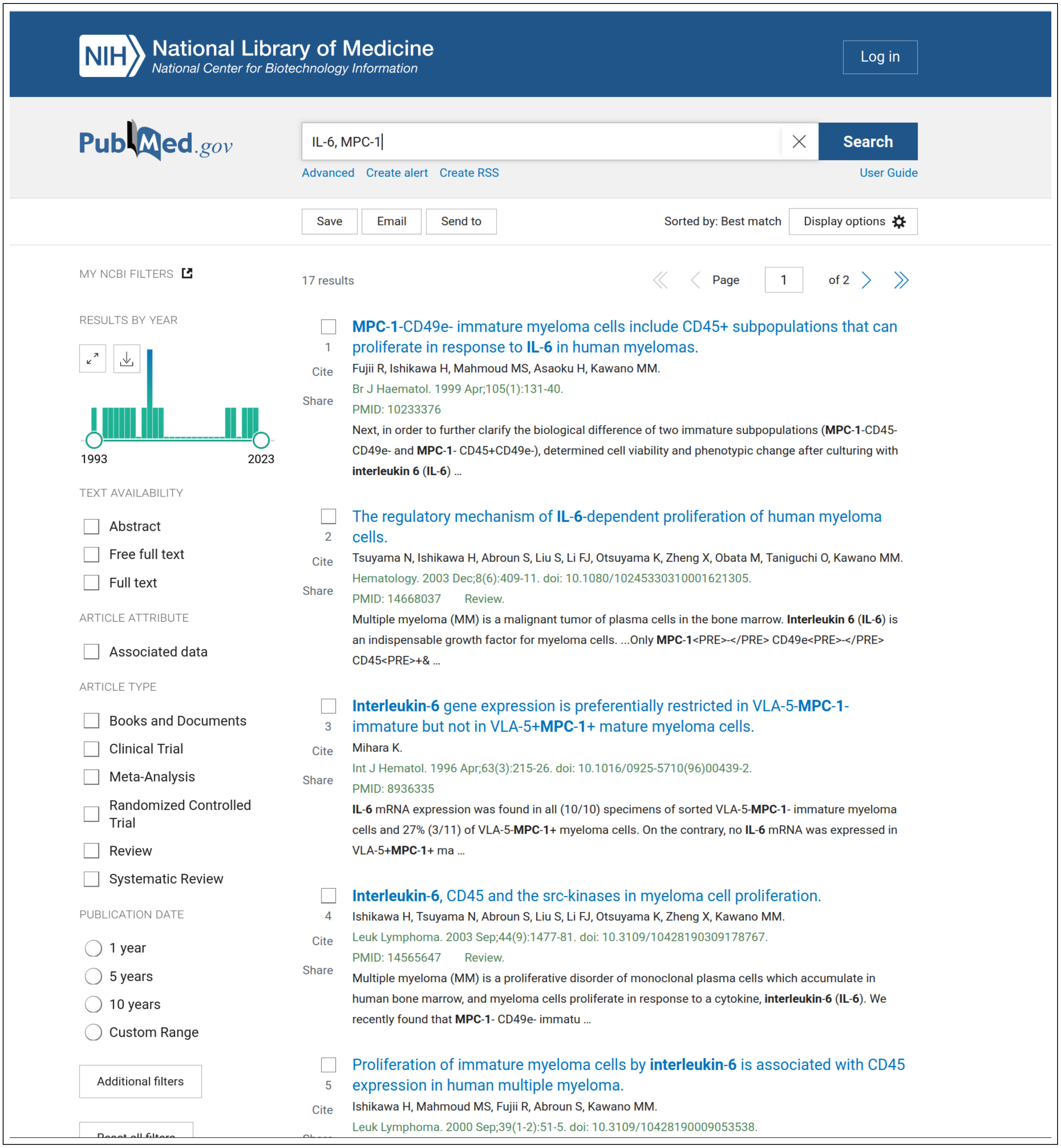
Figure 19: Node-link visualization displaying the reciprocal interactions between MPC-1 and *tumor progression*.



Evidence: Close

- (1.56) PMC6377639: Mechanically , *MPC1* **suppressed** *tumor* progression via interacting with mito-STAT3 , disrupting STAT3 distribution and inhibiting cyto-STAT3 activation .
- (1.56) PMC6377639: Additionally , decreased expression of *MPC1* is **associated** with poor prognosis in various types of *tumors* , including prostate cancer and colon cancer .

Figure 20: Evidence supporting the edge that encodes down-regulations of *tumor progression* by MPC-1 from Figure 19.



IL-6, MPC-1

Sorted by: Best match

17 results

Page 1 of 2

TEXT AVAILABILITY

□ Abstract

□ Free full text

□ Full text

ARTICLE ATTRIBUTE

□ Associated data

ARTICLE TYPE

□ Books and Documents

□ Clinical Trial

□ Meta-Analysis

□ Randomized Controlled Trial

□ Review

□ Systematic Review

PUBLICATION DATE

○ 1 year

○ 5 years

○ 10 years

○ Custom Range

□ 1 **MPC-1**-CD49e- immature myeloma cells include CD45+ subpopulations that can proliferate in response to **IL-6** in human myelomas.
Fujii R, Ishikawa H, Mahmoud MS, Asaoku H, Kawano MM.
Br J Haematol. 1999 Apr;105(1):131-40.
PMID: 10233376
Next, in order to further clarify the biological difference of two immature subpopulations (**MPC-1**-CD45-CD49e- and **MPC-1**- CD45+CD49e-), determined cell viability and phenotypic change after culturing with **interleukin 6** (**IL-6**) ...

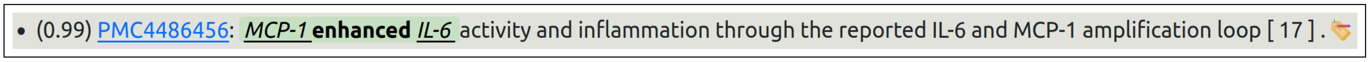
□ 2 The regulatory mechanism of **IL-6**-dependent proliferation of human myeloma cells.
Tsuyama N, Ishikawa H, Abroun S, Liu S, Li FJ, Otsuyama K, Zheng X, Obata M, Taniguchi O, Kawano MM.
Hematology. 2003 Dec;8(6):409-11. doi: 10.1080/10245330310001621305.
PMID: 14668037 Review.
Multiple myeloma (MM) is a malignant tumor of plasma cells in the bone marrow. **Interleukin 6** (**IL-6**) is an indispensable growth factor for myeloma cells. ...Only **MPC-1**<PRE>-</PRE> CD49e<PRE>-</PRE> CD45<PRE>+& ...

□ 3 **Interleukin-6** gene expression is preferentially restricted in VLA-5-**MPC-1**-immature but not in VLA-5+**MPC-1**+ mature myeloma cells.
Mihara K.
Int J Hematol. 1996 Apr;63(3):215-26. doi: 10.1016/0925-5710(96)00439-2.
PMID: 8936335
IL-6 mRNA expression was found in all (10/10) specimens of sorted VLA-5-**MPC-1**- immature myeloma cells and 27% (3/11) of VLA-5-**MPC-1**+ myeloma cells. On the contrary, no **IL-6** mRNA was expressed in VLA-5+**MPC-1**+ ma ...

□ 4 **Interleukin-6**, CD45 and the src-kinases in myeloma cell proliferation.
Ishikawa H, Tsuyama N, Abroun S, Liu S, Li FJ, Otsuyama K, Zheng X, Kawano MM.
Leuk Lymphoma. 2003 Sep;44(9):1477-81. doi: 10.3109/10428190309178767.
PMID: 14565647 Review.
Multiple myeloma (MM) is a proliferative disorder of monoclonal plasma cells which accumulate in human bone marrow, and myeloma cells proliferate in response to a cytokine, **interleukin-6** (**IL-6**). We recently found that **MPC-1**- CD49e- immatu ...

□ 5 Proliferation of immature myeloma cells by **interleukin-6** is associated with CD45 expression in human multiple myeloma.
Ishikawa H, Mahmoud MS, Fujii R, Abroun S, Kawano MM.
Leuk Lymphoma. 2000 Sep;39(1-2):51-5. doi: 10.3109/10428190009053538.

Figure 21: Results returned from searching for “`IL-6, MPC-1`” in PubMed’s search engine.



- (0.99) PMC4486456: *MCP-1* **enhanced** *IL-6* activity and inflammation through the reported IL-6 and MCP-1 amplification loop [17] . 👎

Figure 22: Example of additional relevant interaction and link to its source article found using the full-text search interface of our system to search for “`IL-6, MPC-1`”.