PubTator^3 Ex: Remdesivir NLM API -

PubTator3 Tutorial

About PubTator3

PubTator3 is a web-based system that offers a comprehensive set of features and tools that allow researchers to extensively explore the ever-expanding wealth of biomedical literature for knowledge discovery. It uses advanced text mining and state-of-the-art AI techniques to annotate and unify bio-entities and their corresponding relations for semantic and relation searches and downloads through its online search, APIs and FTP bulk download. PubTator3 is freely accessible to the research community.

PubTator has been serving the research community since 2013 when the widely used PubTator system was first developed and evolved through major upgrades to PubTator3. Compared to its predecessors, PubTator3 has been equipped with new features including:

- Improved entity annotations by newly developed AI tools, such as AIONER, tmVar3, and GNorm2.
- New relation annotations among six bio-entities, made available by BioREx, a cutting-edge transformer-based method for relation extraction.
- New semantic search function powered by query autocomplete.
- New search filters by article section; relation types, etc.
- Significantly more comprehensive visualization, featuring highlights of key bio-entities and their relations.

PubTator publications include:

1. Wei,C.-H., Allot,A., Lai,P.-T., Leaman,R., Tian,S., Luo,L., Jin,Q., Wang,Z., Chen,Q. and Lu,Z. (2024) PubTator 3.0: an AI-powered literature resource for unlocking biomedical knowledge. *Nucleic Acids Research*, 10.1093/nar/gkae235.
2. Wei,C.-H., Allot,A., Leaman,R. and Lu,Z. (2019) PubTator central: automated concept annotation for biomedical full text articles. *Nucleic Acids Research*, 47,

W587–W593.

3. Wei,C.-H., Kao,H.-Y. and Lu,Z. (2013) PubTator: a Web-based text mining tool for assisting Biocuration. *Nucleic Acids Research*, 41, W518–W522.

Data Available on PubTator3

The PubTator3 uses tools developed with advanced text mining and state-of-the-art AI techniques to annotate and normalize six types of bio-entities found in the biomedical literature and extract twelve corresponding relations between these entities. It currently contains over one billion entity and relation annotations across approximately 36 million PubMed abstracts and 6 million full-text articles from the PMC open-access subset and continues to grow with weekly updates.

Entity Annotations

PubTator3 uses a newly developed AI tool, namely AIONER, to annotate the six types of bio-entities including genes, diseases, chemicals, variants, species and cell lines. The annotated entities that appear in different forms in the literature are normalized to unique and standardized concepts in relevant terminologies by tools specific for each entity type. This ensures precise retrieval of relevant articles, irrespective of synonymous terms used in the search. The relevant terminologies used for each entity type are listed in the table as follows.

Table 1. PubTator3 Annotated Entity Types and Corresponding Normalization Terminologies

Entity Type	Terminology
Gene	NCBI Gene
Disease	MeSH (Medical Subject Headings)
Chemical	MeSH (Medical Subject Headings)
Variant	dbSNP, if possible, otherwise HGNV format
Species	NCBI Taxonomy

Cell Line	Cellosaurus

Relation Annotations

PubTator3 extracts a total of twelve types of relation among the six bio-entities using BioREx, a cutting-edge transformer-based method for relation extraction. A list of the extracted relations and their explanations are given in the following table.

Table 2. PubTator3 Extracted Relation Types and Their Description

Relation Type	Description
ASSOCIATE	The associated relation with no specific description. This type applies to various entity pairs.
CAUSE	A positive correlation exists when the status of one entity tends to increase (or decrease) as the other increase (or decreases). This type includes chemical-induced diseases and genetic diseases caused by variants.
COMPARE	The effect comparison of two chemicals/drugs.
COTREAT	It is defined as the use of two or more chemical/drugs administered separately or in a fixed-dose combination.
DRUG_INTERACT	A pharmacodynamic interaction between two chemicals that results in an array of side effects.
INHIBIT	A negative correlation exists when the status of the two entities tends to be opposite. This type includes disease-gene and chemical-variant.
INTERACT	Physical interaction, like protein-binding. This type includes gene-gene, gene-chemical, chemical-variant.
NEGATIVE_CORRELATE	A negative correlation exists when the status of the two entities tends to be opposite. This type includes chemical-gene, chemical co-expression, and gene co-expression.
POSITIVE_CORRELATE	A positive correlation exists when the status of one entity tends to increase (or decrease) as the other increase (or decreases). This type includes chemical-gene, chemical co-expression, and gene co-expression.
PREVENT	A negative correlation exists when the status of the two entities tends to be opposite. This type includes variant-disease.

STIMULATE	A positive correlation exists when the status of one entity tends to increase (or decrease) as the other increase (or decreases). This type includes disease-gene and disease-variant.
TREAT	A chemical/drug treats a disease.

Online Search

PubTator3 offers a powerful online search interface to provide users with advanced search capabilities and enriched visualization, enable large-scale analyses, streamlining many complex information needs. On the PubTator3 website, users can search for relevant articles through keywords, bio-entities, and relations. The search results are displayed in a ranked order with enriched visualization features, such as highlights, statistics, filters. Users can also save the search results for their future needs. Details of each article can also be viewed in an enriched display.

Keyword Search

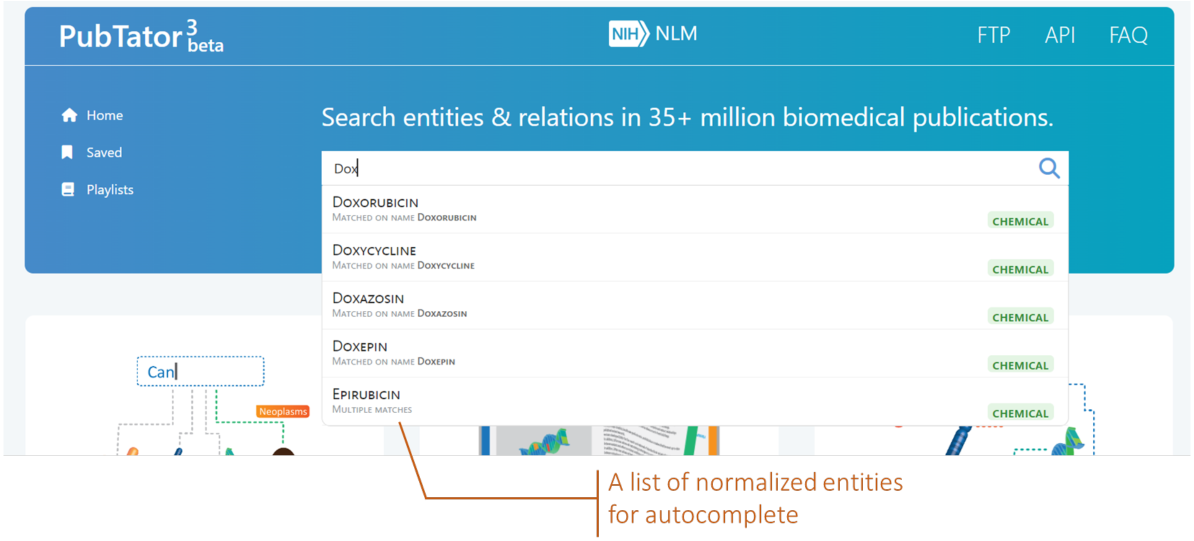
All PubMed abstracts and PMC open-access full-text articles used in PubTator3 are well indexed by keywords. Users can use keywords, such as "breast cancer" to search for articles containing the searched keywords, like searching in Google.

Entity Search

All PubMed abstracts and PMC open-access full-text articles are also well indexed by the annotated and normalized bio-entities for search. Users can easily search for articles containing the searched entities. An autocomplete function is implemented to enable entity search with ease. For example, when users search for articles containing "Doxorubicin", a list of normalized bio-entities is provided for the user to select the most appropriate one for search when the user is typing in the search box, as shown in Figure 1, that translates free-text queries into corresponding semantic concepts ("@CHEMICAL_Doxorubicin").

The identifier of a variant (e.g., "@VARIANT_p.V600E_BRAF_human") is constructed by combining its concept type with its corresponding gene and the most commonly used HGVS (Human Genome Variation Society) nomenclature. Similarly, the identifiers for other entities are formed by combining the concept type (e.g., Chemical) with the official name (e.g., "Metformin"), resulting in identifiers such as

"@CHEMICAL_Metformin".



Search entities & relations in 35+ million biomedical publications.

Dox

A list of normalized entities for autocomplete

Figure 1. The autocomplete function.

Boolean Search

Users can combine keyword and entity search terms using Boolean operators. For example, users can use **AND** to combine terms, as in @DISEASE_COVID_19 AND @GENE_PON1. Users can also use **OR** and **parenthesis**, as in (@DISEASE_COVID_19 AND complications) OR @DISEASE_Post_Acute_COVID_19_Syndrome.

Relation Search

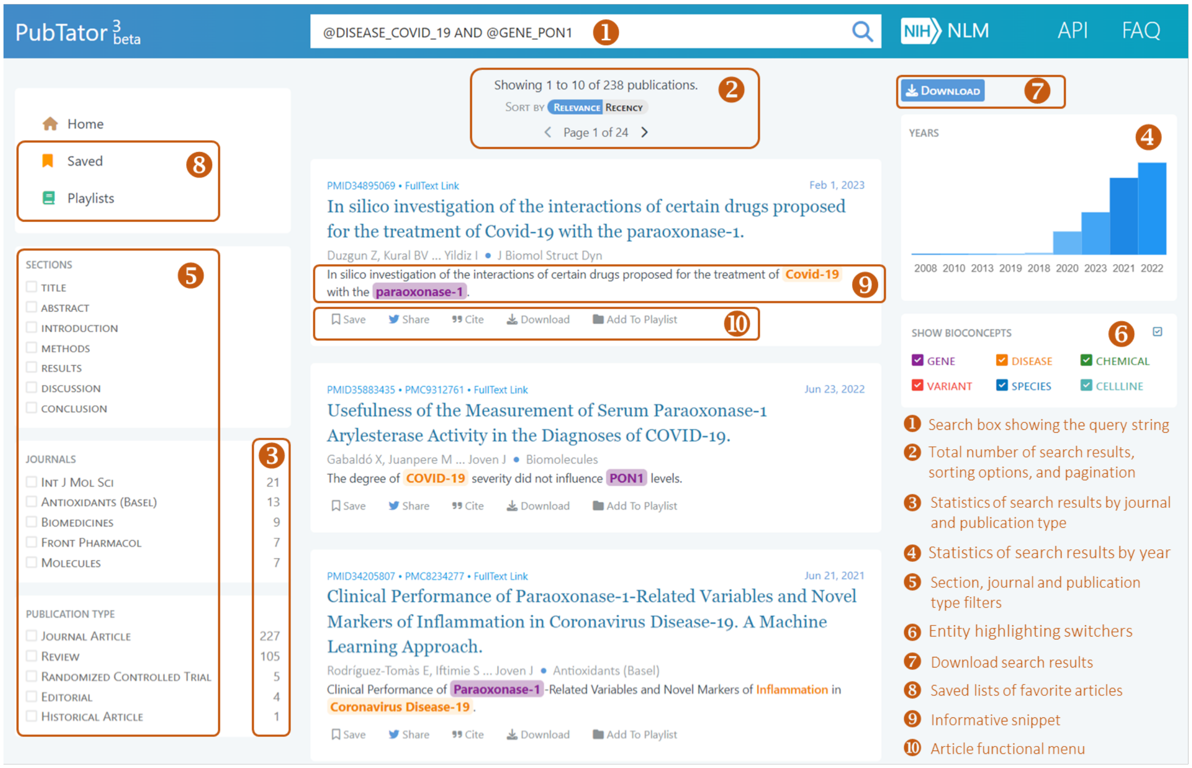
Users can also search for articles containing relations between a pair of entities on PubTator3. Relation search can be performed in different relation query formats, e.g.,

- Use relations:associate|@DISEASE_COVID_19|@GENE_PON1 to search for articles containing the **associate** relation between **@DISEASE_COVID_19** and **@GENE_PON1**
- Use relations:associate|@DISEASE_COVID_19|GENE to search for articles containing the **associate** relation between **@DISEASE_COVID_19** and any **gene** entities
- Use relations:associate|@DISEASE_COVID_19|ANY to search for articles containing the **associate** relation between **@DISEASE_COVID_19** and any other entities

- User can also replace the **associate** with **ANY** in the previous queries to search for articles containing any relations between **@DISEASE_COVID_19** and **@GENE_PON1**, between **@DISEASE_COVID_19** and any **gene** entities, and between **@DISEASE_COVID_19** and any other entities

Search Results

PubTator3 returns a list of articles relevant to the user's query as search results and displays the search results in a sorted order. Each article in the search results contains the title and an informative snippet. Users can click on the title to read and get more details of the article. The user's query is highlighted in the informative snippet. A set of informative features, such as total number of articles in the search results, statistics, and useful tools is provided for the user to navigate, refine, and save the search results according to their specific needs. Figure 2 shows the search results for the query of "@DISEASE_COVID_19 AND @GENE_PON1" as an example. Details of search results are given in the following sections.



@DISEASE_COVID_19 AND @GENE_PON1

Showing 1 to 10 of 238 publications.

SORT BY RELEVANCE RECENCY

Page 1 of 24

SECTIONS	
□ TITLE	
□ ABSTRACT	
□ INTRODUCTION	
□ METHODS	
□ RESULTS	
□ DISCUSSION	
□ CONCLUSION	
JOURNALS	
□ INT J MOL SCI	21
□ ANTIOXIDANTS (BASEL)	13
□ BIOMEDICINES	9
□ FRONT PHARMACOL	7
□ MOLECULES	7
PUBLICATION TYPE	
□ JOURNAL ARTICLE	227
□ REVIEW	105
□ RANDOMIZED CONTROLLED TRIAL	5
□ EDITORIAL	4
□ HISTORICAL ARTICLE	1

PMID34895069 • FullText Link Feb 1, 2023

In silico investigation of the interactions of certain drugs proposed for the treatment of Covid-19 with the paraoxonase-1.

Duzgun Z, Kural BV ... Yildiz I • J Biomol Struct Dyn

In silico investigation of the interactions of certain drugs proposed for the treatment of Covid-19 with the paraoxonase-1.

PMID35883435 • PMC9312761 • FullText Link Jun 23, 2022

Usefulness of the Measurement of Serum Paraoxonase-1 Arylesterase Activity in the Diagnoses of COVID-19.

Gabaldó X, Juanpere M ... Joven J • Biomolecules

The degree of COVID-19 severity did not influence PON1 levels.

PMID34205807 • PMC8234277 • FullText Link Jun 21, 2021

Clinical Performance of Paraoxonase-1-Related Variables and Novel Markers of Inflammation in Coronavirus Disease-19. A Machine Learning Approach.

Rodríguez-Tomàs E, Iftimie S ... Joven J • Antioxidants (Basel)

Clinical Performance of Paraoxonase-1-Related Variables and Novel Markers of Inflammation in Coronavirus Disease-19.

1 Search box showing the query string
2 Total number of search results, sorting options, and pagination
3 Statistics of search results by journal and publication type
4 Statistics of search results by year
5 Section, journal and publication type filters
6 Entity highlighting switchers
7 Download search results
8 Saved lists of favorite articles
9 Informative snippet
10 Article functional menu

Figure 2. Search results for query of "@DISEASE_COVID_19 AND @GENE_PON1".

Ranking of Search Results

The search results returned for a query are sorted in a relevance ranking order by default. PubTator3 uses a relevance algorithm to retrieve the relevant articles for each query, and boots article relevance using the following criteria: articles containing relations related to the user query have the highest boosting weight, articles containing query terms with closer distance have higher boosting weights. Besides displaying articles in relevance ranking order, users can use the sorting option tool to display the articles in an order of publication recency.

Statistics of Search Results

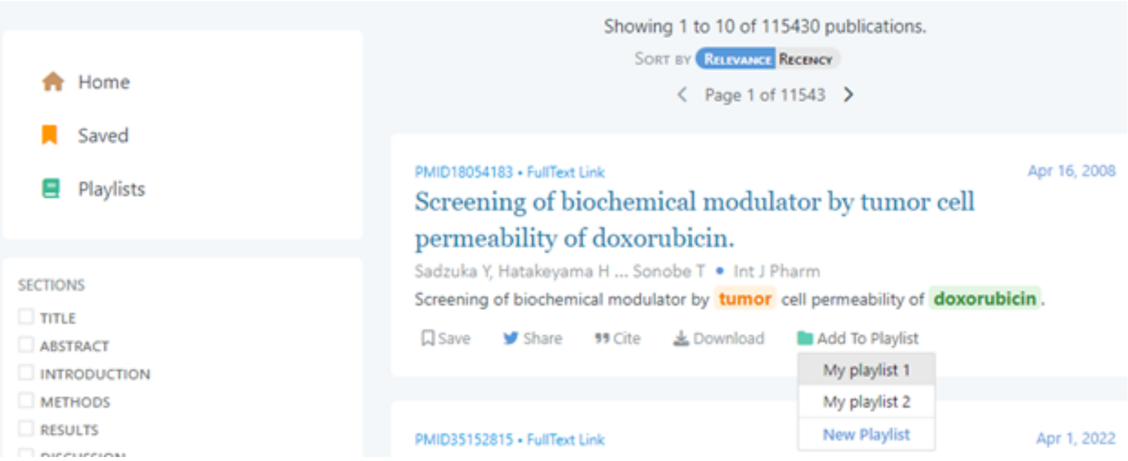
PubTator3 provides users informative summary statistics of the search results, including total number of articles, number of articles by publication year, number of articles by journal, and number of articles by publication type, as shown in Figure 2.

Filtering Search Results

As shown in Figure 2, PubTator3 includes a set of filtering tools that allow users to refine the search results by (1) article sections; (2) journals; and (3) publication type. PubTator3 also provides a filter for users to determine which types of entities to highlight in the search results.

Saving and Downloading Search Results

We have streamlined the article collection feature to make it more user-friendly for users to save searched relevant articles of their favorite for future needs. In PubTator3, users can effortlessly choose to save favorite articles in different collections, download a list of articles in the search results, or download details of the articles. As shown in Figure 2, a set of article functional tools is provided for users to save articles in the temporary Saved list, add articles to any of the playlists stored in the database, and download details of articles. Users can also use the Download tool to download a list of the searched articles. The following figure shows how to add an article to a playlist.



Showing 1 to 10 of 115430 publications.

PMID18054183 • FullText Link Apr 16, 2008

Screening of biochemical modulator by tumor cell permeability of doxorubicin.

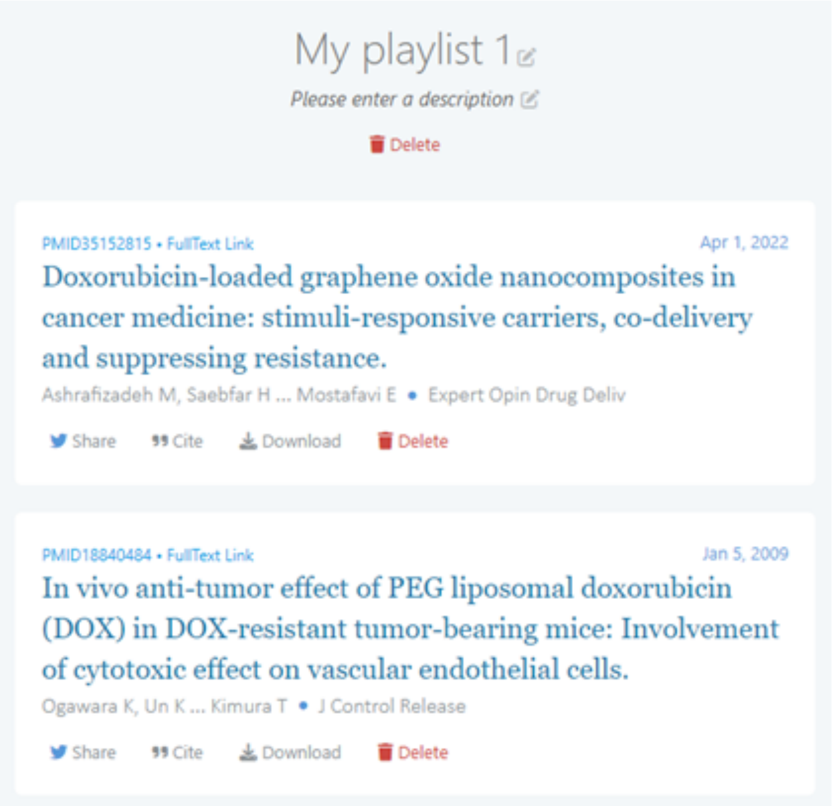
Sadzuka Y, Hatakeyama H ... Sonobe T • Int J Pharm

Screening of biochemical modulator by tumor cell permeability of doxorubicin.

PMID35152815 • FullText Link Apr 1, 2022

Figure 3. Add the favorite article to a playlist.

Users can manage their articles in the playlists on PubTator3 with ease. They can enter a description for each playlist, or remove articles from the playlist, as shown in the following figure.



My playlist 1

Please enter a description

PMID35152815 • FullText Link Apr 1, 2022

Doxorubicin-loaded graphene oxide nanocomposites in cancer medicine: stimuli-responsive carriers, co-delivery and suppressing resistance.

Ashrafizadeh M, Saebfar H ... Mostafavi E • Expert Opin Drug Deliv

PMID18840484 • FullText Link Jan 5, 2009

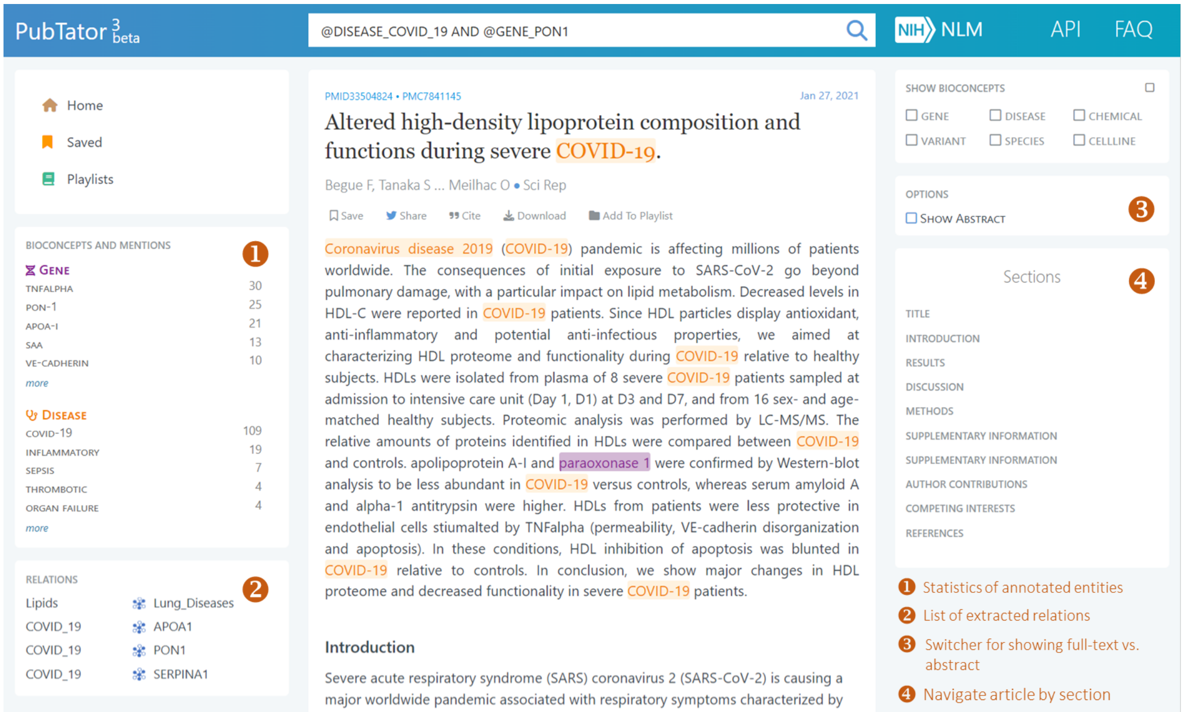
In vivo anti-tumor effect of PEG liposomal doxorubicin (DOX) in DOX-resistant tumor-bearing mice: Involvement of cytotoxic effect on vascular endothelial cells.

Ogawara K, Un K ... Kimura T • J Control Release

Figure 4. Manage a playlist.

Viewing Article Details

By clicking on the title of a selected article, PubTator3 directs users to the publication page, which displays details of an article including the abstract or full text (if available), annotated entities and extracted relations, as shown in Figure 5. A summary of the annotated bio-entities mentioned in the article and a list of extracted relations can be found at the left side of the page. Users can also click on any entity in the summary to highlight mentions of the entity in the article or click on any relation to highlight mentions of the pair of entities in the article. A Show Bioconcepts is also provided for users to highlight different types of entities in the article. When full-text is available, users can also use the Options tool switch display of abstract or full text, and use the Sections tool to navigate the article full-text by section.


PubTator 3 beta

@DISEASE_COVID_19 AND @GENE_PON1

BIOCONCEPTS AND MENTIONS ❶

GENE	
TNFALPHA	30
PON-1	25
APOA-I	21
SAA	13
VE-CADHERIN	10
more	

DISEASE	
COVID-19	109
INFLAMMATORY	19
SEPSIS	7
THROMBOTIC	4
ORGAN FAILURE	4
more	

RELATIONS ❷

Lipids	Lung_Diseases
COVID_19	APOA1
COVID_19	PON1
COVID_19	SERPINA1

PMID33504824 • PMC7841145 Jan 27, 2021

Altered high-density lipoprotein composition and functions during severe COVID-19.

Begue F, Tanaka S ... Meilhac O • Sci Rep

Coronavirus disease 2019 (COVID-19) pandemic is affecting millions of patients worldwide. The consequences of initial exposure to SARS-CoV-2 go beyond pulmonary damage, with a particular impact on lipid metabolism. Decreased levels in HDL-C were reported in COVID-19 patients. Since HDL particles display antioxidant, anti-inflammatory and potential anti-infectious properties, we aimed at characterizing HDL proteome and functionality during COVID-19 relative to healthy subjects. HDLs were isolated from plasma of 8 severe COVID-19 patients sampled at admission to intensive care unit (Day 1, D1) at D3 and D7, and from 16 sex- and age-matched healthy subjects. Proteomic analysis was performed by LC-MS/MS. The relative amounts of proteins identified in HDLs were compared between COVID-19 and controls. apolipoprotein A-I and paraoxonase 1 were confirmed by Western-blot analysis to be less abundant in COVID-19 versus controls, whereas serum amyloid A and alpha-1 antitrypsin were higher. HDLs from patients were less protective in endothelial cells stiumalted by TNFalpha (permeability, VE-cadherin disorganization and apoptosis). In these conditions, HDL inhibition of apoptosis was blunted in COVID-19 relative to controls. In conclusion, we show major changes in HDL proteome and decreased functionality in severe COVID-19 patients.

Introduction

Severe acute respiratory syndrome (SARS) coronavirus 2 (SARS-CoV-2) is causing a major worldwide pandemic associated with respiratory symptoms characterized by

SHOW BIOCONCEPTS □

□ GENE □ DISEASE □ CHEMICAL
□ VARIANT □ SPECIES □ CELLLINE

OPTIONS ❸

□ SHOW ABSTRACT

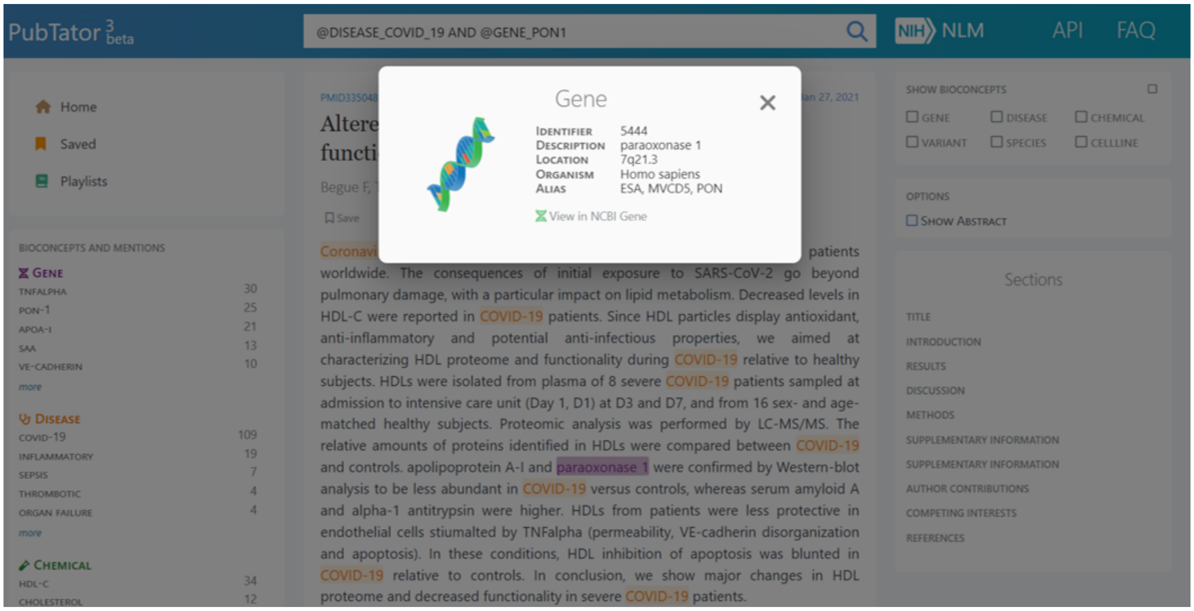
Sections ❹

TITLE
INTRODUCTION
RESULTS
DISCUSSION
METHODS
SUPPLEMENTARY INFORMATION
SUPPLEMENTARY INFORMATION
AUTHOR CONTRIBUTIONS
COMPETING INTERESTS
REFERENCES

❶ Statistics of annotated entities
❷ List of extracted relations
❸ Switcher for showing full-text vs. abstract
❹ Navigate article by section

Figure 5. The publication page.

Clicking on a highlighted mention displays a window summarizing the entity, sourced from publicly available databases or repositories, as shown in Figure 6.


Gene ✕

IDENTIFIER	5444
DESCRIPTION	paraoxonase 1
LOCATION	7q21.3
ORGANISM	Homo sapiens
ALIAS	ESA, MVCDS, PON

View in NCBI Gene

CHEMICAL

HDL-C	34
CHOLESTEROL	12

PubTator 3 beta @DISEASE_COVID_19 AND @GENE_PON1 NIH NLM API FAQ Home Saved Playlists BIOCONCEPTS AND MENTIONS GENE TNFALPHA 30 PON-1 25 APOA-I 21 SAA 13 VE-CADHERIN 10 more DISEASE COVID-19 109 INFLAMMATORY 19 SEPSIS 7 THROMBOTIC 4 ORGAN FAILURE 4 more PMID335048 Jan 27, 2021 Altere functi Begue F, Save Coronavi patients worldwide. The consequences of initial exposure to SARS-CoV-2 go beyond pulmonary damage, with a particular impact on lipid metabolism. Decreased levels in HDL-C were reported in COVID-19 patients. Since HDL particles display antioxidant, anti-inflammatory and potential anti-infectious properties, we aimed at characterizing HDL proteome and functionality during COVID-19 relative to healthy subjects. HDLs were isolated from plasma of 8 severe COVID-19 patients sampled at admission to intensive care unit (Day 1, D1) at D3 and D7, and from 16 sex- and age-matched healthy subjects. Proteomic analysis was performed by LC-MS/MS. The relative amounts of proteins identified in HDLs were compared between COVID-19 and controls. apolipoprotein A-I and paraoxonase 1 were confirmed by Western-blot analysis to be less abundant in COVID-19 versus controls, whereas serum amyloid A and alpha-1 antitrypsin were higher. HDLs from patients were less protective in endothelial cells stiumalted by TNFalpha (permeability, VE-cadherin disorganization and apoptosis). In these conditions, HDL inhibition of apoptosis was blunted in COVID-19 relative to controls. In conclusion, we show major changes in HDL proteome and decreased functionality in severe COVID-19 patients. SHOW BIOCONCEPTS □ □ GENE □ DISEASE □ CHEMICAL □ VARIANT □ SPECIES □ CELLLINE OPTIONS □ SHOW ABSTRACT Sections TITLE INTRODUCTION RESULTS DISCUSSION METHODS SUPPLEMENTARY INFORMATION SUPPLEMENTARY INFORMATION AUTHOR CONTRIBUTIONS COMPETING INTERESTS REFERENCES

Figure 6. The summarization of the selected entity.

APIs

To assist the programmatically access to PubTator, we released our search function via

API. Users can access PubTator's query function and its returned articles. Users can also use curl and our online API to automatically annotate raw text. More information can be found at: https://www.ncbi.nlm.nih.gov/research/pubtator3/api.

FTP Bulk Download

PubTator allows to download annotation in three popular formats (i.e., PubTator, BioC-XML, and BioC-JSON) from our ftp site: https://ftp.ncbi.nlm.nih.gov/pub/lu/PubTator3.

The information available on this website is freely accessible to all. More details about accessibility of this website are available at NLM Accessibility.

Disclaimer

This tool shows the results of research conducted in the Computational Biology Branch, NCBI/NLM.

The information produced on this website is not intended for direct diagnostic use or medical decision-making without review and oversight by a clinical professional. Individuals should not change their health behavior solely on the basis of information produced on this website. NIH does not independently verify the validity or utility of the information produced by this tool. If you have questions about the information produced on this website, please see a health care professional.

More information about NCBI/NLM's disclaimer policy is available.

NLM/NCBI BioNLP Research Group

Powered by

AIONER
BioREx
TaggerOne
tmVar3
GNorm2
NLM-Chem

Contact

Zhiyong Lu, PhD
Chih-Hsuan Wei, PhD
Shubo Tian, PhD

National Center for Biotechnology Information, U.S. National Library of Medicine 8600 Rockville Pike, Bethesda MD, 20894 USA