TIN-X Version 3: Update with expanded dataset and modernized architecture

Vincent T. Metzger[1], Daniel C. Cannon[2], Jeremy J. Yang,[1] Stephen L. Mathias[1], Cristian G. Bologa[1], Anna Waller[3], Stephan C. Schürer[4], Dušica Vidović[4], Lars Juhl Jensen[5], Christophe G. Lambert[1], Tudor I. Oprea[1], and Jeremy S. Edwards[6]

[1]University of New Mexico, Albuquerque, NM, USA; [2]Elevato Digital, Columbia, MO, USA; [3]Center for Molecular Discovery, University of New Mexico Comprehensive Cancer Center, University of New Mexico, Albuquerque, NM, USA; [4]Department of Molecular and Cellular Pharmacology, Miller School of Medicine, University of Miami, Miami, Florida, USA; [5]Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark; [6]Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, New Mexico, USA;

Overview

Target **I**mportance and **N**ovelty e**X**plorer (**TIN-X**)

An Interactive web-based visualization tool for illuminating associations between diseases and drug targets,[1] TIN-X uses natural language processing to identify disease and protein mentions within PubMed content. Two important metrics, **novelty** and **importance**, are computed from this data, and when plotted as log(importance) vs log(novelty), help users visually explore the novelty of drug targets and their associated importance to diseases. The primary data sources are TCRD/Pharos[2] and PubMed content. TIN-X is publicly available:

Version 2.0: ***newdrugtargets.org***

Version 3.0 Demo: ***unmtid-devapps.net***

Informatics Workflow

REST API

Public Web Application

New API, UI, and Database Improvements

TINX-API

- Upgraded the API from Python 2 to Python 3.8
- Includes support for Amazon RDS database
- Identified & fixed bugs

TINX-UI

- New table-view option where plot data is presented in a sortable, searchable table.
- Improving visibility & accessibility
- Fixed bugs and browser compatibility issues

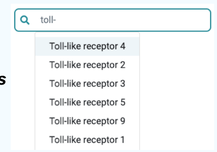
TINX Database

- Now uses Cloud storage w/ Amazon RDS instead of mySQL
- Expanded Dataset to include full-text PubMed content[3]
- Improving process of updating data

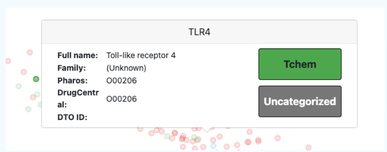
TIN-X User Interface: Includes New Features

Browse Diseases or Browse Targets

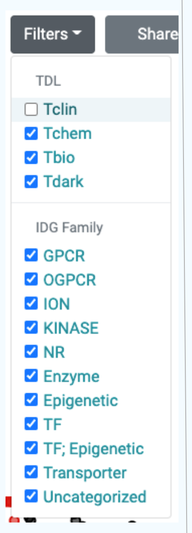
*A search for a specific **Target** associated with **bacterial sepsis** generates autocomplete suggestions:*



Hovering over a data point brings up details:



Share & Export data

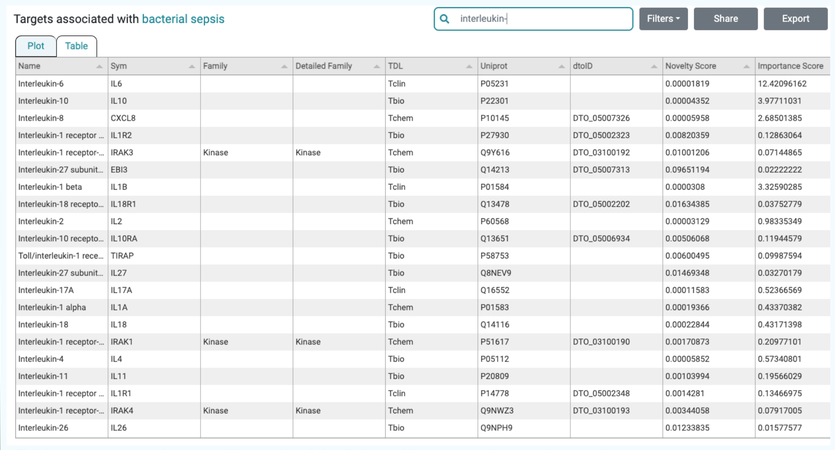


Powerful filters allow the User to select or deselect a variety of TDL and IDG Family categories. Upon selection, the plot is automatically updated to reflect the current status of the Filter.

New Table View

Search table content; rows that don't contain the search query disappear.

Users can apply the same TDL and IDG Family filters available within the View Plot mode

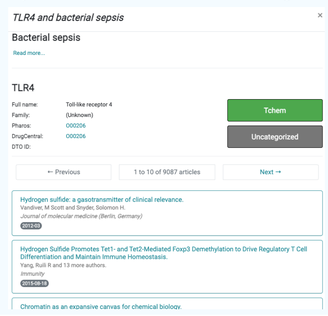


Name	Sym	Family	Detailed Family	TDL	Uniprot	dtoID	Novelty Score	Importance Score
Interleukin-6	IL6			Tclin	P05231		0.00001819	12.42096162
Interleukin-10	IL10			Tbio	P22301		0.00004352	3.97711031
Interleukin-8	CXCL8			Tchem	P10145	DTO_05007326	0.00005958	2.68501385
Interleukin-1 receptor ...	IL1R2			Tbio	P27930	DTO_05002323	0.00820359	0.12863064
Interleukin-1 receptor-...	IRAK3	Kinase	Kinase	Tchem	Q9Y616	DTO_03100192	0.01001206	0.07144865
Interleukin-27 subunit...	EBI3			Tbio	Q14213	DTO_05007313	0.09651194	0.02222222
Interleukin-1 beta	IL1B			Tclin	P01584		0.0000308	3.32590285
Interleukin-18 recepto...	IL18R1			Tbio	Q13478	DTO_05002202	0.01634385	0.03752779
Interleukin-2	IL2			Tchem	P60568		0.00003129	0.98335349
Interleukin-10 recepto...	IL10RA			Tbio	Q13651	DTO_05006934	0.00506068	0.11944579
Toll/interleukin-1 rece...	TIRAP			Tbio	P58753		0.00600495	0.09987594
Interleukin-27 subunit...	IL27			Tbio	Q8NEV9		0.01469348	0.03270179
Interleukin-17A	IL17A			Tclin	Q16552		0.00011583	0.52366569
Interleukin-1 alpha	IL1A			Tchem	P01583		0.00019366	0.43370382
Interleukin-18	IL18			Tbio	Q14116		0.00022844	0.43171398
Interleukin-1 receptor-...	IRAK1	Kinase	Kinase	Tchem	P51617	DTO_03100190	0.00170873	0.20977101
Interleukin-4	IL4			Tbio	P05112		0.00005852	0.57340801
Interleukin-11	IL11			Tbio	P20809		0.00103994	0.19566029
Interleukin-1 receptor ...	IL1R1			Tclin	P14778	DTO_05002348	0.0014281	0.13466975
Interleukin-1 receptor-...	IRAK4	Kinase	Kinase	Tchem	Q9NWZ3	DTO_03100193	0.00344058	0.07917005
Interleukin-26	IL26			Tbio	Q9NPH9		0.01233835	0.01577577

Table contents can be sorted by each of the fields, either ascending or descending.

Detailed View

When clicking on a single datapoint in the Plot View mode or when clicking on a row in the new Table View mode, the User is presented with details on the target/disease association.



This view presents the User with details about the target and an accompanying list of the specific PubMed articles which form the basis of the predicted association between the target and disease. In this example, the target TLR4 is associated with the disease bacterial sepsis. This view includes external links to Pharos, DrugCentral, and disease-ontology.org. The listed publication titles and abstracts link to the corresponding paper. We now prominently feature the publication date.

DTO/DO Disease Ontology

TDL Colors

IDG Families

Above screenshots are from TIN-X Version 3.0, temporarily available at https://unmtid-devapps.net

Development History

2014 *Conceived and Prototyped by Cristian Bologa*

2015 *Work began on what would become TIN-X*

2016

2017 *Version **1.0** is first published online by Cannon et al[1]*

2018

2019 *Version **2.0** of TIN-X released Feb 18, 2019*

2020

2021 *Work on Version **3.0** of TIN-X begins*

2022 Jensen Lab DISEASES 2.0 released[3]

2023 *Version **3.0** launching Spring 2023*

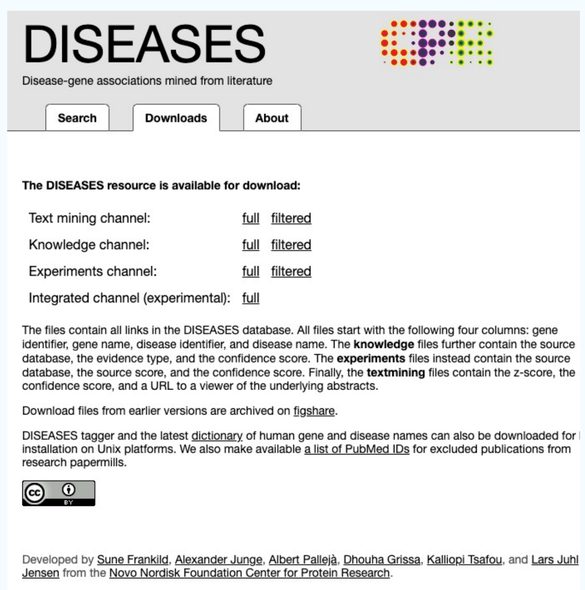
NEW: TIN-X Dataset Expanded via Jensen Lab DISEASES 2.0[3]

TIN-X relies on the DISEASES[3] resource for text-mined PubMed associations

For text-mined associations, the number of disease-gene associations increased by at least **9-fold** at all confidence cutoffs (compared to the previous version of DISEASES)[3]

How is this achieved?

1. Primarily by adding full-text articles to the collection in addition to the titles and abstracts previously included
2. To a lesser extent, improvements to the disease and gene dictionaries used for Named Entity Recognition



Note: Jensen Lab DISEASES resource updated **weekly**

Background & Contributions

TIN-X relies on text mining of PubMed content by JensenLab[3], the Target Central Resource Database (TCRD)[2] for target and Drug Target Ontology (DTO) data, and the DISEASES[3] dictionary for disease ontology. Since 2017, TIN-X has been continually maintained, updated, and improved. The development work highlighted in this poster is being finalized in preparation for a public launch of TIN-X Version 3.0 in Spring 2023, replacing the Version 2.0 which is presently accessible at ***newdrugtargets.org***. Also, TIN-X is among several REST APIs made available at pharos-api.newdrugtargets.org as part of the CFDE Gene Pages Partnership Project. These improvements support the Resource Sharing Plan of KMC, the CFDE, and NIH policies and principles concerning digital resource sharing (e.g. FAIR) as emphasized by the NIH Strategic Plan for Data Science[4].

References:

1) DC Cannon, JJ Yang, SL Mathias, O Ursu, S Mani, A Waller, SC Schürer, LJ Jensen, LA Sklar, CG Bologa, and TI Oprea, "TIN-X: Target Importance and Novelty Explorer." (2017) Bioinformatics, btx200, doi: 10.1093/bioinformatics/btx200

2) Sheils, T., Mathias, S. et al, "TCRD and Pharos 2021: mining the human proteome for disease biology." (2021) Nucl. Acids Res., DOI: 10.1093/nar/gkaa993

3) Dhouha Grissa, Alexander Junge, Tudor I. Oprea, and Lars Juhl Jensen. "DISEASES 2.0: a weekly updated database of disease–gene associations from text mining and data integration." (2022) Database, 1-8; doi/10.1093/database/baac019/6554833

4) NIH Strategic Plan for Data Science, accessed Feb 2022, https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf

This work was supported by US National Institutes of Health (grants CA224370 and U24TR002278), Illuminating the Druggable Genome Knowledge Management Center (IDG KMC), Illuminating the Druggable Genome - Common Fund Data Ecosystem (IDG-CFDE grant number OT2OD030546) and by the Novo Nordisk Foundation (grant number NNF14CC0001).