The Target Importance and Novelty Explorer (TIN-X)

Daniel Cannon[1], Cristian Bologa[1], Lars Jensen[2], and Tudor Oprea[1]

[1]Translational Informatics Division, University of New Mexico Health Sciences Center, Albuquerque, NM
[2]Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark

Overview

The Target Importance and Novelty Explorer (TIN-X) is an interactive web-based visualization tool for discovering interesting targets in the context of a disease.

Background

- New methods are needed to prioritize potential drug targets for further study.
- UNM's Target Central Resource Database (TCRD) catalogs the target development level of 20,202 proteins, including 827 GPCRs, 572 kinases, 342 ion channels, and 48 nuclear receptors.
- The Jensen Lab at the Center for Protein Research has used text mining to compile a database of protein and disease mentions in PubMed abstracts.

Hypotheses

- A target that is mentioned in many abstracts that also mention a specific disease is likely to be of importance to that disease.
- A target that is mentioned in fewer abstracts is more novel and less understood.
- Abstracts which mention only a few targets and diseases are more specific and should be given greater weight than those which mention many.

Approach

- Using the disease and protein mentions compiled by the Jensen lab we assigned weighted **novelty** and disease-specific **importance** scores to the targets in TCRD.

Novelty

- **Novelty** measures the scarcity of publications about a *target*.
- Each paper is assigned a fractional target (FT) score of one divided by the number of targets mentioned in it.

  Example: A paper mentions 5 targets. Its FT score is $1/5$.
- The novelty score of each target is one divided by the sum of the FT scores of all papers which mention the target.

  Example: Two papers mention target T. One mentions 5 targets while the other mentions 3 targets. Target T's novelty score is $1/\left(1/5 + 1/3\right)$.

Importance

- **Importance** measures the strength of the association between a *target* and a *disease*.
- Each paper is assigned a fractional disease-target (FDT) score of one divided by the number of targets and by the number of diseases mentioned in the paper.

  Example: A paper mentions 5 targets and 2 diseases. Its FDT score is $1/(5 \times 2)$.
- The importance score for a target in the context of a given disease is the sum of the FDT scores of all papers which mention both the disease and the target.

Database Contents

- 275,002 abstracts which mention an IDG family target and a disease
- 1,344 targets with disease associations
- 3,726 diseases with target associations

Screenshots

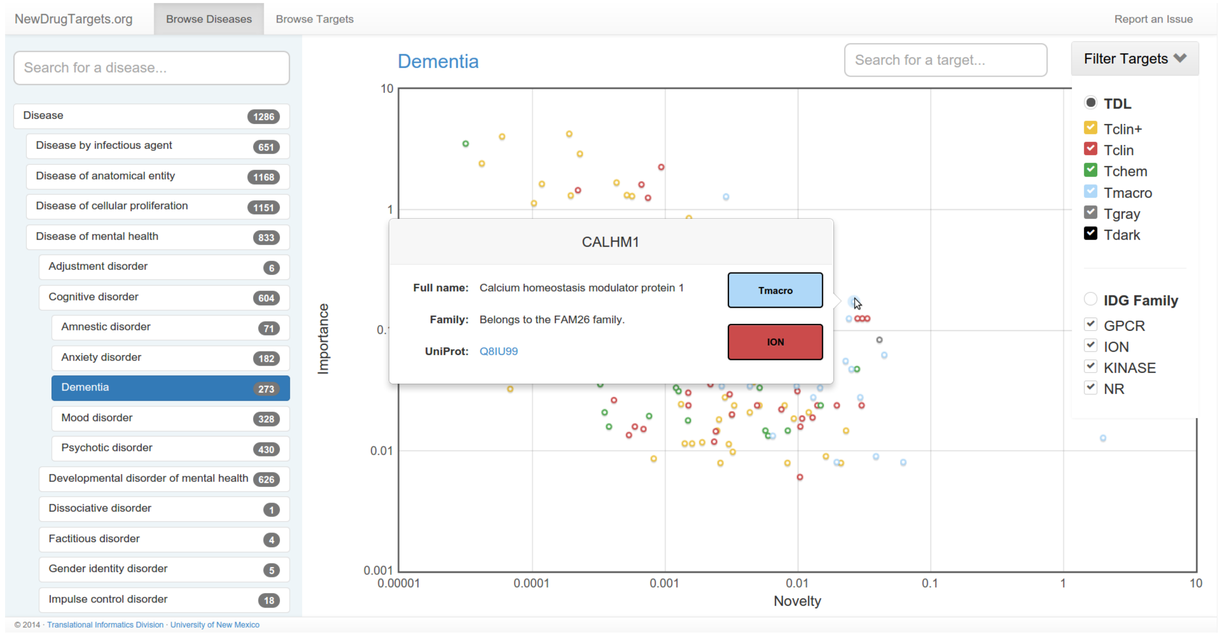

Figure 1: After selecting a disease, TIN-X displays a plot of the novelty and importance of all targets associated with the disease. Hovering over a point displays information about the target. Here CALHM1 is highlighted in the context of dementia.

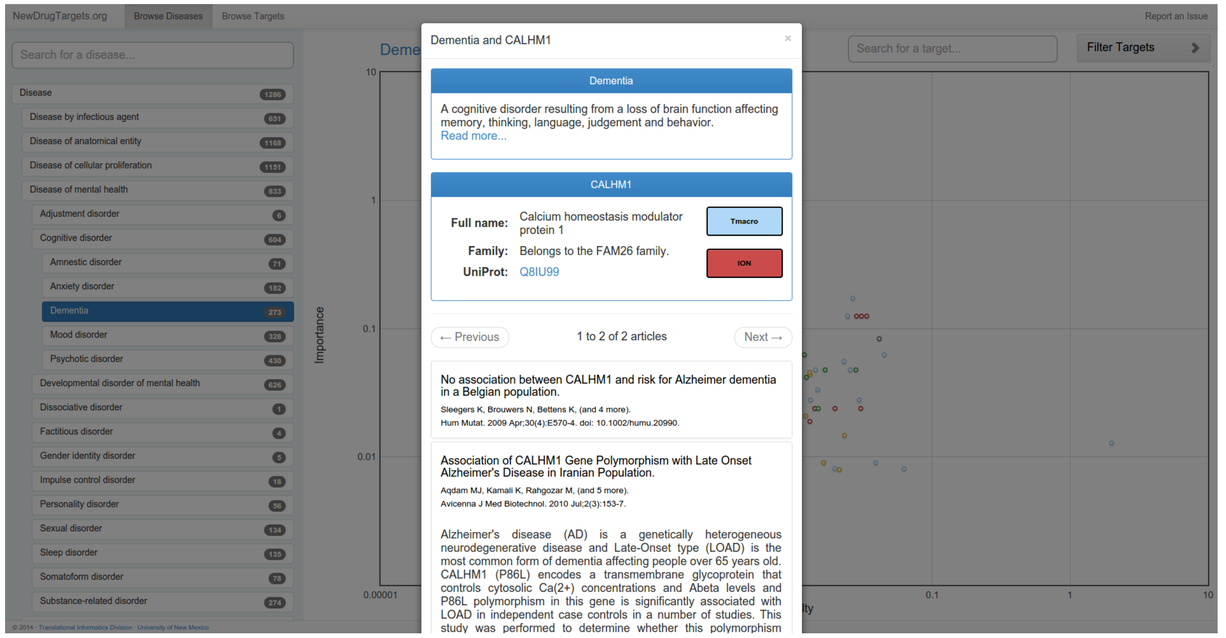

Figure 2: When a point is clicked, another dialog displays information about the target and the disease along with a ranked list of the abstracts used to determine the association between the two. Clicking on an article title expands the full abstract.

Explore New Drug Targets at NewDrugTargets.org