

# Exploring biomedical and chemical data sources with TargetMine, a data warehouse system for target discovery

Yi-An Chen<sup>1,2</sup>, Lokesh P. Tripathi<sup>1</sup>, Kenji Mizuguchi<sup>1,2,\*</sup>

<sup>1</sup> National Institute of Biomedical Innovation, 7-6-8 Saito Asagi, Ibaraki, Osaka, 567-0085, Japan.

<sup>2</sup> Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamadaoka, Suita, Osaka, 565-0871, Japan.

\*E-mail : [kenji@nibio.go.jp](mailto:kenji@nibio.go.jp)

## Abstract

We have recently developed TargetMine<sup>†</sup>, an integrated data warehouse for target discovery, which is based on the flexible InterMine framework and effectively combines biological data from several public resources. It enables biological data gathering and data analysis in a single user-friendly interface and it is, thus, a useful tool to assist in candidate gene selection. TargetMine has been effectively employed for target selection in a protein interaction network-based investigation of Hepatitis C virus (HCV) pathogenesis. For the purpose of drug discovery, we have now included in TargetMine some more data about chemical compounds and their associations with proteins. We have also incorporated new data sources for biological pathways and interactions. The enhanced data warehouse facilitates more complicated queries and new ideas for prioritizing targets.

The academic version of TargetMine is available at <http://targetmine.nibio.go.jp/>.

## Data sources

Integrated data sources in TargetMine



## Using TargetMine

TargetMine features a sequence ontology based data model with a user-friendly web interface that allows the end users to construct flexible and complex queries. The users may also choose from a library of 'templates' consisting of diverse predefined queries with a simple form and description.

**Gene(s) → Upstream Transcription Factors**  
Given a gene or a list of genes, retrieve all upstream regulatory genes (transcription factors) from the AMADEUS and ORegAnno compiled TF-Target gene relations

Gene  
LOOKUP:  for Organism:

**Compound → Protein → Compound**  
Find other compounds which target the same protein(s) as the querying compound.

CompoundGroup  
LOOKUP:

**Gene → Transcription factor → Compound**  
Given a gene or a list of genes, find the upstream regulatory genes (transcription factors) and the compounds which target these genes (transcription factors).

Gene  
LOOKUP:  for Organism:

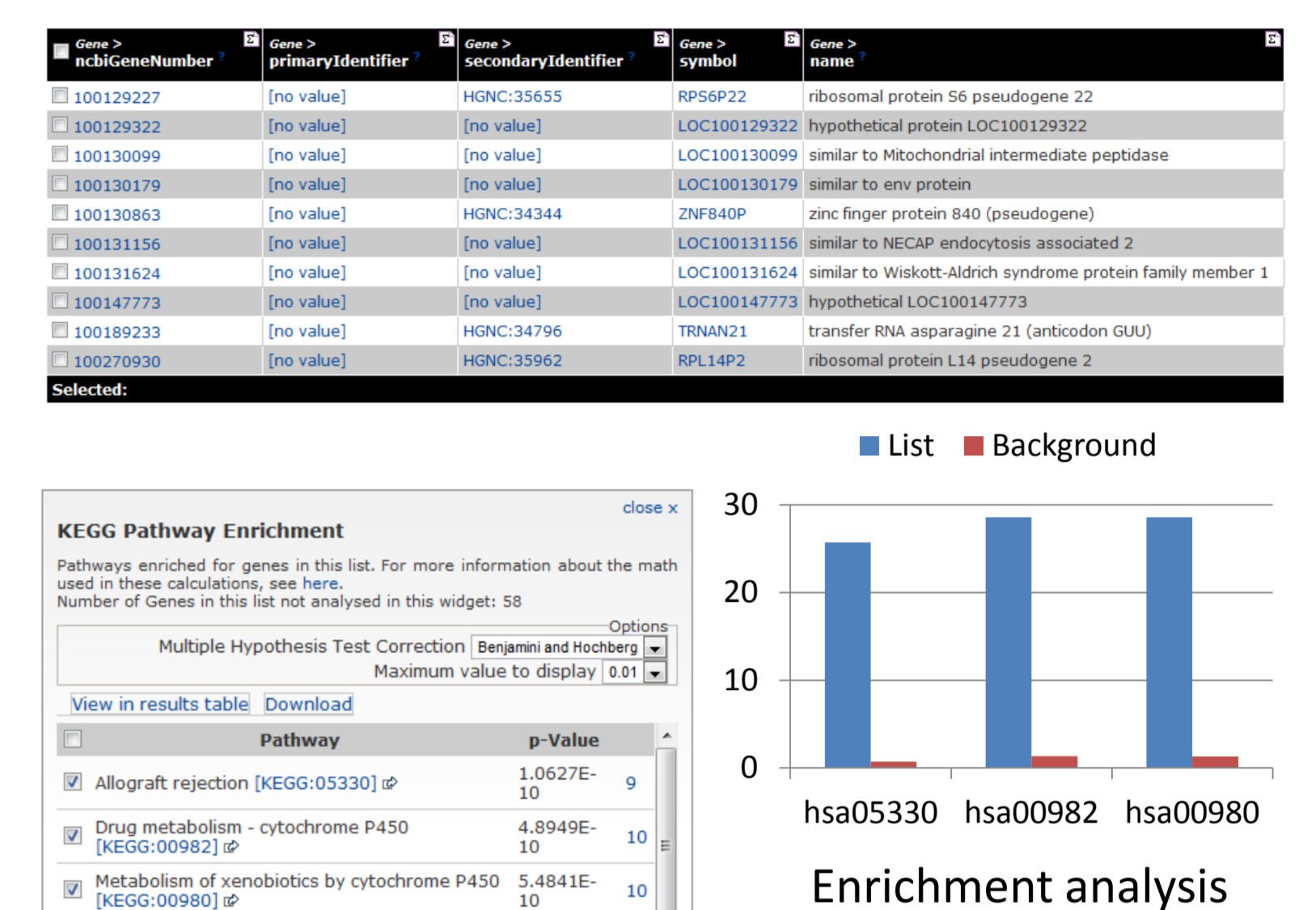
## Gene prioritisation with TargetMine

We propose a protocol for target gene prioritisation using TargetMine.

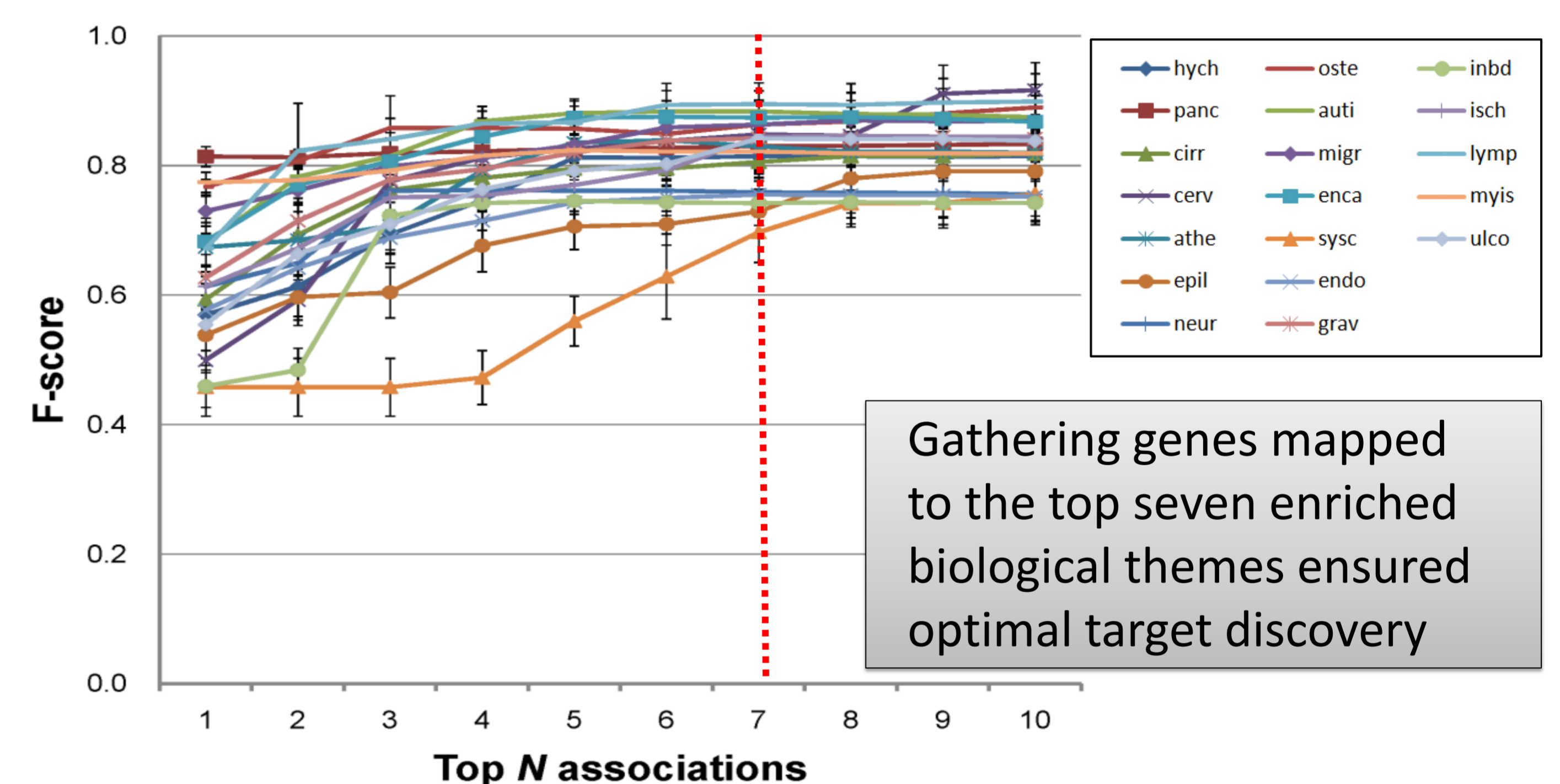
(1) Upload candidate genes to TargetMine

(2) Identify the enriched biological themes

(3) Select genes in the enriched biological themes

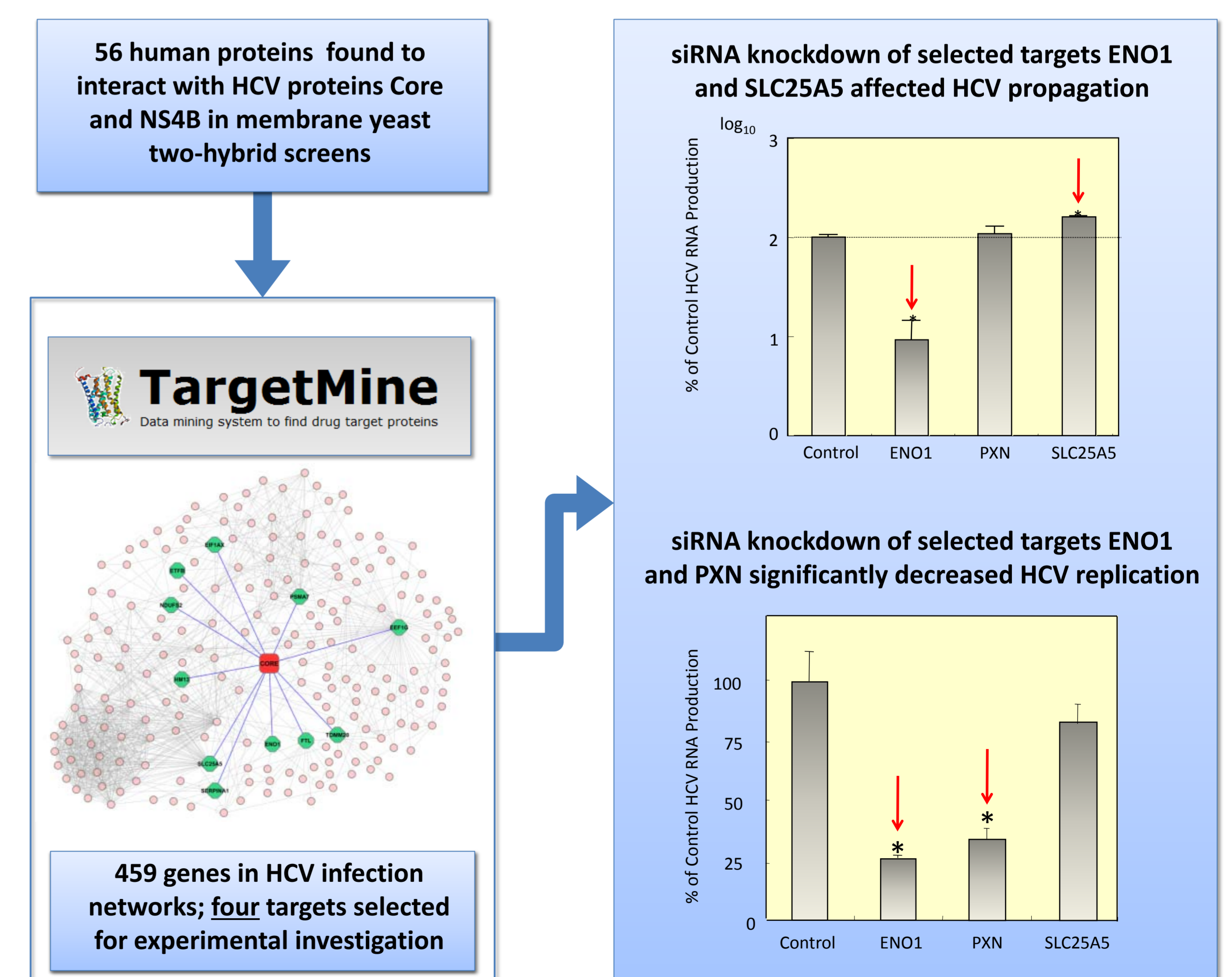


The prioritisation protocol was benchmarked on 19 sets of genes known to be associated with various human diseases.



### Application:

TargetMine has been effectively employed for target selection in the investigation of hepatitis C virus (HCV) pathogenesis<sup>#</sup>.



## Conclusions

TargetMine is an integrated data warehouse that enables complicated searches that are difficult to perform using existing comparable tools and therefore, assists in efficient target prioritisation.

<sup>†</sup>Chen YA, Tripathi LP, Mizuguchi K: TargetMine, an integrated data warehouse for candidate gene prioritisation and target discovery. *PLoS One* 6: e17844, 2011.

<sup>#</sup>Tripathi LP, Kataoka C, Taguwa S, Moriishi K, Mori Y, Matsuura Y, Mizuguchi K: Network based analysis of Hepatitis C virus Core and NS4B protein interactions. *Mol Biosyst* 6, 2539-53, 2010