Gene Information eXtension (GIX): effortless retrieval of gene product information on any website

James D.R. Knight[1], Payman Samavarchi-Tehrani[1], Mike Tyers[2] and Anne-Claude Gingras[1,3*]

[1] Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada

[2] Institute for Research in Immunology and Cancer, Université de Montréal, Montréal, Canada

[3] Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

***Corresponding author**:

Anne-Claude Gingras

Lunenfeld-Tanenbaum Research Institute at Mount Sinai Hospital

600 University Ave, Rm 992A

Toronto, ON M5G 1X5

CANADA

tel: (416) 586-5027

fax: (416) 586-8869

gingras@lunenfeld.ca

Abstract

Retrieving information about genes and gene products is a constant and time-consuming aspect of systems biology research. While there are many high quality and well-designed resources to fulfill this need, they require the user to navigate to different and often complex websites, execute a search, select the desired result and then view retrieved information. This task can be a repetitive, burdensome and disruptive process, for example when exploring the results of large scale genomics or proteomics screens or reading an online article. To address this issue we have developed a browser extension for Google Chrome and Mozilla Firefox called GIX, for Gene Information eXtension, that allows users to retrieve customizable gene product information - especially as it relates to proteins and their expression and functions - directly on a website without having to navigate to another page.

Retrieving information about genes and gene products is a constant and time-consuming aspect of systems biology research. While there are many high quality and well-designed resources to fulfill this need, they require the user to navigate to different and often complex websites, execute a search, select the desired result and then view retrieved information. This task can be a repetitive, burdensome and disruptive process, for example when exploring the results of large scale genomics or proteomics screens or reading an online article.

To address this issue we have developed a browser extension for Google Chrome and Mozilla Firefox called GIX, for Gene Information eXtension. GIX allows users to retrieve customizable gene product information – especially as it relates to proteins and their expression and functions – directly on a website without having to navigate to another page. Simply double-clicking (or alternatively, mouse dragging) on a gene name or supported accession number (Ensembl, Entrez, neXtProt, RefSeq or UniProt) will open an information panel on the current page (**Fig. 1**). This panel includes gene synonyms, the full gene name, alternative names, the size and molecular weight of its canonical protein product, the UniProt description, protein domains and regions, GO terms, protein localization, RNA tissue expression, associated diseases or phenotypes, pathways, protein interactors (from BioGRID and IntAct) and links to external resources (Ensembl, NCBI and Uniprot for all species, and organism-specific databases: dictyBase, FlyBase, MGI, neXtProt, PomBase, SGD, TAIR, WormBase, Xenbase and ZFIN). Gene Info also offers an alternative tooltip mode that simply provides links to these external resources. The extension is fully customizable, allowing the user to control the

information they see, and supports queries for *Homo sapiens* and ten model organisms: *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Danio rerio*, *Dictyostelium discoideum*, *Drosophila melanogaster*, *Gallus gallus*, *Mus musculus*, *Saccharomyces cerevisiae, Schizosaccharomyces pombe* and *Xenopus laevis*. The extension provides a search bar for entering queries manually and an online "workspace" for pasting gene lists from desktop applications for quick queries with GIX. While double-clicking to retrieve results is not possible on websites with embedded content such as Google Docs or PDFs, querying with the search bar does work on such webpages.

GIX collates data from BioGRID (https://thebiogrid.org[1]), Compartments (https://compartments.jensenlab.org[2]), GO (http://www.geneontology.org[3, 4]), HUGO Gene Nomenclature Committee (https://www.genenames.org[5]), Human Protein Atlas (https://www.proteinatlas.org[6, 7]), IntAct (https://www.ebi.ac.uk/intact[8]), OMIM (https://www.omim.org[9]), Pfam (https://pfam.xfam.org[10]), Reactome (https://reactome.org/)[11] and UniProt (https://www.uniprot.org[12]). The GIX database is updated monthly to incorporate changes from these resources. GIX is available for free without restriction at the Chrome Web Store and the Firefox Add-on site. Download links, documentation, a tutorial video and source code can be found at https://gene-info.org.

ACKNOWLEDGEMENTS

We are grateful to all members of the Gingras lab for feedback on the extension. We acknowledge funding from the Governments of Canada and Ontario through Genome Canada, Ontario Genomics and the Ontario Research Fund (OGI-139 to A.-C.G. and RE08-065 to A.-C.G.), the Canadian Institutes of Health Research (Foundation grant FDN143301 to A.-C.G) and the National Institutes of Health Office of Research Infrastructure Programs (R01OD010929 to M.T. and K. Dolinski). This research was enabled in part by support provided by Compute Canada (www.computecanada.ca). A.-C.G. is the Canada Research Chair in Functional Proteomics and the Lea Reichmann Chair in Cancer Proteomics; M.T. is the Canada Research Chair in Systems and Synthetic Biology.

AUTHOR CONTRIBUTIONS

J.D.R.K., P.S.T. and A.-C.G. conceived of the extension. J.D.R.K. wrote the code. M.T. provided input on the extension. J.D.R.K. and A.-C.G. wrote the manuscript with input from P.S.T. and M.T.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

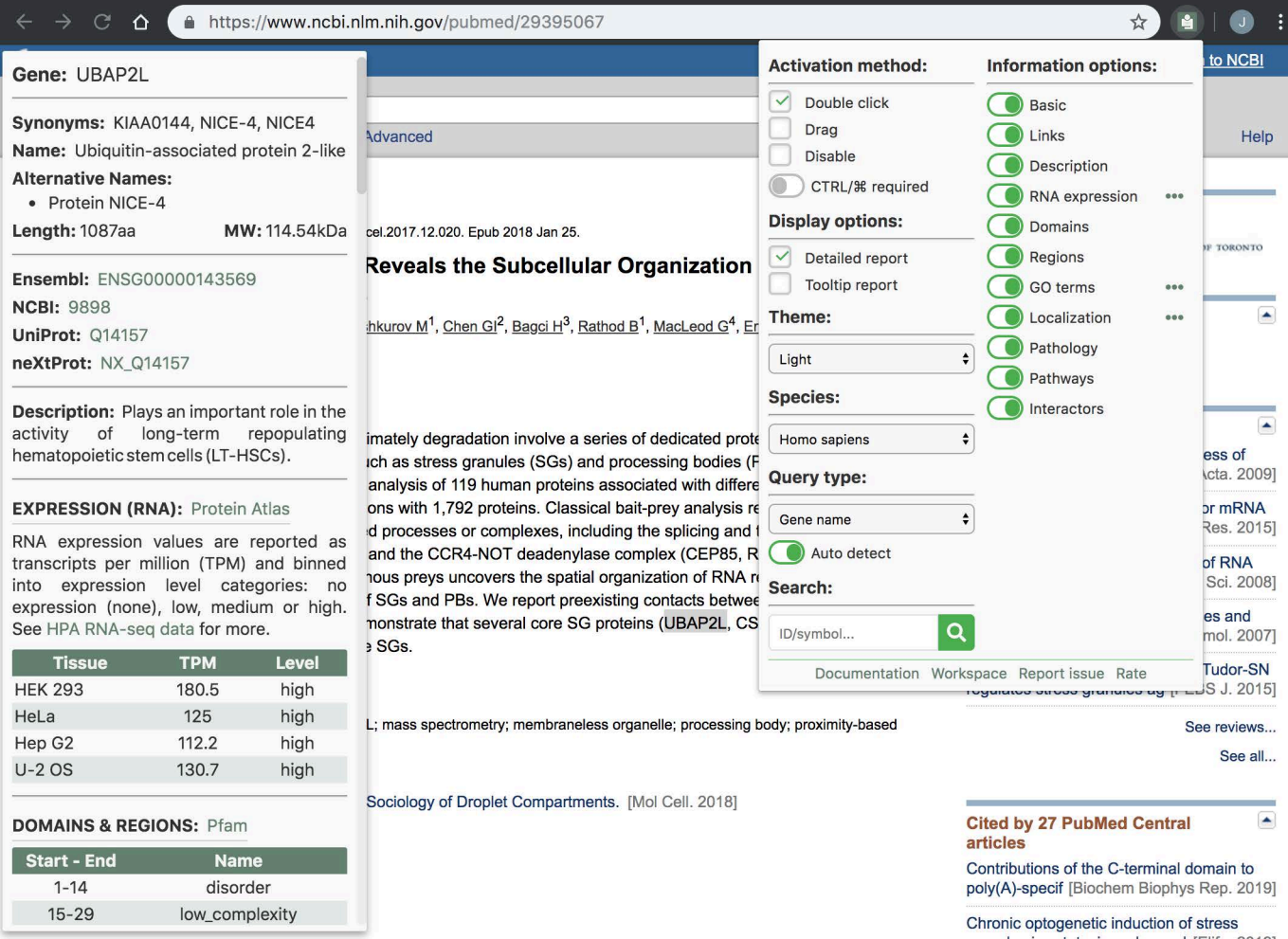
REFERENCES

1. Oughtred, R. et al. The BioGRID interaction database: 2019 update. *Nucleic Acids Res* **47**, D529-D541 (2019).

2. Binder, J.X. et al. COMPARTMENTS: unification and visualization of protein subcellular localization evidence. *Database (Oxford)* **2014**, bau012 (2014).
3. The Gene Ontology, C. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res* **45**, D331-D338 (2017).
4. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).
5. Yates, B. et al. Genenames.org: the HGNC and VGNC resources in 2017. *Nucleic Acids Res* **45**, D619-D625 (2017).
6. Thul, P.J. et al. A subcellular map of the human proteome. *Science* **356** (2017).
7. Uhlen, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
8. Orchard, S. et al. The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* **42**, D358-363 (2014).
9. Amberger, J.S., Bocchini, C.A., Scott, A.F. & Hamosh, A. OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res* **47**, D1038-D1043 (2019).
10. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res* **47**, D427-D432 (2019).
11. Fabregat, A. et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res* **46**, D649-D655 (2018).
12. UniProt Consortium, T. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **46**, 2699 (2018).

FIGURE LEGENDS

Figure 1. Screenshot of GIX in the Chrome browser. Double-clicking on a gene name (in this case UBAP2L) opens an information panel (left side) displaying information about the query. The extension has a number of settings that can be customized by clicking on its icon in the browser toolbar (right side).


https://www.ncbi.nlm.nih.gov/pubmed/29395067
to NCBI
Advanced
Help
cel.2017.12.020. Epub 2018 Jan 25.
Reveals the Subcellular Organization
hkurov M[1], Chen GI[2], Bagci H[3], Rathod B[1], MacLeod G[4], En
imately degradation involve a series of dedicated prote
ich as stress granules (SGs) and processing bodies (P
analysis of 119 human proteins associated with differe
ons with 1,792 proteins. Classical bait-prey analysis re
d processes or complexes, including the splicing and t
and the CCR4-NOT deadenylase complex (CEP85, R
hous preys uncovers the spatial organization of RNA r
f SGs and PBs. We report preexisting contacts betwee
monstrate that several core SG proteins (UBAP2L, CS
e SGs.
L; mass spectrometry; membraneless organelle; processing body; proximity-based
Sociology of Droplet Compartments. [Mol Cell. 2018]
Gene: UBAP2L
Synonyms: KIAA0144, NICE-4, NICE4
Name: Ubiquitin-associated protein 2-like
Alternative Names:
• Protein NICE-4
Length: 1087aa MW: 114.54kDa
Ensembl: ENSG00000143569
NCBI: 9898
UniProt: Q14157
neXtProt: NX_Q14157
Description: Plays an important role in the activity of long-term repopulating hematopoietic stem cells (LT-HSCs).
EXPRESSION (RNA): Protein Atlas
RNA expression values are reported as transcripts per million (TPM) and binned into expression level categories: no expression (none), low, medium or high. See HPA RNA-seq data for more.
Tissue TPM Level
HEK 293 180.5 high
HeLa 125 high
Hep G2 112.2 high
U-2 OS 130.7 high
DOMAINS & REGIONS: Pfam
Start - End Name
1-14 disorder
15-29 low_complexity
Activation method:
Double click
Drag
Disable
CTRL/⌘ required
Display options:
Detailed report
Tooltip report
Theme:
Light
Species:
Homo sapiens
Query type:
Gene name
Auto detect
Search:
ID/symbol...
Documentation Workspace Report issue Rate
Information options:
Basic
Links
Description
RNA expression
Domains
Regions
GO terms
Localization
Pathology
Pathways
Interactors
ess of
Acta. 2009]
or mRNA
Res. 2015]
of RNA
Sci. 2008]
es and
mol. 2007]
Tudor-SN
S J. 2015]
See reviews...
See all...
Cited by 27 PubMed Central articles
Contributions of the C-terminal domain to poly(A)-specif [Biochem Biophys Rep. 2019]
Chronic optogenetic induction of stress