Gene Info: A tool for facile 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

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 its information. This 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 Chrome browser 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. 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, 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, 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 select which type of information they would like to see, and supports queries for *Homo sapiens* and nine model organisms: *Arabidopsis thaliana*,

Caenorhabditis elegans, *Danio rerio*, *Dictyostelium discoideum*, *Drosophila melanogaster*, *Gallus gallus*, *Mus musculus*, *Saccharomyces cerevisiae* and *Xenopus laevis*. The extension also comes equipped with a search bar for entering queries manually if desired. 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.

The Gene Info extension collates data from BioGRID (https://thebiogrid.org[1]), Compartments (https://compartments.jensenlab.org[2]), GO (http://www.geneontology.org[3, 4]), Human Protein Atlas (https://www.proteinatlas.org[5, 6]), IntAct (https://www.ebi.ac.uk/intact[7]), Pfam (https://pfam.xfam.org[8]) and UniProt (https://www.uniprot.org[9]). The database is updated monthly to incorporate changes from these resources. Gene Info is available for free without restriction at the Chrome Web Store. The download link, documentation, a tutorial video and source code can be found at https://gene-info.lunenfeld.ca

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). 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. Stark, C. et al. *Nucleic Acids Res* **34**, D535-539 (2006).
2. Binder, J.X. et al. *Database (Oxford)* **2014**, bau012 (2014).
3. The Gene Ontology, C. *Nucleic Acids Res* **45**, D331-D338 (2017).
4. Ashburner, M. et al. *Nat Genet* **25**, 25-29 (2000).
5. Thul, P.J. et al. *Science* **356** (2017).
6. Uhlen, M. et al. *Science* **347**, 1260419 (2015).

7. Orchard, S. et al. *Nucleic Acids Res* **42**, D358-363 (2014).
8. El-Gebali, S. et al. *Nucleic Acids Res* **47**, D427-D432 (2019).
9. UniProt Consortium, T. *Nucleic Acids Res* **46**, 2699 (2018).

FIGURE LEGENDS

Figure 1. Screenshot of Gene Info Chrome extension. 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

Gene: UBAP2L
Synonyms: KIAA0144, NICE4
Name: Ubiquitin-associated protein 2-like
Alternative Names:
- Protein NICE-4

Length: 1087aa **MW:** 114.54kDa
Ensembl: ENSG00000143569
NCBI: 9898
neXtProt: NX_Q14157
UniProt: Q14157
Description: Plays an important role in the activity of long-term repopulating hematopoietic stem cells (LT-HSCs).

DOMAINS: Pfam

Start - End	Name
495-526	DUF3697

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

Activation method:
- [x] Double click
- [ ] Drag
- [ ] Disable

Display options:
- [x] Detailed report
- [ ] Tooltip report

Species: Homo sapiens

Query type: Gene name

Auto detect

Search: Accession/ID/name...

Information options:
- Basic
- Description
- Domain
- GO terms
- Interactors
- Links
- Localization
- RNA expression

About Documentation Report issue Rate

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 regulatory structures and
f SGs and PBs. We report preexisting contacts between most core SG
monstrate that several core SG proteins (UBAP2L, CSDE1, and PRRC2C) are
e SGs.

L; mass spectrometry; membraneless organelle; processing body; proximity-based

Sociology of Droplet Compartments. [Mol Cell. 2018]

nked FUS-
S A. 2018]

Human MARF1 is an endoribonuclease that interacts with the I [Nucleic Acids Res. 2018]

Review Post-transcriptional Regulatory Functions of Mammali [Trends Genet. 2018]

See all...

Related information

Articles frequently viewed together

Gene

MedGen