**PGDB: Generation of ENSEMBL-based proteogenomics databases**

Husen M. Umera, Yafeng Zhub, Janne Lehtiöa, Rui Brancaa,\* and Yasset Perez-Riverolc,\*

a Department of Oncology‐Pathology, Science for Life Laboratory, Karolinska Institutet, Solna, Stockholm 17121, Sweden

b Department of Genetics, Harvard Medical School, Boston, MA 02115 USA

c European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton,, Cambridge, CB10 1SD, UK

# Abstract

# Introduction

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Tool | URL | Portability | Downloader | File format supported | Translation events supported | Decoy generation |
| customProDB[[1]](#footnote-1) | <http://bioconductor.org/packages/devel/bioc/html/customProDB.html> | * Bioconductor * Conda * BioContainers * Workflows | x | * VCF * BAM * BED | * VCF to proteins * Novel junction peptides | x |
| MutationDB and spliceDB[[2]](#footnote-2) | <http://proteomics.ucsd.edu/software-tools/cancer-proteogenomics-4/> | * Java application | x | * VCF | * VCF to proteins | x |
| pgdb and py-pgatk | <https://github.com/nf-core/pgdb>  <https://github.com/bigbio/py-pgatk> | * Bioconductor * Conda * BioContainers * Workflows | * ENSEMBL sequences (ncRNA, canonical) * ENSEMBL variants * COSMIC mutations * Comsmic cell line mutations * CBiopoertal studies | * VCF * GTF * BED * FASTA | * Transcripts to Proteins * ENSEMBL Alternative splicing * COSMIC and cBioPortal variant translations | DecoyPrat[[3]](#footnote-3) |
| QUILTS | [http://quilts.fenyolab.org](http://quilts.fenyolab.org/) | * Web tool | x | * VCF * BED |  | x |

**References**:

Wang, Xiaojing, and Bing Zhang. “CustomProDB: An R Package to Generate Customized Protein Databases from RNA-Seq Data for Proteomics Search.” *Bioinformatics (Oxford, England)* 29, no. 24 (December 15, 2013): 3235–37. https://doi.org/10.1093/bioinformatics/btt543.

Woo, Sunghee, Seong Won Cha, Gennifer Merrihew, Yupeng He, Natalie Castellana, Clark Guest, Michael MacCoss, and Vineet Bafna. “Proteogenomic Database Construction Driven from Large Scale RNA-Seq Data.” *Journal of Proteome Research* 13, no. 1 (January 3, 2014): 21–28. https://doi.org/10.1021/pr400294c.

Wright, James C., and Jyoti S. Choudhary. “DecoyPyrat: Fast Non-Redundant Hybrid Decoy Sequence Generation for Large Scale Proteomics.” *Journal of Proteomics & Bioinformatics* 9, no. 6 (June 27, 2016): 176–80. https://doi.org/10.4172/jpb.1000404.

1. Wang and Zhang, “CustomProDB.” [↑](#footnote-ref-1)
2. Woo et al., “Proteogenomic Database Construction Driven from Large Scale RNA-Seq Data.” [↑](#footnote-ref-2)
3. Wright and Choudhary, “DecoyPyrat.” [↑](#footnote-ref-3)