# TVTB: The VCF Tool Box

Kevin Rue-Albrecht [1,2]

1 MRC WIMM Centre for Computational Biology, MRC Weatherall Institute of Molecular Medicine,
University of Oxford, Oxford, UK **2** Department of Medicine, Imperial College London, Hammersmith
Campus, United Kingdom

Summary

TVTB is an R/Bioconductor package that offers a toolkit for the filtering, summarisation, and
visualisation of genetic variation data stored in Variant Call Format (VCF) files pre-processed
by the Ensembl Variant Effect Predictor (VEP). In particular, TVTB extends core Bioconductor
infrastructure in the `VariantAnnotation` and `S4Vectors` packages to define classes of filtering
rules applicable to the diverse fields of information recorded in the VCF format. An interactive
web-application, the Shiny Variant Explorer, provides an interface to demonstrate the package
functionality in a programming-free environment.

Statement of need

The Variant Call Format (VCF) provides infrastructure for storing genetic variation data,
including core information such as position, reference, and alternate alleles, alongside optional
information such as consequences predicted by the Ensembl Variant Effect Predictor (McLaren
et al., 2016).

Downstream analyses involve iterative filtering

Earlier Bioconductor infrastructure provide functionality for efficiently reading and writing VCF
files (Obenchain et al., 2014), as well as defining classes of rules re-usable across data sets
(Pagès et al., 2024).

... including single nucleotide polymorphisms (SNPs), insertions, deletions, and structural
variants ...

Computational analysis ... exploration ... visualisation ...

The plethora of information stored in Variant Call Format (VCF) files

Analyses of genetic variation data produce a plethora of information

Gala is an Astropy-affiliated Python package for galactic dynamics. Python enables wrap-
ping low-level languages (e.g., C) for speed without losing flexibility or ease-of-use in the
user-interface. The API for Gala was designed to provide a class-based and user-friendly
interface to fast (C or Cython-optimized) implementations of common operations such as
gravitational potential and force evaluation, orbit integration, dynamical transformations, and
chaos indicators for nonlinear dynamics. Gala also relies heavily on and interfaces well with
the implementations of physical units and astronomical coordinate systems in the `Astropy`
package (**?**) (`astropy.units` and `astropy.coordinates`).

Gala was designed to be used by both astronomical researchers and by students in courses
on gravitational dynamics or astronomy. It has already been used in a number of scientific
publications (**?**) and has also been used in graduate courses on Galactic dynamics to, e.g.,
provide interactive visualizations of textbook material (**?**). The combination of speed, design,

and support for Astropy functionality in `Gala` will enable exciting scientific explorations of forthcoming data releases from the *Gaia* mission (**?**) by students and experts alike.

Mathematics

Single dollars ($) are required for inline mathematics e.g. $f(x) = e^{\pi/x}$

Double dollars make self-standing equations:

$$\Theta(x) = \left\{ \begin{array}{l} 0 \text{ if } x < 0 \\ 1 \text{ else} \end{array} \right.$$

You can also use plain LaTeXfor equations

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(x)e^{i\omega x}dx \tag{1}$$

and refer to Equation 1 from text.

Citations

Citations to entries in paper.bib should be in rMarkdown format.

If you want to cite a software repository URL (e.g. something on GitHub without a preferred citation) then you can do it with the example BibTeX entry below for (**?**).

For a quick reference, the following citation commands can be used: - `@author:2001` -> "Author et al. (2001)" - `[@author:2001]` -> "(Author et al., 2001)" - `[@author1:2001;` `@author2:2001]` -> "(Author1 et al., 2001; Author2 et al., 2002)"

Figures

Figures can be included like this: Caption for example figure. and referenced from text using section .

Figure sizes can be customized by adding an optional second parameter: Caption for example figure.

Acknowledgements

We acknowledge support from Prof. Martin R Wilkins during the genesis of this project.

References

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., Flicek, P., & Cunningham, F. (2016). The ensembl variant effect predictor [Journal Article]. *Genome Biol*, *17*(1), 122. https://doi.org/10.1186/s13059-016-0974-4

Obenchain, V., Lawrence, M., Carey, V., Gogarten, S., Shannon, P., & Morgan, M. (2014). VariantAnnotation: A bioconductor package for exploration and annotation of genetic variants [Journal Article]. *Bioinformatics*, *30*(14), 2076–2078. https://doi.org/10.1093/bioinformatics/btu168

Pagès, H., Lawrence, M., & Aboyoun, P. (2024). *S4Vectors: Foundation of vector-like and list-like containers in bioconductor*. https://bioconductor.org/packages/S4Vectors