

# Representation and Manipulation of Genomic Tuples in R

Peter F Hickey

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

17 May 2016

**Paper DOI:** <http://dx.doi.org/10.21105/joss.00020>

**Software Repository:** <https://github.com/PeteHaitch/GenomicTuples>

**Software Archive:** <http://dx.doi.org/10.5281/zenodo.53186>

## Summary

**GenomicTuples** is an R/Bioconductor package (R Core Team 2016; Wolfgang Huber et al 2015) that defines general purpose containers for storing and manipulating *genomic tuples*. A genomic tuple of size `m` is of the form `chromosome:strand:{pos_1, pos_2, ..., pos_m}` where `pos_1 < pos_2 < ... < pos_m` are positions along the chromosome. The difference between a genomic tuple and a genomic range/interval is like that of a difference between an ordered set and an interval. For example, the genomic 2-tuple `chr3:+:{65, 77}` differs from the genomic range `chr3:+:[65, 77]` by not including any of the intervening loci, `chr3:+:66` to `chr3:+:76`.

**GenomicTuples** aims to provide functionality for manipulating tuples of genomic co-ordinates that are analogous to those available for genomic ranges in the popular **GenomicRanges** R/Bioconductor package (Lawrence et al. 2013). To that end, the **GenomicTuples** API mimics that of **GenomicRanges**. By extending classes defined in the **GenomicRanges** package, objects from the **GenomicTuples** package may be used as drop-in replacements for objects from the **GenomicRanges** package. This ensures easy interoperability with other popular Bioconductor packages, such as **SummarizedExperiment** (Morgan et al. 2016), and the availability of common operations, such as finding overlaps between genomic tuples and genomic features of interest.

## References

- Lawrence, Michael, Wolfgang Huber, Hervé Pagès, Patrick Aboyoun, Marc Carlson, Robert Gentleman, Martin Morgan, and Vincent Carey. 2013. “Software for Computing and Annotating Genomic Ranges.” *PLoS Computational Biology* 9 (8). doi:10.1371/journal.pcbi.1003118.
- Morgan, Martin, Valerie Obenchain, Jim Hester, and Hervé Pagès. 2016. *SummarizedExperiment: SummarizedExperiment Container*.
- R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wolfgang Huber et al. 2015. “Orchestrating High-Throughput Genomic Analysis with Bioconductor.” *Nature Methods* 12 (2): 115–21. doi:10.1038/nmeth.3252.