Walter+Eliza Hall
Institute of Medical Research
DISCOVERIES FOR HUMANITY

May 8, 2018

Dear Editor,

We would like to lodge a presubmission inquiry for our article entitled *'plyranges: a grammar of genomic data transformation'* to PLoS Computational Biology for consideration as a Software article.

While there is a lot of software for analysing genomics data, they require a familiarity with a multitude of UNIX tools and proficiency with a several programming languages. Our software, *plyranges*, provides a coherent and consistent interface for analysing genomic data with the beginner user in mind.

plyranges develops a grammar for transforming genomics data with the aim of facilitating interactive and reproducible data analysis. Our software contributes a relational genomics algebra for reasoning about data measured on genomic regions. As a consequence of our grammar, *plyranges* can be used to explore data from a wide variety of sequencing assays.

Our software is an open source R/Bioconductor package that has recently been accepted for the next release of Bioconductor (available: https://bioconductor.org/packages/release/bioc/html/plyranges.html).

This means that it has gone under extensive code review and can be run on Linux, Windows and Mac operating systems. As it is a Bioconductor package, *plyranges* is fully interoperable with other packages in that ecosystem.

We believe that our manuscript will be of much interest to your readers, especially those interested in reproducible and accessible computational biology workflows.

If you have any questions about this work, please feel free to contact me.

Yours sincerely,

Stuart Lee

PhD Candidate - Monash Econometrics and Business Statistics

Walter and Eliza Hall Institute

Email: lee.s@wehi.edu.au

Presubmission Inquiry Checklist

1. The software name and the address of the repository used for the software: *plyranges* is the name of our package. The development version is available at https://github.com/sa-lee/plyranges and the stable release version is available at https://bioconductor.org/packages/bioc/release/html/plyranges.html
2. The license used for the software: Artistic 2.0
3. A description of the software input, output, and the method(s) it implements, including citations for previous publications of the method or software. Input: BAM, BED, BigWig, WIG, narrowPeak Methods included in the pipeline: statistical summaries, restrictions, genomic arithmetic, overlap and nearest neighbour finding. Output: same as input except for BAM
4. Between two and four keywords: Data Analysis, Bioconductor, Genomics, Grammar
5. The number of users who tested the software outside of the authors group. The software has undergone extensive code review at Bioconductor and users have reported bugs at our issues page. We will have more accurate estimates of user numbers once the next Bioconductor version is released.
6. The number of inputs analysed during testing. Our software has high test coverage and has an extensive test suite that checks plyranges can read, process and write genomics data from a variety of formats.
7. The number of examples (input files) provided with the software. Most functions in the package have examples, there is a also an extensive vignette detailing how to use *plyranges*.
8. Any other similar or related methods (with links and citations). The *plyranges* grammar is built on top of the *GenomicRanges* package (doi = 10.1371/journal.pcbi.1003118). Another closely related command line tool is BEDtools (doi = 10.1093/bioinformatics/btq033).
9. *A description of continuous integration of the software, a bug reporting URL (such as bugzilla or github issues), and a mailing list or chat rooms, if these things exist.* The github repository has continuous integration handled via travis. Continuous integration is also handled by the Bioconductor projects build system. Support for users is available at the github repository or via the Bioconductor support site (https://support.bioconductor.org).