**Software paper for submission to MEDSCI 736 University of Auckland**

To complete this template, please replace the blue text with your own. The paper has three main sections: (1) Overview; (2) Availability; (3) Reuse potential.

Please submit the completed paper to: f.kubke@auckland.ac.nz

**(1) Overview**

Title

“ePygenetics: An extraction tool for epigenetic data”.

Paper Authors

1. *Chalmers, Callum*

2. *O’Sullivan, Justin*

*3. Pireva, Kreshnik*

Paper Author Roles and Affiliations

1. Development Team, Faculty of Medical and Health Sciences UoA

2. Product Owner, Liggins Institute UoA

3. Scrum Master, UoA

Abstract

ePygenetics is a python module that retrieves epigenetic data from .wig files. It allows the user to input the data they require from the files, finds the data within the file and outputs the data to a database. This allows geneticists to easily extract the epigenetic status of a single nucleotide polymorphism (SNP) in a particular cell line so this can be considered in any analyses. It is licensed under the Attribution-ShareAlike 3.0 New Zealand license and the source code is freely available for use and modification at: <https://github.com/UOA-MEDSCI-736/CallumChalmers29-crispy-disco>.

Keywords

*Software, Python, Genetics, Epigenetics, SNPs*

Introduction

An overview of the software, how it was produced, and the research for which it has been used, including references to relevant research articles. A short comparison with software which implements similar functionality should be included in this section. Please refer to <http://openresearchsoftware.metajnl.com/articles/> for exemplars.

ePygenetics is a python module that allows the user to extract epigenetic data from .wig files. The module was created to allow geneticists to quickly and easily access the epigenetic status of single nucleotide polymorphisms (SNPs) from DNAse hypersensitivity data.

Currently, epigenetic data is not considered in most genome-wide association studies, potentially leading to a reduction in the power of this method [1]. While .wig files can be viewed in programmes such as the UCSC Genome Browser, this software is not effective at looking at multiple SNPs at once or comparing multiple cell lines. ePygenetics can perform both these functions by exporting the data as a filterable database which can be easily viewed or integrated into another programme. This allows epigenetic status to be considered when doing genetic analyses, potentially providing the missing piece of the puzzle.

**Implementation and architecture**

How the software was implemented, with details of the architecture where relevant. Use of relevant diagrams is appropriate. Please also describe any variants and associated implementation differences.

**Quality control**

Detail the level of testing that has been carried out on the code (e.g. unit, functional, load etc.), and in which environments. If not already included in the software documentation, provide details of how a user could quickly understand if the software is working (e.g. providing examples of running the software with sample input and output data).

**(2) Availability**

***Operating system***

Please include minimum version compatibility.

***Programming language***

Please include minimum version compatibility.

***Additional system requirements***

E.g. memory, disk space, processor, input devices, output devices.

***Dependencies***

E.g. libraries, frameworks, incl. minimum version compatibility.

***List of contributors***

Please list anyone who helped to create the software (who may also not be an author of this paper), including their roles and affiliations.

***Software location:***

***Archive*** (e.g. institutional repository, general repository) (required – please see instructions on journal website for depositing archive copy of software in a suitable repository)

***Name:*** The name of the archive

***Persistent identifier:*** e.g. DOI, handle, PURL, etc.

***Licence:*** Open license under which the software is licensed

***Publisher:*** Name of the person who deposited the software

***Version published:***The version number of the software archived

***Date published:*** dd/mm/yy

**Code repository** (e.g. SourceForge, GitHub etc.) (required)

***Name:*** The name of the code repository

***Identifier:*** The identifier (or URI) used by the repository

***Licence:*** Open license under which the software is licensed

***Date published:*** dd/mm/yy

**Emulation environment** (if appropriate)

***Name:*** The name of the emulation environment

***Identifier:*** The identifier (or URI) used by the emulator

***Licence:*** Open license under which the software is licensed here

***Date published:*** dd/mm/yy

***Language***

Language of repository, software and supporting files (including links to documentation)

**(3) Reuse potential**

Please describe in as much detail as possible the ways in which the software could be reused by other researchers both within and outside of your field. This should include the use cases for the software, and also details of how the software might be modified or extended (including how contributors should contact you) if appropriate. Also you must include details of what support mechanisms are in place for this software (even if there is no support).

**Acknowledgements**

Please add any relevant acknowledgements to anyone else who supported the project in which the software was created, but did not work directly on the software itself.

**Funding statement**

If the software resulted from funded research please give the funder and grant number.

**Competing interests**

If any of the authors have any competing interests then these must be declared. The authors’ initials should be used to denote differing competing interests. For example: “BH has minority shares in [company name], which part funded the research grant for this project. All other authors have no competing interests."

If there are no competing interests, please add the statement:

“The authors declare that they have no competing interests.”

**References**

Please enter references in the Harvard style and include a DOI where available, citing them in the text with a number in square brackets, e.g.

[1] Hardison, R.C., 2012. Genome-wide epigenetic data facilitate understanding of disease susceptibility association studies. *Journal of Biological Chemistry*, *287*(37), pp.30932-30940. doi: [10.1074/jbc.R112.352427](https://dx.doi.org/10.1074%2Fjbc.R112.352427)

**Copyright Notice**

Authors who publish with this journal agree to the following terms:

Authors retain copyright with the work simultaneously licensed under a [Creative Commons Attribution License](http://creativecommons.org/licenses/by/3.0/) that allows others to share the work with an acknowledgement of the work's authorship and initial publication in this journal.

Authors are able to enter into separate, additional contractual arrangements for the non-exclusive distribution of the journal's published version of the work (e.g., post it to an institutional repository or publish it in a book), with an acknowledgement of its initial publication in this journal.

By submitting this paper you agree to the terms of this Copyright Notice, which will apply to this submission if and when it is published by this journal.