

Corresponding Author:
Dr Andrew W. George
Data61, CSIRO
Ecosciences Precinct
41 Boggo Road, Qld 4102, Australia
Tel: +61 7 3833 5543
Email: andrew.george@csiro.au

16/10/18

Re: Submission of manuscript to *Nature Genetics*

Dear Editors,

We would like to submit our manuscript entitled “Eagle: Making multi-locus association mapping on a genome-wide scale routine” for consideration in *Nature Genetics*. Our goal is to transform the way in which genome-wide association study (GWAS) data are being analyzed. We aim to shift the animal, plant, and human genetic communities from using single-locus (or locus-by-locus) association mapping to more powerful multi-locus association mapping.

In this year alone, over 1400 papers, 19 in *Nature Genetics*, have been published on GWASs (Web of Science). Of these, only 21 featured multi-locus methods. Single-locus methods are by far the predominant method-of-choice when analyzing GWAS data. We believe the poor uptake of multi-locus association mapping is due to a couple of problems. First, multi-locus methods do not scale well, computationally, with study size. Second, many of the multi-locus implementations have not moved beyond prototype code, making them difficult to use in practice. We found the multi-locus programs/packages to be poorly documented, perform almost no error checking, provide a user with little feedback when things don’t work, and require a user to have a high level of computing prowess. With *Eagle*, we have solved these problems.

In our paper, through a simulation study, we show that the *Eagle* method has greater power to find true SNP-trait associations while rejecting false associations than competing single-locus and multi-locus methods. It also has a run time that is equal to the fastest single-locus association mapping program. *Eagle*’s increased power over single-locus association mapping was most evident when we analyzed real data from a published mouse study (Nicode *et al.* 2016, *Nature Genetics*). *Eagle* found over two times more (true) results and no false results than the competing (single-locus) method-of-choice.

We believe the greatest impact of this work will come from our computer implementation of the *Eagle* method. We have implemented *Eagle* as an R package of the same name. However, it is very different to other R implementations. We have specifically designed *Eagle* to harness multi-threaded computation. *Eagle* can handle data larger than the memory capacity of a computer. To use *Eagle*, a knowledge of R

is not required. *Eagle* comes with a GUI for easy use. We have also built a website (<http://eagle.r-forge.r-project.org>) to accompany the *Eagle* package. Here, users can find videos, “Quick Start” guides, tutorials, answers to FAQs, installation instructions, and even a link to our server where users can “try out” *Eagle* on some test data before installing.

We anticipate that this work will impact a large proportion of the readership of *Nature Genetics*. We have developed the statistical methodology and implemented it to the highest standard. For the first time, geneticists have access to a multi-locus association mapping tool that makes the analysis of GWAS data routine.

Yours sincerely,

Andrew George
CSIRO