

Matthew Lyon  
NIHR Bristol Biomedical Research Centre &  
MRC Integrative Epidemiology Unit  
Population Health Sciences, University of Bristol  
Oakfield House, Oakfield Grove  
Bristol BS8 2BN  
Tel: +44 (0) 7851589455  
Email: [matt.lyon@bristol.ac.uk](mailto:matt.lyon@bristol.ac.uk)  
Monday, 11th May 2020

Dear Editor

We are submitting to you our letter "The variant call format provides efficient and robust storage of GWAS summary statistics".

Genome-wide association study (GWAS) summary statistics are a vital resource for a range of secondary research applications. However, no common storage format has been widely adopted hindering tool development and data sharing, analysis and integration. Existing formats are computationally inefficient and lack robust approaches to storing genetic variation and essential study metadata required for reproducible research.

To address these issues, we have developed the following resources described in our letter:

- Specification for storing GWAS summary statistics and study metadata in the variant call format (GWAS-VCF)
- Open-source software for reading, writing and integrating these data with existing analysis packages
- Simulation studies demonstrating GWAS-VCF is 9-46x faster than tabular alternatives when extracting variants by genomic position
- Full summary statistics on over 10,000 complete studies in GWAS-VCF

We feel these resources will be of interest to a diverse range of practitioners investigating genetics of common disease and therefore well placed in Nature Genetics. We appreciate your time in considering this paper.

Yours sincerely

Matthew Lyon, on behalf of all authors