


RESEARCH ARTICLE

# Evaluation and application of summary statistic imputation to discover new height-associated loci

Sina Rüeger<sup>1,2</sup>, Aaron McDaid<sup>1,2</sup>, Zoltán Kutalik<sup>1,2\*</sup>

**1** Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, 1010, Switzerland, **2** Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland

 These authors contributed equally to this work.

\* [zoltan.kutalik@unil.ch](mailto:zoltan.kutalik@unil.ch)



## Abstract

As most of the heritability of complex traits is attributed to common and low frequency genetic variants, imputing them by combining genotyping chips and large sequenced reference panels is the most cost-effective approach to discover the genetic basis of these traits. Association summary statistics from genome-wide meta-analyses are available for hundreds of traits. Updating these to ever-increasing reference panels is very cumbersome as it requires reimputation of the genetic data, rerunning the association scan, and meta-analysing the results. A much more efficient method is to directly impute the summary statistics, termed as *summary statistics imputation*, which we improved to accommodate variable sample size across SNVs. Its performance relative to *genotype imputation* and practical utility has not yet been fully investigated. To this end, we compared the two approaches on real (genotyped and imputed) data from 120K samples from the UK Biobank and show that, *genotype imputation* boasts a 3- to 5-fold lower root-mean-square error, and better distinguishes true associations from null ones: We observed the largest differences in power for variants with low minor allele frequency and low imputation quality. For fixed false positive rates of 0.001, 0.01, 0.05, using *summary statistics imputation* yielded a decrease in statistical power by 9, 43 and 35%, respectively. To test its capacity to discover novel associations, we applied *summary statistics imputation* to the GIANT height meta-analysis summary statistics covering HapMap variants, and identified 34 novel loci, 19 of which replicated using data in the UK Biobank. Additionally, we successfully replicated 55 out of the 111 variants published in an exome chip study. Our study demonstrates that *summary statistics imputation* is a very efficient and cost-effective way to identify and fine-map trait-associated loci. Moreover, the ability to impute summary statistics is important for follow-up analyses, such as Mendelian randomisation or LD-score regression.

## OPEN ACCESS

**Citation:** Rüeger S, McDaid A, Kutalik Z (2018) Evaluation and application of summary statistic imputation to discover new height-associated loci. PLoS Genet 14(5): e1007371. <https://doi.org/10.1371/journal.pgen.1007371>

**Editor:** Michael P. Epstein, Emory University, UNITED STATES

**Received:** December 1, 2017

**Accepted:** April 18, 2018

**Published:** May 21, 2018

**Copyright:** © 2018 Rüeger et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was supported by the Leenaards Foundation (<http://www.leenaards.ch>), the Swiss National Science Foundation [31003A-143914, 31003A-169929]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.