Dear Editor,

Please consider our proposal for a review tentatively titled “**A practical guide to filtering in the era of large genomic data sets.**” In this manuscript, we propose to provide the most thorough review of the approaches used for filtering genomic sequencing data and their consequences on downstream analyses to date.

Filtering is a critical step in *every* project that uses genomic data to address essentially any research or applied question. However, there are currently no clear, comprehensive guidelines for researchers who find themselves faced with the massive array of potentially confusing and project options available for filtering, many of which can have serious downstream statistical implications that can translate to substantial differences in biological inferences. Should a researcher use a minor allele frequency (MAF) filter? If so, what should the frequency cutoff be, and should that cutoff depend on the research question? For that matter, should they calculate MAFs globally, or within populations? Should they employ a Linkage Disequilibrium (LD) filter? Should low quality loci or individuals be filtered (removed) first? How many individuals should each genotype be called in to make the cut? And, most importantly, **what will the effects of these choices be on the often-critical downstream applications of their research?** As stated above,researchers currently answer these questions without comprehensive guidelines and must synthesize information from dozens of different papers, coupled with their “gut feelings,” and often do so with little understanding of the possible outcomes of their decisions.

A review of filtering is therefore timely and desperately needed by the genomics community. In this paper, we propose to provide the “go-to-guide” to filtering, where researchers can find at a glance the collected consensus of filtering approaches (when it exists) and quickly grasp what the downstream results of those choices might be. We will provide a flow chart to filtering best practices and a list of software, scripts, and/or workflows to help researchers quantify filtering effects on downstream results (including re-filtering tools and simulation tools to assess filtering effects).

We are confident this paper will be widely cited and **helpful to a broad audience** of population geneticists, ecologists, conservationists, agricultural geneticists, medical researchers, and others using genomics for research or management. Thank you for considering our review (outline attached) for publication in Nature Reviews Genetics. Given the urgent need for this review, we will quickly prepare the manuscript for review and publication.

Sincerely,

Dr. William Hemstrom, Dr. Gordon Luikart, and Dr. Mark Christie