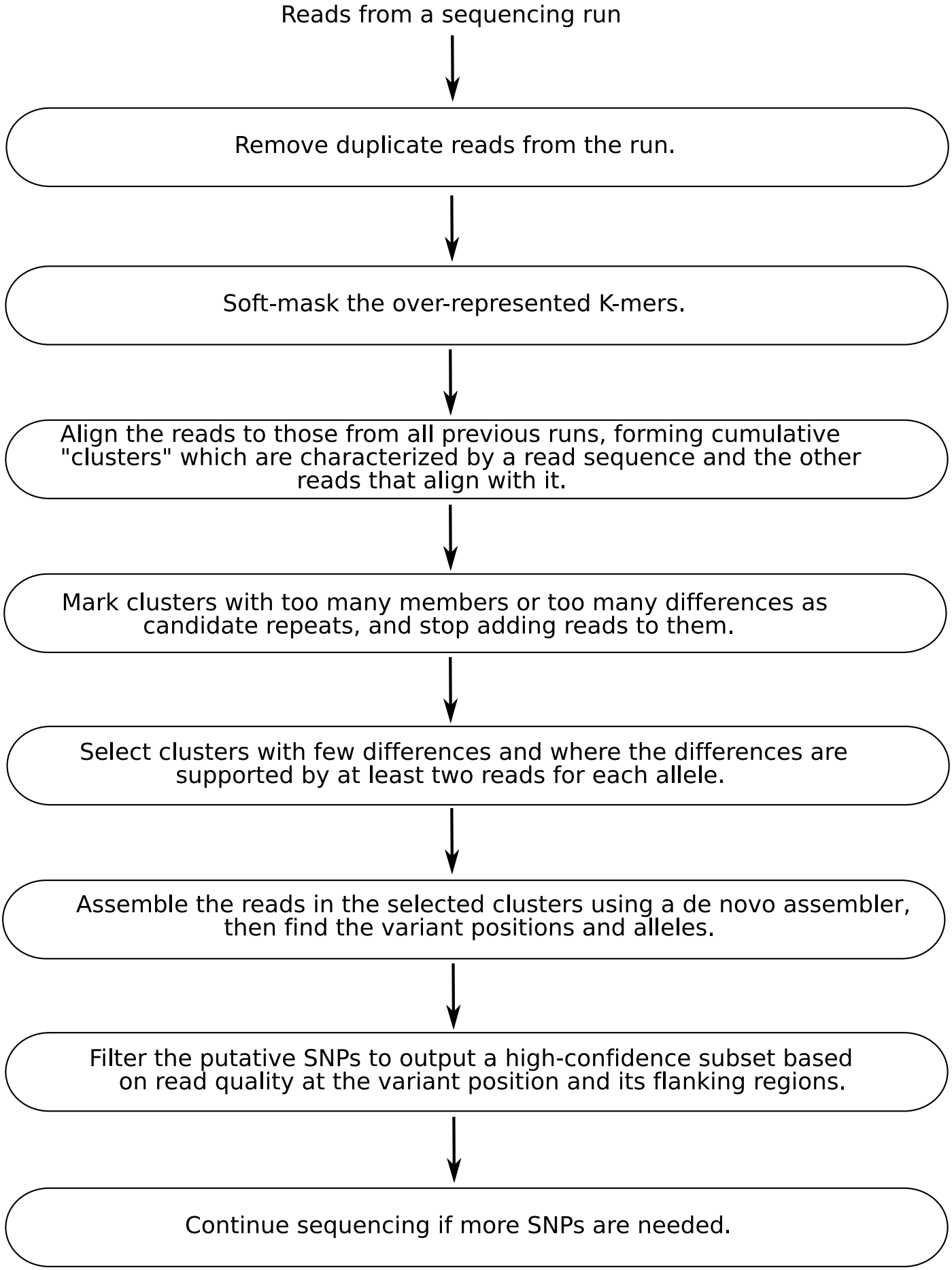
Reads from a sequencing run
Remove duplicate reads from the run.
Soft-mask the over-represented K-mers.
Align the reads to those from all previous runs, forming cumulative "clusters" which are characterized by a read sequence and the other reads that align with it.
Mark clusters with too many members or too many differences as candidate repeats, and stop adding reads to them.
Select clusters with few differences and where the differences are supported by at least two reads for each allele.
Assemble the reads in the selected clusters using a de novo assembler, then find the variant positions and alleles.
Filter the putative SNPs to output a high-confidence subset based on read quality at the variant position and its flanking regions.
Continue sequencing if more SNPs are needed.