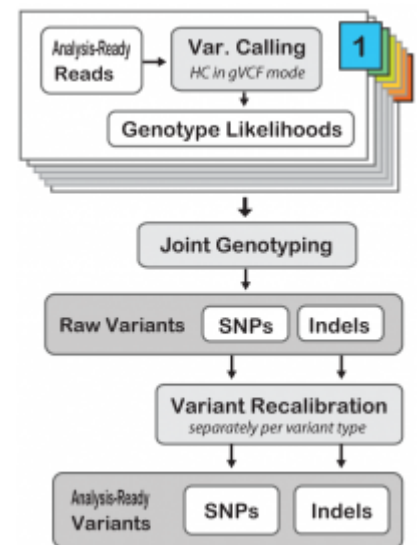


The process to discover new variants consists in several steps, as detailed in GATK's Best Practices for Variant Discovery Section [BP2.0](#).



Once you've pre-processed your data according to our recommendations, you are ready to undertake the variant discovery process, i.e. identify the sites where your data displays variation relative to the reference genome, and calculate genotypes for each sample at that site. Unfortunately some of the variation you observe is caused by mapping and sequencing artifacts, so the greatest challenge here is to balance the need for sensitivity (to minimize false negatives, i.e. failing to identify real variants) vs. specificity (to minimize false positives, i.e. failing to reject artifacts). We have found that it is very difficult to reconcile these objectives in a single step, so instead we decompose the variant discovery process into separate steps: variant calling (performed per-sample), joint genotyping (performed per-cohort) and variant filtering (also performed per-cohort). The first two steps are designed to maximize sensitivity, while the filtering step aims to deliver a level of specificity that can be customized for each project.

Notes on which tools to use The GATK includes two variant calling tools, HaplotypeCaller and UnifiedGenotyper.

The HaplotypeCaller is a more recent and sophisticated tool than the UnifiedGenotyper, and we recommend using HaplotypeCaller in all cases, with only a few exceptions (see FAQs below).

For best results, the variant filtering should be done with the Variant Quality Score Recalibration (VQSR) tools. In some cases (small datasets, non-human organisms) this is not possible and must be done by applying hard filters instead (see FAQs below).

From:  
<https://cortafuegos.fundacioace.com/wiki/> - **Detritus Wiki**

Permanent link:  
[https://cortafuegos.fundacioace.com/wiki/doku.php?id=genetica:bioinf\\_process:discovery](https://cortafuegos.fundacioace.com/wiki/doku.php?id=genetica:bioinf_process:discovery)

Last update: **2020/08/04 10:58**

