Standardised benchmarking of NGS variant calling

Andrew Bond[1], Aled Jones[1], Joo Wook Ahn[2*]
1 Genome Informatics, Genetics Laboratories, Viapath, London, 2 Clinical Bioinformatics, Guy's & St Thomas' NHS Foundation Trust, London
*joowook.ahn@nhs.net

The need for (and challenges of) standardised benchmarking

ACGS best practice guidelines[2] require that validation of a pipeline is performed using reference material such as the NIST Genome in a Bottle (GIAB) NA12878 sample to assess the performance of an NGS assay. Detected variants are compared with previously characterised variants in the reference, however, due to challenges including multiple ways of representing variants and different methods for calculating performance metrics, results may not be directly comparable between centres.

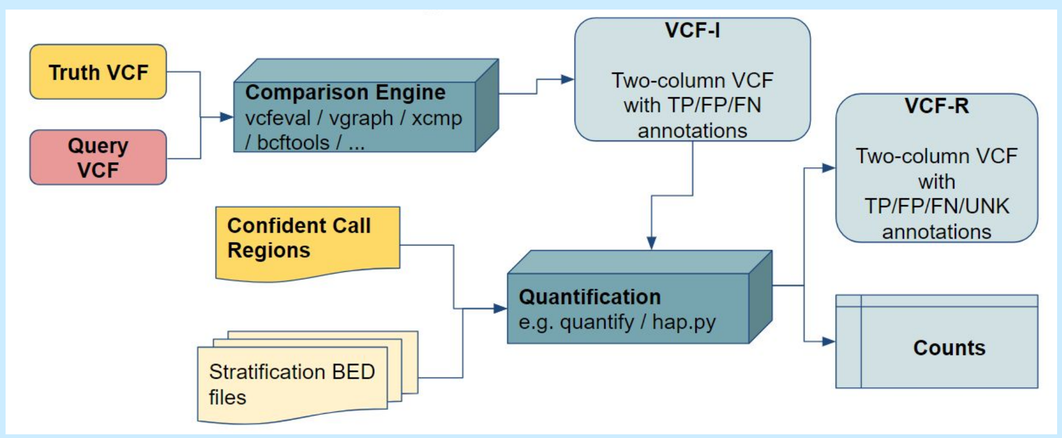
Standardised benchmarking is essential to enable development, optimisation, and demonstration of performance for sequencing and analysis tools. Global Alliance for Genomics and Health (GA4GH) Benchmarking Team have published best practice guidelines for benchmarking[1] including (figure 1):

- Use of sophisticated variant comparison tools
- Use of reference data with both high confidence variant calls as well as high confidence regions
- Reporting precision and recall with confidence intervals.

We have made available a verified instance of a benchmarking tool following GA4GH best practice recommendations.

This is publically available and can be used to compare the results of NA12878 GIAB reference sample using any panel, pipeline or sequencing technology.

Figure 1 – GA4GH suggested benchmarking tool architecture



How the Benchmarking tool follows GA4GH guidelines

vcfeval[3] is used to compare variants.
hap.py[4] is used to calculate performance metrics
The tool has been implemented on DNA Nexus cloud compute platform and has been fully verified.

Usage

The tool can be accessed at: https://genomics.viapath.co.uk/benchmark

Inputs (figure 2)

- An email address to send the results
- A VCF
- BED file (optional)

Output is an email containing:

- Recall (sensitivity) and precision (PPV) for SNP and INDEL, including 95% confidence intervals
- A link to the hap.py html report (figure 3)
- A link to download all files produced by hap.py

Limitations

Currently only supports build GRCh37 (GRCh38 support will be added in the future).
Supports only NA12878 reference sample

Summary

This tool facilitates standardised benchmarking.
Since June 2017 this tool has been used more than 200 times by multiple diagnostic genetics laboratories

Figure 2. Benchmarking tool input

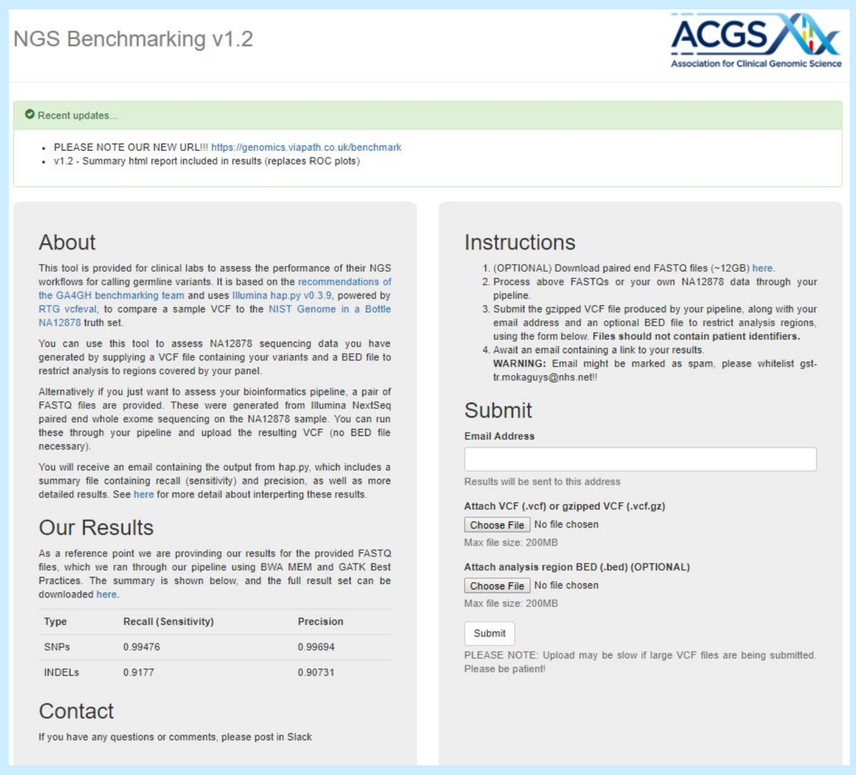


Figure 3. hap.py html report

hap.py creates a html report containing true positive, false negative and false positive counts.

Recall and PPV are plotted for SNV and INDELs.

References

1. Best Practices for Benchmarking Germline Small Variant Calls in Human Genomes. Krusche et al 2018 (biorxiv)
2. Guidelines for development and validation of software, with particular focus on bioinformatics pipelines for processing NGS data in clinical diagnostic laboratories. Whiffin, Brugger, Ahn 2016 (PeerJ)
3. https://github.com/RealTimeGenomics/rtg-tools
4. https://github.com/Illumina/hap.py