

Standardised benchmarking of NGS variant calling

Andrew Bond¹, Aled Jones¹, 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 guidelines² 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 benchmarking¹ 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

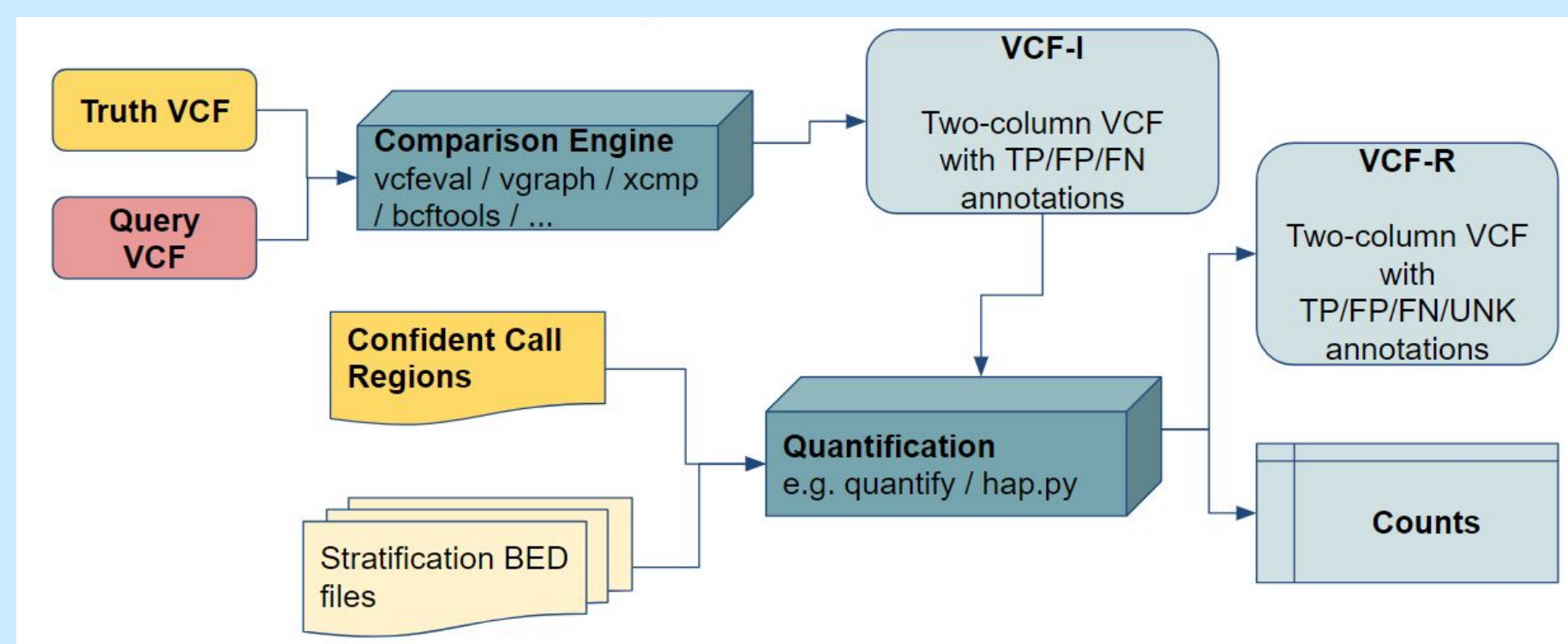


Figure 2. Benchmarking tool input

NGS Benchmarking v1.2

Recent updates.

- PLEASE NOTE OUR NEW URL!! <https://genomics.viapath.co.uk/benchmark>
- v1.2 - Summary.html report included in results (replaces ROC plots)

About

This tool is provided for clinical labs to assess the performance of their NGS workflows for calling germline variants. It is based on the recommendations of the GA4GH benchmarking team and uses Illumina HiSeq v3.0.9, powered by RTO, versioned, to compare a sample VCF to the NIST Genome in a Bottle NA12878 truth set.

You can use this VCF to assess NA12878 sequencing data you have generated by supplying a VCF file containing your variants and a BED file to restrict analysis to regions covered by your panel.

Alternatively if you just want to assess your bioinformatics pipeline, a pair of FASTQ files are provided. These were generated from Illumina HiSeq paired end whole exome sequencing on the NA12878 sample. You can run these through your pipeline and upload the resulting VCF (no BED file necessary).

You will receive an email containing the output from hap.py, which includes a summary file containing recall (sensitivity) and precision, as well as more detailed results. See [here](#) for more detail about interpreting these results.

Instructions

- (OPTIONAL) Download paired end FASTQ files (~12GB) [here](#).
- Process above FASTQs or your own NA12878 data through your pipeline.
- Submit the gipped VCF file produced by your pipeline, along with your email address and an optional BED file to restrict analysis regions, using the form below. Files should not contain patient identifiers.
- Await an email containing a link to your results.

WARNING: Email might be marked as spam, please whitelist gsit.makaguy@nhs.net!

Our Results

As a reference point we are providing our results for the provided FASTQ files, which we ran through our pipeline using BWA MEM and GATK Best Practices. The summary is shown below, and the full result set can be downloaded [here](#).

Type	Recall (Sensitivity)	Precision
SNPs	0.99476	0.99694
INDELs	0.9177	0.90731

Contact

If you have any questions or comments, please post in Slack

How the Benchmarking tool follows GA4GH guidelines

vcfeval³ is used to compare variants.

hap.py⁴ 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.viopath.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

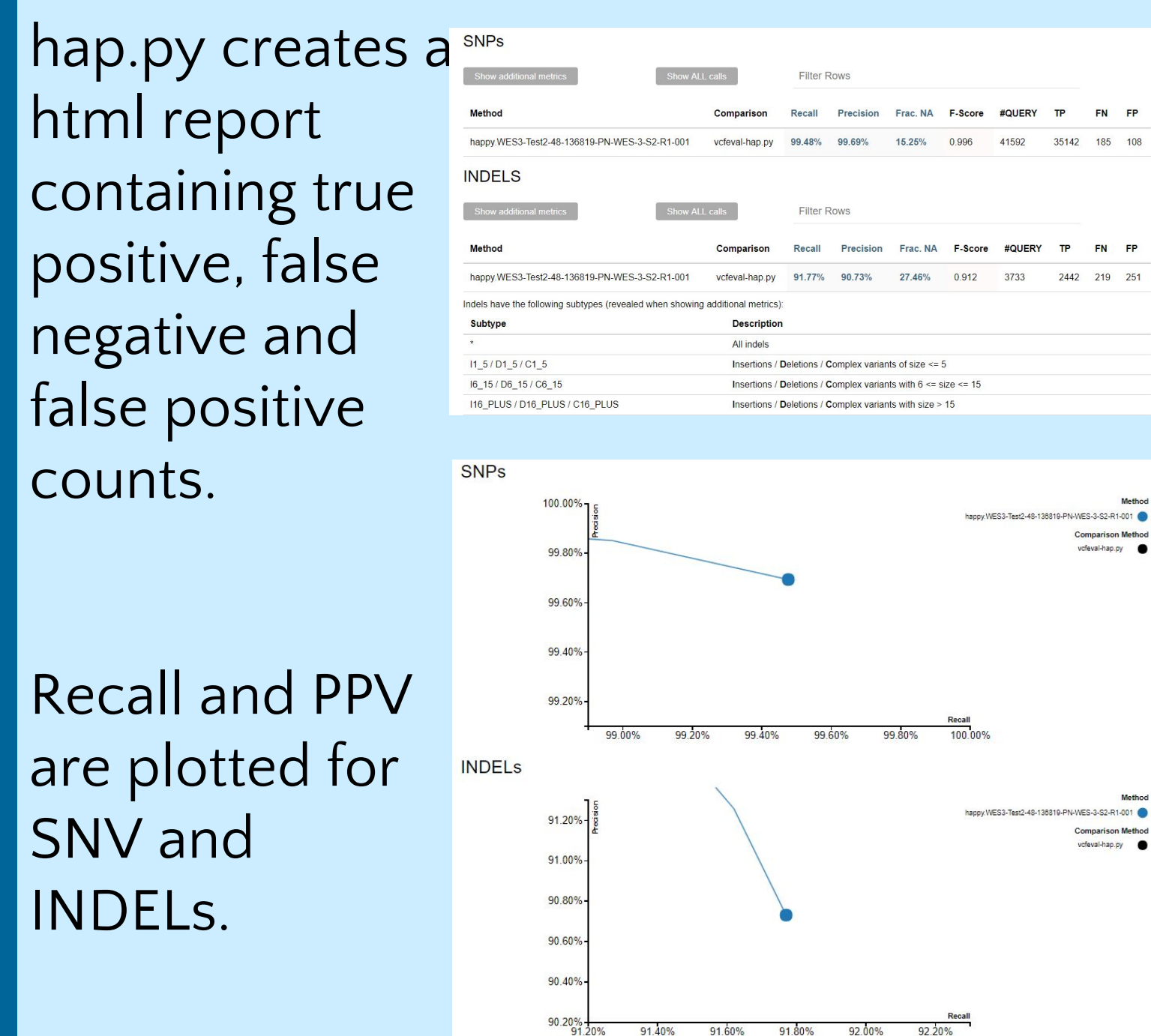
Currently only supports build GRCh37 (GRCh38 support will be added in the future).
Supports only NA12878 reference sample

Summary

This tool facilitates standardised benchmarking.

Since June 2017 this tool has been used more than 200 times by multiple diagnostic genetics laboratories

Figure 3. hap.py html report



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)
<https://github.com/RealTimeGenomics/rtg-tools>
3. <https://github.com/Illumina/hap.py>
- 4.