Assuring Data Quality with Variant Benchmarking tools from GIAB and GA4GH

N. D. Olson[1], J. McDaniel[1], J. Wagner[1], J. M. Zook[1], the GA4GH Benchmarking Team, and the Genome In A Bottle Consortium
[1]Biosystems and Biomaterials Division, Materials Measurement Laboratory, NIST

National Institute of Standards and Technology
U.S. Department of Commerce

Abstract

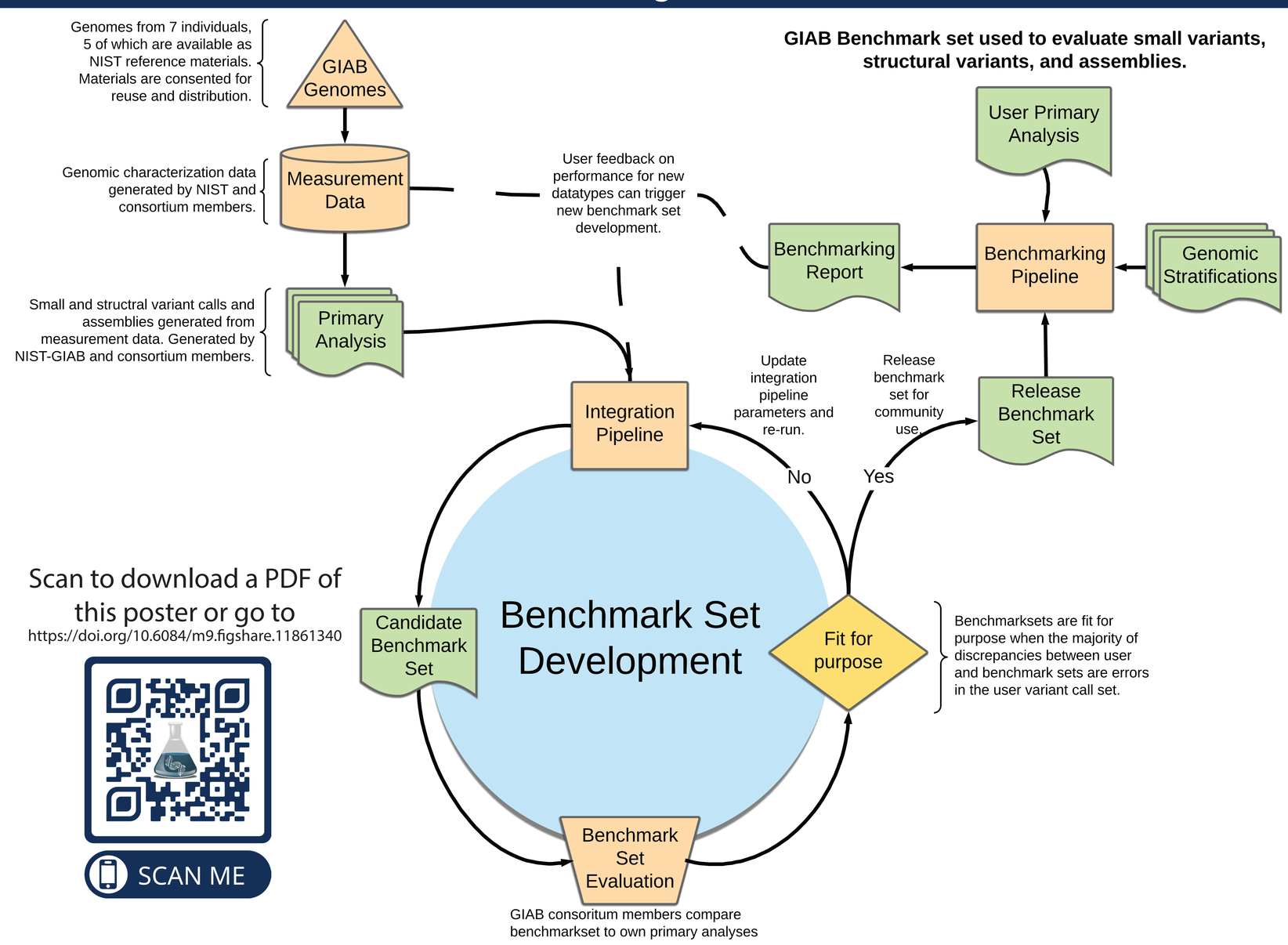
Integrating variant calls from diverse laboratories and sequencing methods requires a comprehensive understanding of false positive and false negative rates for different types of variants and genome contexts. Similarly, the clinical use of genomics requires validation of genome sequencing and variant calling methods. Clinical laboratories can validate their sequencing and variant calling methods using the GA4GH benchmarking tool and the NIST human genome reference materials developed with the Genome In A Bottle (GIAB) Consortium. The GIAB reference materials are authoritatively characterized for differences to the human reference genome and are provided as benchmark sets. The benchmark sets include benchmark calls, well-characterized sequence differences from the reference genome, and benchmark regions, positions that agree with the reference genome or are a benchmark call. The benchmarking tool compares a user-provided variant call set to the GIAB benchmark set in a manner that accounts for variant representation differences and disagreement types. The tool calculates performance metrics stratified by genomic context. However, the benchmarking tool's capabilities are limited due to; a lack of resources for benchmarking against GRCh38, difficult to interpret output, and a limited number of challenging variants in the GIAB benchmark sets. To address these limitations, we developed new resources for benchmarking against GRCh38 and are developing a summary report to simplify interpretation of the benchmarking results. We are also expanding the genomic coverage of the GIAB characterization to include more challenging regions such as low complexity regions and segmental duplications for both GRCh37 and GRCh38. The GIAB consortium is also developing benchmark sets and benchmarking methods for validating structural variants, allowing for high confidence structural variant detection. Expanding the GIAB genome characterization and improving the interpretability of the GA4GH benchmarking tool will inevitably help clinical genomics reach its full potential.

Genome In A Bottle

Started in 2012, over the last eight years the Genome in a Bottle Consortium has developed and characterized the first set of human whole genome reference materials. GIAB has developed small variant benchmark sets for 7 genomes (HG001 - HG007), NA12878 and two trios from the Personal Genome Project. Approximately 90% of non-gapped bases, primarily easy to characterize regions, of the genome are covered by the small variant benchmark sets.

- **2012** GIAB Consortium formed – no human genome Reference Materials
- **2014** Small variant genotypes for ~77% of pilot genome NA12878
- **2015** NIST releases first human genome Reference Material
- **2016** Small variants for 90% of 7 genomes for GRCh37/38
- **2018** Draft SV Benchmark
- **2019+** Characterizing difficult variants and regions

Benchmarking with GIAB



Scan to download a PDF of this poster or go to
https://doi.org/10.6084/m9.figshare.11861340

SCAN ME

Small Variant Benchmarking

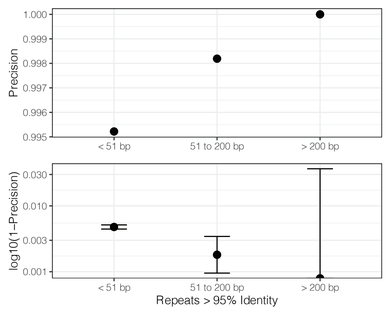
See Justin Zook's Poster for GIAB Small Variant Benchmark set V4.1 details.

- GA4GH Benchmarking team best practices include;
  - Comparison methods that account for variant representation differences,
  - Matching and performance metric definitions, and
  - Stratifying results by variant type and context.
- Best practices summarized in Krusche et al. 2019 (doi.org/10.1038/s41587-019-0054-x) Table S1.
- Best practices implementation (github.com/Illumina/hap.py).
  - Web application available on precisionFDA (precision.fda.gov)
  - precisionFDA accounts required for access available upon request.

Small Variant Benchmarking Underdevelopment
- Updated GRCh37 and GRCh38 stratifications.
- Improved benchmarking report to aid interpretation.

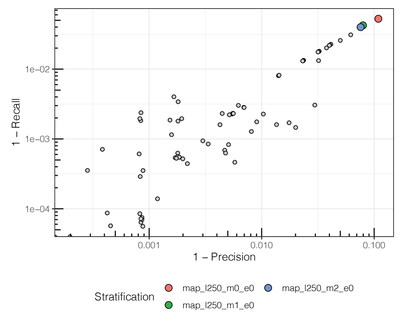
Interpreting Benchmarking Results

For better separation - plot 1 minus Metric on a log_{10} scale.
Use confidence intervals for more informative stratification comparisons.



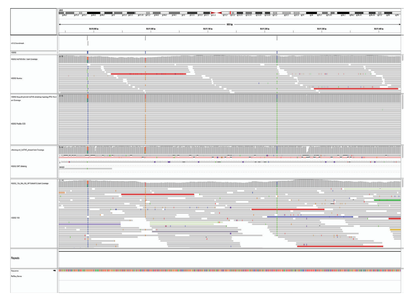
Comparison of benchmarking results on different scales and with confidence intervals for improved interpretation.

1- Precision and 1-Recall scatter plots help identify of poor performing stratifications.



Precision-Recall scatter plots. Points represent different stratifications. Poor performing stratifications are in the top right corner of the plots. Colors used to indicate low performing stratifications.

Manually curating a subset of FPs and FNs ensures they are errors in the query set and helps determine biases responsible for the error.

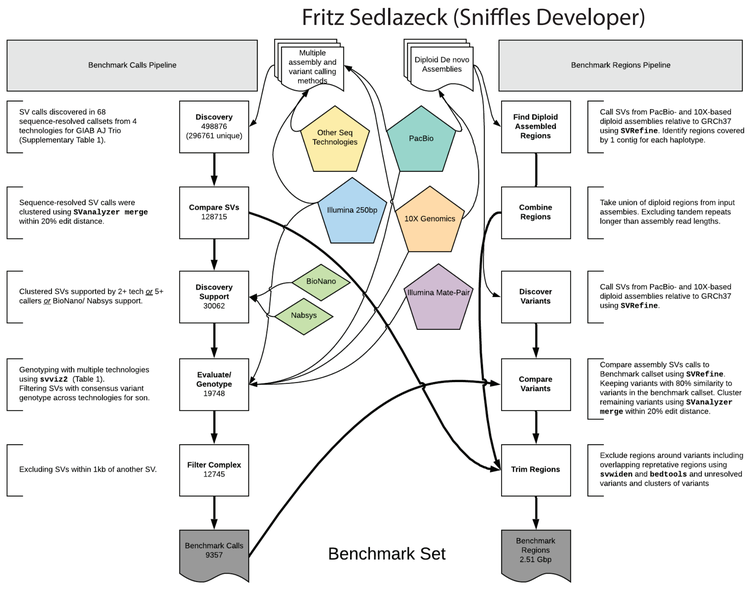


Example IGV session for manual curation.

Structural Variant (SV) Benchmarking

SV benchmark set developed using variant calls and assemblies from multiple technologies (Zook et al. 2019, doi.org/10.1101/664623).
- Benchmarking set integration pipeline diagram below.
- Structural variant benchmarking tools:
  - TRUVARI (github.com/spiralgenetics/truvari)
  - SVanalyzer (svanalyzer.readthedocs.io/en/latest/)
- V0.6 Benchmark set used in the development of Sniffles one of the most popular long read SV callers (Sedlazeck et al 2018, doi.org/10.1038/s41592-018-0001-7).

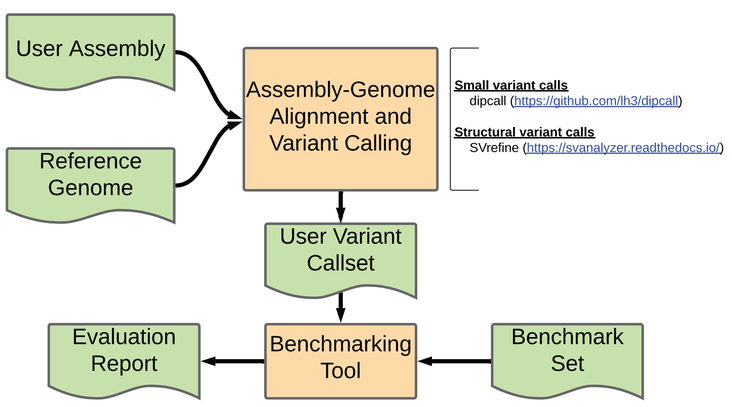
> *"The GIAB SV callset helped me to improve Sniffles especially for the large and very small SVs and improve the overall genotyping accuracy of my method"*
>
> Fritz Sedlazeck (Sniffles Developer)



V1.0 SV Benchmark Sets Under Development
- GRCh37 and GRCh38 GIAB other genomes
- Utilize ONT UL and PacBio HiFi data
- Rewriting code base for reproducibility and improved transparency.

Assembly Benchmarking

GIAB small and structural benchmarking set along with benchmarking methods complements existing assembly evaluation methods.
Assembly based variant calling methods are under active development.



GIAB Benchmark sets assembly benchmarking examples
- ***Garg et al 2019 (doi.org/10.1101/810341)***
- ***Shafin et al. 2019 (doi.org/10.1101/715722)***

Assembly Benchmarking Under Development
- Developing pipeline for assembly benchmarking and report to summarize results.

Contact Us

Want to benchmark your variant callset or assemblies but don't know how?

Have data you want to contribute?

Want to get involved in GIAB?

Email Us!

Justin Zook GIAB Co-Leader: justin.zook@nist.gov
Nate Olson NIST-GIAB Team Member: nolson@nist.gov

Better yet!

Want to join the NIST-GIAB team!

Ask about available Post-Doc opportunities.

Find out more about GIAB @
www.genomeinabottle.org

Acknowledgments

We would like to thank our collaborators on the GA4GH Benchmarking Team and Genome in a Bottle Consortium.

References

- Krusche, Peter, et al. "Best practices for benchmarking germline small-variant calls in human genomes." Nature biotechnology 37.5 (2019): 555.
- Zook, Justin M., et al. "An open resource for accurately benchmarking small variant and reference calls." Nature biotechnology 37.5 (2019): 561.
- Zook, Justin M., et al. "A robust benchmark for germline structural variant detection." BioRxiv (2019): 664623.