# nf-core/ variantbenchmarking

# GHGA
THE GERMAN HUMAN GENOME-PHENOME ARCHIVE

# A Robust Variant Benchmarking Pipeline within the GHGA Framework

Kübra Narci[1], Florian Heyl[1], Sameesh Kher[1], Ata Jadidharari[2], Nicolas Vannieuwkerke[3], Maxime U Garcia[4], Daniel Hübschmann[1], nf-core community, and the GHGA consortium

1 Deutsches Krebsforschungszentrum, Heidelberg, Germany
2 Technical University of Munich, Garching, Germany
3 Ghent University, Ghent, Netherlands
4 Seqera Labs, Barcelona, Spain

Variant calling is crucial for identifying genetic variations in next-generation sequencing (NGS) data, but differences in tools, technologies, and variant representations cause inconsistencies. To address this, we developed the **nf-core/variantbenchmarking** pipeline within the **German Human Genome-Phenome Archive** (GHGA) project. It benchmarks **small** variants, **indels**, **structural** variants, and **copy number** variations for both **germline** and **somatic** samples. The pipeline uses normalization methods to resolve allelic ambiguities and provides metrics like precision, recall, and F1 scores. Built with **Nextflow** under **nf-core repository**, it is scalable, reproducible, and integrates securely with GHGA, supporting the development of accurate variant-calling workflows for precision medicine.

## ✅ Multiple input types can run in parallel

- Multi-sample VCFs
- Germline or somatic samples
- Regions BED
- Bad formatted VCFs

## ✅ Variant preprocessing ensures standardization

- **Liftover** tools
- Structural Variant **reformatting**
- **Decomposition** of complex variants
- Filtering

## ✅ Optional benchmarking tools and parameters

- Small variants
  - **hap.py** or **som.py** and **rtgtools**
- Structural variants
  - **truvari, svanalyzer** and **witty.er**
- Copy number variations
  - truvari, witty.er and intersection

## ✅ Key features

- Performs both germline and somatic benchmarking
- **nf-core** compliant
  - Open-source & community-driven
  - Modular design
  - **Reproducible**
- FAIR-aligned within GHGA
- Generates **publication-ready** reports
  - Benchmarking metrics & plots



run with

v1.2.0
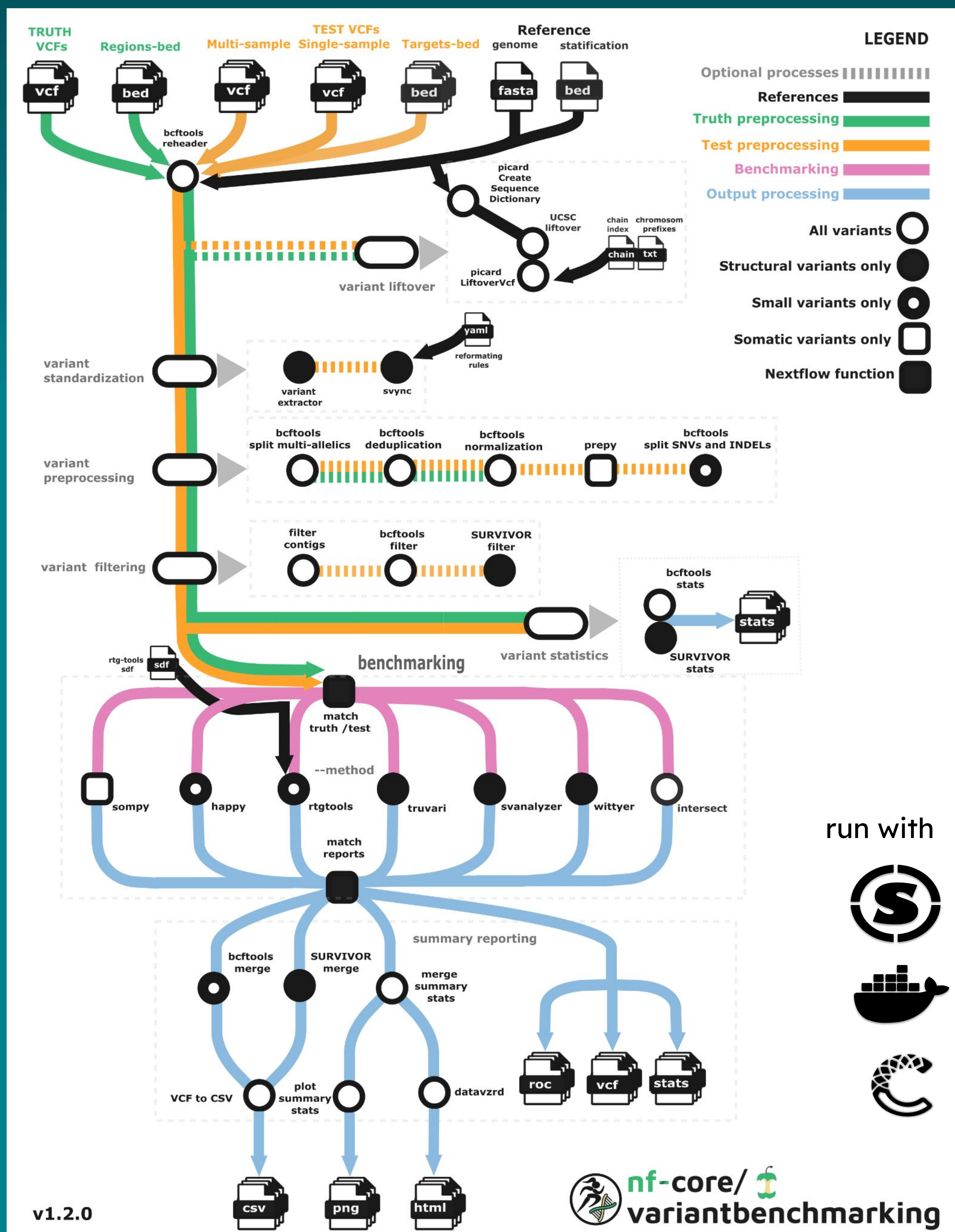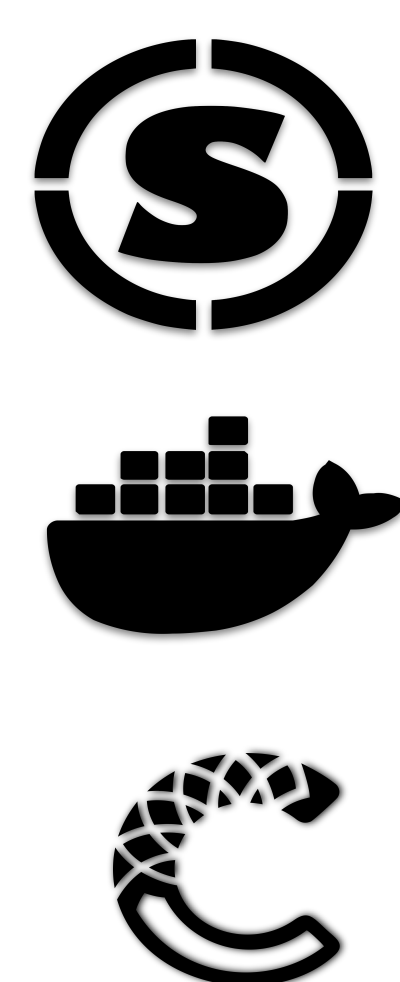
nf-core/ variantbenchmarking

slack nf-core #variantbenchmarking

zenodo DOI:10.5281/zenodo.15116431

Find the pipeline https://github.com/nf-core/variantbenchmarking

Questions? send an email to kuebra.narci@dkfz-heidelberg.de

dkfz. DEUTSCHES KREBSFORSCHUNGSZENTRUM IN DER HELMHOLTZ-GEMEINSCHAFT

de.NBI GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE