

The German Human Genome-Phenome Archive

Standardizing and harmonizing NGS analysis workflows to create a unified omics data resource for Germany

Zehra Hazal Sezer^{1,2}, Florian Heyl, Kübra Narci, Paul Menges, Luiz Gadelha, Sameesh Kher, Drew Behrens, Florian Hölzlwimmer, Shounak Chakraborty, Jens Krueger, Julien Gagneur and Christian Mertes

¹High Performance and Cloud Computing Group and Applied Bioinformatics, Eberhard Karls Universität Tübingen, Germany, ²German Human Genome-Phenome Archive (GHGA), Tübingen, Germany

GHGA is developing a scalable and secure IT infrastructure for Germany, an ethico-legal framework to handle omics data in a data-protection-compliant but open and **FAIR** manner, a harmonized metadata schema, and standardized workflows to process all incoming omics data. Our main objectives are:

- Build upon the **nf-core** and **nextflow** community to build NGS analysis workflows for all incoming data modalities
- Enable the creation of harmonized and **standardized NGS** resources across datasets
- Maintain and develop workflows, create runtime configurations for each data modality with stable identities, and continuously evaluate performance

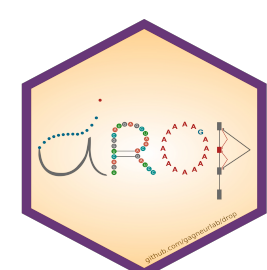
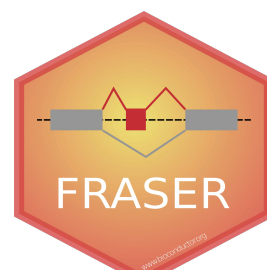
• Variant Calling analysis with WGS, WES and TS



• Single Cell analysis



• RNA-seq analysis



• Nanopore analysis



• Somatic variant analysis



• Benchmarking



* Pipelines tested and used by GHGA Workflows

In collaboration with the **nf-core** community and beyond, GHGA maintains and co-develops workflows. To evaluate the performance of the workflows, GHGA together with the German NGS Competence Network (NGS-CN) developed a continuous benchmarking framework: NCBench. GHGA is creating a runtime configuration to uniformly process NGS data while guaranteeing the highest standards and quality of the workflows. Harmonized resources will enable cross-analysis of projects and population-scale studies, promote new collaborations and research projects, and establish the foundation for developing a German-based variant frequency database.

