c a a c a d a c a a NGS da a

Friederike Hanssen[1], Maxime Garcia[2], Lasse Folkersen[3], Susanne Jodoin[1], Oskar Wacker[1], Anders Sune Pedersen[4], Edmund Miller[5], Francesco Lescai[6], Nick Smith[7], nf-core community, Gisela Gabernet[1], Sven Nahnsen[1,8]

Quantitative Biology Center, University of Tübingen, Tübingen SciLifeLab, Karolinska Institutet, Stockholm Nucleus Genomics Ltd., New York Danish National Genome Center, Copenhagen University of Texas, Dallas Department of Biology and Biotechnology, University of Pavia, Pavia German Human Genome-Phenome Archive Biomedical Data Science, Department of Computer Science, University of Tübingen, Tübingen

1. Introduction

Somatic variant calling studies often include many patients with dataset sizes varying widely between oncopanel, whole-exome, and whole-genome sequencing data. nf-core/sarek[1] is an es

2. Implementation Details

Ported to DSL2:
- Using nf-core/modules for 80 of 82
- Related steps are bundled into subworkflows

Extensive CI using pytest:
- Testing for docker, conda, and singularity
- Md5sum or file content checks
- Tests for a variety of use cases

Usage of CRAM files

Automatic testing on AWS on release with:
- Germline: GIaB HG0001
- Somatic: HCC1395

→ Both have a ground truth set available and can be used for validation