

# nf-core/sarek: a pipeline for efficient germline, tumor-only, and somatic analysis of NGS data on different compute infrastructures

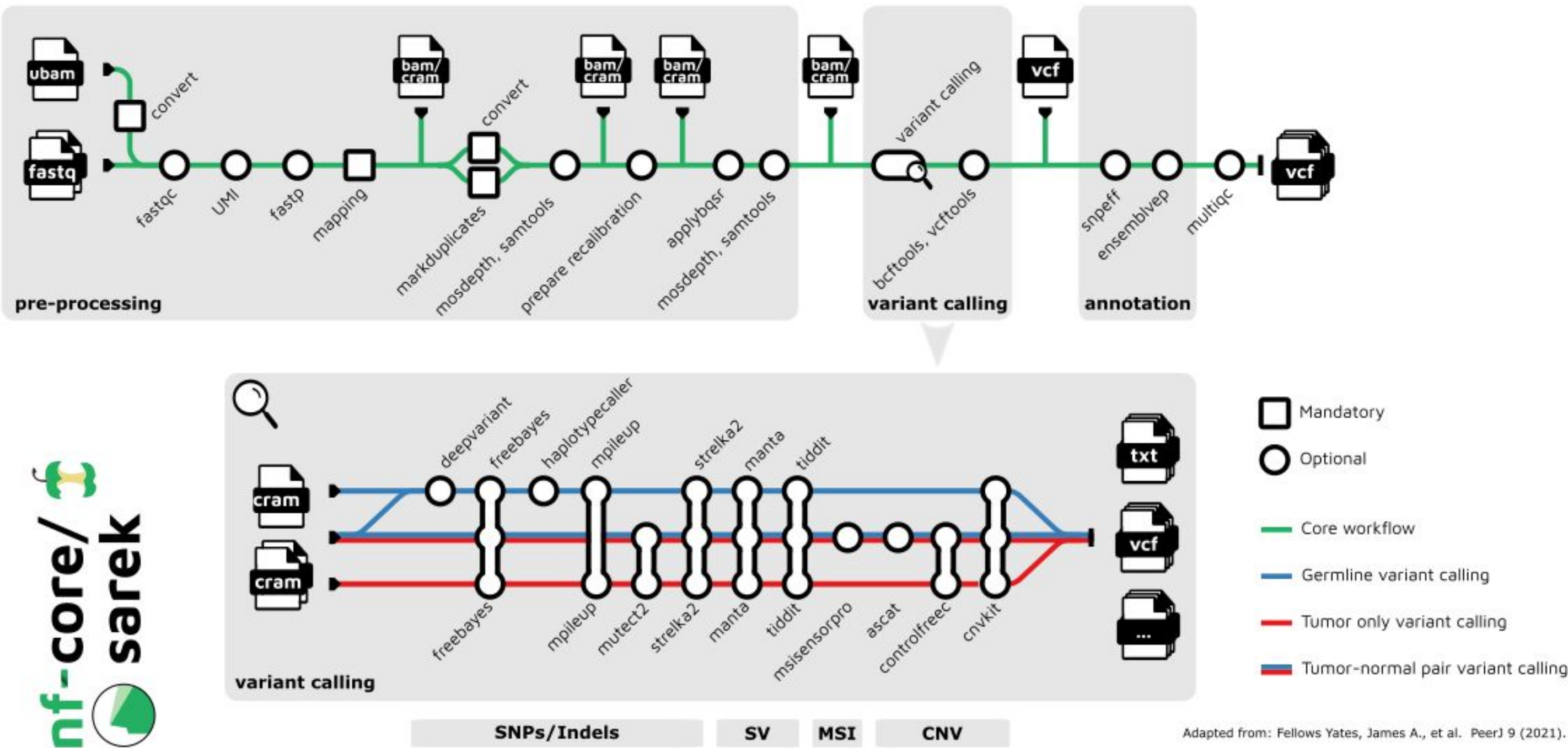
Friederike Hanssen<sup>1,2</sup>, Maxime U. Garcia<sup>3</sup>, Lasse Folkersen<sup>4</sup>, Susanne Jodoin<sup>1</sup>, Oskar Wacker<sup>1</sup>, Anders Sune Pedersen<sup>5</sup>, Edmund Miller<sup>6</sup>, Francesco Lescai<sup>7</sup>, Nick Smith<sup>8</sup>, nf-core community, Gisela Gabernet<sup>1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

<sup>1</sup>Quantitative Biology Center, University of Tübingen, Tübingen <sup>2</sup>Biomedical Data Science, Department of Computer Science, University of Tübingen, Tübingen <sup>3</sup>Seqera Labs, Barcelona <sup>4</sup>Nudeus Genomics Ltd., New York <sup>5</sup>Danish National Genome Center, Copenhagen <sup>6</sup>University of Texas, Dallas <sup>7</sup>Departement of Biology and Biotechnology, University of Pavia <sup>8</sup>German Human Genome-Phenome Archive

## Overview

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<sup>1</sup> is an established pipeline for exploring single-nucleotide variants, structural variation, microsatellite instability, and copy-number alterations of germline, tumor-only, and paired tumor-normal short-reads. nf-core/sarek is part of nf-core<sup>2</sup>, a community project which provides an infrastructure to create reproducible, scalable, and portable open-source Nextflow<sup>3</sup>-based pipelines. Here, we show the latest updates including improvements to the data flow and tool selection reducing time, compute resources, and cloud computing costs, as well as modularization improving code maintainability.

Preprint available at: [biorxiv.org/content/10.1101/2023.07.19.549462v1](https://www.biorxiv.org/content/10.1101/2023.07.19.549462v1)



Pipeline metro map showing a high-level view of the different analysis steps. The pipeline can be started from six different entry points and run through all subsequent tasks. All optional tools can be selected in any combination. This allows to recompute and extend the results throughout a project's duration.

## Scatter/gathering speeds up analysis

