Introducing

# nf-core/ phaseimpute

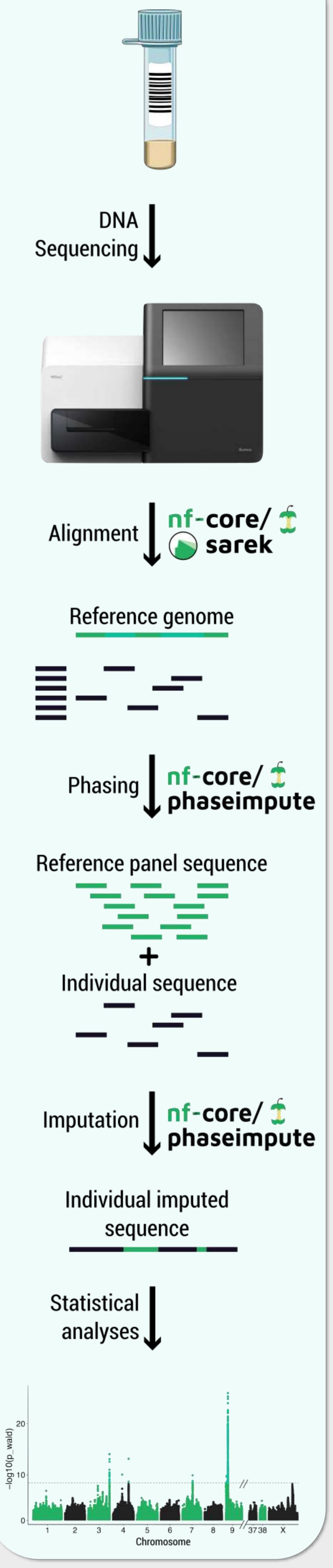
from idea to release

**Louis LE NÉZET**[1], **Anabella TRIGILA**[2], **Pascale QUIGNON**[1], **Catherine ANDRÉ**[1]

[1] Université de Rennes, CNRS, Institut de Génétique & Développement de Rennes, UMR 6290 CNRS - Rennes, France
[2] ZS Discovery, Argentina

---

## WORKFLOW

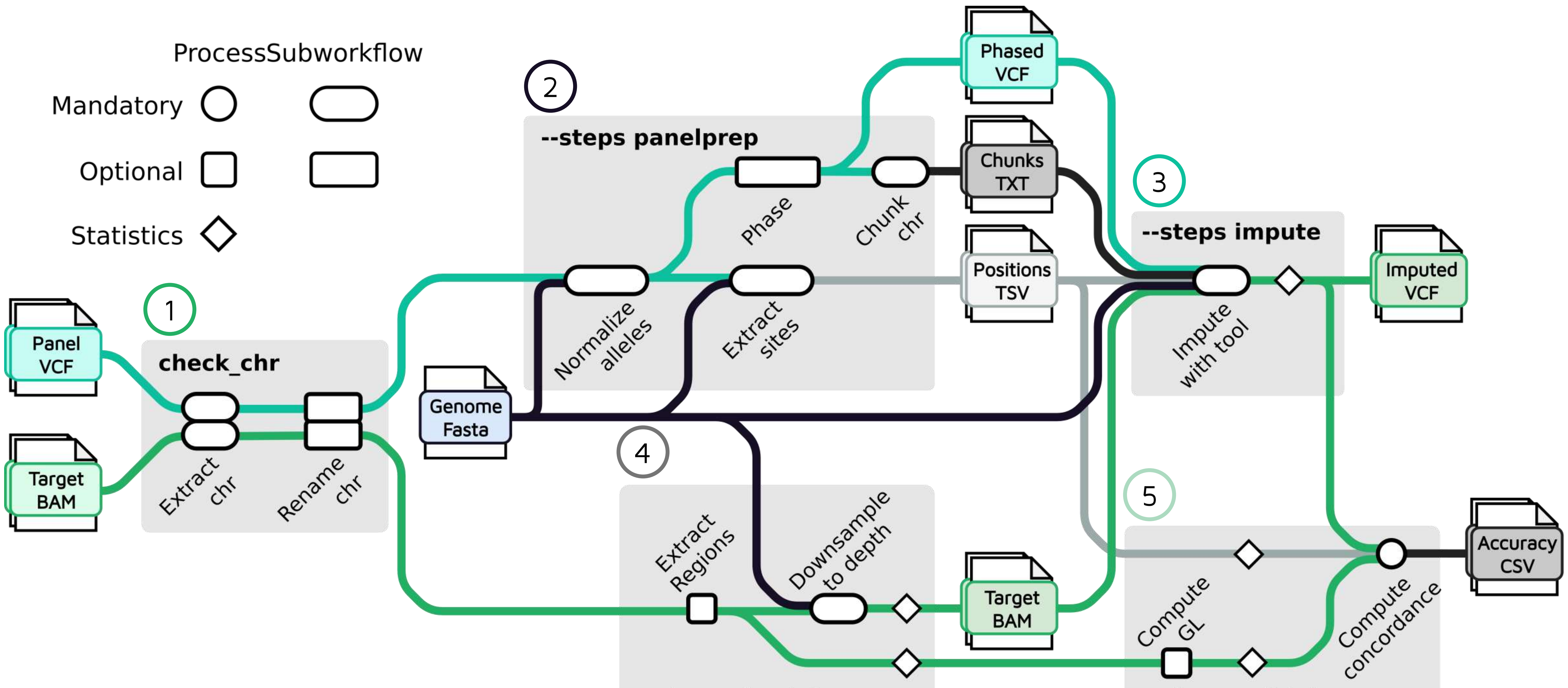


## INTRODUCTION

Genome imputation is a statistical technique that enhances the resolution of genotyping arrays and low-pass sequencing (<1x) by filling missing data with information from reference panels. While existing pipelines primarily focus on the imputation step and in the human species, crucial steps such as panel preparation, phasing, and imputation assessment are often overlooked.

To address this gap, we introduce **nf-core/phaseimpute**, a comprehensive pipeline to perform **panel preparation**, **genomic data simulation**, **imputation**, and **tool assessment**.

## NF-CORE/PHASEIMPUTE DESIGN



① Check presence/absence of \`chr\` prefix
② Normalize, extract and phase reference variants
③ Impute target BAM files with ≠ tools <glimpse1, glimpse2, stitch, quilt>
④ Downsample BAM files to specific depth
⑤ Compute imputation accuracy between imputed files and truth set

## DEVELOPMENT

Back and forth with **nf-core/ modules** repository

**nf-tests** for every module, function and subworkflow

CI/CD with GitHub actions for **different use cases**

**Mega-tests** with Seqera Platform / AWS

New imputation modules in **multiQC**

## FUTURE PERSPECTIVES

- Handling of **SNP chip data**
- Allow **batch** imputation
- Imputation of **chrX**
- Add **new imputation tools**
- Cost and environment impact with **nf-CO2footprint**

## JOIN THE DISCUSSION !

**#phaseimpute**

## ACKNOWLEDGMENTS

- nf-core community
- With the financial support of the ACGAO and the Visio fundation

ZS Discovery
Elucidate. Innovate. Accelerate.

cnrs igdr Université de Rennes

LES CHIENS GUIDES D'AVEUGLES DE L'OUEST
Ensemble voyons loin

FONDATION VISIO
L'INNOVATION NOUS GUIDE

Fondation pour la Recherche Médicale