



Karolinska  
Institutet

# BALSAMIC: Bioinformatic Analysis Pipeline for Somatic Mutations in Cancer

Hassan Foroughi-Asl<sup>1,2</sup>, Ashwini Jeggari<sup>1</sup>, Vadym Ivanchuk<sup>1</sup>, Khurram Maqbool<sup>1</sup>, Annick Renevey<sup>1</sup>, Maria Ropat<sup>1</sup>, Sarath Murugan<sup>1</sup>, Keyvan Elhami<sup>1</sup>, Valtteri Wirta<sup>1</sup>

<sup>1</sup>Department of Microbiology, Tumour and Cell Biology, Science for Life Laboratory, Karolinska Institutet, <sup>2</sup>Department of Molecular Medicine and Surgery, Karolinska Institutet



SciLifeLab

## Motivation

Massively parallel sequencing technologies are used to identify genetic alterations in cancer. However, their detection remains an open topic in bioinformatics, as there are no single tool capable of characterising the full spectrum of DNA-based tumour mutations for different case scenarios. To address this problem, we present here the implementation of BALSAMIC.



## Pipeline

BALSAMIC is a **Snakemake**-based pipeline that integrates read alignment, quality control (QC) assessment, variant calling, filtering, and annotation of whole genome (**WGS**), whole exome (**WES**), and targeted gene sequencing (**TGS**) data. It can process tumour-only or tumour-normal sample pairs from **humans** or **dogs**, and assays with unique molecular identifiers (**UMIs**).

# BALSAMIC

## Whole Genome Sequencing

### ALIGNMENT

fastp

BWA-MEM

Sentieon

sam2bam

Dedup

### QUALITY CONTROL

Picard

Sentieon

MultiQC

WGS metrics

## COMMON WORKFLOW

CONFIG

### REFERENCE GENOME

HG19

HG38

CanFam3

### ANALYSIS TYPE

SINGLE

PAIRED

tumour-only

tumour-normal

### RUN MODE

PERSONAL

COMPUTER

COMPUTER

CLUSTER

### VARIANT CALLERS

SELECT

DEACTIVATE

## Targeted Gene / Whole Exome Sequencing

### ALIGNMENT

fastp

BWA-MEM

SAMtools

UMI

Sentieon

BWA-MEM

UMI

### QUALITY CONTROL

Picard

MultiQC

UMI

Mosdepth

Sambamba

## VARIANT CALLING

### SOMATIC

Manta

Delly

AscatNGS

TIDDIT

CNVpytor

SVDB

merge variants

SV

Sentieon

TNscope

SNV

### GERMLINE

SV

Manta

SNV

Sentieon

DNAscope

VCF

### ANNOTATION

VEP

vcfanno

### FILTERING

Loqusdb

GENMOD

## ANNOTATION & FILTERING

## Results

- Analysis **report**
- **Quality control** indicators
- Filtered and annotated **VCFs**
- Supporting **figures** and **graphs**

## VARIANT CALLING

### SOMATIC

Manta

Delly

CNVkit

VarDict

SVDB

merge variants

SV

SNV

SNV

Sentieon

TNscope

UMI

## Conclusions

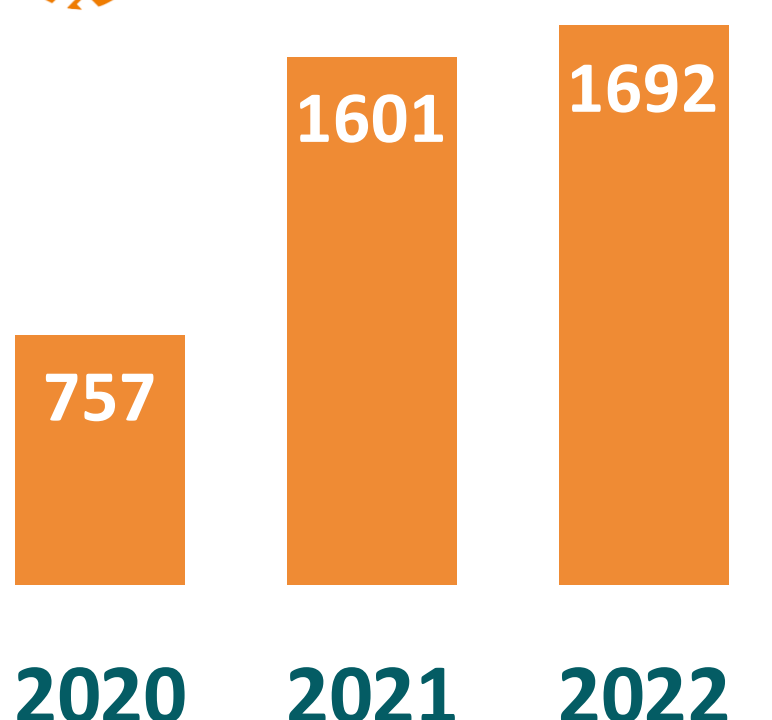
BALSAMIC covers all the analysis steps from raw sequencing data to identification of the entire range of DNA-based alterations. It helps researchers and clinicians detect somatic and germline mutations with improved sensitivity and specificity by combining multiple variant calling algorithms and bioinformatic tools.

**BALSAMIC is freely available to the research community.**

## Production setting

Clinical Genomics

The **Clinical Genomics** unit at SciLifeLab, in collaboration with the **Swedish healthcare system** represented by several clinicians and clinic-affiliated researchers, works on integrating genomics-based tools into routine medical care.



Analysed cancer samples per year