BALSAMIC: Bioinformatic Analysis Pipeline for Somatic Mutations in Cancer

Hassan Foroughi-Asl[1,2] , Ashwini Jeggari[1], **Vadym Ivanchuk**[1], Khurram Maqbool[1], Annick Renevey[1], Maria Ropat[1], Sarath Murugan[1], Keyvan Elhami[1], Valtteri Wirta[1]

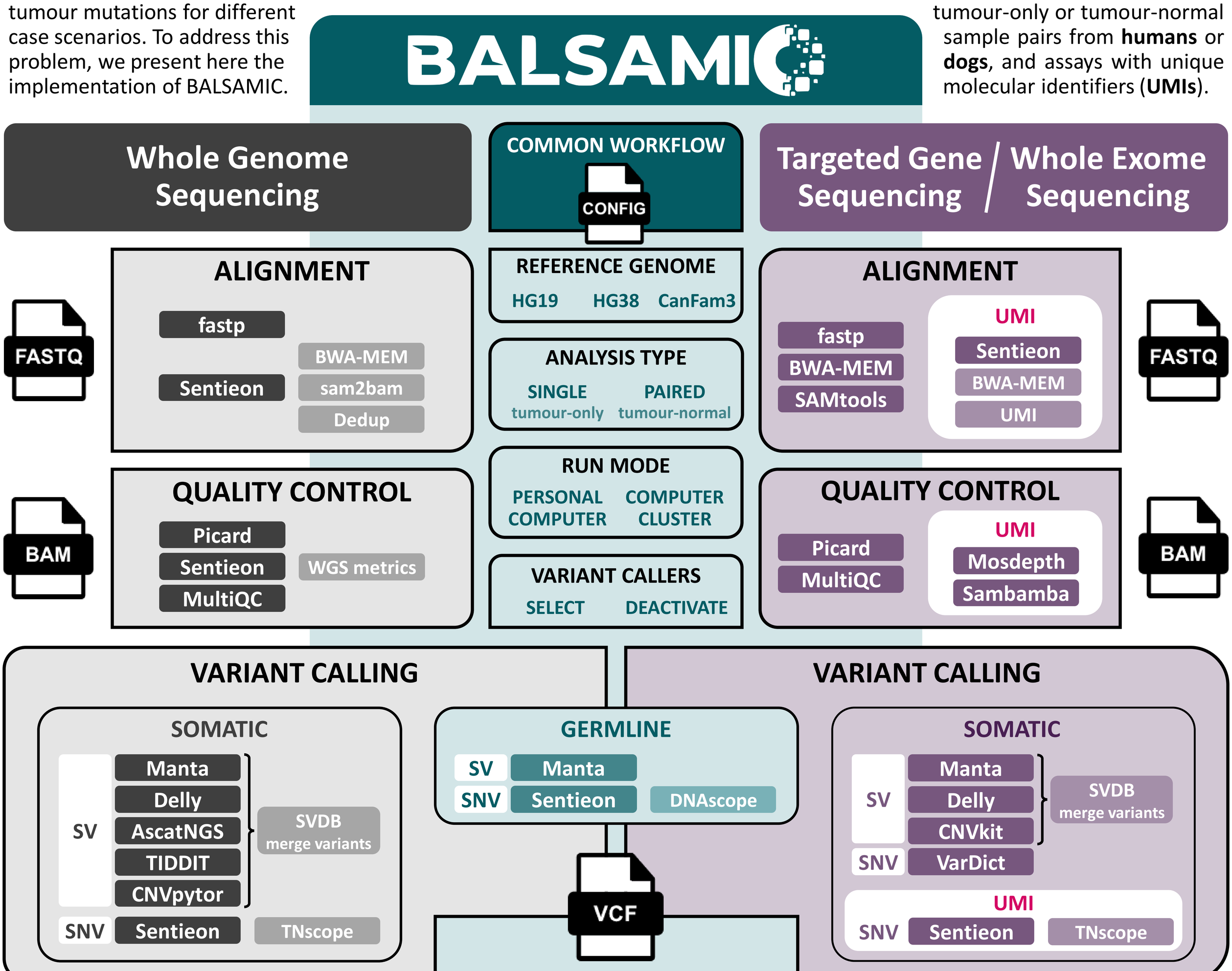
[1]Department of Microbiology, Tumour and Cell Biology, Science for Life Laboratory, Karolinska Institutet,
[2]Department of Molecular Medicine and Surgery, Karolinska Institutet

Motivation

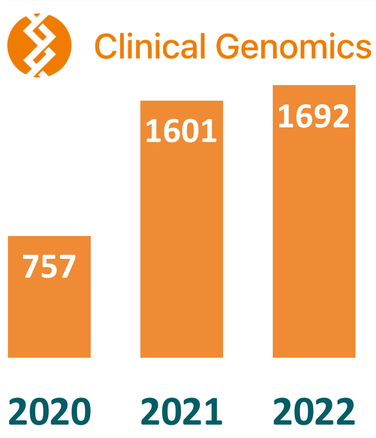
Massively parallel sequencing technologies are used to identify genetic alterations in cancer. However, their detection remains an open topic in bioinformatics, as there are no single tool capable of characterising the full spectrum of DNA-based tumour mutations for different case scenarios. To address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



Production setting



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.

Results

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

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.