

nf-core¹ for tailored clinical research data analysis – challenges and use cases

Anne Bertolini, Linda Grob, Vladislav Grigorjev, Irene Keller, Franziska Singer, and NEXUS team
 NEXUS Personalized Health Technologies, ETH Zurich; SIB Swiss Institute of Bioinformatics

Introduction

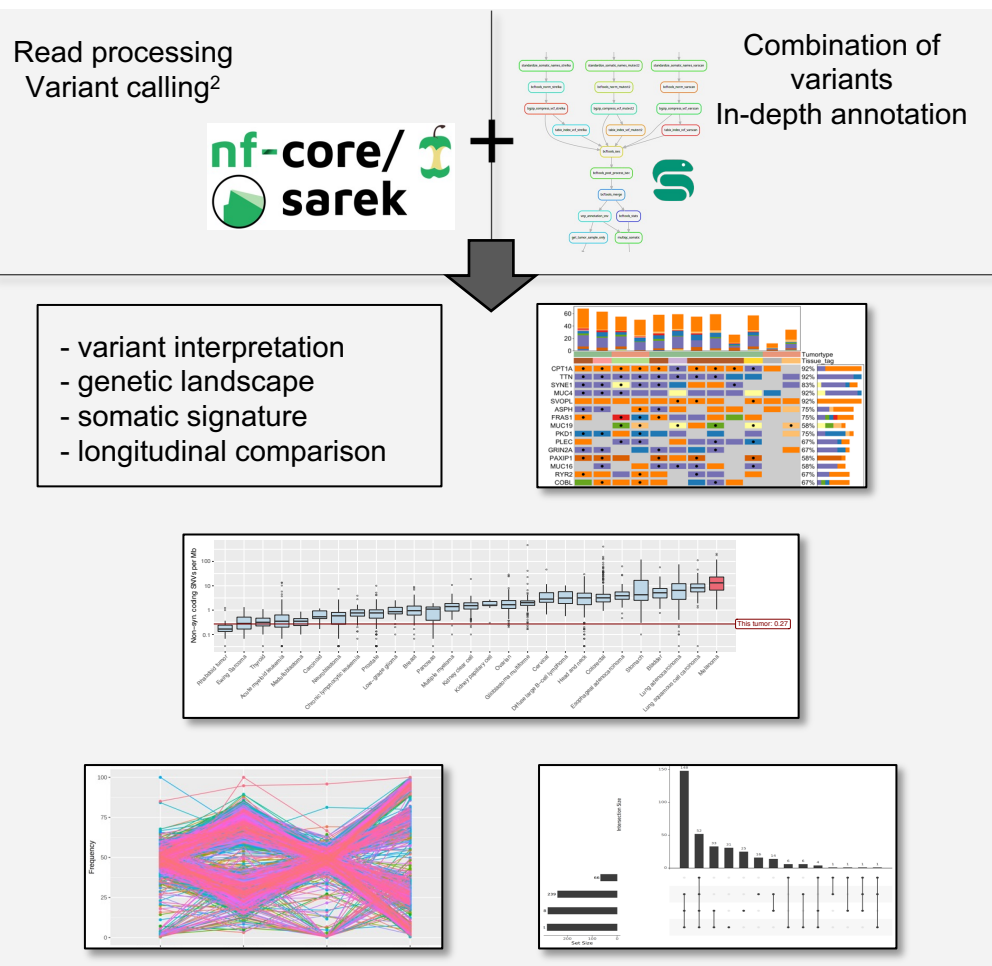
As a core facility for biomedical research and translation, our analyses involve clinical data that poses additional requirements on data processing and management. E.g., we are restricted to non-cloud-based analysis and have to use self-contained local runs of the pipelines, once the initial setup is completed.

In addition, the project goals typically require customized follow-up analyses after the standardized workflows for basic data processing.

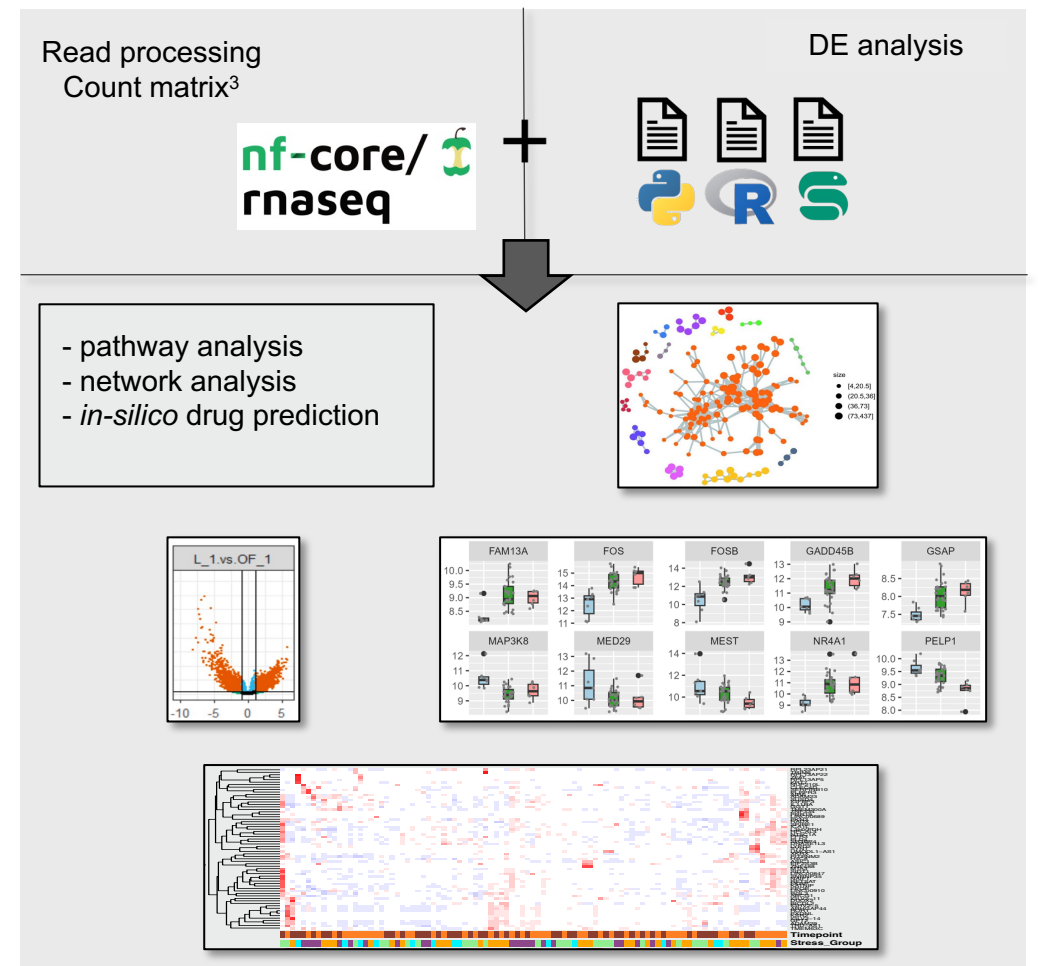
Therefore, we typically use a hybrid of Nextflow and other workflow management systems, including Snakemake⁴, or extend pipelines with customized downstream analysis scripts.

Here, we show a typical clinical data analysis where nf-core pipeline and tailored downstream analyses are intertwined and illustrate the privacy-ensuring computing and software installation setup.

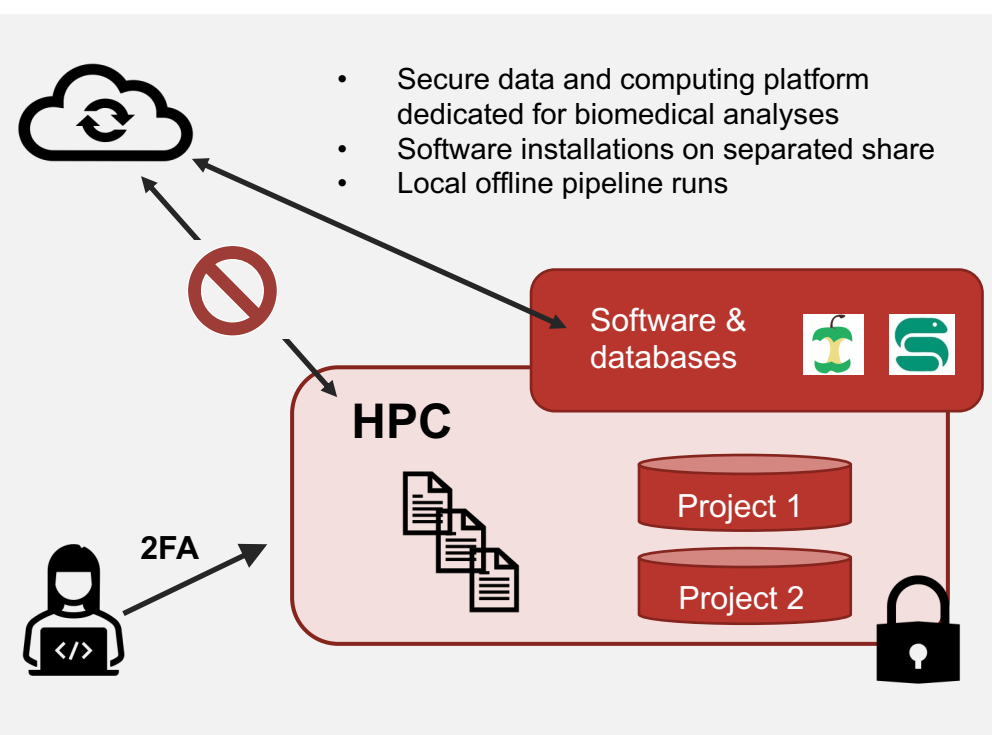
Example DNA – variant calling



Example RNA-seq analysis



Infrastructure Overview



Conclusion

- Robust Foundation:** nf-core pipelines using Nextflow provide a great foundation for reproducible bioinformatics data analyses.
- Setup Challenges:** Navigating a challenging setup with restricted internet access, prioritizing data protection.
- Integration Challenges:** Fragmented analyses using Nextflow pipelines, Snakemake, and custom scripts to achieve tailored results.
- Outlook:** Strive to unify and streamline analyses, e.g., using Nextflow and the nf-core environment.

References

- Ewels, Philip A., et al. "The nf-core framework for community-curated bioinformatics pipelines." *Nature biotechnology* 38.3 (2020): 276-278.
- Garcia, Maxime, et al. "Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants." *F1000Research* 9 (2020).
- Patel, Harshil, et al. "nf-core rnaseq pipeline" doi: [10.5281/zenodo.14400710](https://doi.org/10.5281/zenodo.14400710)
- Mölder, Felix, et al. "Sustainable data analysis with Snakemake." *F1000Research* 10 (2021).