

Computational rescue strategy for orphan codebases: nf-core/ismquant

Carolin Schwitalla^{1,2,3}, Niclas Grote^{1,2,3}, Luis Kuhn Cuellar^{1,2,3}, nf-core community, Jason Stein⁴, Sven Nahnsen^{1,2,3}

¹Quantitative Biology Center, Eberhard-Karls University of Tübingen, Germany.

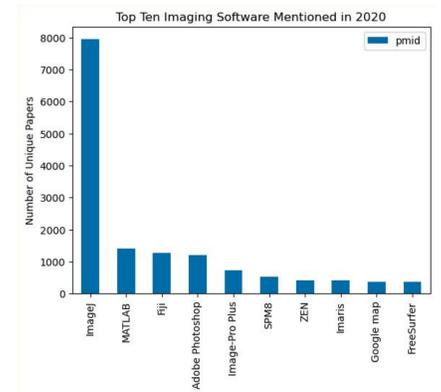
²The M3 Research Center, University Hospital Tübingen, Germany.

³Biomedical Data Science, Department of Computational Science, Eberhard-Karls University of Tübingen, Germany.

⁴Department of Genetics & Neuroscience Center (Stein lab), University of North Carolina at Chapel Hill.

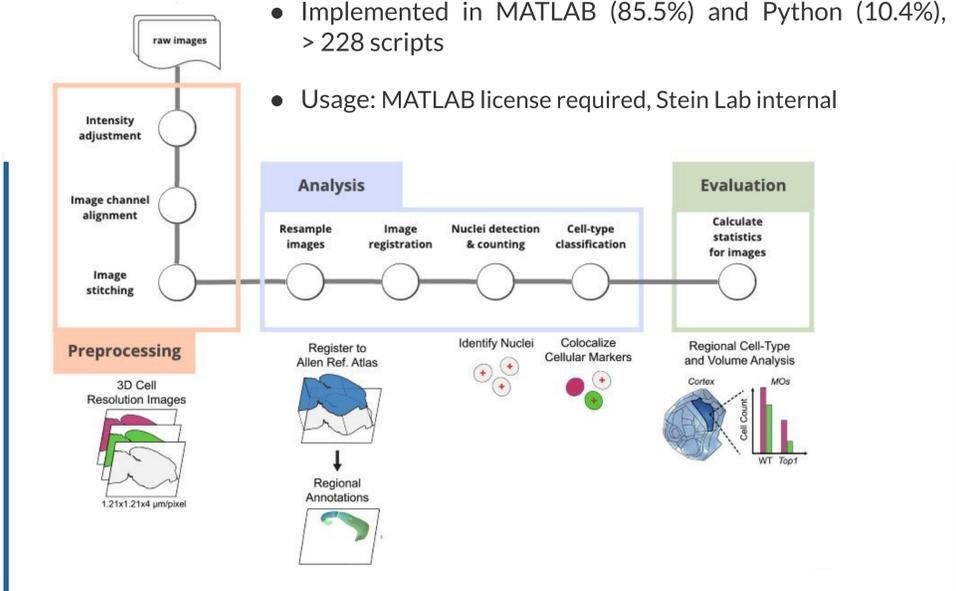
INTRODUCTION

In the diverse field of biological imaging, where datasets can be in the terabyte range and image acquisition gets faster, data analysis demonstrates special challenges. Scientists mostly use GUIs like ImageJ and FIJI for their analysis, which involves a lot of manual inspections or rely on proprietary software like MATLAB and Imaris. There is a lack of FAIR open-source end-to-end processing tools (Istrate, Ana-Maria et al., 2022) that can efficiently process such large, high-dimensional datasets. Here, we demonstrate the rescue of the “orphan” pipeline NuMorph (Krupa et al., Cell reports 2021) an extensive pipeline for processing and analysis of light-sheet microscopy datasets in the terabyte range. This pipeline is considered as “orphan” because, since its publication, no activity on the repository could be observed indicating it's only used by the research group of the authors. We aim to enable the reuse of this extensive research software by implementing it with nextflow and integrating it into the nf-core community. With this, we want to contribute to a more sustainable research software development process, as well as introduce one of the first best-practice pipelines for fluorescence image processing and analysis into the nf-core community thereby moving towards standardized and reproducible analysis in this field of research.



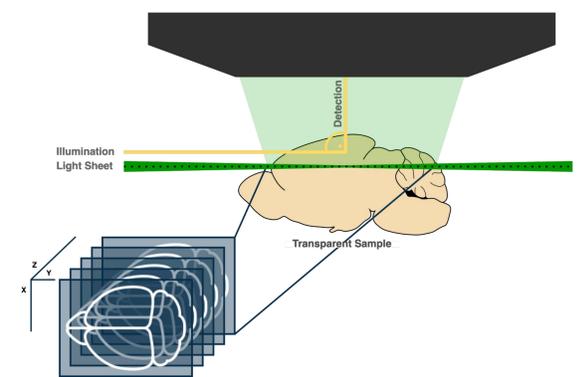
NuMorph (Nuclear Morphometry) (Krupa et al., Cell reports 2021)

- Preprocessing and analysis of whole mouse brain images (approx. 1 TB/ sample)
- Implemented in MATLAB (85.5%) and Python (10.4%), > 228 scripts
- Usage: MATLAB license required, Stein Lab internal



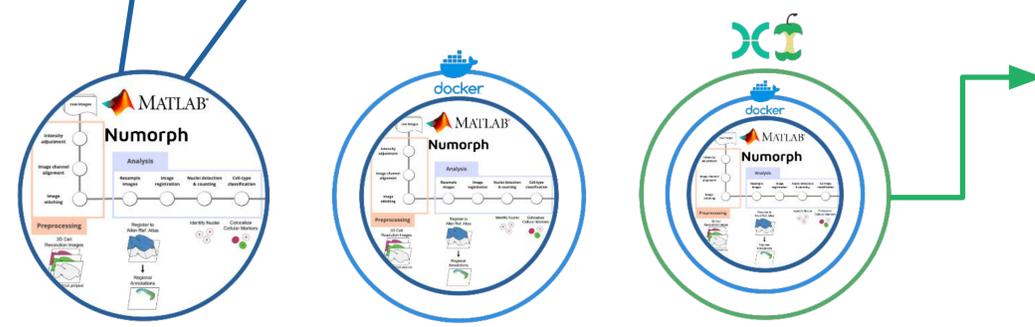
Light-sheet fluorescence microscopy

- Selective illumination of the focal plane with a laser light-sheet
- Reduces photodamage and stress on living samples
- Scanning with a plane of light speeds up image acquisition at high resolution
- 3D and 4D images in the TB range



FAIRification workflow

AIM
Reproducible, usable, extendable, open-source processing and analysis of TB sized images



MATLAB standalone application

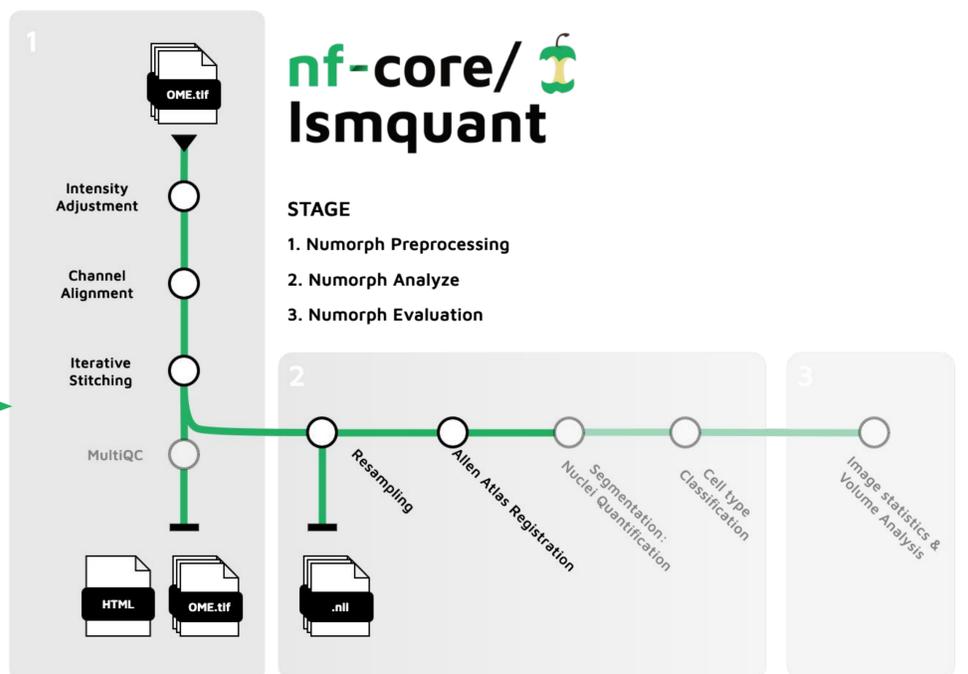
- No proprietary license requirements
- Open file formats for in- and output

Containerisation

- Reproducibility
- Portability

Nextflow / nf-core implementation

Nextflow pipeline



Outlook

- Implementation of complete NuMorph functionalities
- Gradual extending functionalities for efficient and generalizable preprocessing
- Improve cell segmentation for dense tissue samples

Acknowledgements

We would like to acknowledge funding from the Carl Zeiss Foundation and the Reinhard-Frank Stiftung. We are grateful to the nf-core and nextflow communities for their support during the development.

Get in Touch:



nf-core/ismquant



qbic-pipelines/ismquant