CREATIS Medical Imaging Research Laboratory
www.creatis.insa-lyon.fr

UNIVERSITÉ DE LYON

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

Inserm

cnrs

UCB Lyon 1

UNIVERSITÉ JEAN MONNET SAINT-ÉTIENNE

Reproductibilité computationnelle en imagerie médicale

Sorina POP
CREATIS

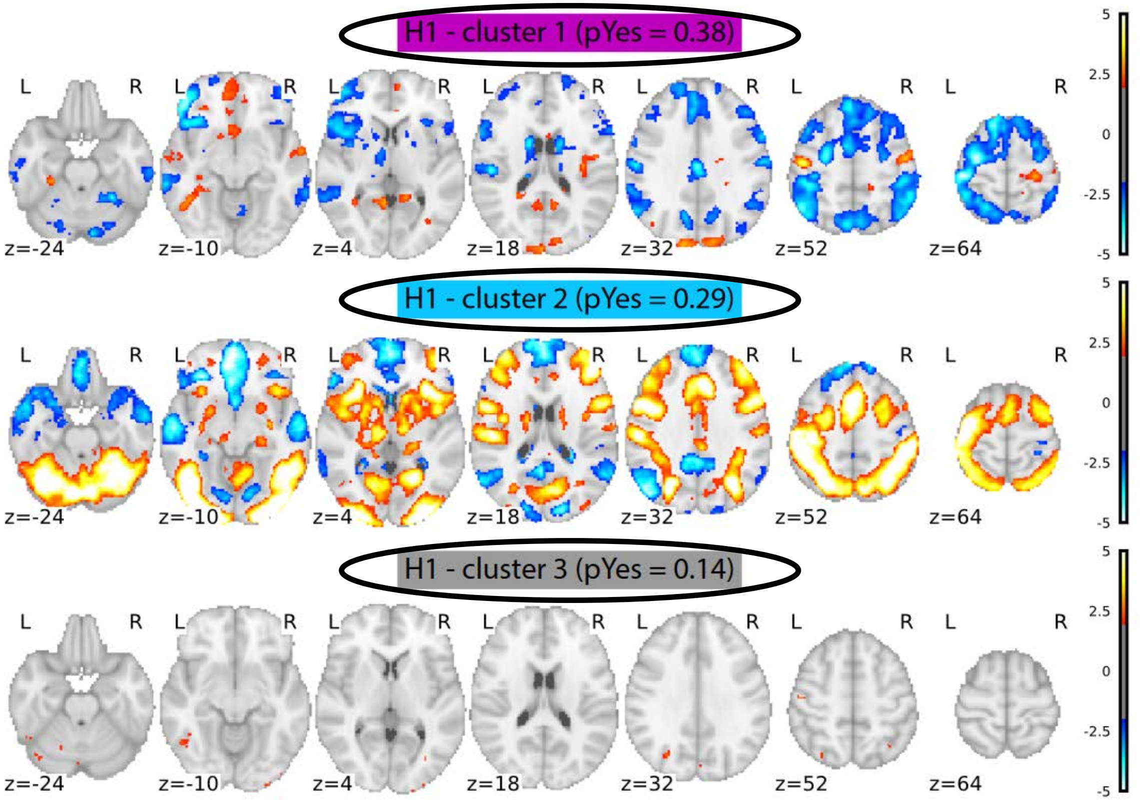
Journée Reproductibilité en Sciences
04/04/2024

Outline

- What is reproducibility?
- Computational reproducibility
- ReproVIP
- Broader overview
- Conclusions and discussions

Reproducibility Issues in Medical Imaging

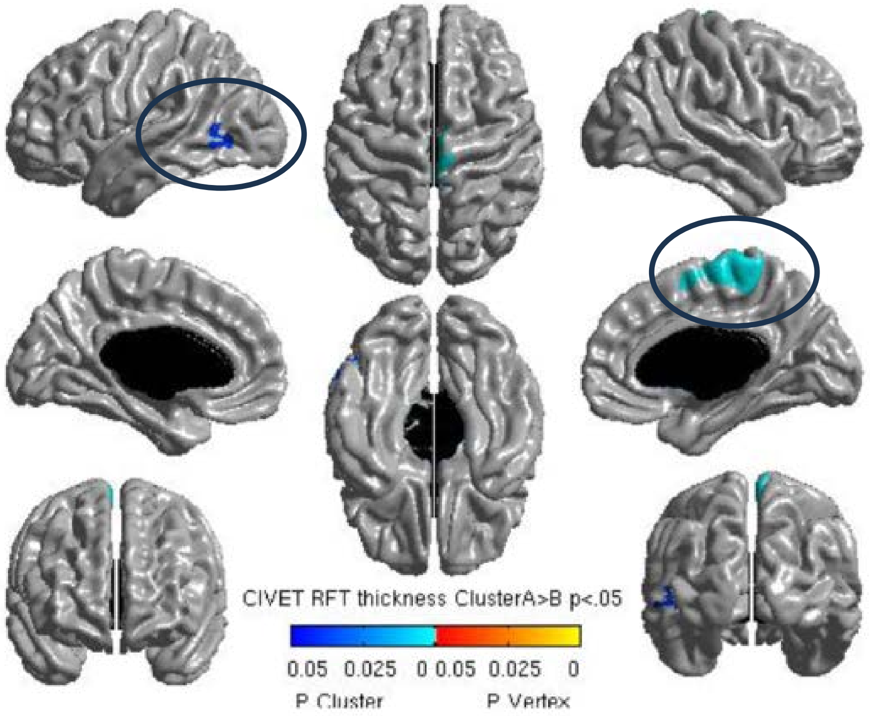
- R. Botvinik-Nezer *et al.*, « Variability in the analysis of a single neuroimaging dataset by many teams », *Nature* 2020
  - ➢ 70 research teams studied the same fMRI dataset, testing 9 hypotheses
  - ➢ 20% of the teams on average gave different results from the rest



Average statistical maps (thresholded at uncorrected z > 2.0) for three groups of teams testing the same hypothesis.

Reproducibility Issues in Medical Imaging

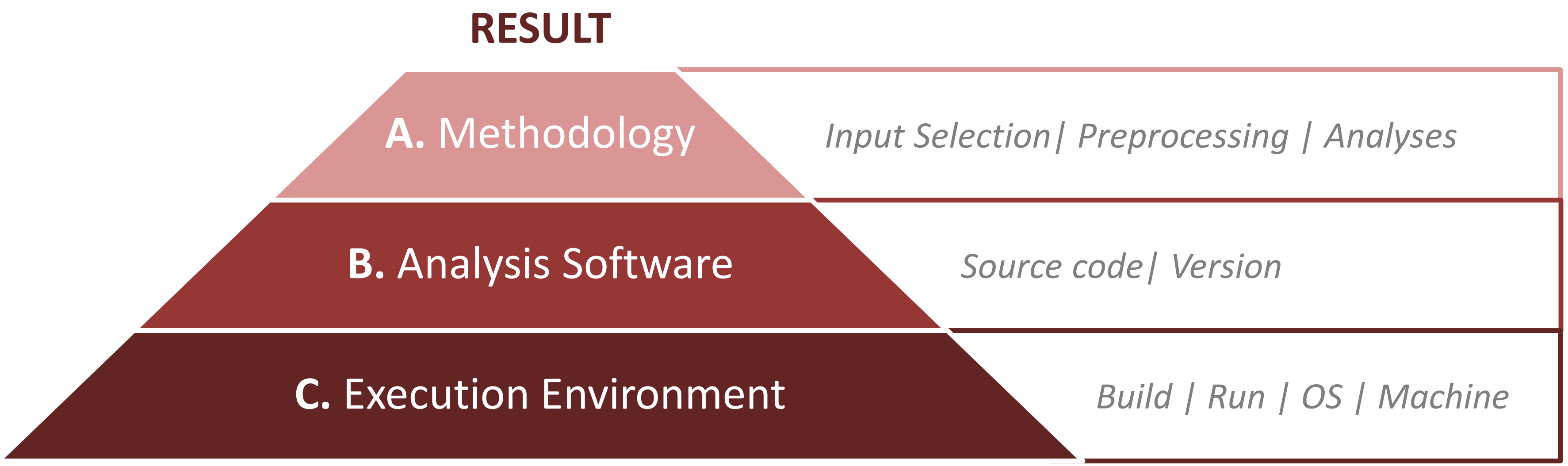
- T. Glatard *et al.*, « Reproducibility of neuroimaging analyses across operating systems », *Front. Neuroinform.* 2015
    - ➢ Same analyzes (*e.g.* cortical thickness) run on 2 computing clusters
    - ➢ Significant differences found between both clusters



Sources of Variability in Medical Imaging

- Botvinik-Nezer et al, 2020: *research teams*
- Glatard et al, 2015: *computing clusters*
- …

➢ Distinct **sources of variability** in medical imaging



Definitions

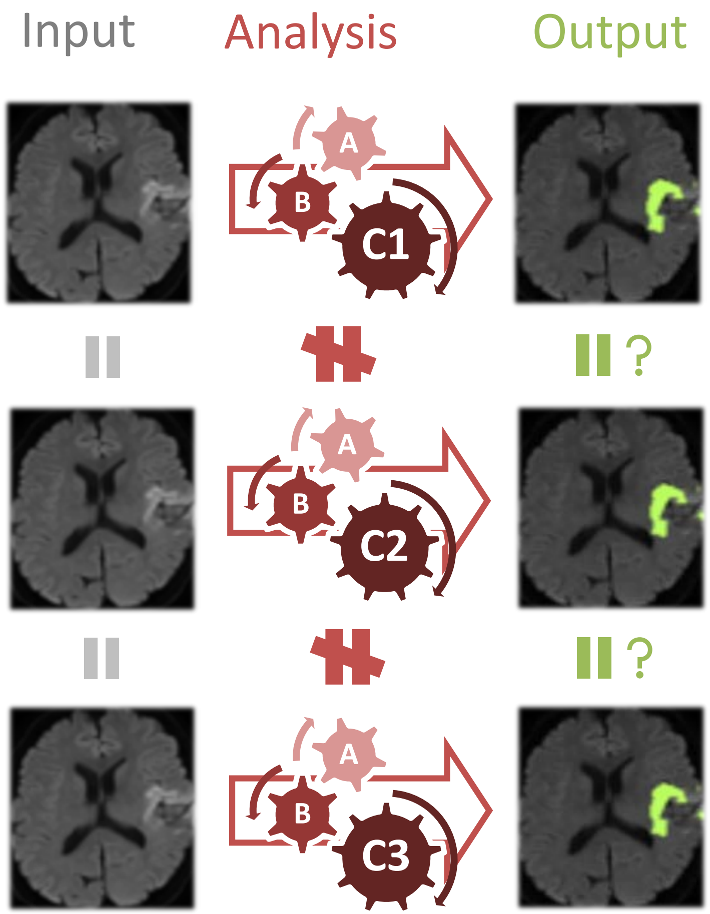
- Spectrum of concerns/terms
  - Lorena A. Barba « Terminologies for Reproducible Research », 2018
- Reproducible research
  - Authors provide all the necessary data and the computer codes to run the analysis again, re-creating the results

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

- Replication
  - A study that arrives at the same scientific findings as another study, collecting new data (possibly with different methods) and completing new analyses

What does «same result» mean?

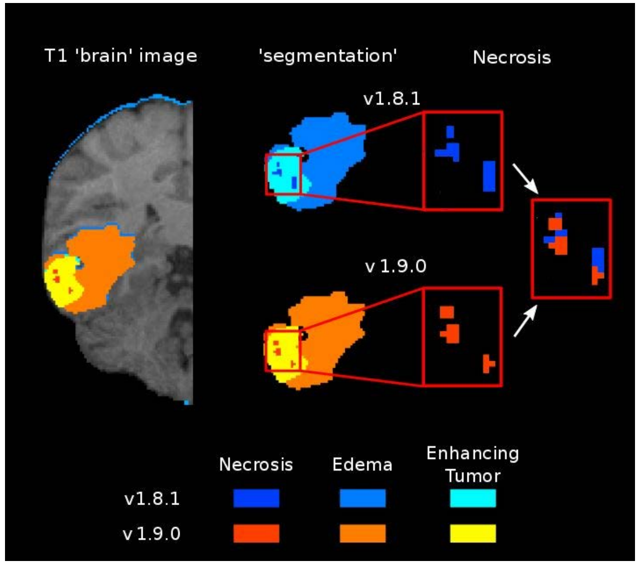
- A published study, e.g. hypothesis testing
  - Figures, conclusions
- A result in particular, e.g. binary file
  - Bitwise reproducibility: checksum
  - Statistical reproducibility: p<0,05
  - Other specific metrics



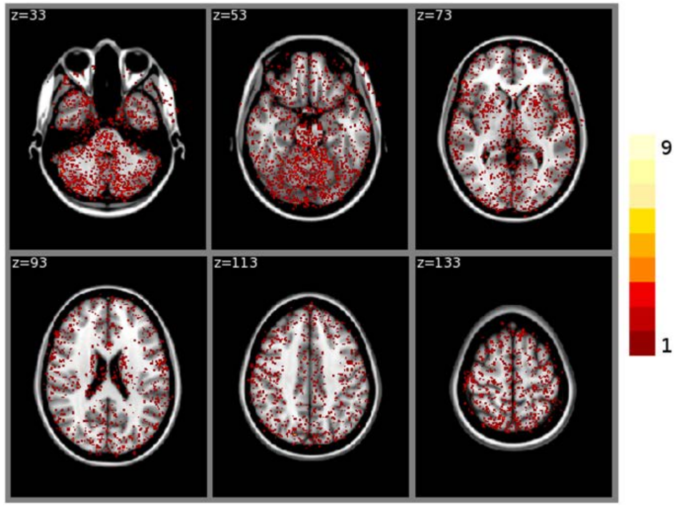
Outline

- What is reproducibility?
- **Computational reproducibility**
- ReproVIP
- Broader overview
- Conclusions and discussions

Sources of variability at the computational level



Differences in tumor segmentation outputs obtained with two different versions of Brain Tumor Segmentation (BraTS) pipeline on the same input image, as presented in [desligneris2023]



expf(1.54051852226257324218750000000)
=4.66700935363769531250000 (glibc 2.5)
expf(1.54051852226257324218750000000)
=4.66700983047485351562500 (glibc 2.18)

Sum of binarized differences between cortical tissue classifications obtained on cluster A (CentOS) and cluster B (Fedora) (FSL FAST, build 1, n = 150 subjects). Credits: Tristan Glatard, https://www.frontiersin.org/articles/10.3389/fninf.2015.00012/full

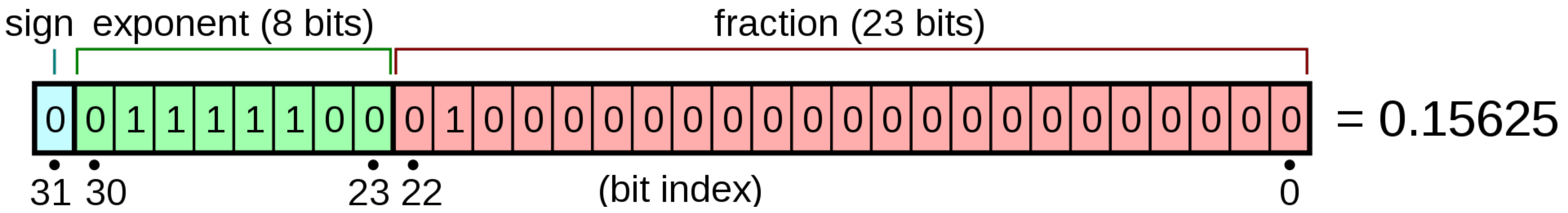
- Software, dependencies and their evolution over time
- Numerical instability related to floating point arithmetic

Floating Point Arithmetic

- Subsets of real numbers using an integer with a fixed precision, called the significand (mantissa), scaled by an integer exponent of a fixed base

$$12.345 = \underbrace{12345}_{\text{significand}} \times \underbrace{10}_{\text{base}}{}^{\overbrace{-3}^{\text{exponent}}}$$

- An example of a layout for 32-bit floating point is



Source: https://en.wikipedia.org/wiki/Floating-point_arithmetic

Floating-Point Operations and Rounding Errors

- Floating-point representation: approximate real numbers within a limited precision => rounding errors

```python
import numpy as np

# Large number
a = 1e16

# Slightly different large number
b = 1e16 + 1

# Expected difference
expected_difference = 1

# Actual difference due to floating-point arithmetic
actual_difference = b - a

print(f"Expected Difference: {expected_difference}")
print(f"Actual Difference: {actual_difference}")
```

Standardization

- IEEE 754 standardizes the representation and behavior of floating-point numbers
- Ensure consistency across different computing platforms and programming languages
- No programming language gives direct access to IEEE 754 operations
- Optimisations can modify the result
- Compiler and compiler options need to be taken into account

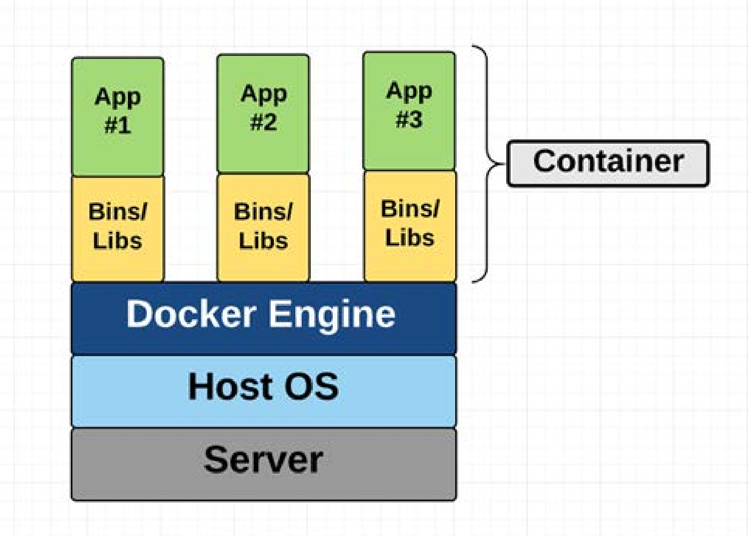
Software dependencies

- Software versions evolve
- Software may have multiple dependencies
- A software = the result of one source code transformed into binary by another software (e.g., compiler)
  - The compiler, also binary, is thus obtained from source by other software, so called built-time dependencies

  ⇒recursive stack of binaries
- Existing solutions for reproducible environments
  - Containers, functional package managers

What are containers?

- A container = an entire runtime environment
  - An application + all its dependencies, libraries and other binaries, and configuration files needed to run it, bundled into one package
  - Differences in OS distributions and underlying infrastructure are abstracted
- Containers and Virtual Machines (VMs) are similar in their goals
  - isolate an application and its dependencies into a self-contained unit
- Docker has become synonymous with container technology, but
  - Container technology is not new
  - Other containers exist: Singularity/Apptainer, Charliecloud …



Guix

- Containers can become black boxes (lacking transparency)
- Guix captures the whole computational environment, controls the complete recursive stack and is able to redeploy anytime
  - https://guix.gnu.org
  - Scheme language
  - https://www.nature.com/articles/s41597-022-01720-9
- Two components
  - the list of software effectively used
  - an identifier committing the complete stack
- Guix can also pack Docker or Singularity images

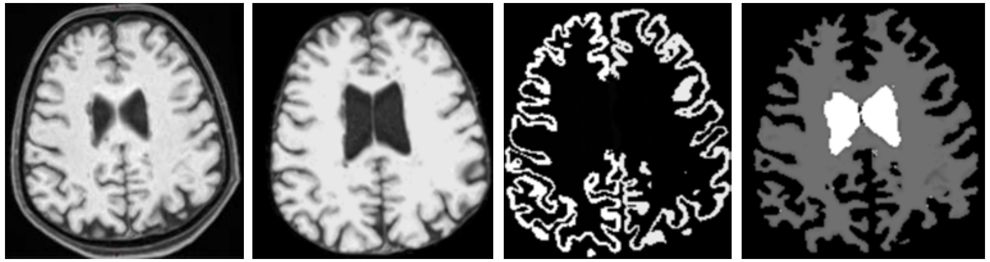
Outline

- What is reproducibility?
- Computational reproducibility
- **ReproVIP**
- Broader overview
- Conclusions and discussions

The Virtual Imaging Platform (VIP)

- Scientific applications as a Service
  - More than 20 applications publicly available
  - https://vip.creatis.insa-lyon.fr/home.html
- Transparent access to computing resources
  - 40 CPU years (EGI biomed VO) used in 2022
- Large community
  - More than 1500 registered users
  - 75 publications since 2011
- Open and reproducible science
  - Zenodo, DOIs, Containers, Boutiques

https://vip.creatis.insa-lyon.fr

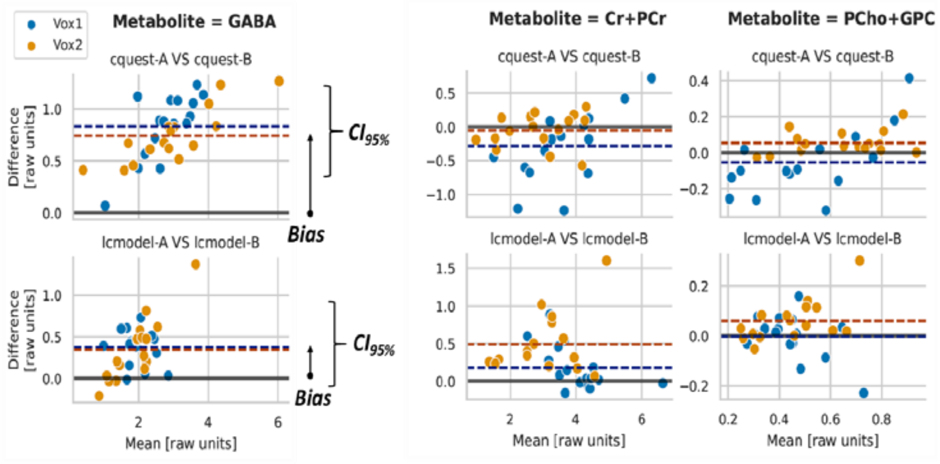


Example of white/grey matter brain segmentation with Freesurfer on VIP
Credits : Berardino Barile and Dominique Sappey-Marinier, Creatis

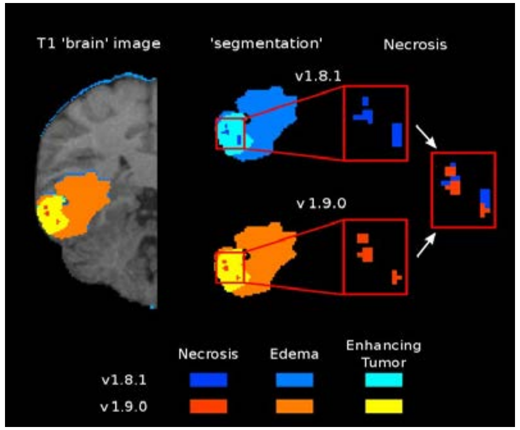
ReproVIP

- Ongoing ANR JCJC project
    - Partners: CREATIS, IPHC, Concordia University
- Main objectives
    - Evaluate and improve the reproducibility of scientific results: same result when the code is executed with the same set of inputs
    - Provide an integrated, end to end solution, allowing to launch reproducible executions in a transparent manner
    - Evaluate the proposed methods and tools on two studies
        - Optimization of the MRI acquisition protocol
        - Optimization of a processing pipeline for brain cancer prediction

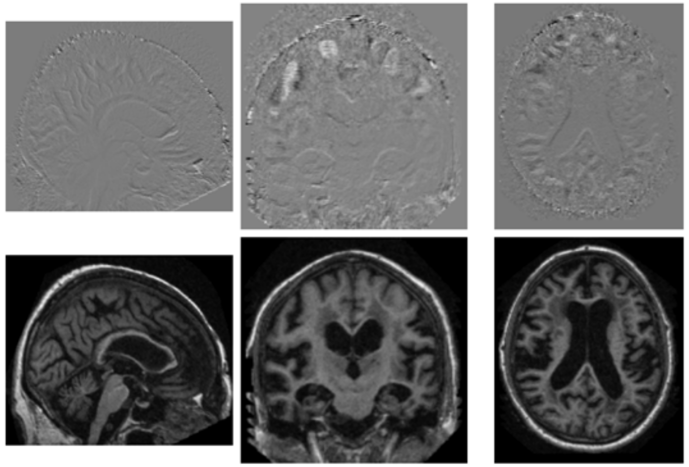
Evaluate and improve the reproducibility of scientific results



Bland-Altman plots showing between-parameter variations in the quantification results of the same software (LCModel, cQUEST) for three metabolites. [Computational Reproducibility in Metabolite Quantification Applied to Short Echo Time in Vivo MR Spectroscopy](), ISBI 2023.



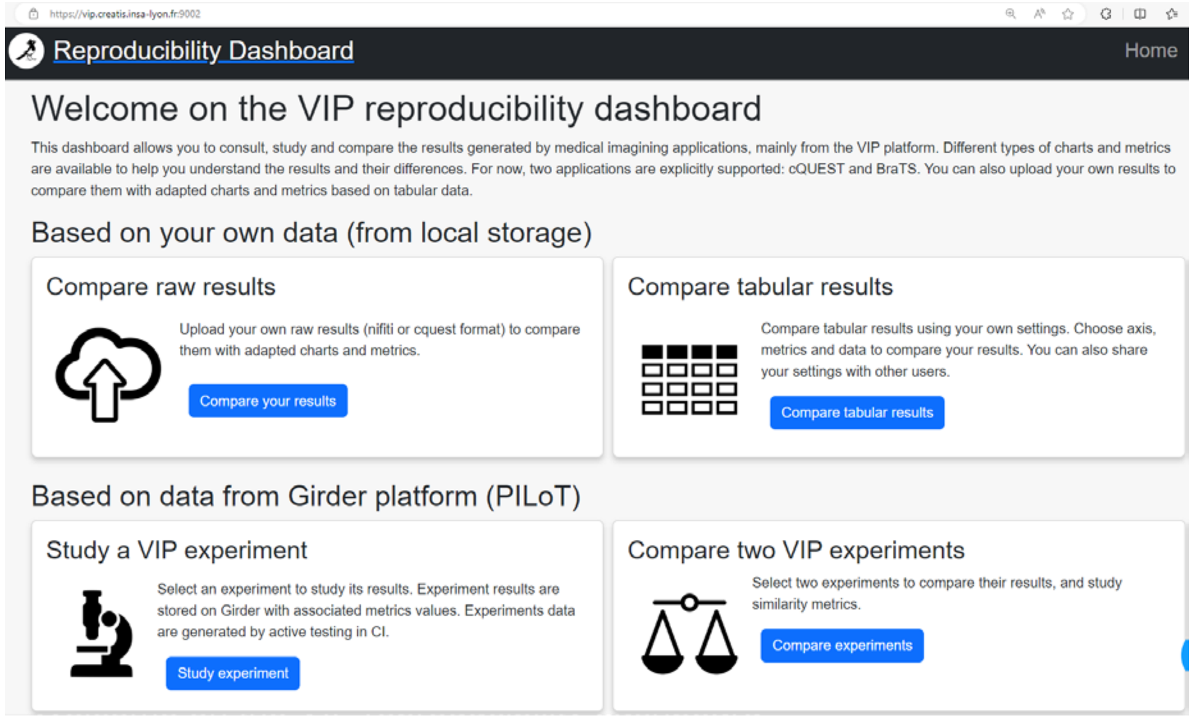
Differences in segmentation results obtained on VIP with two different version of the [BRATS pipeline](). [Reproducibility of Tumor Segmentation Outcomes with a Deep Learning Model, ISBI 2023]().



Sagittal, coronal and axial views of differences in results (at the top) along with the corresponding input file (bottom). [The Impact of Hardware Variability on Applications Packaged with Docker and Guix: a Case Study in Neuroimaging](), submitted in ACM REP'24

Integrated end to end solution

- VIP portal
  - Applications as a service
  - Execution sharing (Zenodo)
- Automation
  - Jupyter Notebooks (templates)
  - Python client, REST API
- Reproducibility Dashboard
  - https://vip.creatis.insa-lyon.fr:9002
- Continuous Integration (CI)
- Integration with storage platforms
  - Girder, Shanoir



ReproVIP reproducibility dashboard

Outline

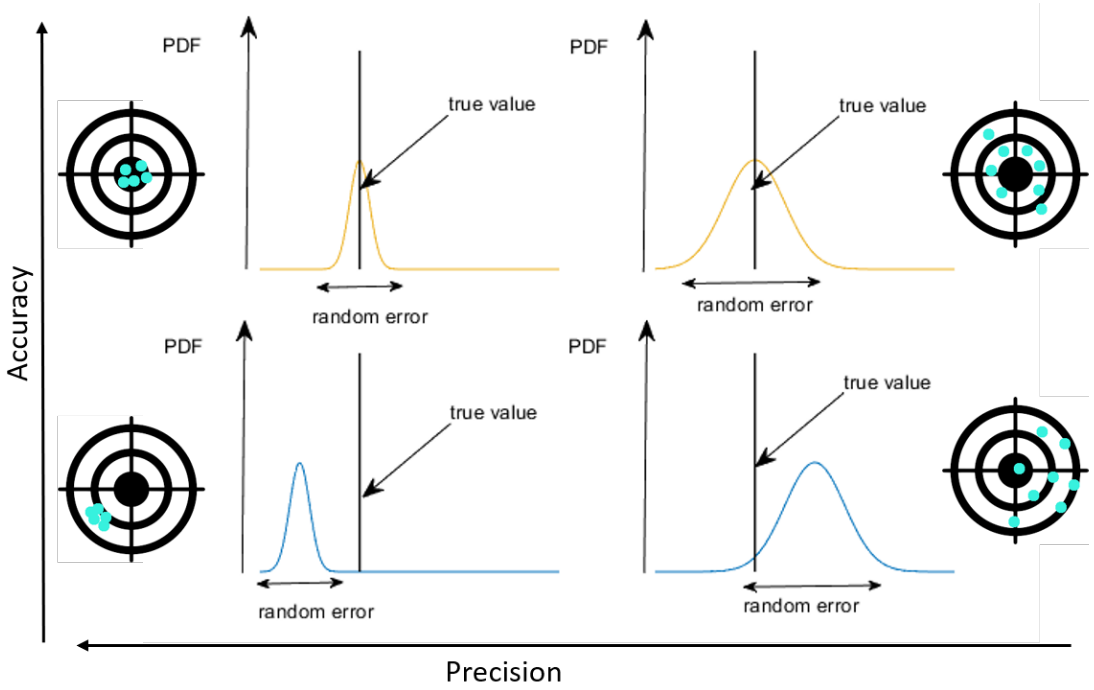
- What is reproducibility?
- Computational reproducibility
- ReproVIP
- **Broader overview**
- Conclusions and discussions

Transparency

- The one practice that can be universally commended is the transparent and complete reporting of all facets of a study, allowing a critical reader to evaluate the work and fully understand its strengths and limitations [Nichols at al, “Best practices in data analysis and sharing in neuroimaging using MRI”, 2017]
- Guidelines
  - Document choices and analyses
  - Use version control systems, such as Git
  - Share code and data whenever possible
- Challenges
  - Ethical and legal problems

Validation

- Continuous (never ending) process
  - Evolving software
  - New databases
- Guidelines
  - Define clear validation objectives
  - Define/use formalised and transparent validation procedures
  - Use standardized open datasets



Outline

- What is reproducibility?
- Computational reproducibility
- ReproVIP
- Broader overview
- **Conclusions and discussions**

Sum-up on computational reproducibility

- Containers help mitigate the extent of environment-introduced variability
  - May lose trace of the build environment
- Reproducible builds with Guix
  - Software development practices that create an independently-verifiable path from source to binary code
- Parallelization or hardware may still lead to different results
- In the long term, software and infrastructures cannot be frozen
- Variability sources need to be taken into account, evaluated and addressed

Take-home messages

- Computational reproducibility
  - Challenging and often over-looked
  - Various, possibly complex solutions
- VIP provides an integrated, end-to-end solution for reproducible executions of scientific applications available in VIP
  - Playground for reproducible experiments
- Reproducible and generalisable software solutions
  - Computational reproducibility is only a small aspect of a larger issue
  - Transparence and validation are also essential

Additional info

- French network
  - http://www.recherche-reproductible.fr
- Fun Mooc
  - https://www.fun-mooc.fr/en/courses/reproducible-research-methodological-principles-transparent-scie/
  - https://www.fun-mooc.fr/en/courses/reproducible-research-ii-practices-and-tools-for-managing-comput/
- Reproducibility tutorials
  - https://www.creatis.insa-lyon.fr/miccai2023 (Hands-on material)
  - https://miccai2023-reproducibility-tutorial.github.io/
- The turing way
  - https://the-turing-way.netlify.app/index.html
- French book « Vers une recherche reproductible »
  - https://hal.science/hal-02144142

Credits and acknowledgements

- ReproVIP parteners anr
  - CREATIS : Hélène Ratiney, Carole Frindel, Claire Mouton, Axel Bonnet, Frédéric Cervenansky, Gaël Vila, Alexandre Cornier, Hippolyte Blot
  - IPHC : Emmanuel Medernach, Jérôme Pansanel
  - Concordia University : Tristan Glatard, Yohan Chatelain
- Reproducibility presentations and tutorials
  - Arnaud Legrand
  - Konrad Hinsen
  - Guix team

Thank you for your attention!
Questions?