Reproducibility in Scientific Research

We're doing it wrong

Marco Esposito

What are we talking about?

- Repeatability: same team, same setting
- Reproducibility: different team, same setting
- Replicability: different team, different setting, can be re-implemented

What reproducibility can do for the others

- Colleagues/students
- Reviewers
- The research community
- The world!

What reproducibility can do for you

- Better control over your workflow
- Higher confidence in your results
- Quicker prototyping
- Think of yourself in five years' time as *a different person*

Tricks of the trade

- Git: Commit *before* experiments and tag the results
- Build a script cascade
  - One "root" `run.sh` script
  - Several child scripts (`figure1.sh`, `exp1.sh`, ...)
- Provide a Docker container image (please!)
- TELL ME YOUR RANDOM SEED
- And infinitely many others
  - `pip` / `conda` / `maven`
  - notebooks
  - Automatic LaTeX report generation

Conclusions

- *Design* your workflow to be reproducible
- Don't be lazy
- YOU FORGOT THE RANDOM SEED

One more thing...

A

Tumor Size Change (%)

100

50

0

-50