Why does reproducibility matter?

Childhood Cancer Data Lab

Have you ever had problems reproducing…

- Someone else's research?
- Your research?
- Both?

Reproducibility vs replication

- Reproducibility
  - Authors provide all the necessary data and computer code to re-run the analysis and re-create the results
  - The exact same data/code are used to re-derive the exact same results
- Replication
  - A separate study arrives at the same scientific findings as another study
  - New data/code and analyses are performed that identify consistent results with previous work

	Same data	**Different data**
Same methods	Reproducibility	Replicability
Different methods	Robustness	Generalizability

Adapted from https://www.repro4everyone.org/resources/intro-reproducibility

Why do we like reproducible science?

- Reproducibility supports…
  - You!
  - Your collaborators and team!
  - Your community!
  - The scientific endeavor!

- Reproducibility makes your funders and journals happy

What are your barriers to reproducible science?

Some barriers others have noticed…

- Technical factors
  - Natural variation
  - Batch effects and sensitivity to equipment or experimental conditions
- Study design and statistics
  - Selective reporting and "P-hacking"
- Human factors
  - Confirmation bias
  - Poor record keeping and/or sharing
- Rewards and incentives
  - Hypercompetition and "publish or perish" culture
  - Paywalls

Adapted from https://www.repro4everyone.org/resources/intro-reproducibility

Lab notebooks in computational biology

- It feels weird to write about what you coded!
  - But it doesn't feel weird to write about your pipetting steps.
  - *But why should we treat these differently?*

Always 👏 Be 👏
Documenting

- Science on the computer is just like science at the bench - you need to *record* your steps, even if they seem obvious at the time!
  - If you'd write it down in a lab notebook, you should write it down for your coding as well
  - Write down things that didn't work too!
  - How do you write it down?

Future You is unimpressed with your code

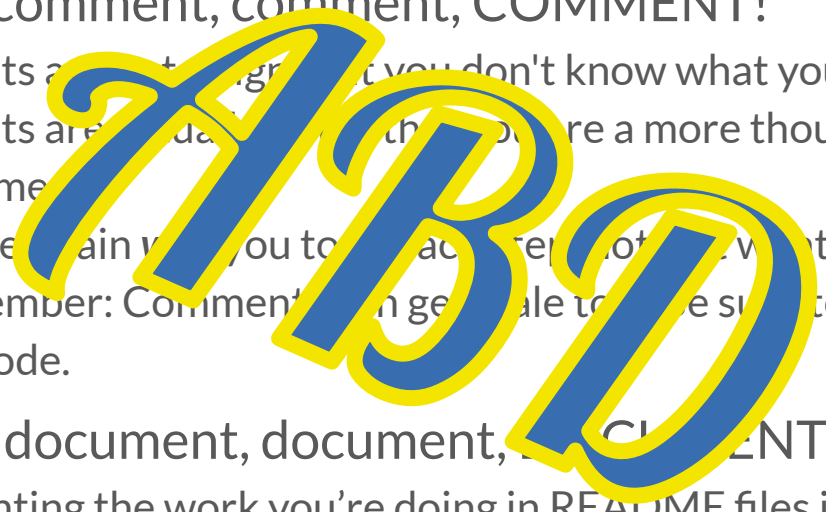
We tend to feel code we are writing is clear and obvious *in the moment we are writing it*.

Future You does not agree! They have no idea what you were thinking and why you were thinking it.

And if Future You doesn't understand what's going on, odds are others won't either.

How can you appease Future You?

- Comment, comment, comment, COMMENT!
  - Comments a[illegible] you don't know what your doing
  - Comments a[illegible]re a more thoughtful (and advanced?) programme[illegible]
  - Strive to e[illegible]ain w[illegible]ou to[illegible]ac[illegible]ep[illegible]ot[illegible]t the step was
  - But remember: Comment[illegible]n ge[illegible]ale to[illegible]e su[illegible]to update them when you update code.
- Document, document, document, [illegible]CU[illegible]ENT!
  - Documenting the work you're doing in README files is (part of) your lab notebook
  - Documenting your code can also help you gain insight into your code and find ways to improve it



Every little bit helps

- Everyone has to start somewhere!
- Nobody's code is perfect!

WORLD VIEW *A personal take on events*

Publish your computer code: it is good enough

Freely provided working code — whatever its quality — improves programming and enables others to engage with your research, says **Nick Barnes**.

https://doi.org/10.1038/467753a

Reproducibility standards for machine learning in the life sciences

To make machine-learning analyses in the life sciences more computationally reproducible, we propose standards based on data, model and code publication, programming best practices and workflow automation. By meeting these standards, the community of researchers applying machine-learning methods in the life sciences can ensure that their analyses are worthy of trust.

Benjamin J. Heil, Michael M. Hoffman, Florian Markowetz, Su-In Lee,
Casey S. Greene and Stephanie C. Hicks

Table 1 | Proposed reproducibility standards

	Bronze	Silver	Gold
Data published and downloadable	x	x	x
Models published and downloadable	x	x	x
Source code published and downloadable	x	x	x
Dependencies set up in a single command		x	x
Key analysis details recorded		x	x
Analysis components set to deterministic		x	x
Entire analysis reproducible with a single command			x

https://doi.org/10.1038/s41592-021-01256-7

ONE CUP AT A TIME CLUB

But remember - It's an honor just to compete!