GRAIL

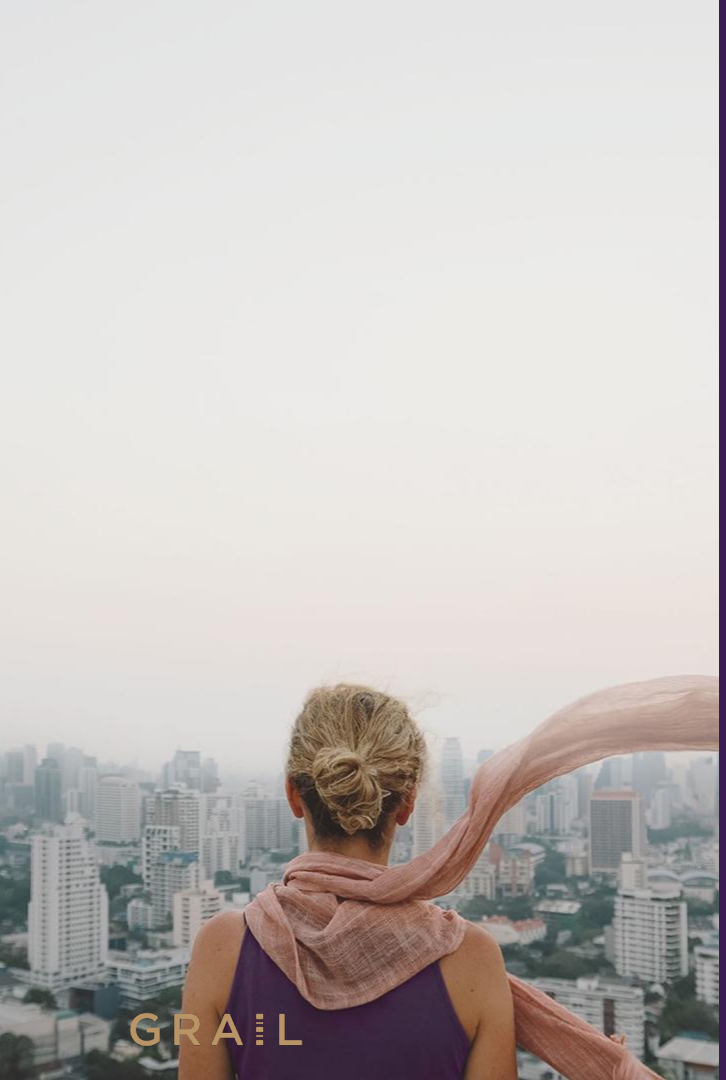
A Revolution in Early Cancer Detection

Interactive and Reproducible Analysis Reports in R

BBSW 2019, Nov 7-8 2019

Daniel Civello
Clinical Data Science

GRAIL



Thank you GRAIL technical staff, current and former, especially:

Siddhartha Bagaria
H John Kim
Razvan Musaloiu-E

Agenda

- Motivation: Communication and Engagement
- Practical Interactive Tools
- Levels of Reproducible Analysis

Motivation

Communication and Peer Engagement.

- Changing how people work together (often at the intersection of software and science) is crucial to building reproducible research
  - More than just tooling or a language

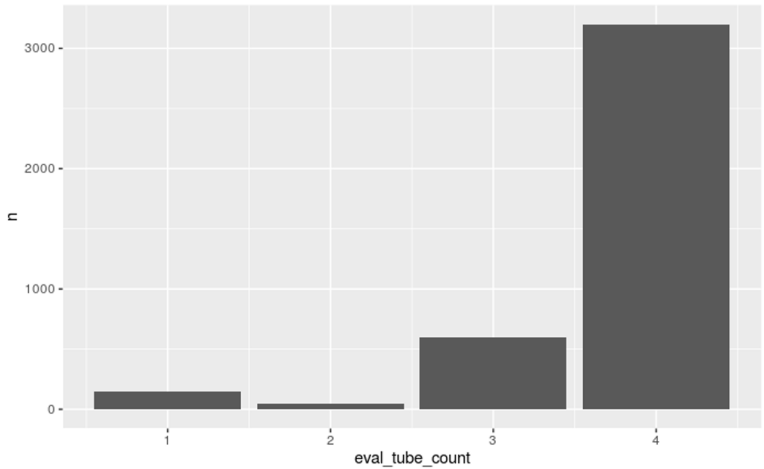
- “Tools that enable going from data to knowledge”

- Adopting tools which streamline how analysis takes place, how decisions are made and how they’re communicated need to be embraced

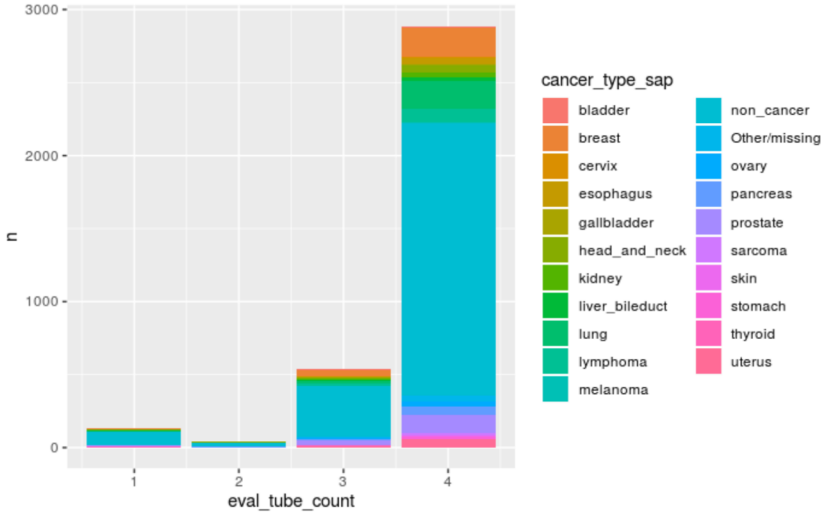
Motivation

Communication and Peer Engagement.

"Can you create a plot of the number of participants in our study and the count of available tubes of blood they have?"



"Okay, great, can you add a breakdown of enrolled cancer status to this?"

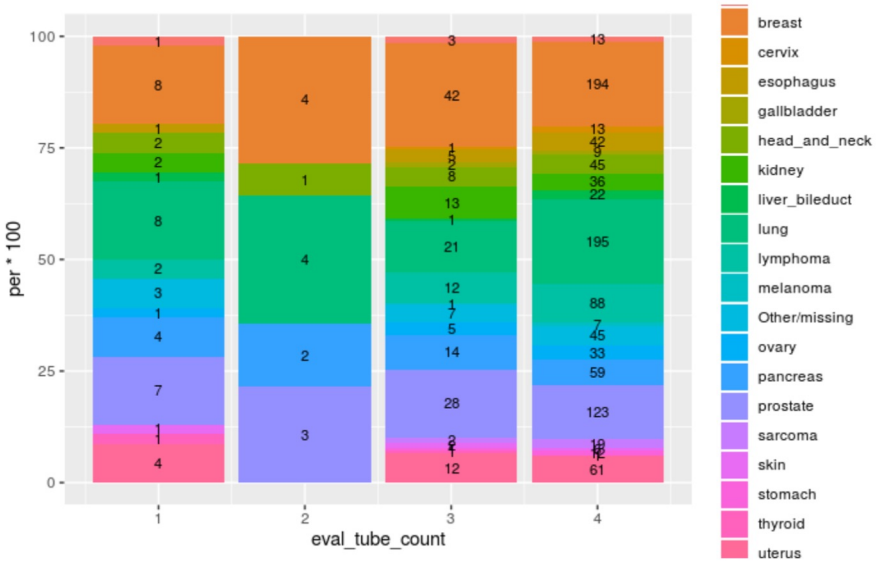


simulated data from GRAIL CCGA study

Motivation

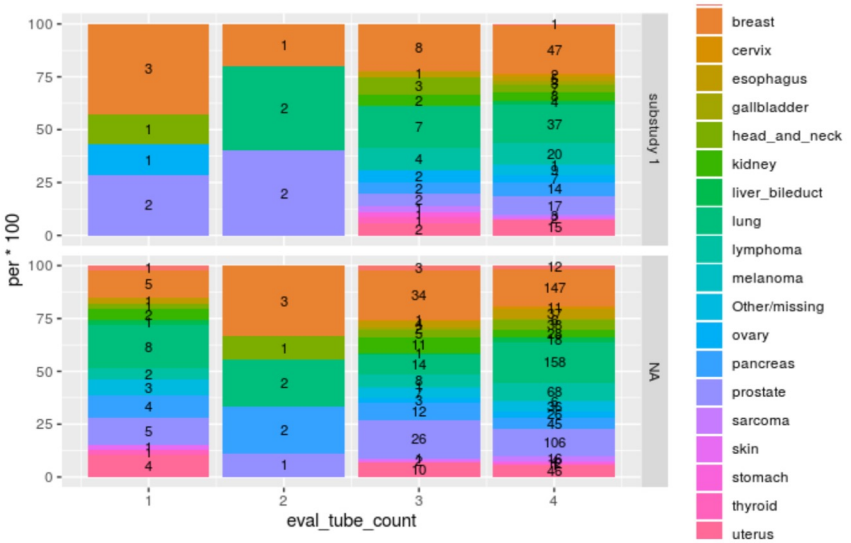
Communication and Peer Engagement.

"I'm not interested in non-cancer, and that one category is too small to see..."



simulated data from GRAIL CCGA study

"Okay, how many of these participants have been analyzed in a previous cohort?"



GRAIL

Motivation

Communication and Peer Engagement.

The Real Question

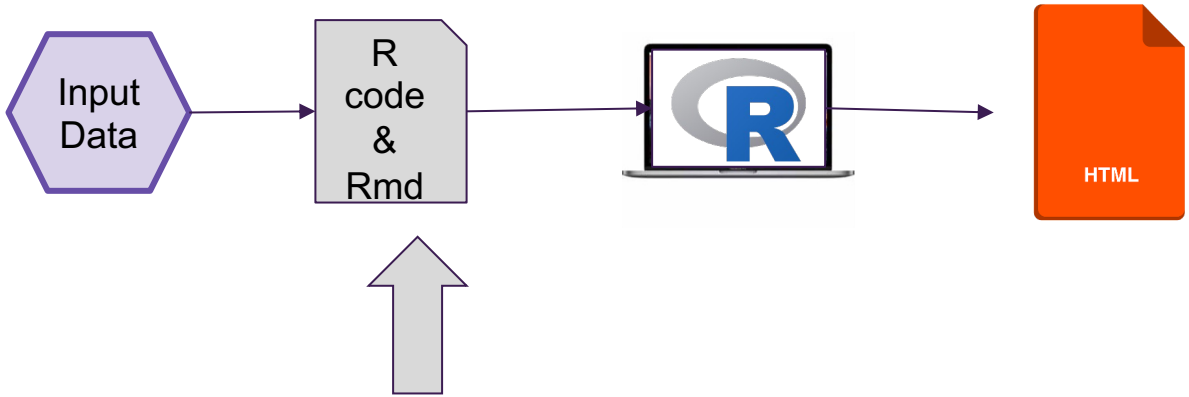
"Is there a correlation between blood collected and the cancer for the participants in our study?"

How far did we get?

Practical Interactive Tools

Interactive Tools

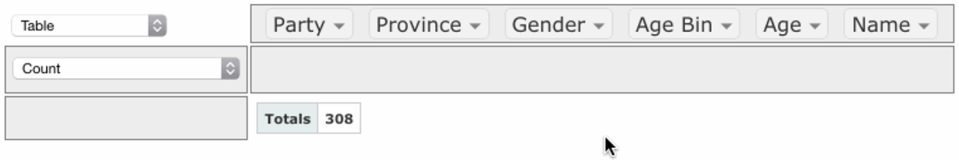
Example workflow for EDA



htmlwidgets

Interactive Tools

Examples rpivotTable, DT, esquisse



Can be embedded into
a markdown!

https://github.com/nicolaskruchten/pivottable/wiki/UI-Tutorial

GRAIL

Interactive Tools

Examples rpivotTable, DT, esquisse

Can be embedded into a markdown!

Blood Shipping Stability Samples with failed OD readings

participant_id	set_id	blood_accession_label
65	A	A00065-A1
69	B	A00069-B2
72	B	A00072-B2
71	B	A00071-B2

Showing 1 to 4 of 4 entries

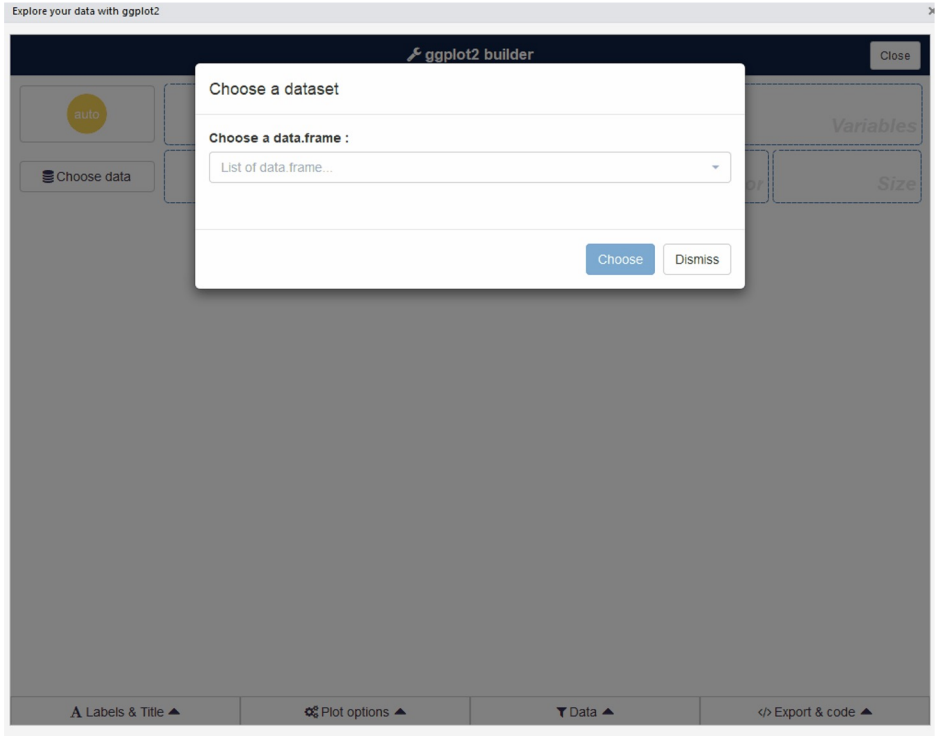
PREVIOUS 1 NEXT

COPY CSV EXCEL

simulated data from GRAIL CCGA study

Interactive Tools

Examples rpivotTable, DT, esquisse



Can be embedded into
a markdown!

https://dreamrs.github.io/esquisse/index.html

GRAIL

Improve Communication and Engagement Through Tools?

Sure!

Shift how people work together (understanding the **why**)

- The correct data can be combined in the first place
- Interactive tools can be used (versus iterative static snapshots) to generate knowledge, which is the ultimate goal.

So maybe now my new title of this talk is:

“An environment fostering collaboration and innovative tooling can turn data into insight”

but I’m forgetting the reproducibility aspect...

Reproducibility
What type of “Reproducibility” do I mean?

Ten Simple Rules for Reproducible Computational Research

1. For Every Result, Keep Track of How It Was Produced
2. Avoid Manual Data Manipulation Steps
3. Archive the Exact Versions of All External Resources Used
4. Version Control All Custom Scripts
5. Record All Intermediate Results, When Possible in Standardized Formats
6. For Analyses That Include Randomness, Note Underlying Random Seeds
7. Always Store Raw Data behind Plots
8. Generate Hierarchical Analysis Output, Allowing Layers of Increasing Detail to Be Inspected
9. Connect Textual Statements to Underlying Results
10. Provide Public Access to Scripts, Runs, and Results

S. Bagaria

https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003285

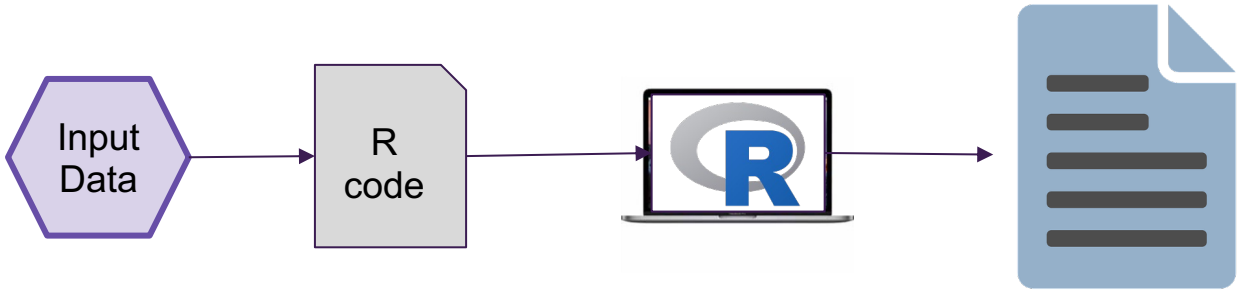
Levels of Reproducible Analysis

Why?

- Not every analysis needs to be publication ready
- Exploratory analysis is OK to be local if working independently
- But, also need a process to "graduate" exploratory analysis
  - To verify findings are reproducible
  - To collaborate internally
  - To present to an internal audience (lab, team, company, etc.)
  - To collaborate externally
  - To publish in journals

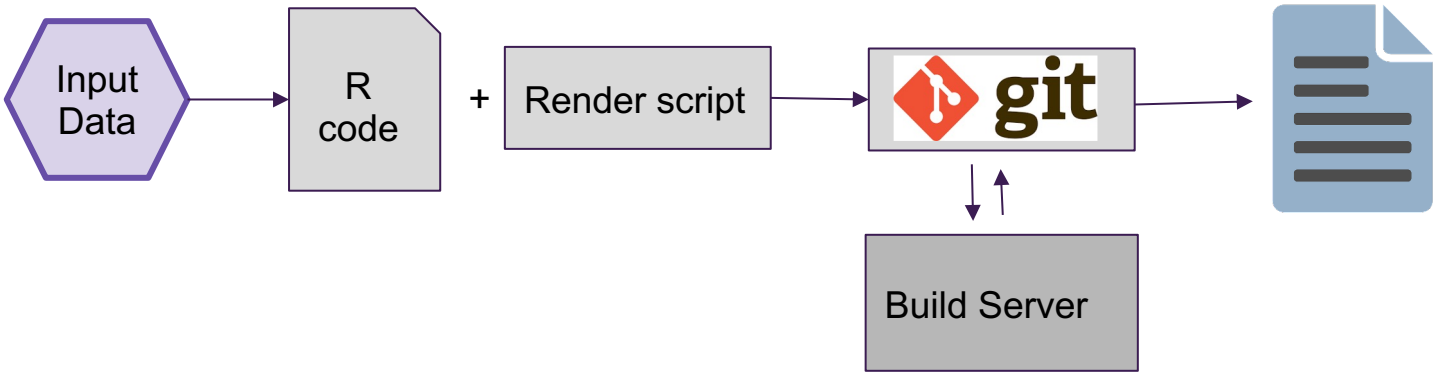
S. Bagaria

Local Rendering is Not Reproducible



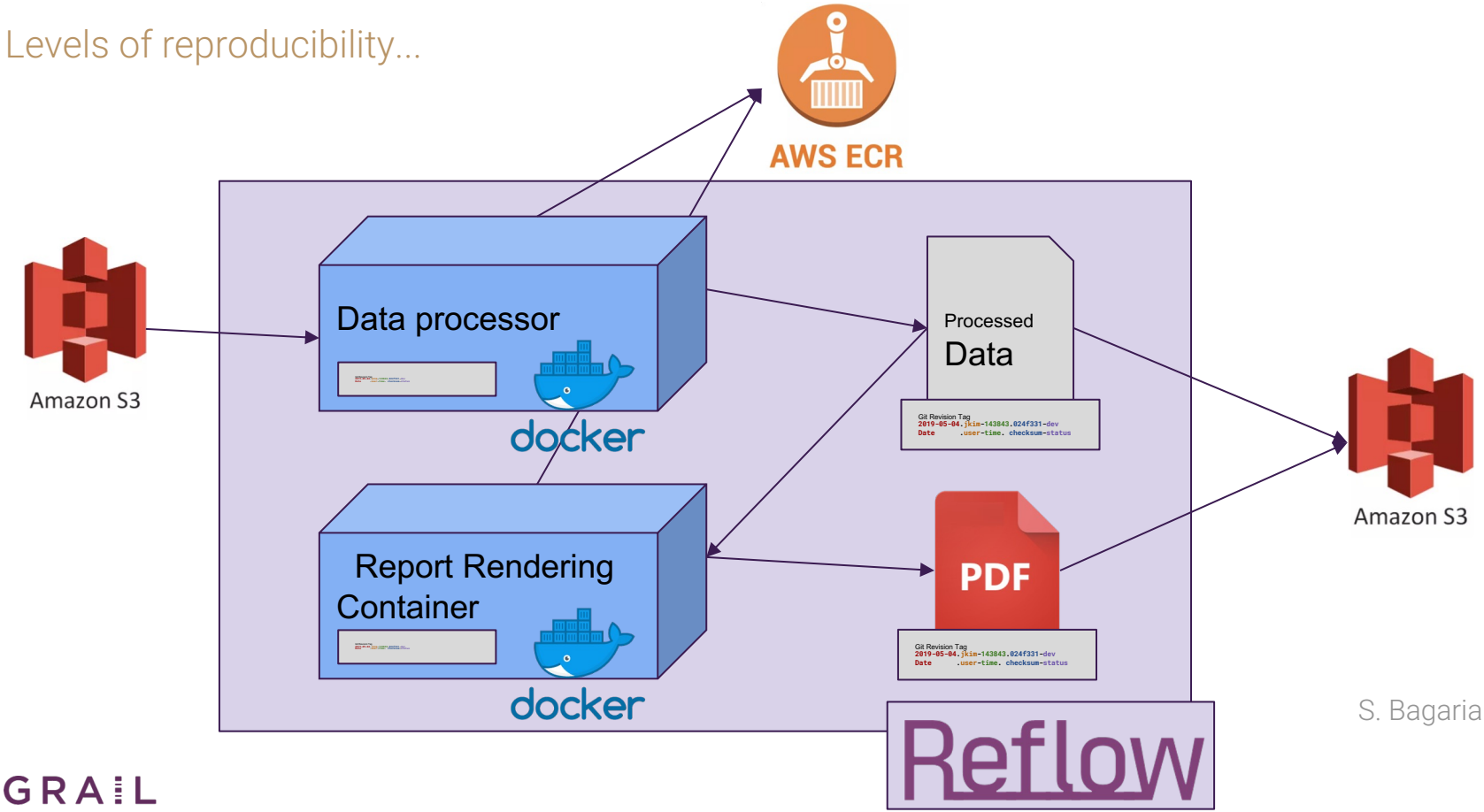
Improving Reproducibility with Tools

- Code hosting
- Build system
  - Common in software development, newer concept to comp. analysis



Full Data Provenance with Reflow™

Levels of reproducibility...



S. Bagaria

Internally at GRAIL

When we use fully reproducible workflows at GRAIL

- For business decisions
- For external publications
- *Considered* for internally presented results (smaller team meetings)
  - When mistakes could be embarrassing or waste time
- *Considered* for continuous maintenance and testing
- *Rarely* for disposable code
  - Analysis never meant to run more than once
  - Exploratory

Summary

If you take away anything from this talk..

1. The science, the analysis, and the engineering all need to partner
2. Tools that are easy to adopt are needed to foster communication
3. Reproducibility needs to be *established at the start* and should not be the exception

Engineering and analysis can be coupled with tools

(and it doesn't add to timelines)

Thank you!

github.com/grailbio

dcivello@grailbio.com