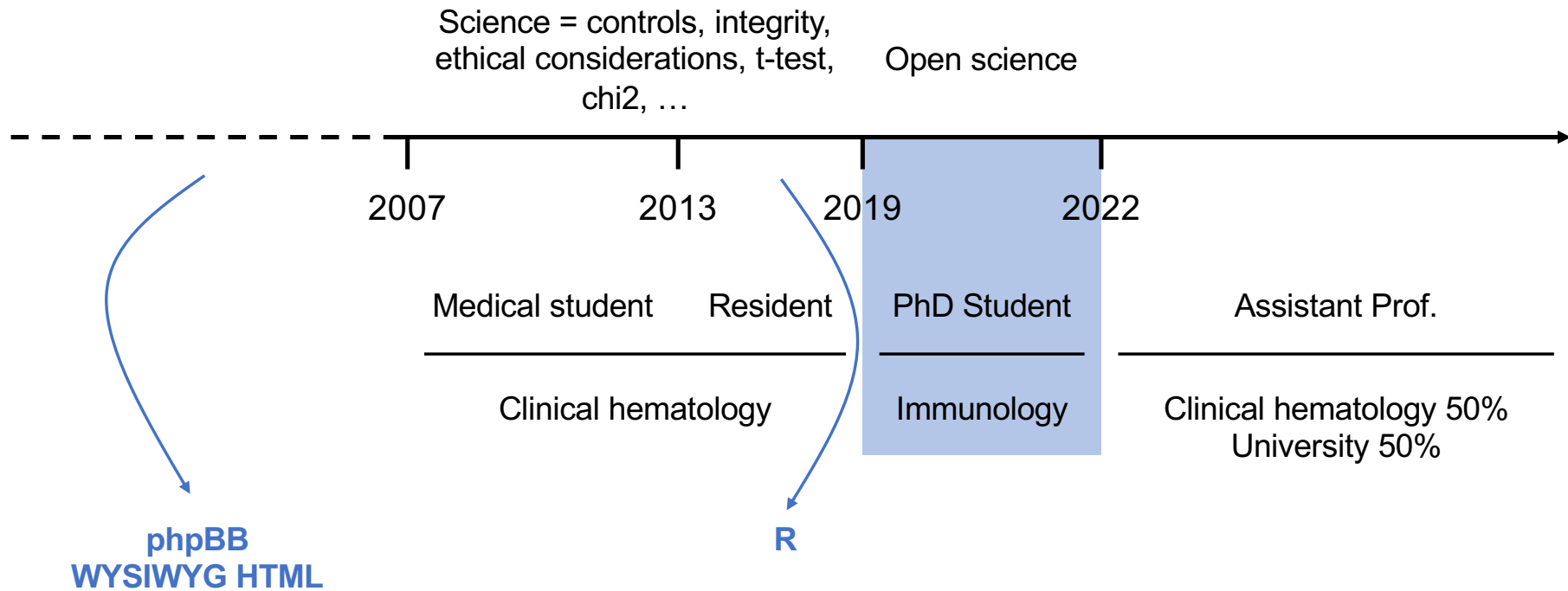




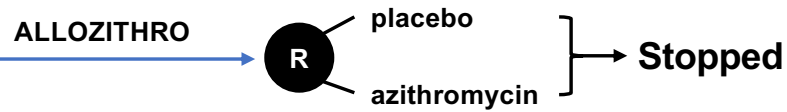
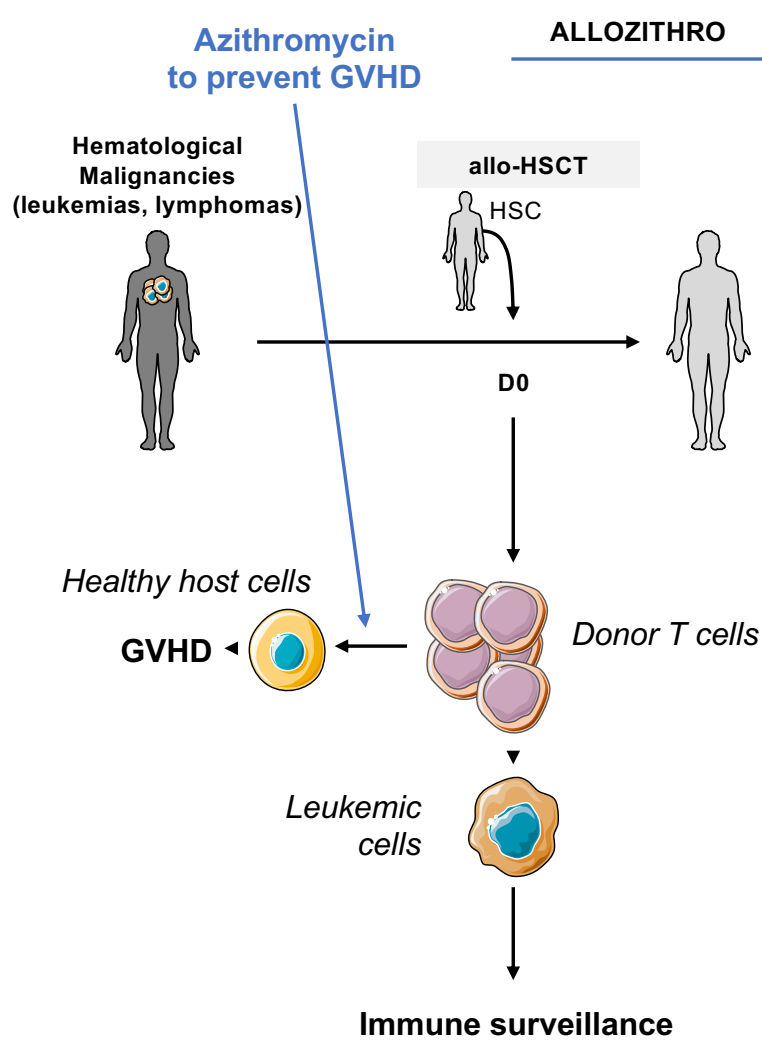
Everyone can learn how to Guix

Dr Nicolas VALLET, MD, PhD
Assistant Professor
Hematology and Cell Therapy Department,
University Hospital Center Tours
Inserm U1069 N2COx

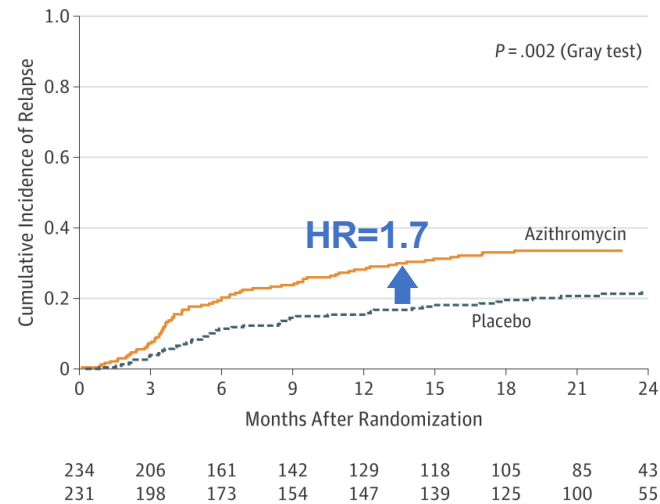
My background



Context



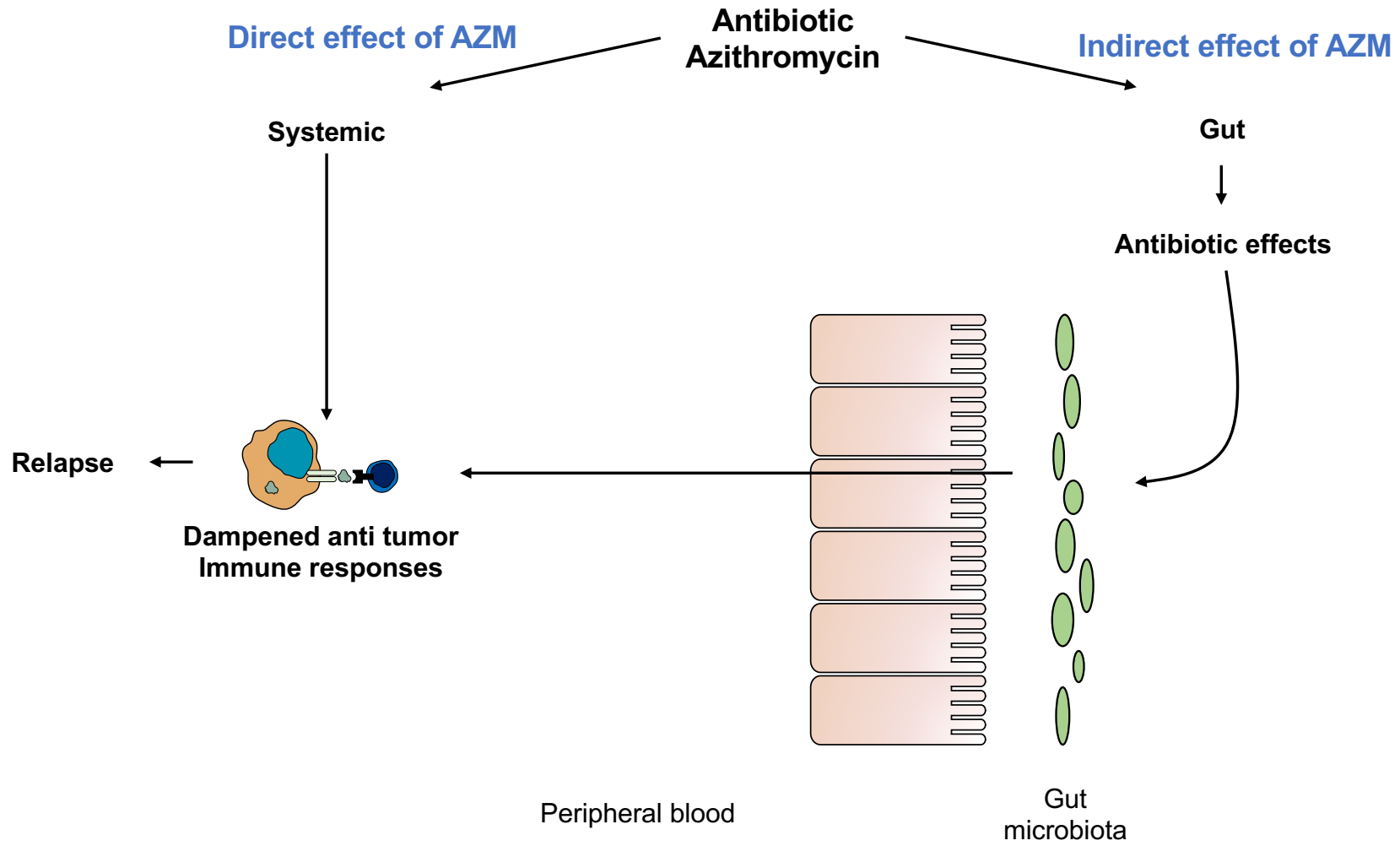
An antibiotic increases risk of relapse



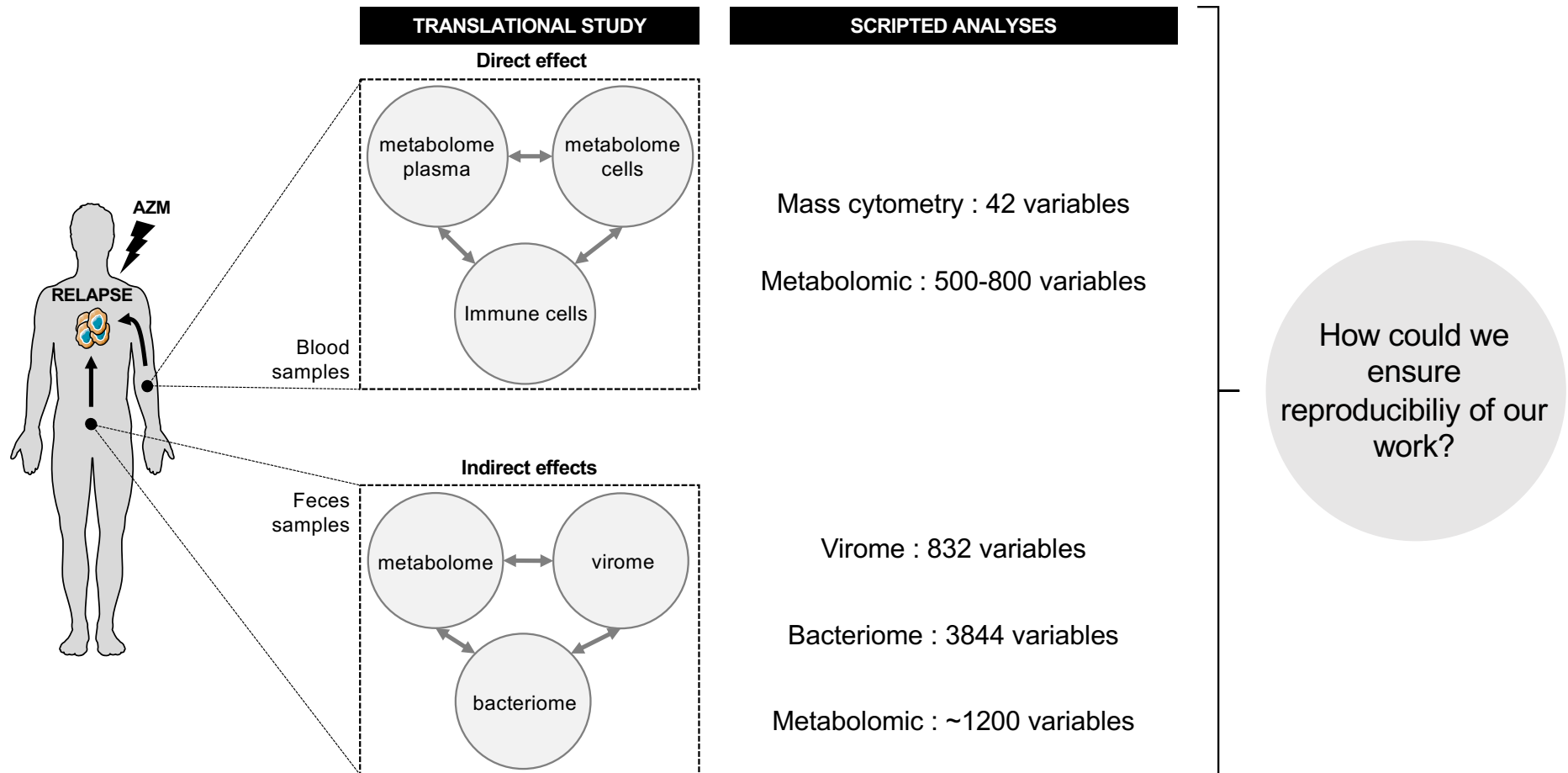
Bergeron A. *et al.* JAMA. 2017

How could an antibiotic dampen antitumor immune responses ?

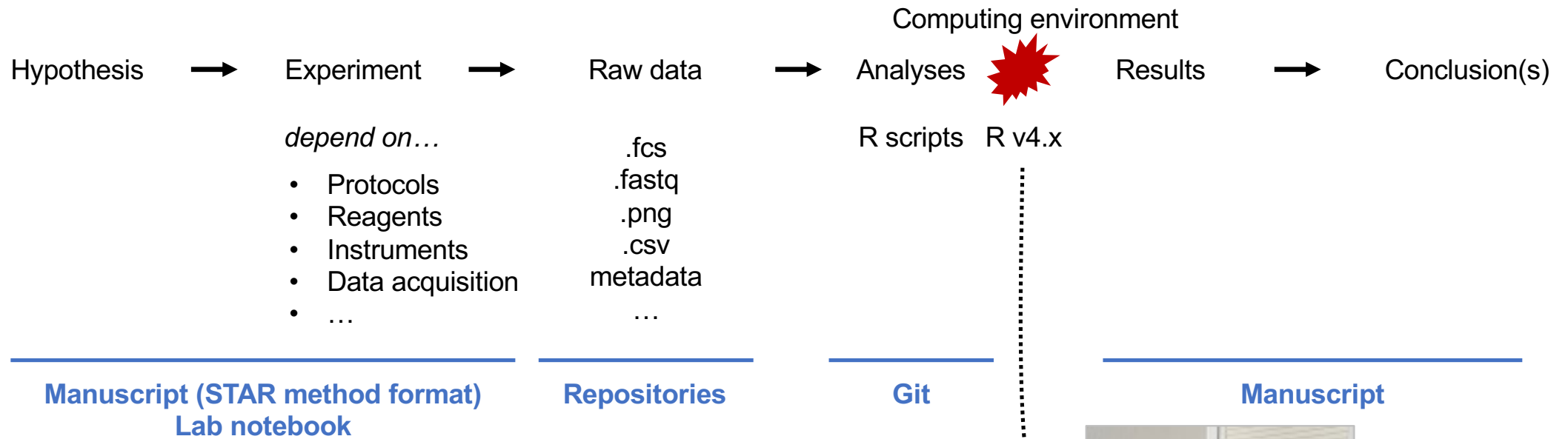
Hypotheses



Methods



Reproducibility planning from the beginning of the project



Is the computing
environment
reproducible/transparent?



Simon Tournier
zimoun

Dependencies hell

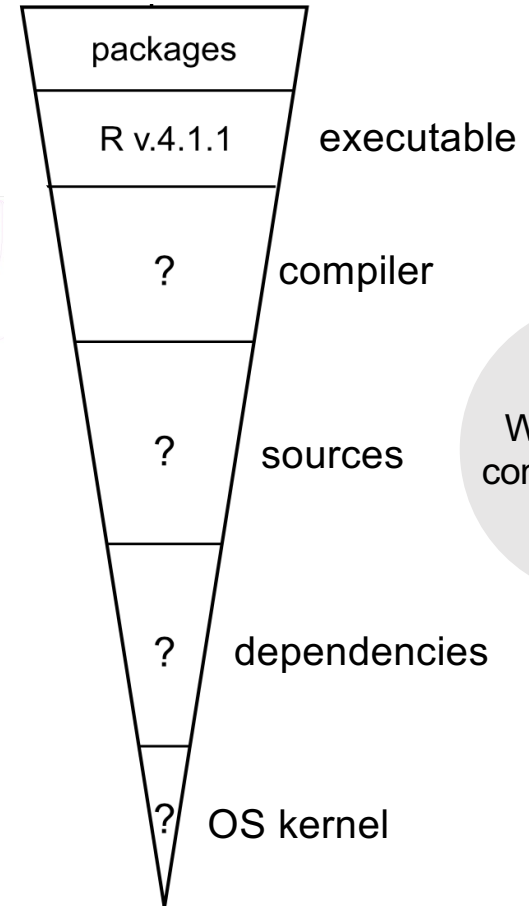
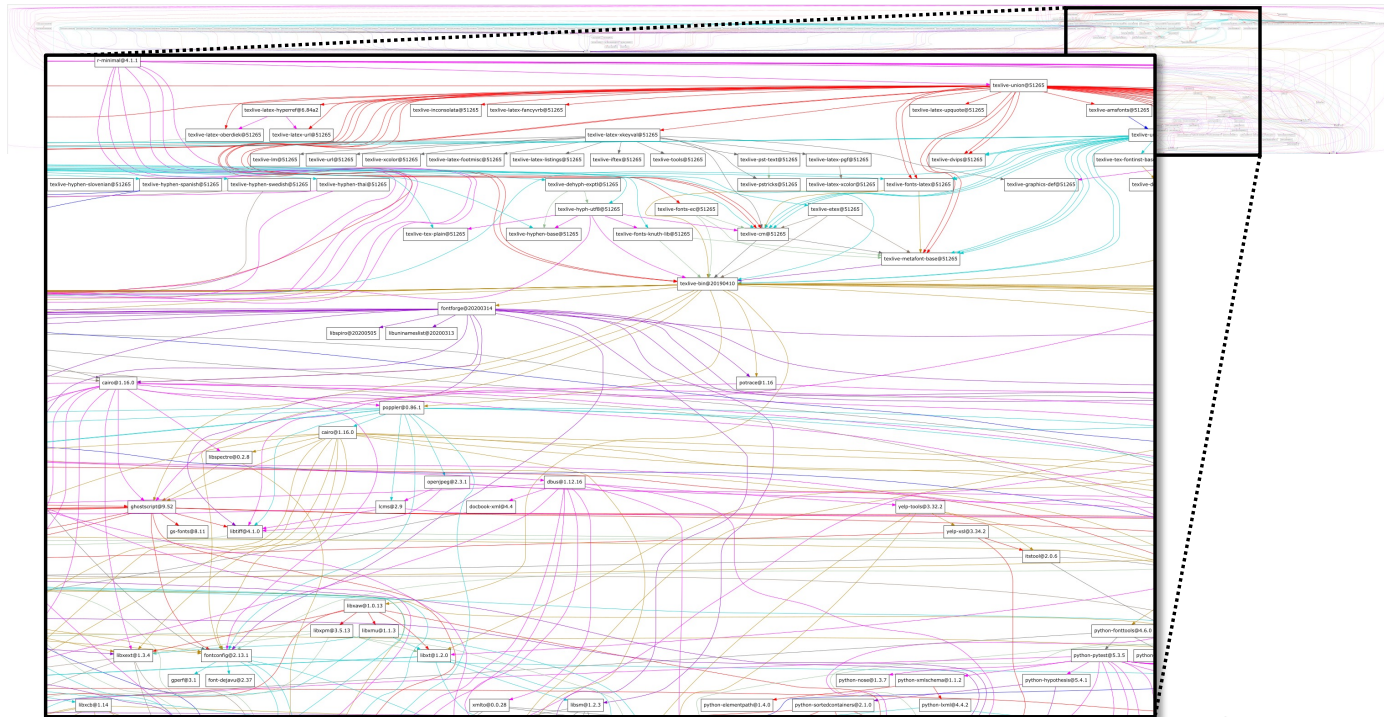
Hypothesis → Experiment → Raw data → Analyses

Computing environment



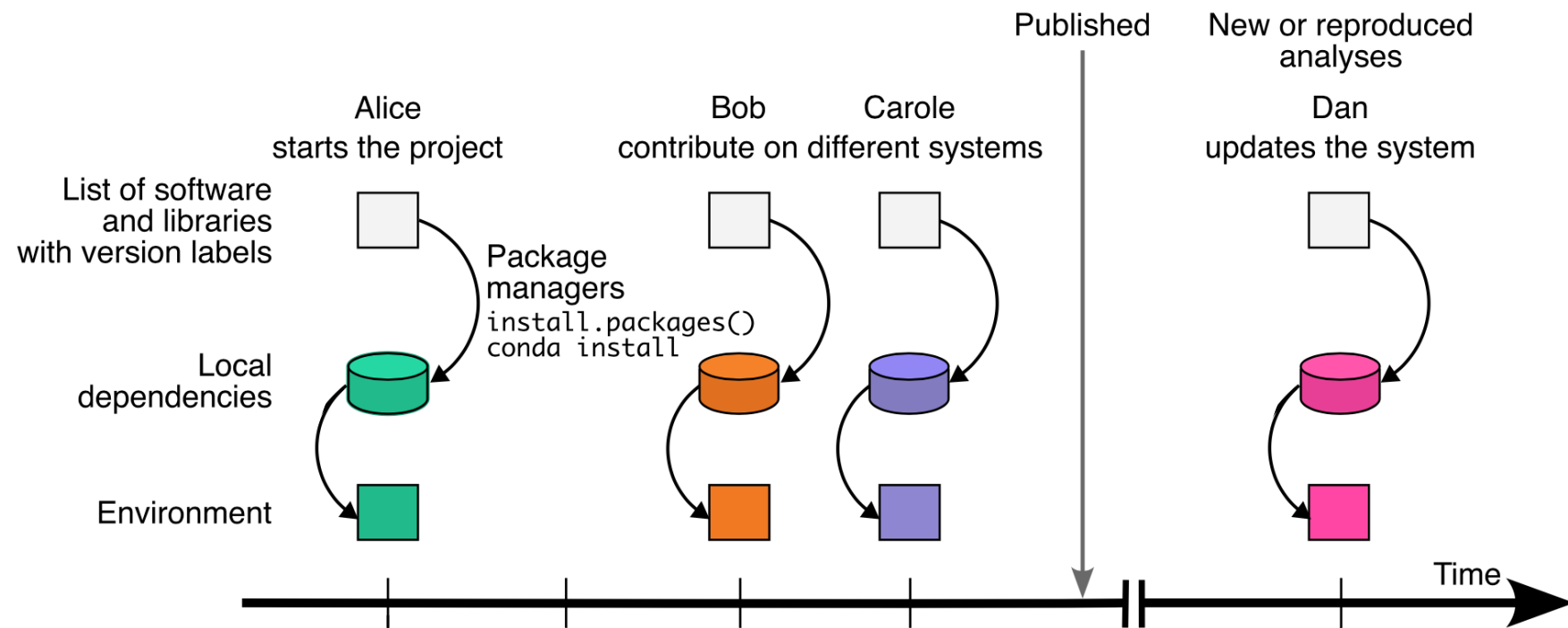
`guix graph r-minimal | dot -Tplain > r-minimal.txt`
`guix graph r-minimal | dot -Tsvg > r-minimal.svg`

R R minimal : 304 dependencies

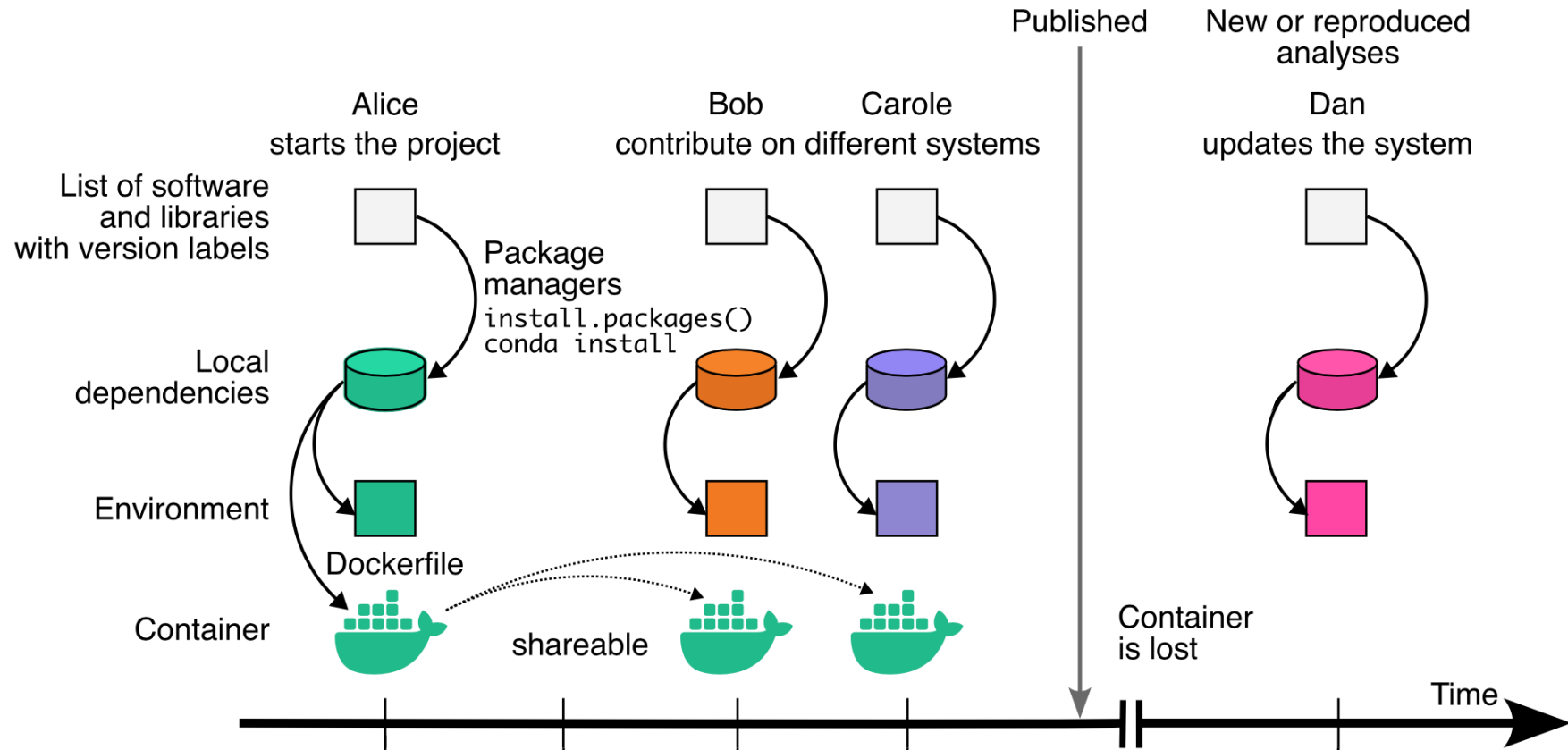


What are the consequences?

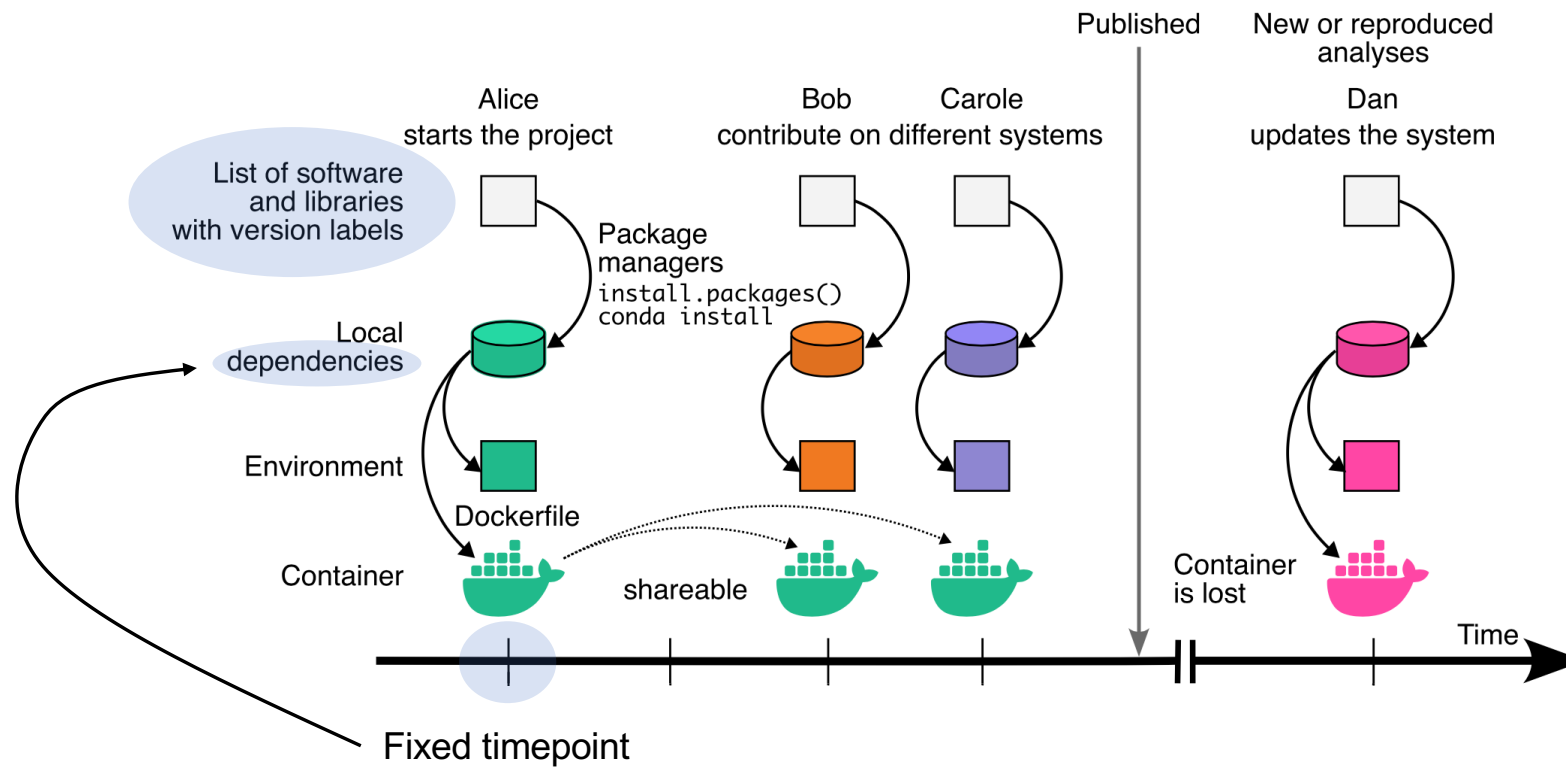
Different computing environments



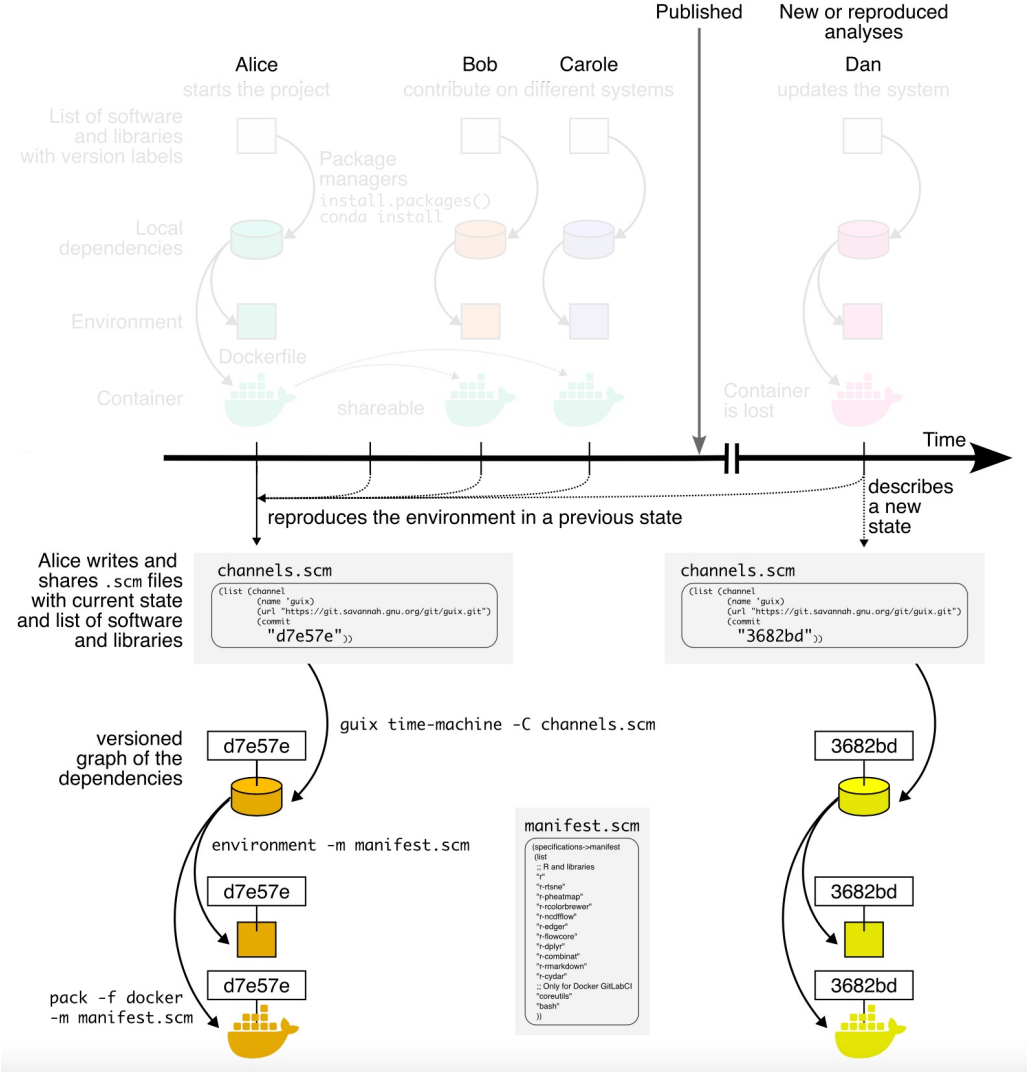
Containers are not enough



Minimum required data to reproduce an environment

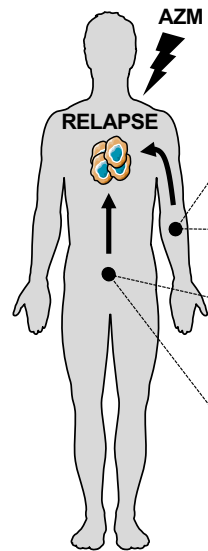


Using Guix to save, share and reproduce the environment



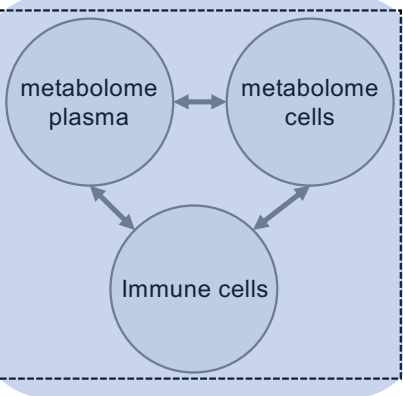
Guix in our workflow

Environment 1

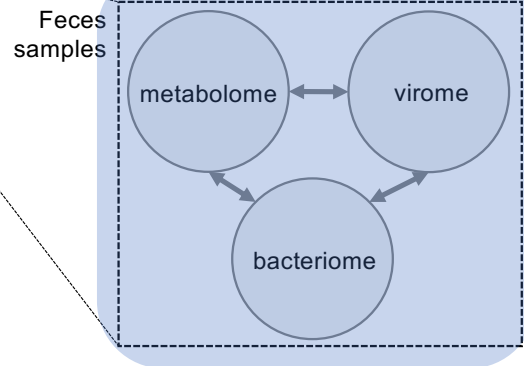


TRANSLATIONAL STUDY

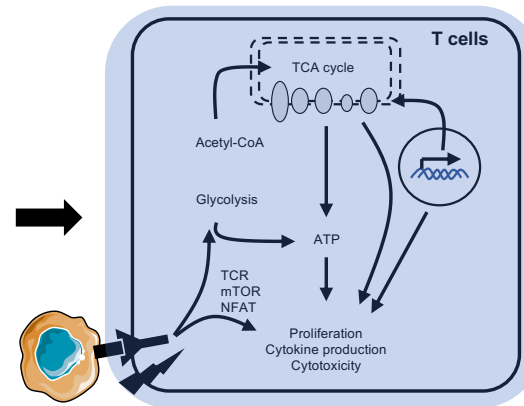
Direct effect



Indirect effects



IN VITRO STUDY



Environment 3

Environment 2

Guix config directories in Git repositories

<https://gitlab.com/nivall/azimutfeces>
<https://gitlab.com/nivall/azimut-in-vitro>
<https://gitlab.com/nivall/azimut-blood>

master

azimutfeces / guixconfig

Historique

Rechercher un fichier

↓

Nom	Dernière validation	Derniè
..		
my-pkgs	coda4microbiome added to guixconfig	
README.md	Add README files	
channels.scm	Updated Guix configuration	
manifest.scm	Added new packages (WIP)	
runguix.sh	first config guix for reproducibility \o/	

README.md

GUIX configuration

This folder contains all files necessary to generate the environment from GUIX See : <https://hpc.guix.info/> and <https://guix.gnu.org/>

- channels.scm : gets the commit number to get reproducible versions of the environment
- manifest.scm : gets the packages stored in GUIX
- my-pkgs.scm : gets the packages with reproducible versions that are not stored in GUIX

Run either the `runguix.sh` file or use the commande line below to load the environment

```
guix time-machine -C channels.scm -- environment -m manifest.scm -- R
```

channels.scm 617 o

```
1 ;; This file point to the Guix channel.
2 (list (channel
3       (name 'guix)
4       (url "https://git.savannah.gnu.org/git/guix.git")
```

1b8d32822952eaf6a1683bb0b6dba0")))

manifest.scm 496 o

```
1 (specifications->manifest
2   (list
3     ;; upstream packages
4     "r"
5     "r-tidyverse"
6     "r-vegan"
7     "r-factoextra"
8     "r-ggplot2"
9     "r-cowplot"
10    "r-ggrepel"
11    "r-factominer"
12    "r-rstatix"
13    "r-rcolorbrewer"
14    "r-ape"
15    "r-phyloseq"
16    "r-complexheatmap"
17    "r-ggraph"
18    "r-igraph"
19    "r-cmprsk"
20    "r-prodlim"
21    "r-survminer"
22    "r-venndiagram"
23    "r-viridis"
24    "r-nnet"
25    ;; defined in my-pkgs
26    "r-khroma"
27    "r-ggalt"
28    "r-gtsummary"
29    "r-gt"
30    "r-visnetwork"
31    "r-riskregression"
32    "r-coda4microbiome"
33  ))
```

126d5d949dc0b8d7c7618142")
roductio
096b48283debdccfea34bad"
print
EA F6A8 0D1D E643 A2A0 6DF2 A33A 54FA")))))

my-pkgs.scm 17.86 Kio

```
1 (define-module (my-pkgs)
2   #:use-module ((guix licenses) #:prefix license:)
3   #:use-module (gnu packages statistics)
4   #:use-module (gnu packages gcc)
5   #:use-module (gnu packages bioconductor)
6   #:use-module (gnu packages bioinformatics)
7   #:use-module (gnu packages cran)
8   #:use-module (gnu packages commencement)
9   #:use-module ((gnu packages compression) #:prefix compression:)
10  #:use-module (gnu packages geo)
11  #:use-module (gnu packages guile)
12  #:use-module (gnu packages pkg-config)
13  #:use-module (guix git-download)
14  #:use-module (guix packages)
15  #:use-module (guix build-system r)
16  #:use-module (guix download)
17  #:use-module (guix licenses)
18  )
19
20 ;; khroma
21 (define-public r-khroma
22   (let ((commit "c7eaf45d5127a3807e3bc7a648dd463d27030361") (revision "1"))
23     (package
24       (name "r-khroma")
25       (version (git-version "1.9.0.9000" revision commit))
26       (source
27         (origin
28           (method git-fetch)
29           (uri (git-reference
30                 (url "https://github.com/tesselle/khroma")
```

Guix in our research papers

Data and code sharing: Raw data are available in the following public repositories: (1) mass cytometry: FlowRepository FR-FCM-Z5ZB and FR-FCM-Z5L7; (2) metabolomic: Metabolights MTBLS406; (3) single-cell RNA sequencing: GEO GSE197658 and GSE208399. Analysis pipelines are available in a Git repository: <https://gitlab.com/nivall/azimut-blood>; <https://gitlab.com/nivall/azimut-in-vitro>; <https://gitlab.com/nivall/azimutscma>.

Long term preservation with Software Heritage

Computational environment

`manifest.scm` and `channels.scm` files are available in `guixconfig` directories of each git repositories to reproduce computational environment of FlowSOM clustering and single cell RNA sequencing analyses with GNU Guix (guix.gnu.org). Software: Manuscript figures were created from R outputs and combined with Inkscape (version 1.1.1) on macOS (version 11.5.1).

Vallet et al. Blood. 2022

Git repositories version at submission time are archived on Software Heritage

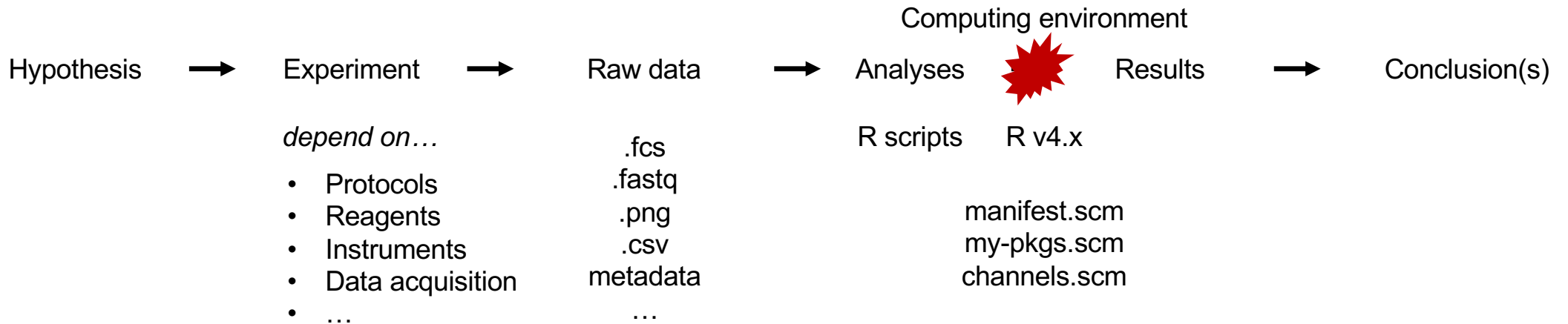
- (i) <https://archive.softwareheritage.org/swh:1:dir:16a829f98f9d8d707343b3b48bc13f5d641998d9>,
- (ii) <https://archive.softwareheritage.org/swh:1:dir:cc8aec0e0c85d7ea70979581e766353d3e2fec22>,
- (iii) <https://archive.softwareheritage.org/swh:1:dir:68ec62f3d8f0cc9f42a15449f75844df12f65934>

Data and code availability

Raw data are available on a public repository: (i) metabolomic, Metabolights: MTBLS406; (ii) bacteriome, Bioproject: PRJNA902819; (iii) virome, Bioproject: PRJNA898682. Source codes, data and CSV files of interim analyses are available on the following Git repositories: <https://gitlab.com/nivall/azimutfeces>. Git repositories version at submission time are archived on Software Heritage: <https://archive.softwareheritage.org/swh:1:dir:9a24ab367cbe9519007f0dd407423fe47def5cfa>. Unless otherwise specified, data analyses were computed using R environment. Raw figures were made with the “ggplot2” R package. Computational environment may be reproduced using GNU Guix with the files “manifest.scm” and “channels.scm” in “guixconfig” directory of the git repository.⁸⁶ Manuscript figures were created from R outputs and combined with Inkscape (version 1.2.1) on macOS (version 12.5).

Vallet et al. Cell Host Microbe. 2023

Reproducibility planning from the beginning of the project



Manuscript (STAR method format)
Lab notebook

Repositories

Git



Manuscript

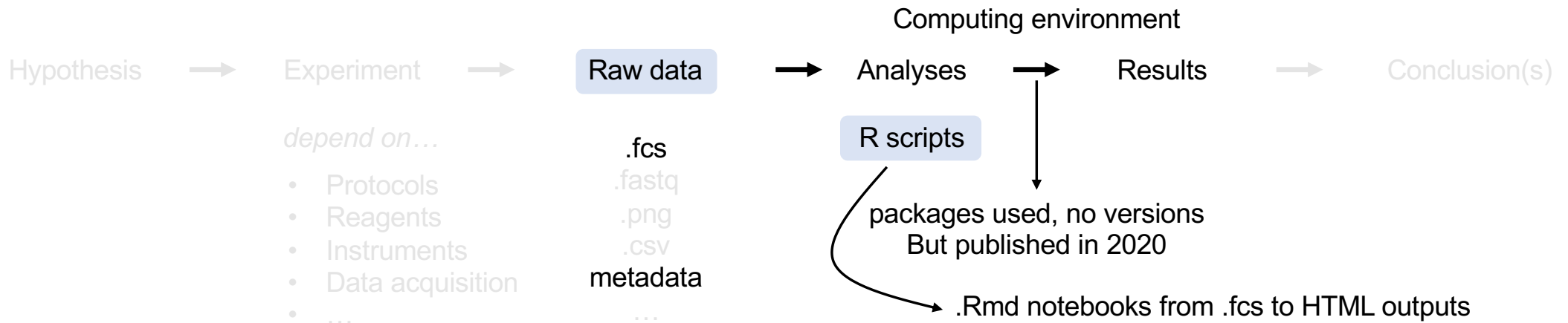
Will Guix effectively
allow us to
reproduce our
analyses in the
future?

Can we provide a
proof of concept?

Aim

- Build an environment that can be deployed anywhere
- AND
- which reproduces analyses

Requirement to provide a proof of concept



Flowrepository
FR-FCM-Z2CP

Github

Research Article

Computational and Systems Biology, Immunology and Inflammation

Stimulation strength controls the rate of initiation but not the molecular organisation of TCR-induced signalling

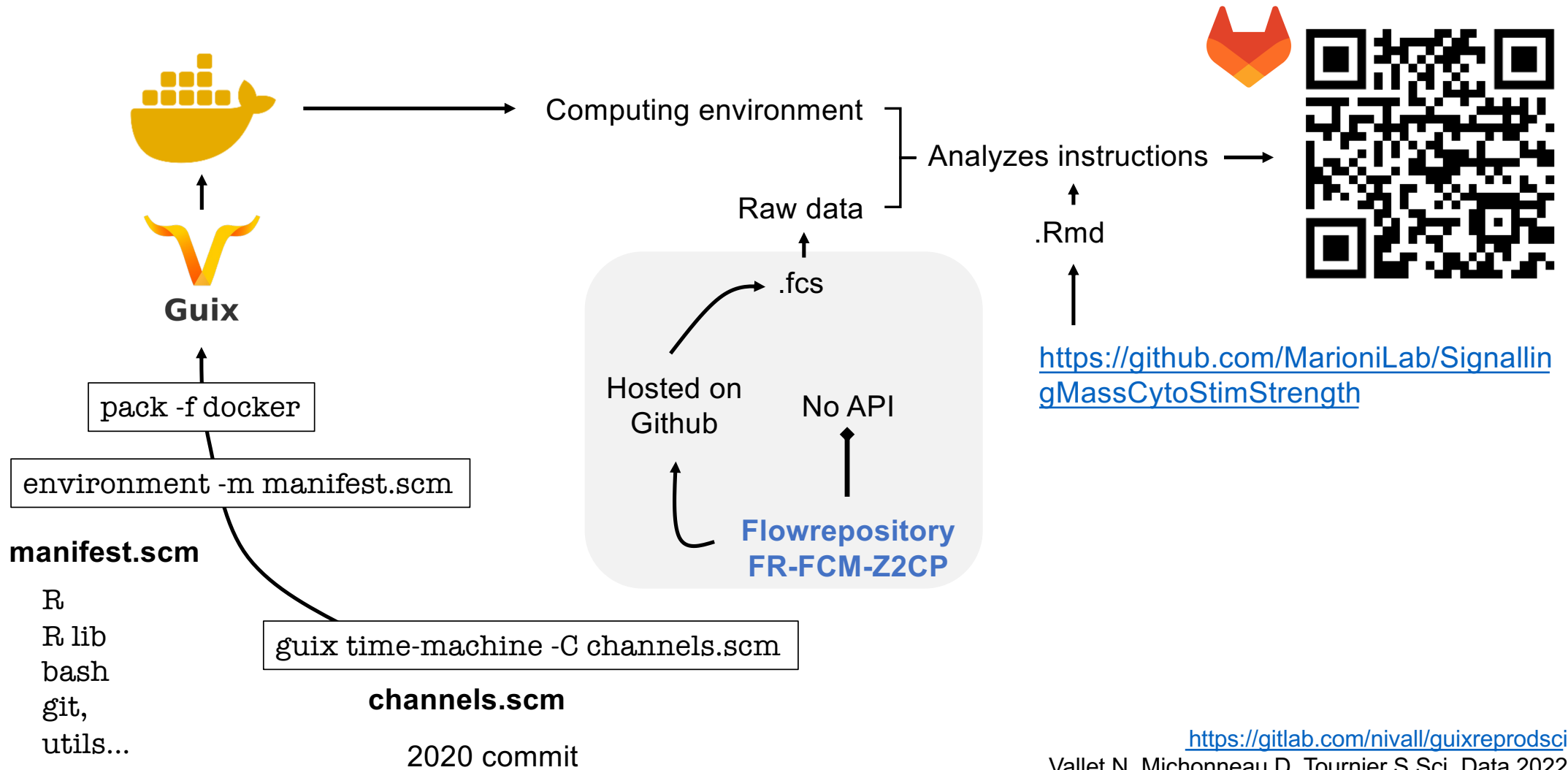
Claire Y Ma, John C Marioni , Gillian M Griffiths , Arianne C Richard 

Cambridge Institute for Medical Research, University of Cambridge, United Kingdom; Cancer Research UK Cambridge Institute, University of Cambridge, United Kingdom; EMBL-European Bioinformatics Institute, Wellcome Genome Campus, United Kingdom; Wellcome Sanger Institute, Wellcome Genome Campus, United Kingdom

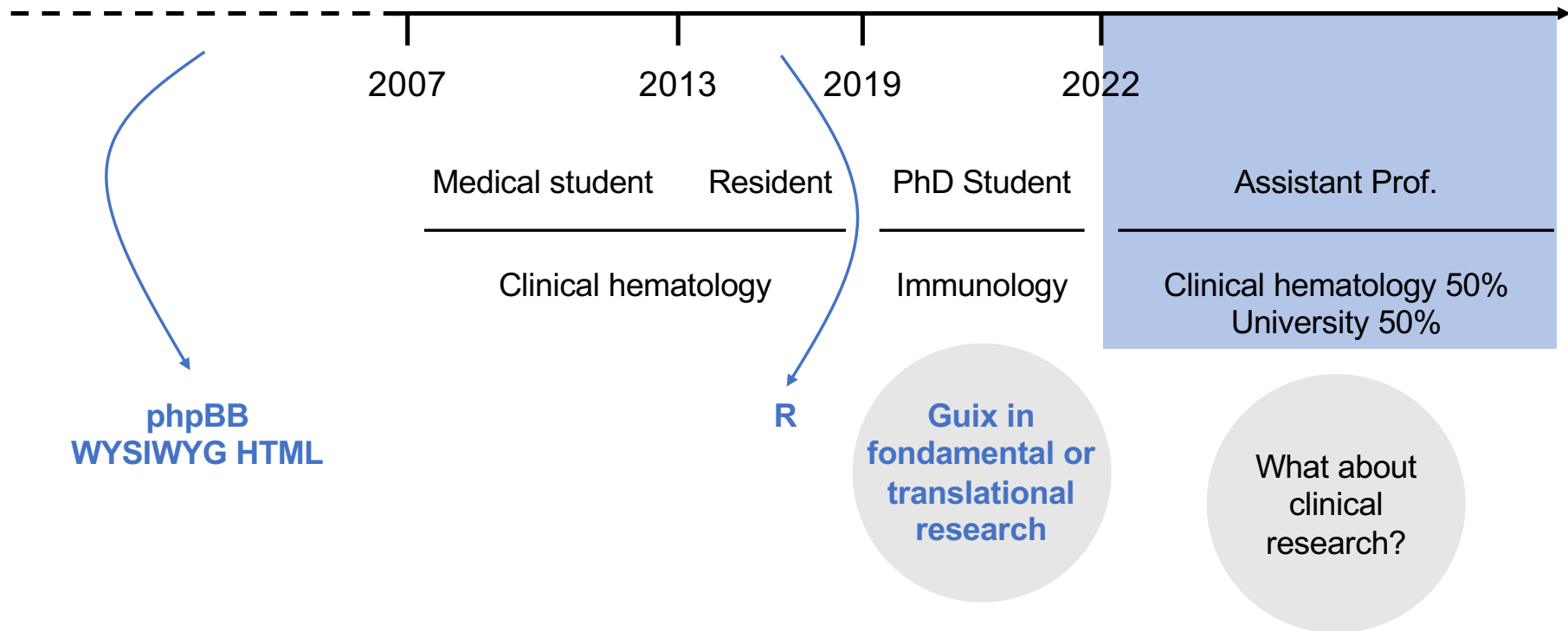
May 15, 2020 · <https://doi.org/10.7554/eLife.53948>  

Let's make these
.Rmd be compiled
on a Gitlab CI
like in 2020

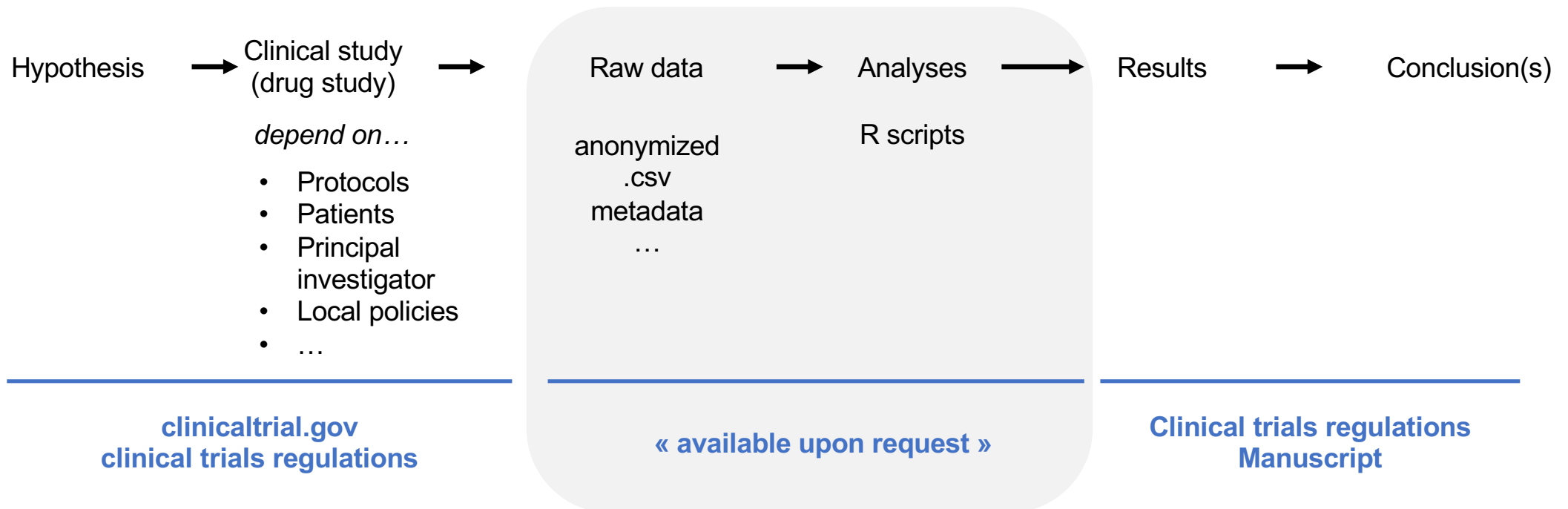
Reproducing the analyzes



The next step with clinical research

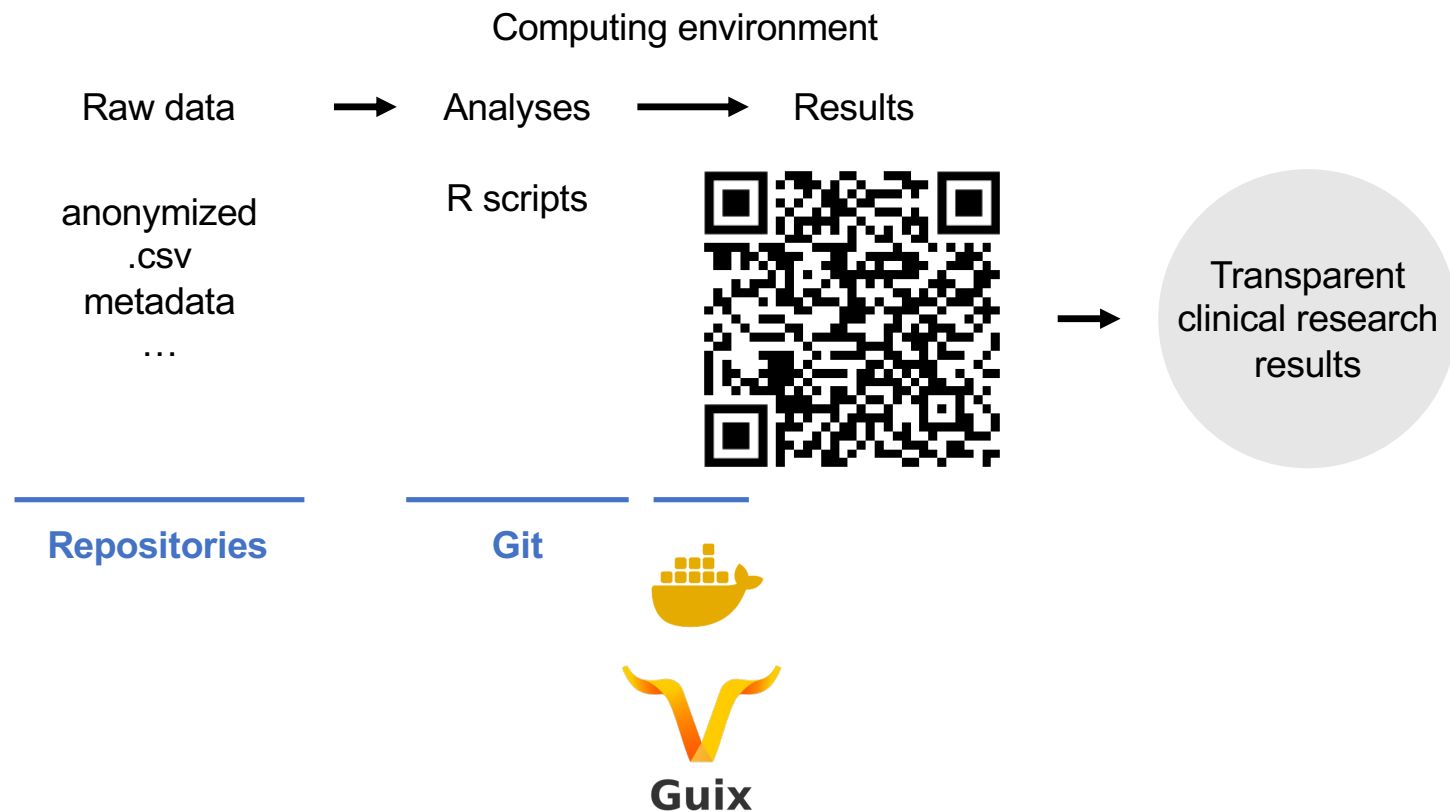


Is clinical research reproducible / transparent



Can we improve our practices toward transparent and reproducible clinical research?

Clinical research



To conclude

Unlikely to be 100% reproduced

Repositories

Hypothesis



Experiment



depend on...

- Protocols
- Reagents
- Instruments
- Data acquisition
- ...

Raw data



Analyses



Results



Conclusion(s)

.fcs
.fastq
.png
.csv
metadata
...

R scripts

**Computing
environment is now
the most reproducible
and transparent step
of biomedical
research**

Modulo the accessibility

Hypothesis



Clinical study
(drug study)



depend on...

- Protocols
- Patients
- Principal investig.
- Local policies
- ...

Raw data



Analyses



Results



Conclusion(s)

.csv
metadata
...

R scripts

Acknowledgments



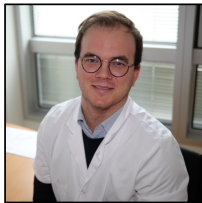
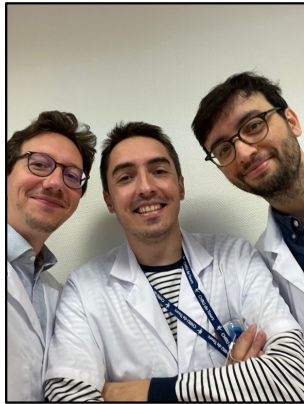
Emmanuel Gyan



Olivier Hérault



Service d'Hématologie et Thérapie Cellulaire



David Michonneau



Collaborations

Jean-Baptiste Delaye, Chantal Barin-Le Guellec, Hélène Blasco
(Service de Biochimie CHRU Tours)
Yoann Devisgnes (DRCI Tours)
Agnès Caille (U1246 SPHERE)
Martine Ropert (Biochimie Rennes)

Patricia Lepage (INRA MICALIS U1319)
Equipe de CyPS (Mass cytometry)
Y. Marie & D. Bouteiller (ICM)
Anne Bergeron (Pneumologie Genève)
Catherine Poirot (GRECOT)



CRYOSTEM

Régis Peffault de Latour



CRYOSTEM

Guix HPC / IRSL

Simon Tournier



Pour leur soutien

France Leucémie Espoir
Association pour Laurène
SOS Oxygène
ITMO Cancer Aviesan / Inserm
Association CANCEN
Novartis

