

# R in Grenoble DATA CHALLENGES

Magali Richard & Florent Chuffart

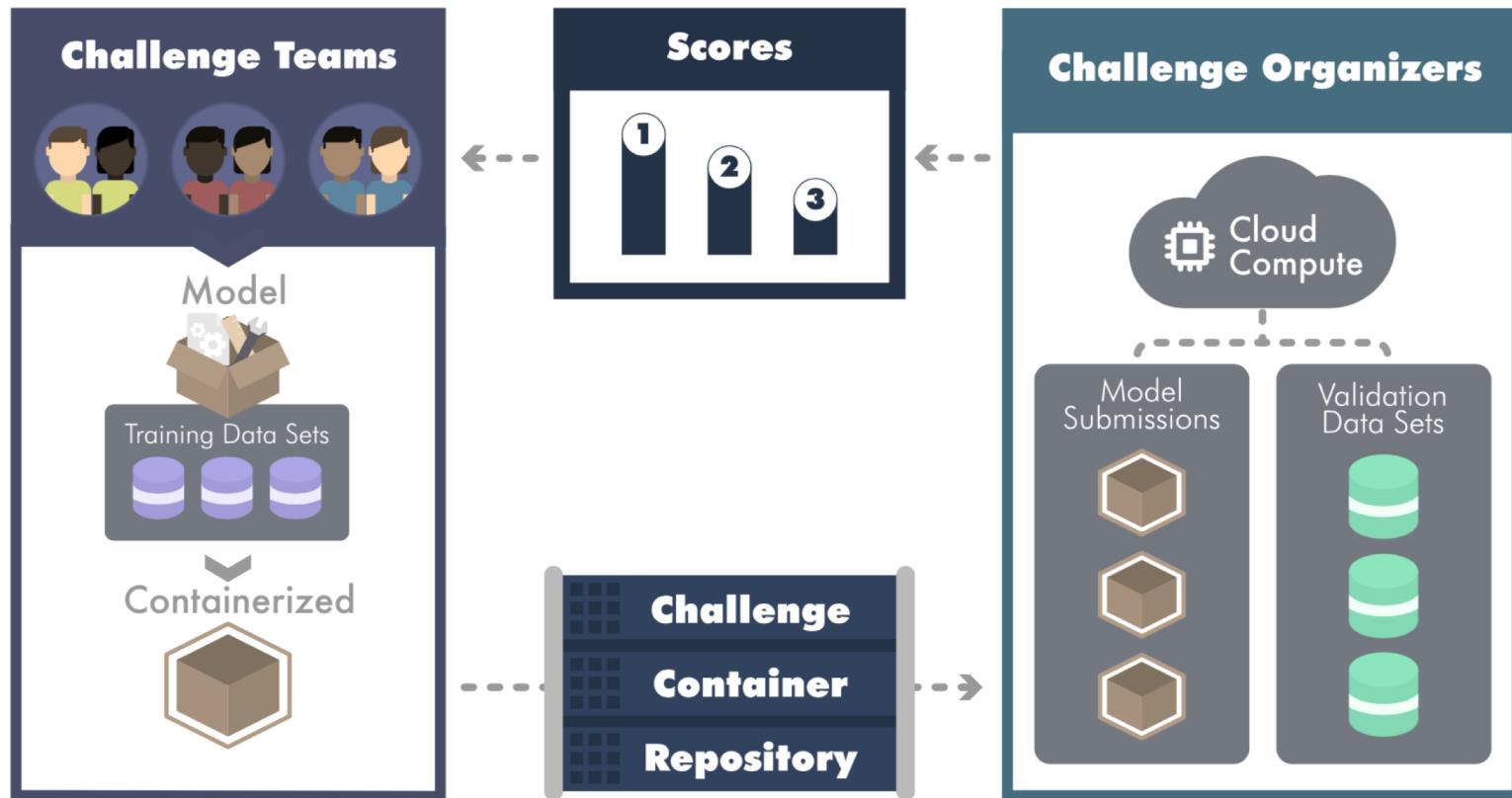
# Introduction

Data challenges in class

Data challenges for scientists

Tutorial

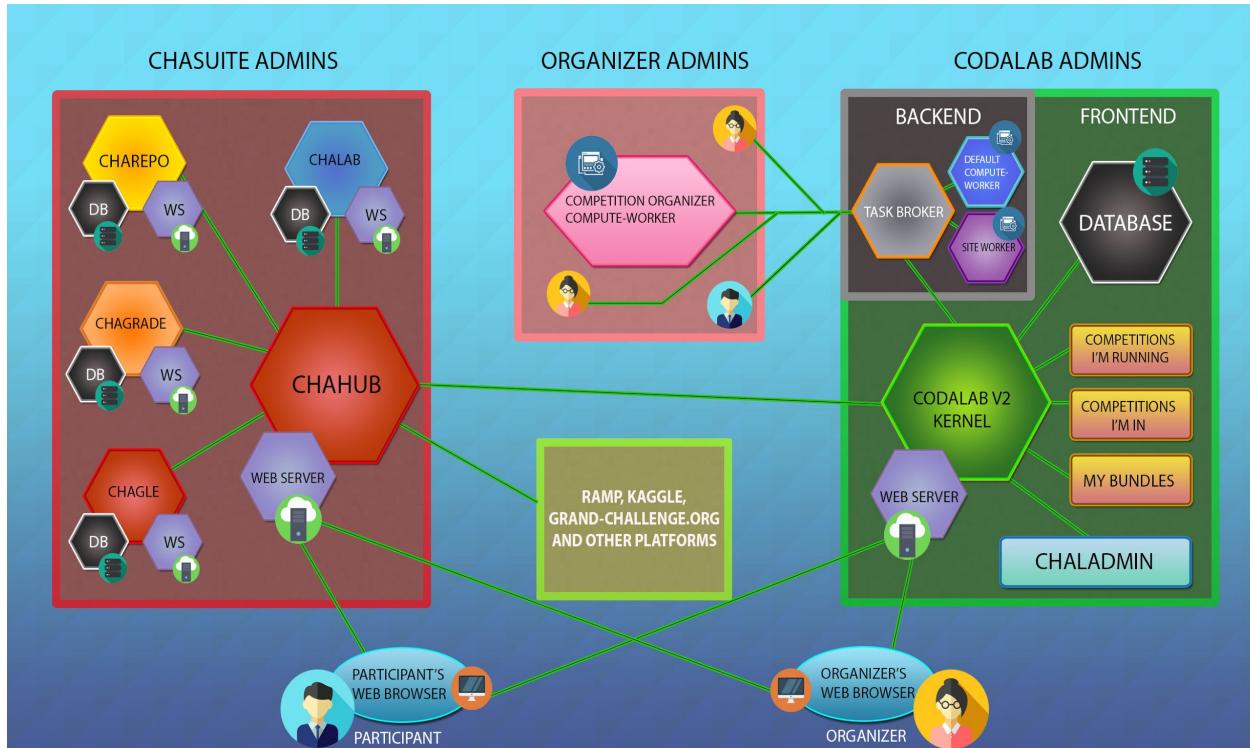
# What is a data challenge?



*Ellrott et al. Genome Biology (2019) 20:195*

# The challenge platform

CodaLab



- Enables participants to submit their codes
- Automatically rank the participants



Alexis Arnaud

# History

MICCAI Multimodal Brain Tumor Segmentation (BRaTS) Challenge  
Organized by kalpathy  
The BRaTS challenge is designed to gauge the current state-of-the-art in automated brain tumor segmentation and to compare between different ...

COCO Detection Challenge  
Organized by richardaen

VQA Real Image Challenge (Open-Ended)  
Organized by vqteam  
This challenge evaluates algorithms on the VQA Open-Ended task for the

Dialog State Tracking Challenge 4 (DSTC4)  
Organized by seokhwan.kim  
Dialog state tracking is one of the key sub-tasks of dialog management, which defines the representation of dialog states and ...

ChaLearn Fast Causation Coefficient Challenge  
Organized by chalearn  
This is an extension of ChaLearn's Cause-Effect Pair Challenge. Score pairs of variables (A, B) with a causation coefficient.

ChaLearn Automatic Machine Learning Challenge (AutoML)  
Organized by automl.chalearn  
Create a fully Automatic Machine Learning solution, capable of building models without ANY human intervention.

ChaLearn Looking at People 2016 - Track 1: Age Estimation  
SemEval-2017 Task 1: Semantic Textual Similarity  
Organized by llopez077  
Official competition site for the SemEval-2017 Task 1: Semantic Textual Similarity

2013: Medical data. Result submission.  
**COMPETITION BUNDLES**

2014: Computer vision, speech, NLP, IR.  
MSCOCO: 361 participants.

2015: **CODE SUBMISSION**  
AutoML: 687 participants  
Hackathons. Coopetitions.

2016: **ChaLab Wizard.**  
**USE IN EDUCATION**

2017: 480 challenges, 10000 users.  
**SCALABILITY, REUSABILITY,  
DYNAMIC COMPETITIONS**  
See.4c: EU prize, 2 million Euros

Introduction

Data challenges in class

Data challenges for scientists

Tutorial

# A novel pedagogic approach

As practical work

## Sex Prediction Challenge (sexpred)

Organized by Alexis\_Arnaud

The challenge provides the data.frame d of tumoral tissues described by genes expression values, histological and clinical attributes. The ...

Nov 15, 2019-Dec 30, 2019

37 participants

As homework

## Histology Prediction Challenge (histpred)

Organized by Alexis\_Arnaud

The challenge provides the data.frame d of tumoral tissues described by genes expression values, histological and clinical attributes. The ...

Nov 21, 2019-Dec 30, 2019

37 participants

As final evaluation



## Virulence Prediction Challenge (virpred)

Organized by FlorentC

The challenge provides the data.frame d of tumoral tissues described by genes expression values, histological and clinical attributes. The ...

Jan 09, 2020-Jan 11, 2020

35 participants

# Chagrade: a dedicated tool in codalab

The screenshot shows the Chagrade homepage. At the top, there's a dark header bar with the Chagrade logo on the left, a help icon (?), and a login button on the right. Below the header, the main content area has a white background. On the left, a large box contains the heading "Welcome to Chagrade" in a large, bold font, followed by the subtext "Here you can create and grade classes attached to challenges". It includes two buttons: "Sign Up" and "Login". To the right of this box is a sidebar titled "By the numbers..." which displays statistics: "CHAGRADE BRINGS TOGETHER" with counts of 162 students, 415 submissions, 159 users, and 13 classes. At the bottom, there's a section titled "About Chagrade" with three columns. The first column, "Learning By Doing", features an icon of a graduation cap and describes how students learn by solving actual machine learning problems. The second column, "Collaborating", features an icon of three people and describes how students can work individually or in teams. The third column, "Gaining Visibility", features a GitHub icon and describes how teachers can provide a starting kit in a GitHub repo for students to clone and modify.

## Welcome to Chagrade

Here you can create and grade classes attached to challenges

[Sign Up](#) [Login](#)

### By the numbers...

CHAGRADE BRINGS TOGETHER

162	415
STUDENTS	SUBMISSIONS
159	13
USERS	CLASSES

#### About Chagrade

**Learning By Doing**  
Challenges make great projects. Students learn by solving actual machine learning problems. They can submit code, which is automatically evaluated by the challenge platform.

**Collaborating**  
Students can work individually or in teams. A class can mix individual and team homework. Homework can include questions on top of challenge performance.

**Gaining Visibility**  
Teachers can provide a starting kit in a GitHub repo that students can clone and modify. They can start building a portfolio with these projects using GitHub.

Introduction

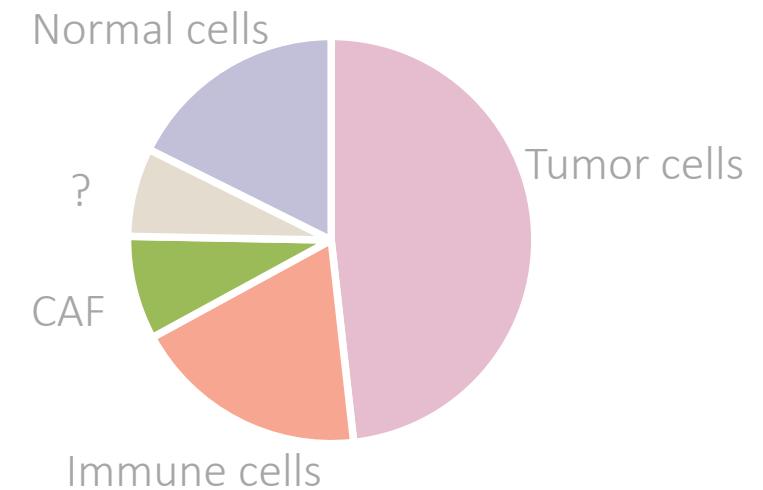
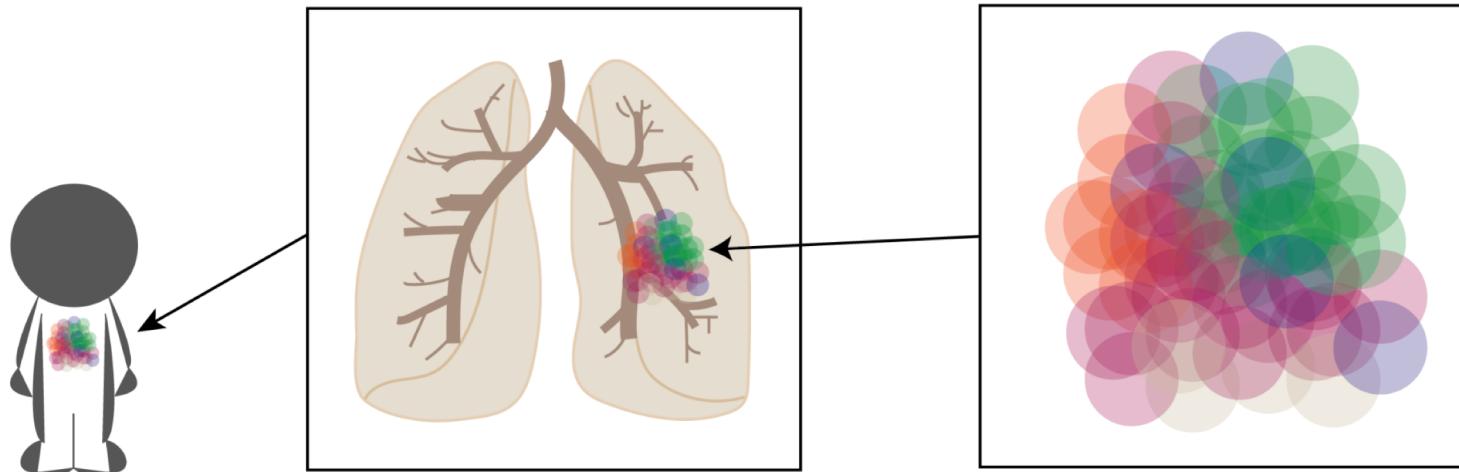
Data challenges in class

Data challenges for scientists

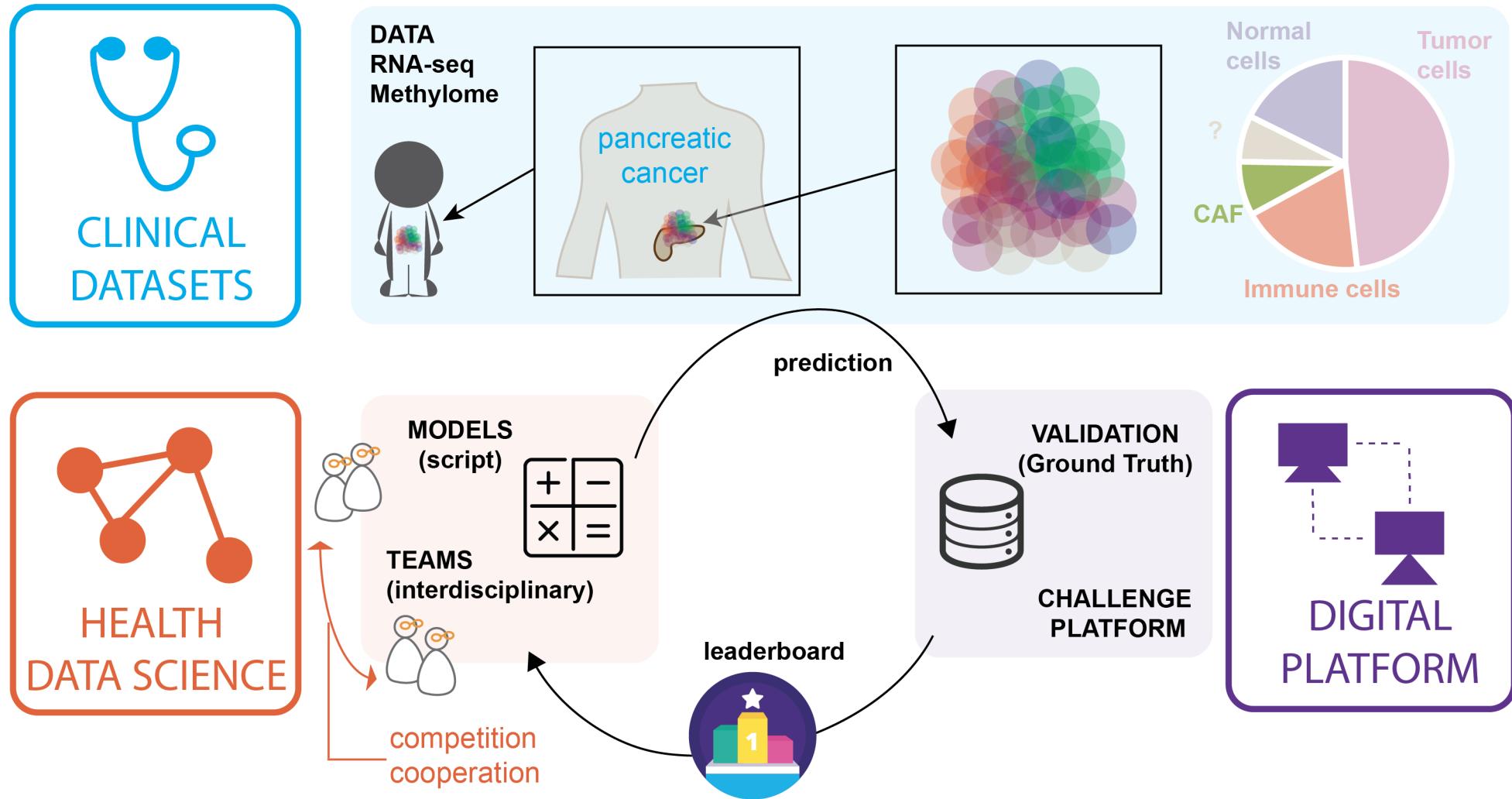
Tutorial

# Goal of our data challenge

- Quantification of tumor heterogeneity



# A challenge for scientists



# 2 editions of the data challenge

## 1<sup>st</sup> edition (2018)



## 2<sup>nd</sup> edition (2019)



- Methylation Data
- One cancer type
- Cell lines

- Methylation and transcriptomic Data
- Several cancer types
- Primary tumors / cell lines

# When and where?

**Health Data Challenge**  
(2<sup>nd</sup> edition)  
2019

Deconvolution methods  
to quantify tumor heterogeneity  
in cancer research.

Registration deadline extension  
September 30

**Invited speakers**

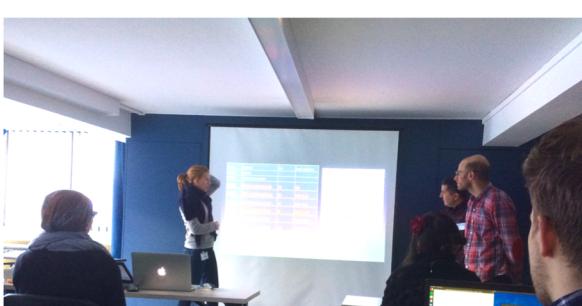
Michael Scherf  
Fraunhofer Institut für Informatik,  
Saarbrücken, Germany;  
Francisco Avila Cobos  
Xavier University, Chest, Belg.  
Jerome Cros  
IAP, Paris, France

**Organizers**

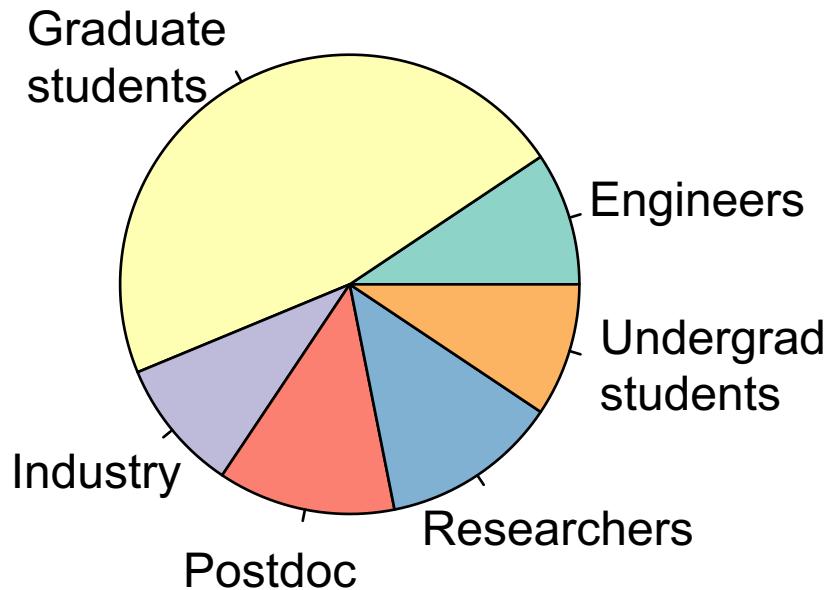
Yuna Blum & Florent Petitprez  
Programme CIT, INCC, Paris, France;  
Magali Richard & Alexis Arnaud  
XNA, Grenoble, France

**MORE INFORMATION & REGISTRATION**  
Webpage: <http://bit.ly/2WYeXYV>  
Contact: [datas.institute@univ.grenoble-alpes.fr](mailto:datas.institute@univ.grenoble-alpes.fr)  
[@granobledata](https://twitter.com/granobledata)

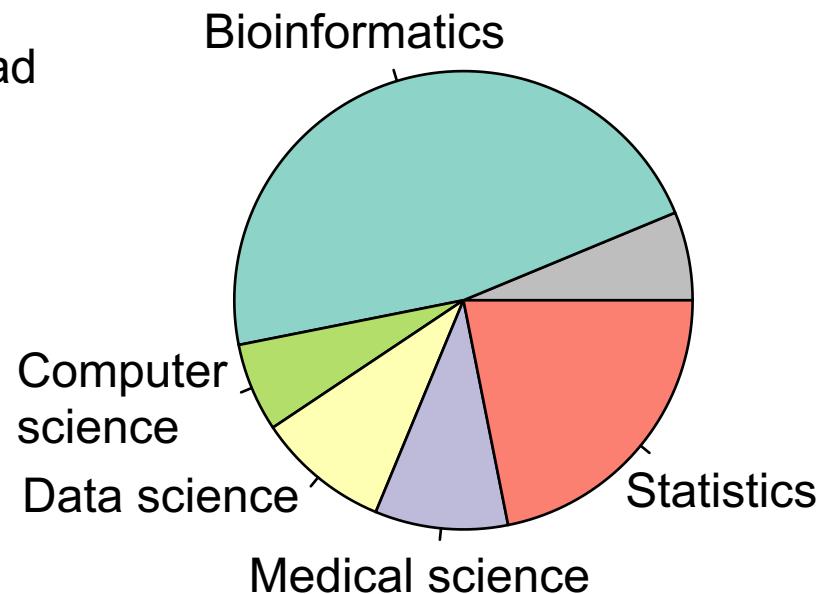
- **When:** 25-29 of November 2019
- **Where:** Aussois (CAES CNRS), French Alps



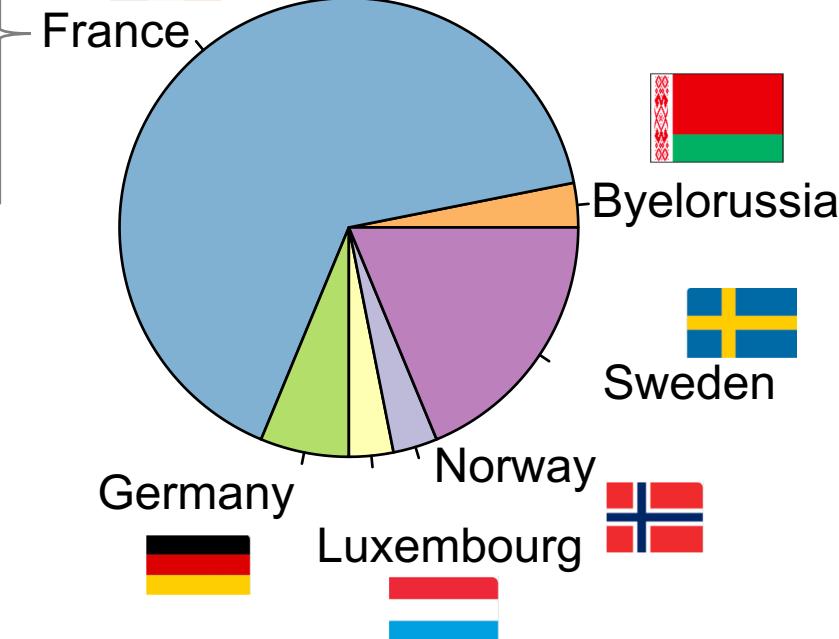
# About the participants (n=34)



10 teams of 3-4 people

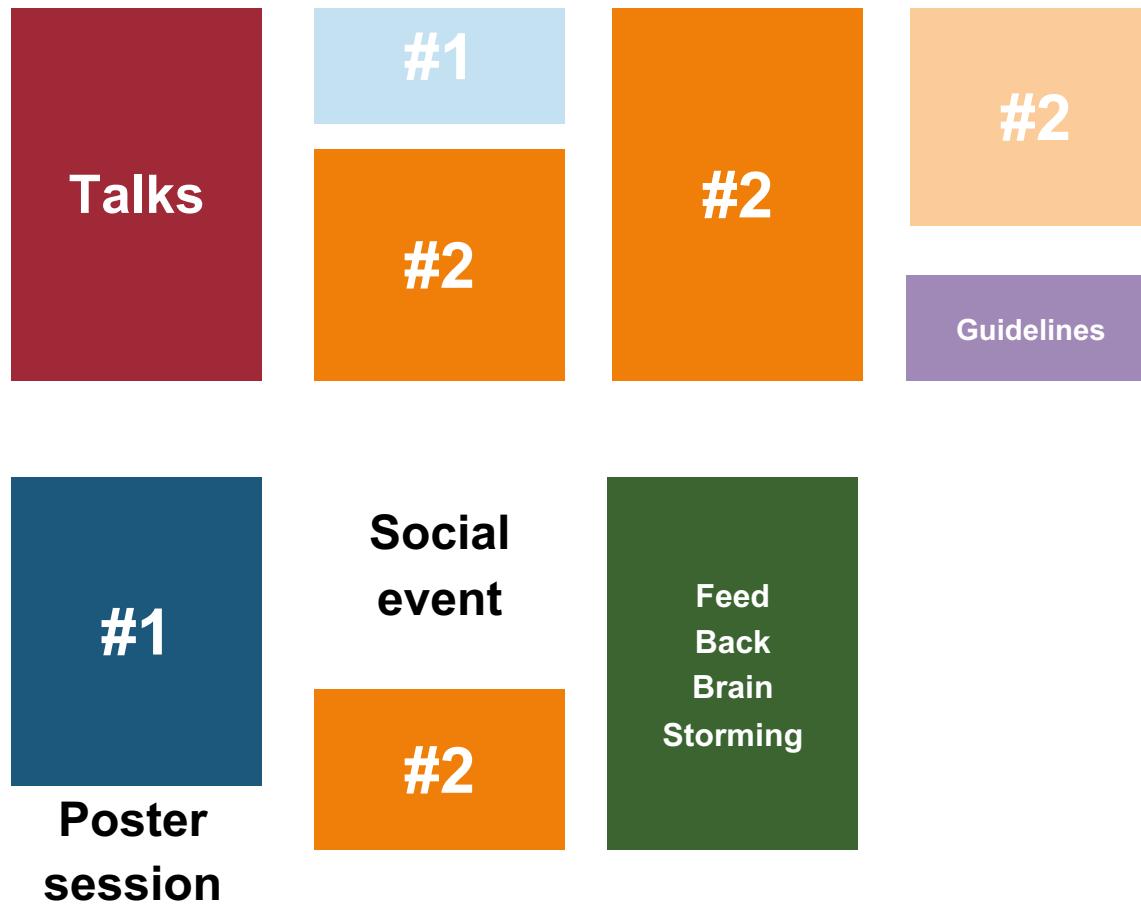


Institut Curie  
UGA Grenoble  
CRC Cordeliers  
INSERM  
CEA  
Innate Pharma  
Verteego  
...



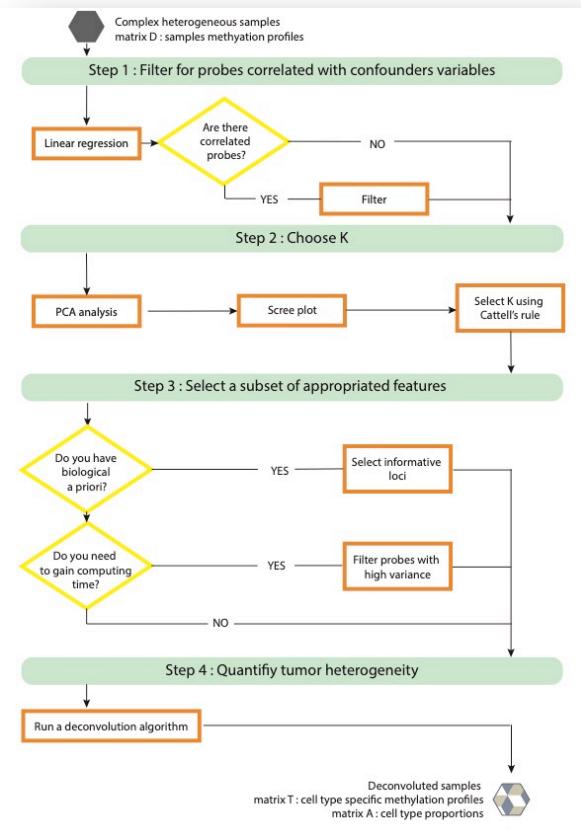
# Agenda

PRACTICALS



# Valorization of the first edition

## ➤ Guidelines



## ➤ Article

**bioRxiv**  
THE PREPRINT SERVER FOR BIOLOGY

In review in BCM bioinformatics

New Results Comment on this paper

**Guidelines for cell-type heterogeneity quantification based on a comparative analysis of reference-free DNA methylation deconvolution software**

Clementine Decamps, Florian Privé, Raphael Bacher, Daniel Jost, Arthur Waguet, HADACA consortium, Eugene Andres Houseman, Eugene Lurie, Pavlo Lutsik, Aleksandar Milosavljevic, Michael Scherer, Michael G.B. Blum, Magali Richard

doi: <https://doi.org/10.1101/698050>

## ➤ R package medepir

<https://rdrr.io/github/bcmuga/medepir/man/medepir-package.html>

M Richard, C Decamps , F Privé, M Blum



## ➤ Blog posts

**Towards Data Science**  
Sharing concepts, ideas, and codes

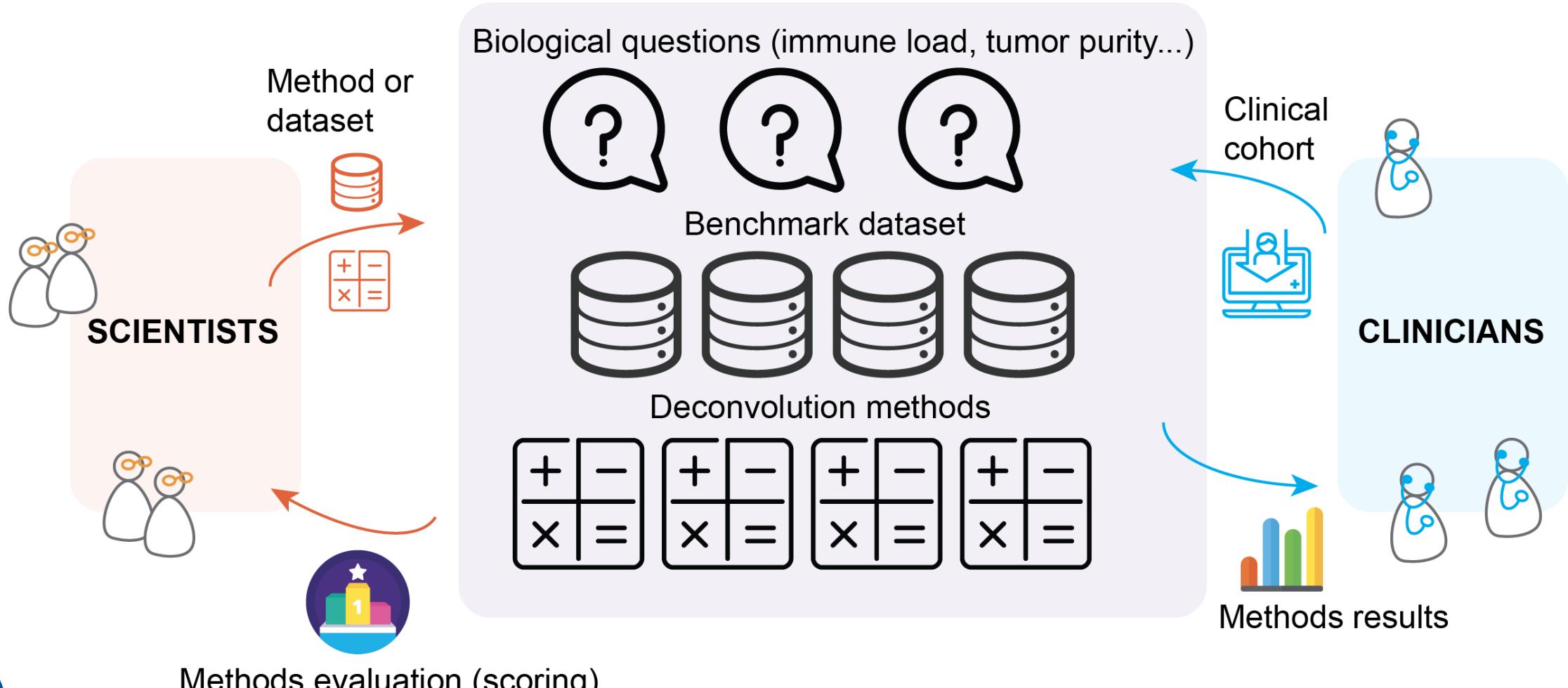
DATA SCIENCE MACHINE LEARNING PROGRAMMING VISUALIZATION

**Health data challenges organization: feedback, comments and recommendations.**

Authors: Elise Amblard, Yuna Blum, Jane Merlevede, Magali Richard

*In preparation*

# From data challenges to benchmarks



Introduction

Data challenges in class

Data challenges for scientists

Tutorial

# On line tutorial by Alexis Arnaud

Codalab >  Bundle\_generation > Details



The screenshot shows a GitHub-like interface for a project named 'Bundle\_generation'. At the top, there's a large R logo icon, the project name 'Bundle\_generation' with a green checkmark icon, and a 'Project ID: 5596'. To the right are buttons for notifications (Bell), stars (Star 0), forks (Fork 0), and cloning (Clone). Below this, a summary bar shows '11 Commits', '1 Branch', '0 Tags', and '748 KB Files'. The main content area is titled 'README.md'.

## README.md

To create a Codalab competition, you have to generate a zip file with the required files inside and upload the archive on the Codalab platform. The *R* script `Bundle_generator_--cmd.Rmd` allows you to automatically create a valid bundle by using the *R* terminal to ask you some questions in order to personalize the competition.

## Generate a Codalab bundle

In a *R* terminal, run the following command : `rmarkdown::render(input = "Bundle_generator_--cmd.Rmd", envir = new.env( ) )`

You can also generate a toy bundle, with the parameters by default, with *Rscript* : `Rscript -e 'rmarkdown::render(input = "Bundle_generator_--cmd.Rmd", envir = new.env( ) )'`

## Test the Codalab bundle

In a *R* terminal, run the following command : `rmarkdown::render(input = "Bundle_test.Rmd", envir = new.env( ) )`

## [WIP] Render a web site to previsualize the competition

It gives you the possibility to visualize your competition with a minimalist website which looks like the Codalab platform.

`rmarkdown::render_site(input = "Web_page_generator.Rmd", envir = new.env( ) )`

It's up to you now!

# Thank you for your attention !

