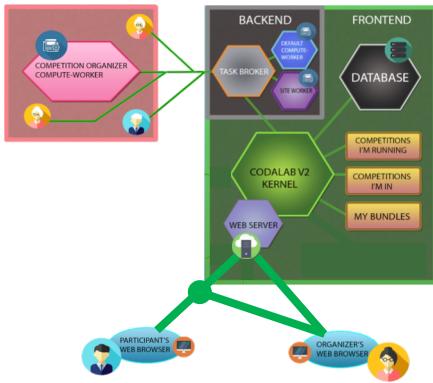


DATA CHALLENGE

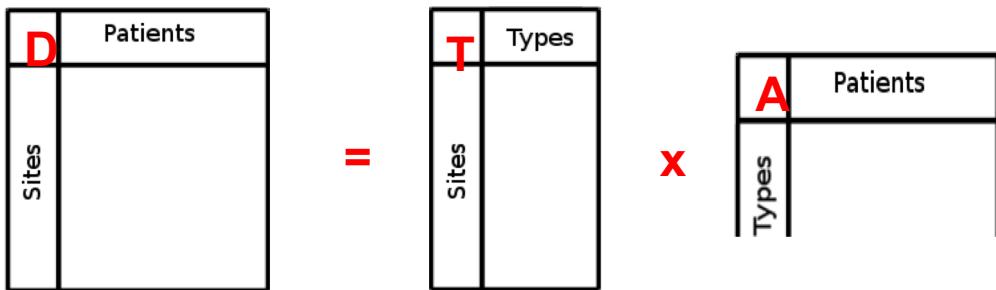
Deconvolution methods to quantify tumor heterogeneity



Introduction to the challenge #1



1) How to use the challenge platform codalab?



2) Challenge #1 Starting kit

CHALLENGE PLATFORM

- **Challenge subscription**

Connect to : <https://competitions.codalab.org/competitions/21949>

- **Get the starting kit and the public data**

- In Participate / Files

- **File submission (code)**

- Zip file with code (*program.R*) and metadata file (*metadata*) at the root → see starting kit example !

- **Specific rules for the CHALLENGE #1**

- ▶ 3 mn max per submission
 - ▶ 20 submissions max per user

CHALLENGE - Description

Learn the Details

Phases

Participate

Results

Public Submissions

Forums ➔

Overview

Evaluation

Terms and
Conditions

Overview

[!] The subscription is only open to people physically attending the challenge in Aussois [!]

How to start ?

- [1] Go on the challenge page, in the **Participate** tab, in the
- [2] On your local machine, unzip the just downloaded zip files
The unzipped strating-kit directory contains now:
- [3] Then open *R* in the **starting_kit** directory,

```
rmarkdown::render("submission_script_p1.Rmd")
```

CHALLENGE - Phases

Learn the Details

Phases

Participate

Results

Public Submissions

Forums 

CHALLENGE #1

Start: Nov. 24, 2019, 11 a.m.

CHALLENGE #2 // Phase 1 // Exploration

Start: Nov. 27, 2019, 8 a.m.

CHALLENGE #2 // Phase 2 // Validation

Start: Nov. 28, 2019, 8 a.m.

Competition Ends

Dec. 1, 2019, midnight

CHALLENGE #1 - Files

[Learn the Details](#)[Phases](#)[Participate](#)[Results](#)[Public Submissions](#)[Forums](#) ➔[Data](#)[Files](#)[Submit / View](#)[Results](#)

	Download	Size (mb)	Phase
	Starting Kit	0.270	#1 CHALLENGE #1
	Public Data	173.181	#1 CHALLENGE #1
	Starting Kit	0.270	#2 CHALLENGE #2 // Phase 1 // Exploration
	Public Data	173.198	#2 CHALLENGE #2 // Phase 1 // Exploration

CHALLENGE #1 - Starting Kit & Public Data

■ What files are in the starting kit :

- ▶ *starting_kit_p1.html* :
- ▶ a summary of the challenge with all the required informations

- ▶ *submission_script_p1.Rmd* :
- ▶ a Rmd file to modify with your code
- ▶ this file generates the zip file to upload on Codalab

■ What files are in public data :

- ▶ *DC1_D_met.rds* : methylome data
- ▶ *DC1_D_rna.rds* : transcriptome data

CHALLENGE #1 – Submit Results

Learn the Details Phases **Participate** Results Public Submissions

Forums →

Data **CHALLENGE #1** CHALLENGE #2 // Phase 1 // Exploration

Files CHALLENGE #2 // Phase 2 // Validation

Submit / View Results

Phase description
Estimate the proportion matrix A from a mixture matrix D (either DNAm or RNA-seq). Submit your script (program) on the platform.
(Data type correspondence : 1.0 = met, 2.0 = rna)

Max submissions per day: 20

Max submissions total: 20

Click the Submit button to upload a new submission.

*Team name (20 characters max):

*Method name (20 characters max):

*Method description:

Submit

CHALLENGE #1 – Submitted Results

■ What files are in the submitted zip file :

► *program.R* → a R file with :

- your code inside the function ‘program’
- the variable ‘dataType’ = met or rna (based on your team name)

► *metadata* → a file required by Codalab

- use the one generated by the starting kit and don’t modify it !

CHALLENGE #1 – Leaderboard

Learn the Details Phases Participate Results

Public Submissions Forums ➔

CHALLENGE #1

CHALLENGE #2 // Phase 1 // Exploration

CHALLENGE #2 // Phase 2 // Validation

Phase description

Estimate the proportion matrix A from a mixture matrix D (either DNAm or RNA-seq). Submit your script (program) on the platform. (Data type correspondence : 1.0 = met, 2.0 = rna)

Max submissions per day: 20

Max submissions total: 20

[Download CSV](#)

[Download all submissions on leaderboard](#)

Results

#	User	Entries	Date of Last Entry	Team Name	MAE ▲	MAE 1 ▲	MAE 2 ▲	Data Type ▲	Detailed Results
---	------	---------	--------------------	-----------	-------	---------	---------	-------------	------------------

CHALLENGE – Public Submissions & Forum

Public Submissions

[Go Back to Competition Homepage](#)

[CHALLENGE #1](#) [CHALLENGE #2 // Phase 1 // Exploration](#)

[CHALLENGE #2 // Phase 2 // Validation](#)

No submissions have been made public!

Forums

Health Data Challenge 2019 Forum

[Go back to competition](#) [Start a new topic](#)

Title	Started by	Date created	Latest post	Posts	
↗ FAQ	Alexis_Arnaud	Nov 21, 2019	3 days, 5 hours	1	

CHALLENGE #1: STARTING KIT

■ Your goal: estimate the matrix A

- The matrix A represents the proportion of each cell type in each patient.

You have this

↓

D	Patients
Sites	

=

T	Types
Sites	

x

You search this

↓

A	Patients
Types	

CHALLENGE #1: STARTING KIT

■ Data format - simulations

- ▶ RNAseq **or** DNA methylation
- ▶ 21,566 genes or 772,316 probes
- ▶ 30 patients
- ▶ ? cell types

D	Patients
Sites	

=

T	Types
Sites	

x

A	Patients
Types	

CHALLENGE #1: STARTING KIT

■ Two D matrix

- Your D: T is a **median cell profile**

D	Patients
Sites	
A	

=

T	Types
Sites	
X	

X

A	Patients
Types	

CHALLENGE #1: STARTING KIT

■ Two D matrix

- ▶ Your D: T is a **median cell profile**
- ▶ Second D to test your method: **T is different for each patient** (more realistic of the inter-tumor heterogeneity but more complex)
- **Your score is computed on both matrices !**

D	Patients
Sites	

=

T	Types
Sites	

x

A	Patients
Types	

