

15-16 February 2021

# COMETH Training course

From omics data

to tumor heterogeneity quantification

EIT Health is supported by the EIT,  
a body of the European Union



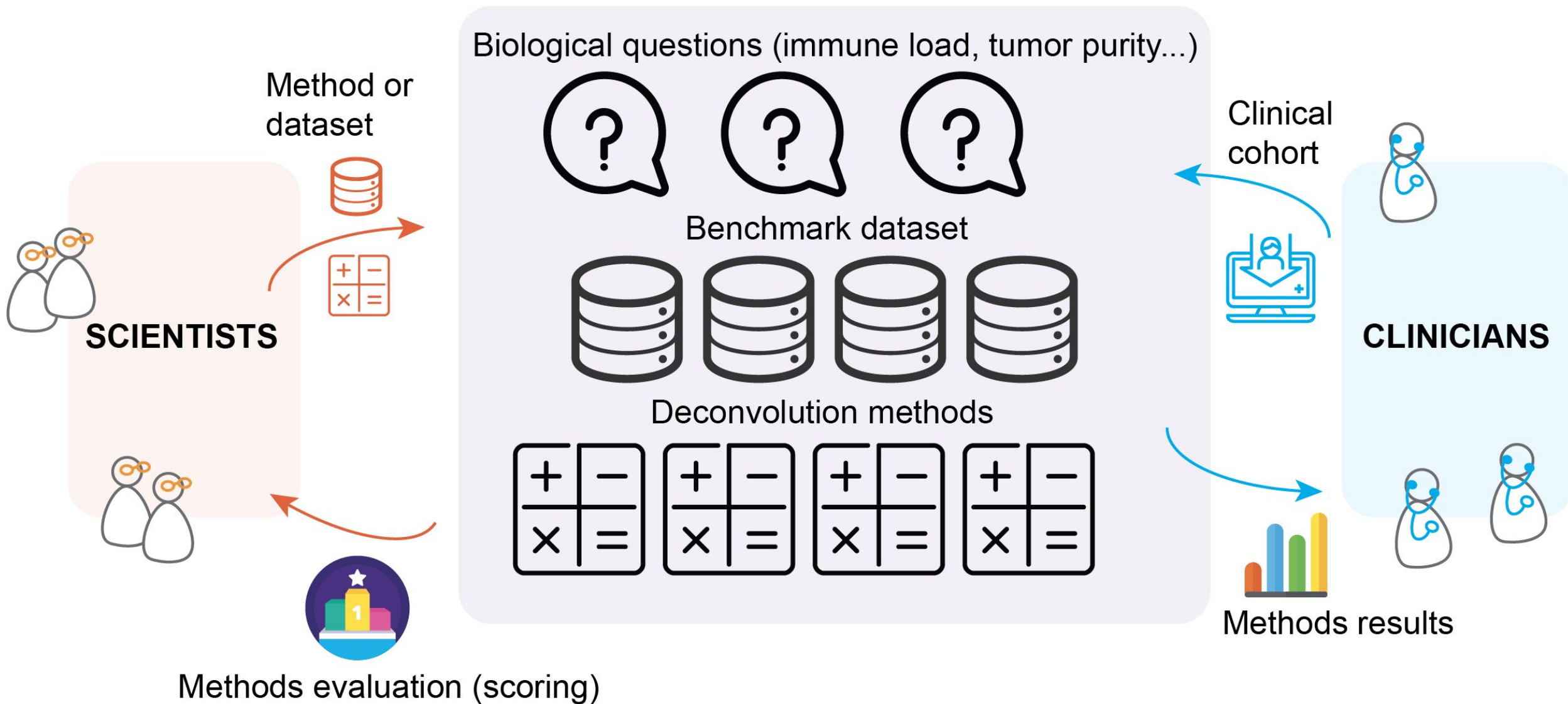
15 February 2021

# Digital tools

**Magali Richard**



# The COMETH program



# The COMETH interfaces

DAY 1-2

COMETH Data Challenge interface (for data scientists)  
Aim: Benchmark new computational methods



Codabench



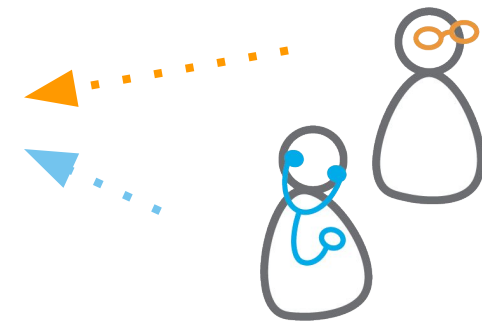
COMETH user friendly web interface  
Aim: Choose and apply computational methods

Cometh



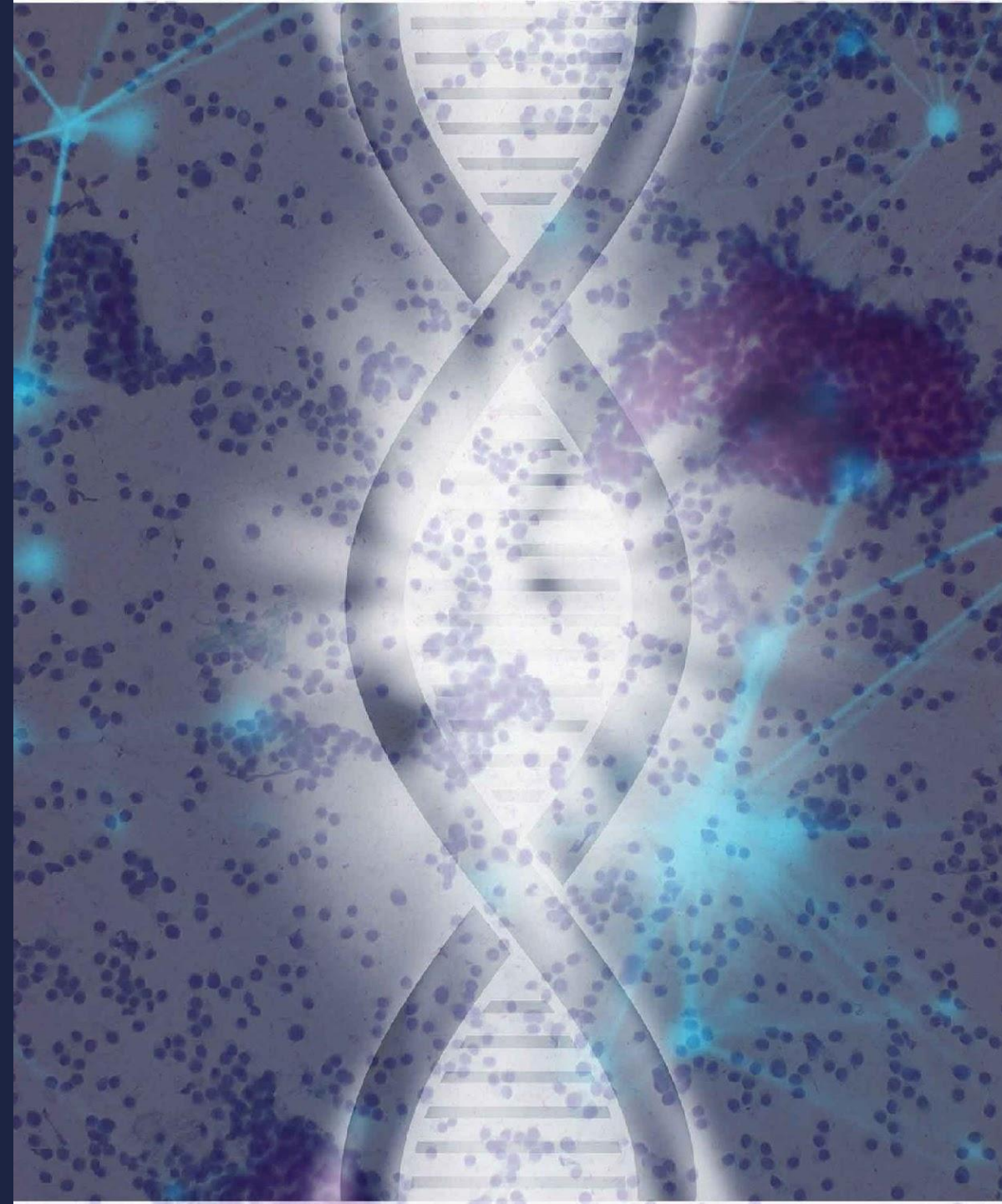
DAY 2

COMETH shiny app  
Aim: Visualise the results



# How to evaluate a computational method?

Compare your estimate with a ground truth  
Evaluate the accuracy of the estimation



# The Ground Truth datasets used in COMETH project



**Complex omic  
data on bulk  
samples**



Lung  
adenocarcinoma  
(luad)

METHYLOME  
(27K)



Colorectal  
adenocarcinoma  
(coad)

TRANSCRIPTOME  
(RNA-seq)



Pancreatic  
adenocarcinoma  
(paad)

TRANSCRIPTOME  
(RNA-seq)

METHYLOME  
(Meth-EPIC)



Breast invasive  
carcinoma  
(brca)

TRANSCRIPTOME  
(RNA-seq)

METHYLOME  
(450K)

# The Ground Truth datasets used in COMETH project



Lung adenocarcinoma (luad)



Colorectal adenocarcinoma (coad)



Pancreatic adenocarcinoma (paad)



Breast invasive carcinoma (brca)

**Ground Truth  
(real tumor  
composition in  
cell types  
proportions)**



In silico  
simulations

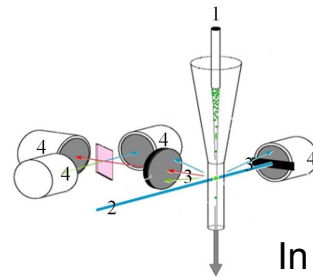


In vitro  
mixtures

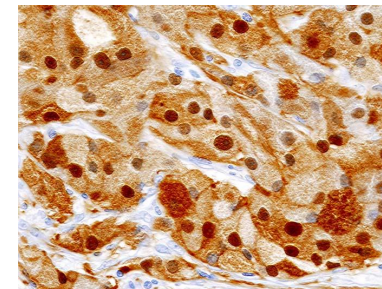


**More to come...**

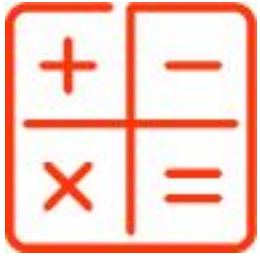
In vivo FACS  
countings



In vivo immuno



# The computational methods used in the COMETH project (transcriptome)



To quantify all cell types

To quantify immune cell types

Unsupervised

NMF-based (MT2 & MT19)  
ICA-based (MT1 & MT14)

NMF-based (MT19)  
ICA-based (MT14)

Semi supervised

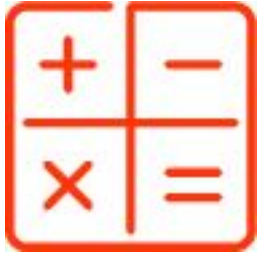
NMF-based (MT16, MT17 & MT18)

Supervised

Cibersort (MT8)  
EPIC (MT9)  
Quantiseq (MT11)



# The computational methods used in the COMETH project (methylome)



To quantify all cell types

To quantify immune cell types

Unsupervised

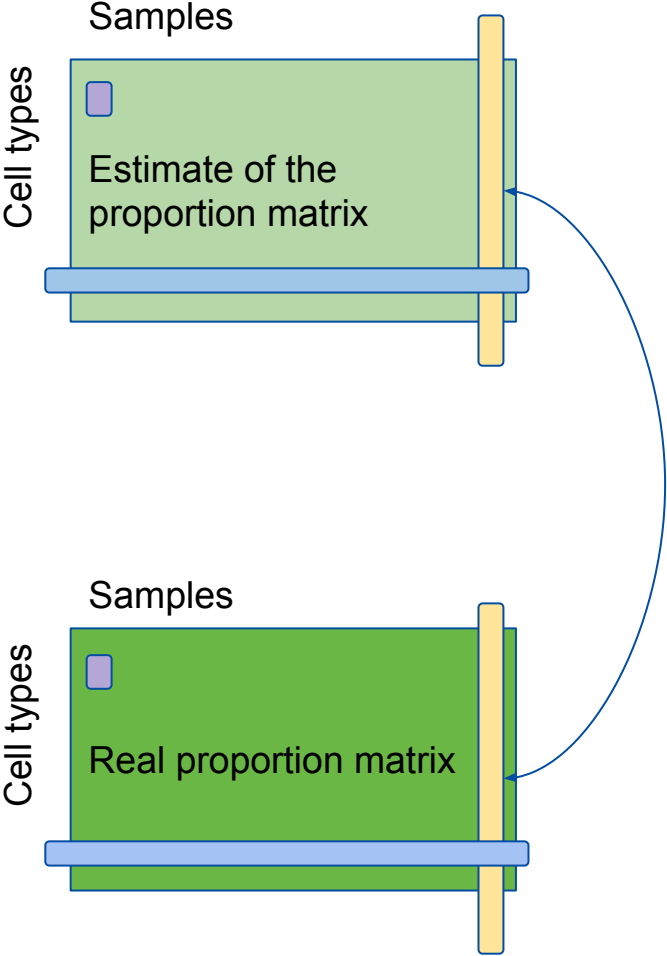
Edec (MT3)  
NMF-based (MT19)  
ICA-based (MT14)

Semi supervised

Supervised

EdiDISH (MT6)

# Evaluation of method performances (scoring)



Multi-criteria comparison

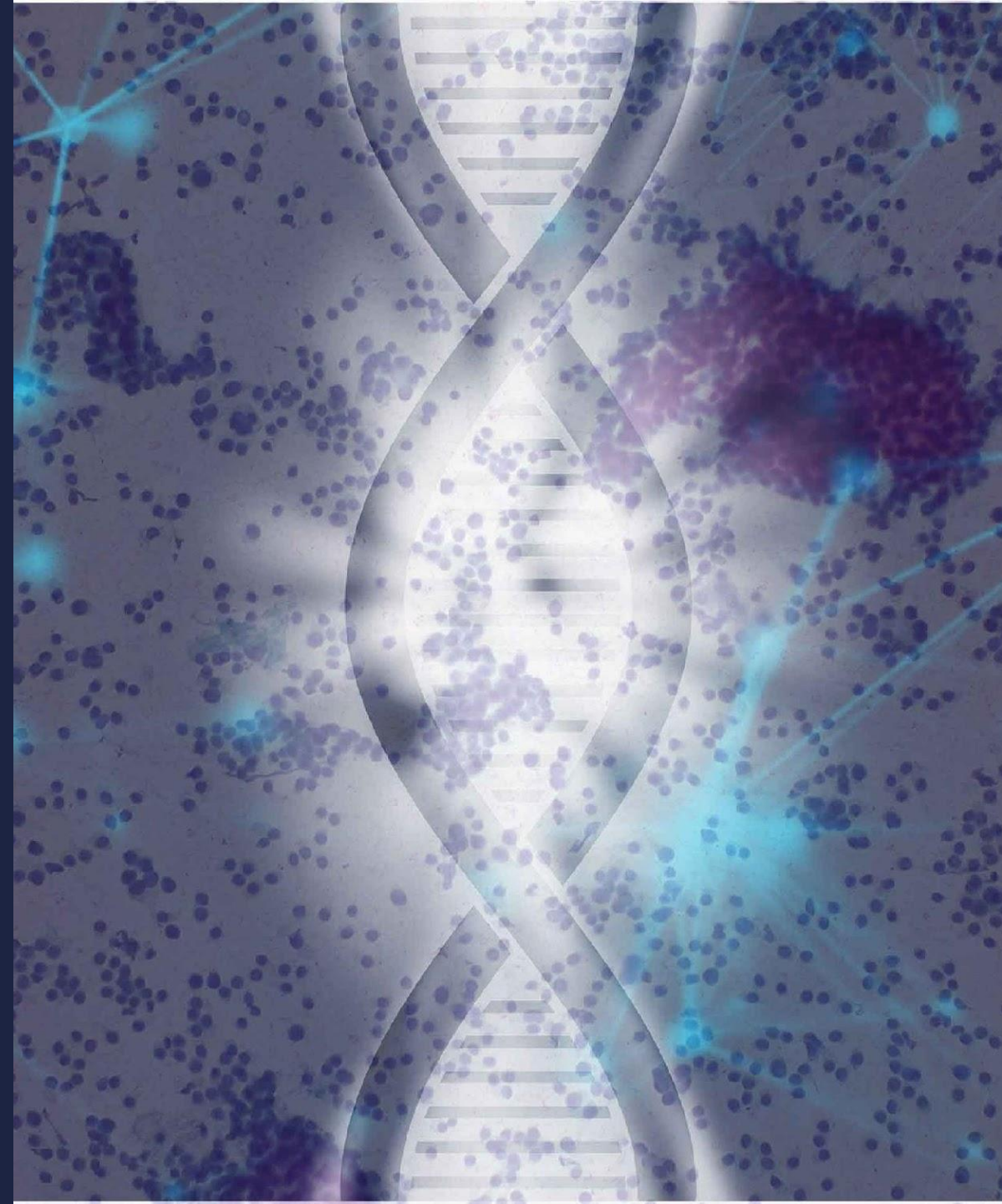
- Correlation on rows
- Correlation on columns
- Mean Absolute Error



**FINAL ACCURACY SCORE**  
(min = 0, max = 1)

# Which digital interfaces?

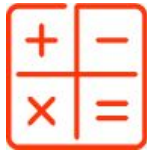
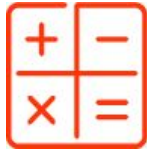
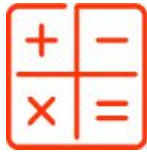
The COMETH data challenge interface  
The COMETH user-friendly web application  
The COMETH shiny app for visualisation



# Systematic method evaluation



**Data scientists**  
Develop methods  
Run it on the benchmark  
Get evaluated



SCORE

## Codabench Leaderboard

		Results						
Task:		Fact Sheet Answers	MT1_v4			MT2_v4		
#	Participant	Submission ID?	Accuracy_mean	Accuracy_sd	Time	Accuracy_mean	Accuracy_sd	Time
1	magrichardtest	DT1 pancreas	n/a	n/a	n/a	0.9721809186	0.0057113517	0.00000000
2	magrichardtest	DT pancreas linear	0.7299364238	0.0309803372	0.0000000000	n/a	n/a	n/a
3	magrichardtest	DT8 brca log 32	0.2888423247	0.0000000000	0.0000000000	n/a	n/a	n/a
4	magrichardtest	DT9 brca lin 32	0.4431279213	0.0000000000	0.0000000000	n/a	n/a	n/a
5	magrichardtest	DT11 breast linear	n/a	n/a	n/a	0.4200006067	0.2359621818	0.00000000
6	magrichardtest	DT10 breast log	n/a	n/a	n/a	n/a	n/a	n/a
7	magrichardtest	DT4	n/a	n/a	n/a	0.5660189192	0.0000000000	0.00000000
8	magrichardtest	DT5	0.8315607989	0.0000000000	0.0000000000	n/a	n/a	n/a
9	magrichardtest	DT1 pancreas	n/a	n/a	n/a	n/a	n/a	n/a



## COMETH BENCHMARK

Edit Participants Submissions Dumps Migrate

ORGANIZED BY: Magrichardtest

CURRENT PHASE ENDS: Never

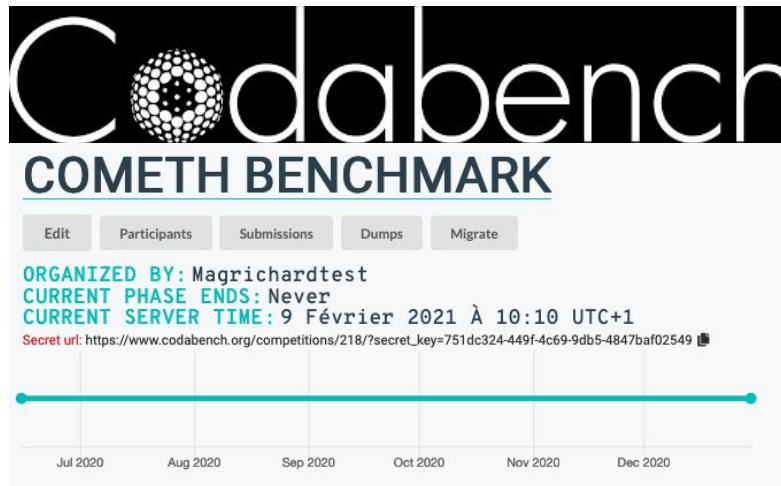
CURRENT SERVER TIME: 9 Février 2021 À 10:10 UTC+1

Secret url: [https://www.codabench.org/competitions/218/?secret\\_key=751dc324-449f-4c69-9db5-4847baf02549](https://www.codabench.org/competitions/218/?secret_key=751dc324-449f-4c69-9db5-4847baf02549)

Jul 2020 Aug 2020 Sep 2020 Oct 2020 Nov 2020 Dec 2020

# Access current leaderboard

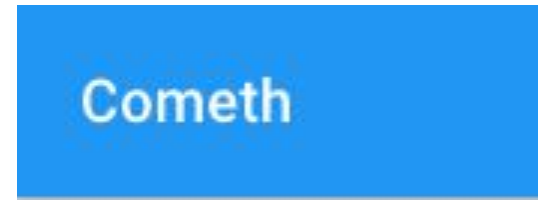
## Apply methods on clinical dataset



1) Send leaderboard

2) Run methods

3) Send results

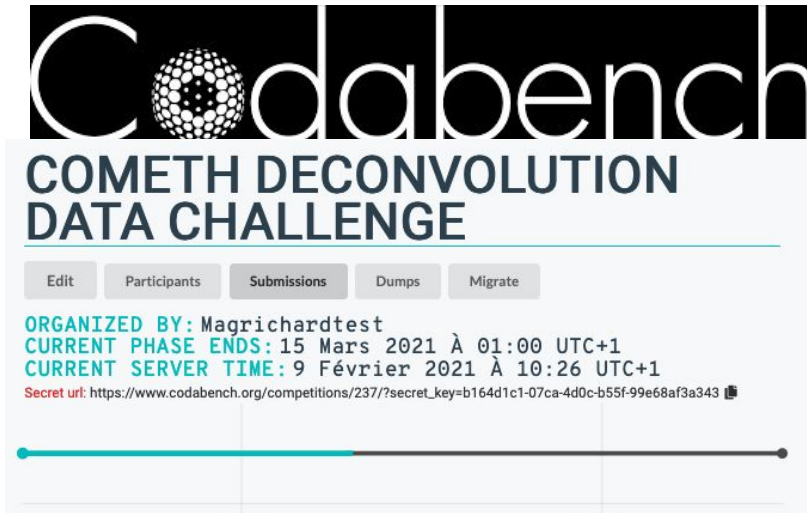


Method	DT8			DT9		
	Score (mean)	Score (sd)	Execution time	Score (mean)	Score (sd)	Execution time
MT17				0.366662	0	0
MT18	0.105368	0	0	0.555236	0	0
MT19	0.155399	0	0	0.0911333	0	0
MT14	0.2862	0	0	0.443478	0	0
MT16	0.432079	0	0	0.223464	0	0



**Clinicians**  
 Upload a dataset  
 Check leaderboard  
 Choose a method to run  
 Get the results

# In practical during the COMETH training



Computational group DAY 1-2

Learn how to contribute to the codabench benchmark using a toy data challenge



Medical group DAY 1-2

Learn how to use the user-friendly COMETH web application to run methods on toy TCGA clinical datasets

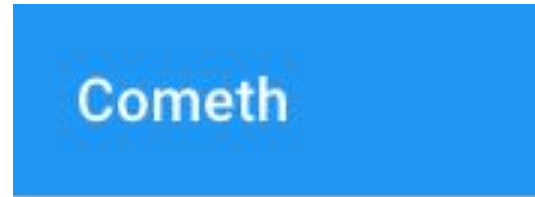
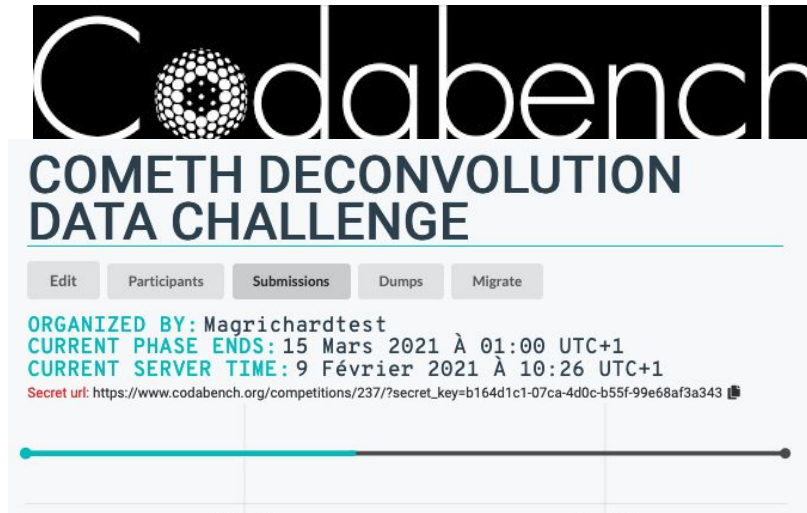


DAY 2

Learn how to biologically interpret the results of the methods



# Important Note



All digital platform develop within COMETH program are now at **staging state** of development  
The URLs provided during the training course are TEMPORARY  
Please fill up the final questionnaire and indicate your interest in the program to be informed when all the tools will be at **production state** (final public URLs)

LET'S START THE PRACTICAL WORK ON...

zoom

breakrooms



discord





UNIVERSITAT DE  
BARCELONA



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

Yuna Blum, Ligue contre le Cancer

Jérôme Cros, APHP

Clémentine Decamps, Uni Grenoble Alpes

Carl Herrmann, Medical Faculty Heidelberg

Slim Karkar, Uni Grenoble Alpes

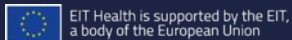
Yasmina Kermezli, Uni Grenoble Alpes

Magali Richard, Uni Grenoble Alpes

Ashwini Sharma, Uni Grenoble Alpes

[https://cancer-heterogeneity.github.io/cometh\\_training.html](https://cancer-heterogeneity.github.io/cometh_training.html)

[www.eithealth.eu](http://www.eithealth.eu) | [info@eithealth.eu](mailto:info@eithealth.eu)



EIT Health is supported by the EIT,  
a body of the European Union

# Practical session medical group

