What about this afternoon?

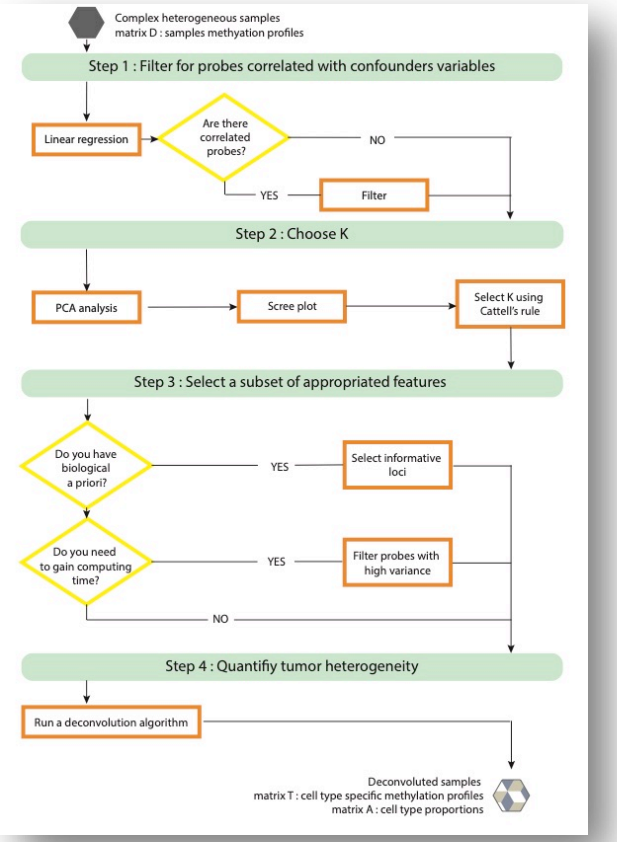
	TEAM #1, #2, #3, #4	TEAM #5, #6, #7, #8
2pm – 3pm	Pedagogy Room la scolette	Meeting report Mezzanine
3pm – 4pm	Meeting report Mezzanine	Pedagogy Room la scolette

BREAK

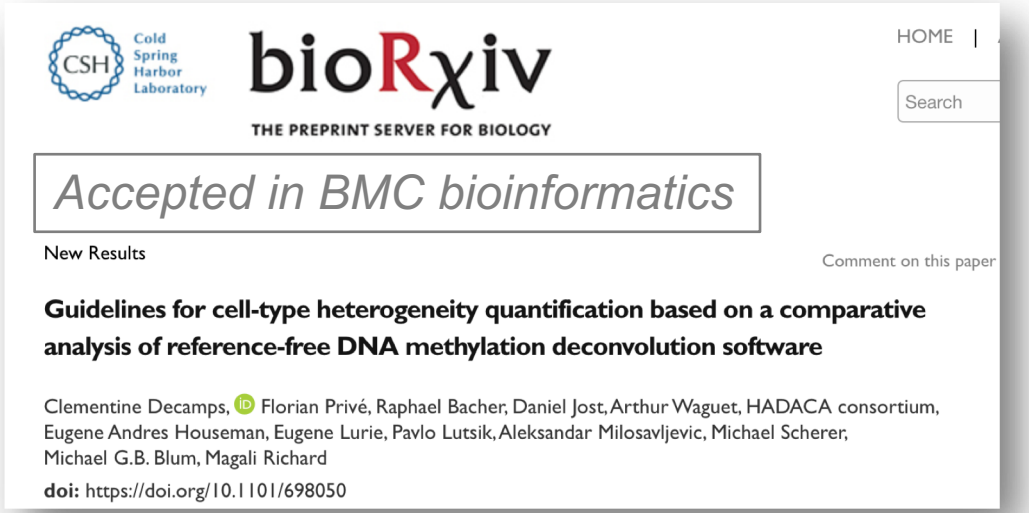
	???	???
4.30pm – 6pm	Brainstorming Biological interpretation Room la scolette	Brainstorming Benchmark dataset Mezzanine
6pm – 6.30pm	Restitution all together	Restitution all together

What came out of the first edition

- **Guidelines**



- **Article**



- **R package** *medepir*

  https://rdrr.io/github/bcm-uga/medepir/man/medepir-package.html

  M Richard, C Decamps , F Privé, M Blum

- **Blog posts**

Health data challenges organization: feedback, comments and recommendations.

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

In preparation

- **Collaborations**

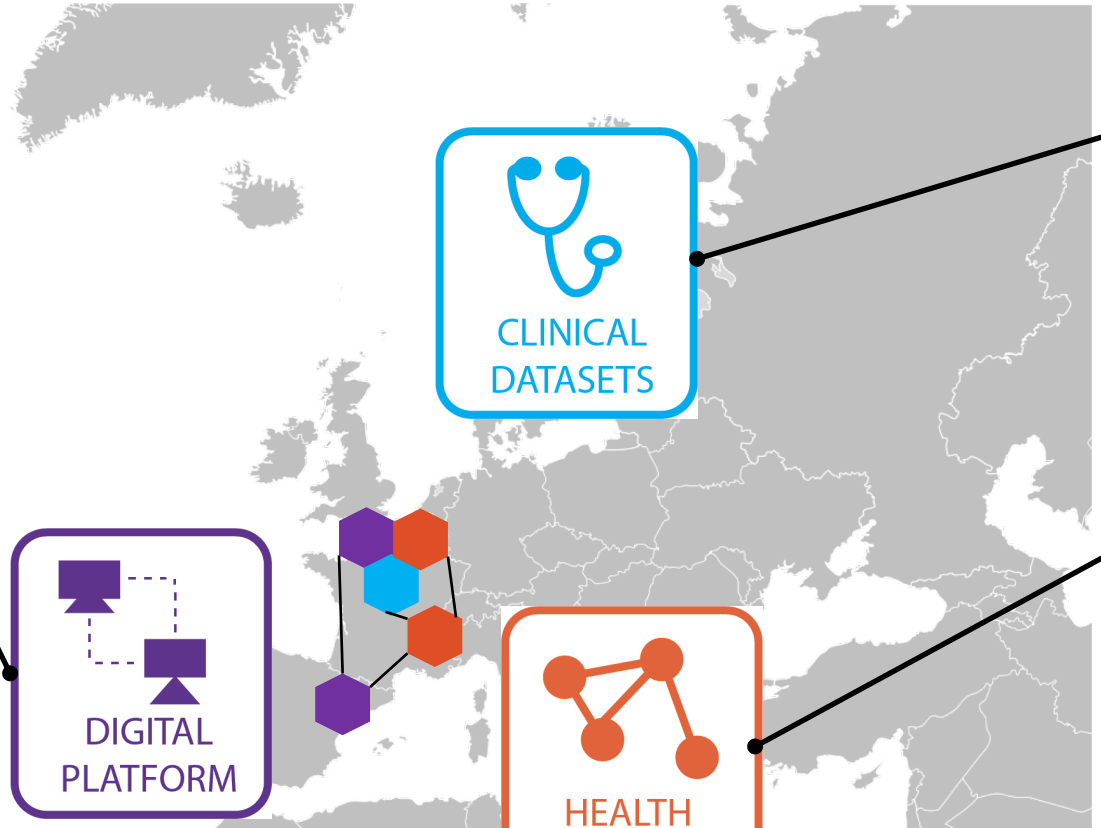
COMETH – COmputational METhods in Health

EIT Health call 2020 (600k€)

Isabelle Guyon

Sergio Escalera

Carl Herrmann



Jérôme Cros

Magali Richard

Y Blum

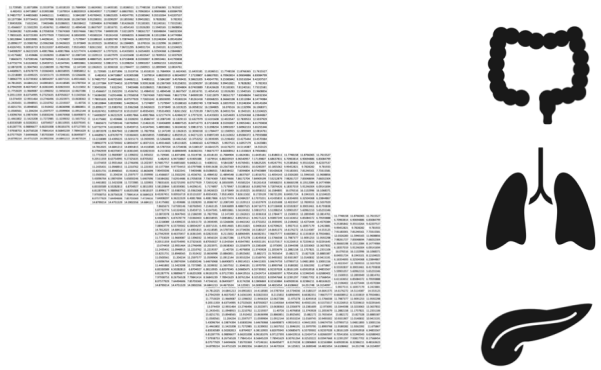
F Petitprez

COMETH – COmputational METHods in Health

EIT Health call 2020 (600k€)

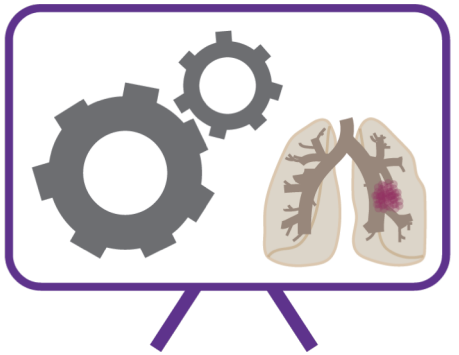
Unbiased evaluation of computational methods

- Generation of high quality benchmarking datasets



Various technologies and cancer types

- Development of a dedicated benchmarking platform



CodaLab

Fig1: Scheme of the benchmarking platform

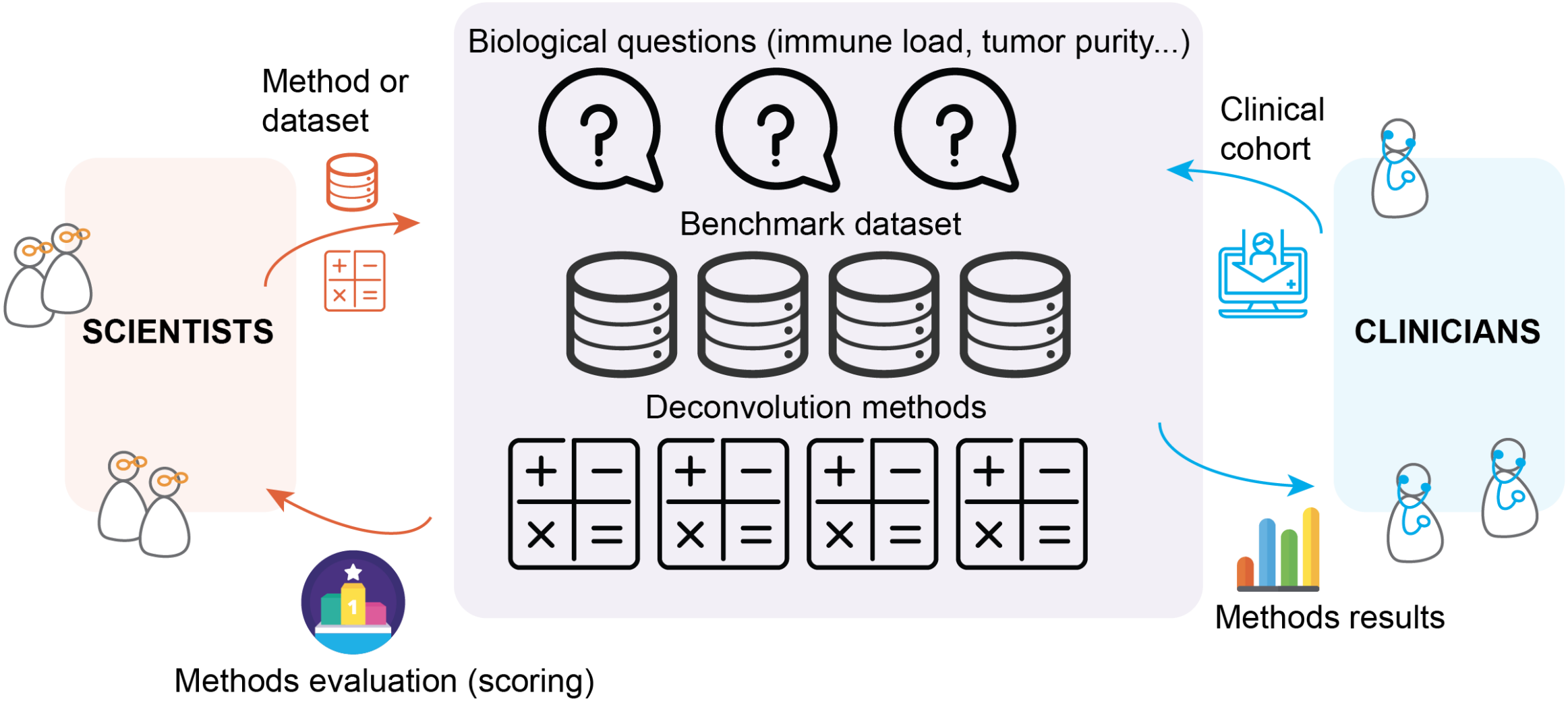


Fig2: Dataset. An example of pancreatic cancer dataset

? **What is the cellular composition of PDAC tumours?**

Estimation of the error between estimated proportion matrix and true proportion matrix

Pure Cell Type isolation

Omic Data

Data Pre-processing

Mixture

Model

PDX

GTEx

Cultured cells

PDX

Cell sorting (from FFPE)

RNAseq

Methylation

Normalization	**edgeR**
Transformation	Log2 +1
Aggregation	median

Cell types (genes) => T_{RNA} cell type profile matrix

Normalization	**Lumi**
Transformation	β-value
Aggregation	median

Cell types (probes) => T_{MET} cell type profile matrix

In silico Realistic Dirichlet distribution

samples (genes) / *Cell types* => Common A cell proportion matrix

$T_{RNA} * A$ => D_{RNA}

samples (genes)

$T_{MET} * A$ => D_{MET}

samples (probes)

Fig3: Benchmark result (example)

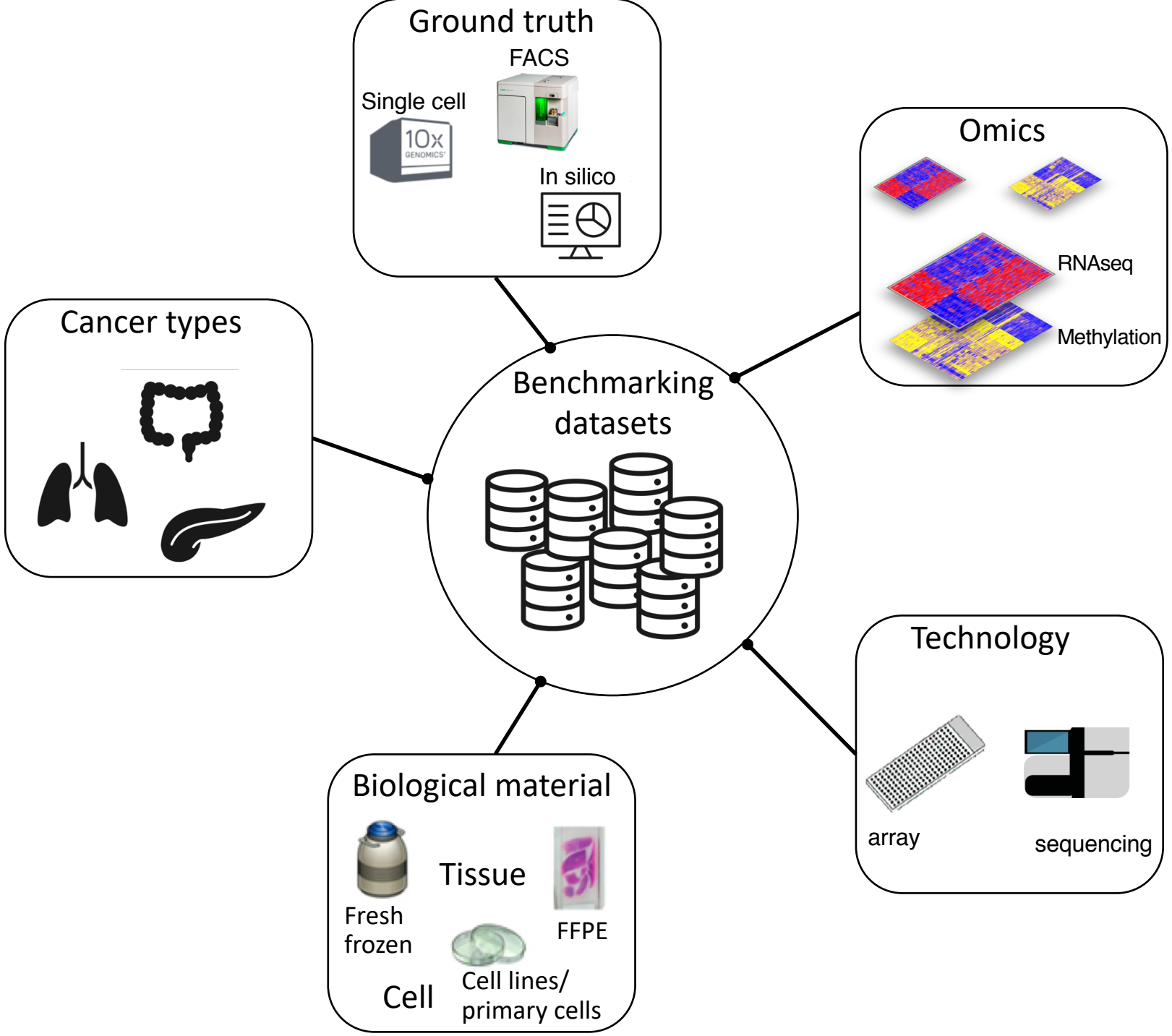

Ground truth
FACS
Single cell
10x GENOMICS
In silico
Omics
RNAseq
Methylation
Cancer types
Benchmarking datasets
Technology
array
sequencing
Biological material
Tissue
Fresh frozen
FFPE
Cell
Cell lines/ primary cells

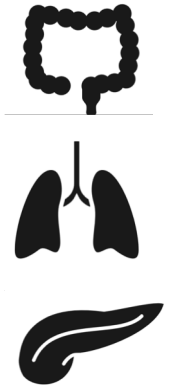
Fig5: Evaluation metrics

MAE error on estimated A (in silico simulations, FACS counting…)
RMSE error on estimated A (in silico simulations, FACS counting…)
Correlation with Immune Cell types (in silico simulations, FACS counting…)

BUDGET : Data generation

reflecting the diversity of samples clinicians may be confronted with

- Different cancer types
- Variety of sample types
- Different type of molecular data
- Innovative technologies



			Molecular level		
			Gene Expression		DNA methylation
Type of cancers	**Type of Samples**	Nb of samples	**3'RNA-seq 100€/sample**	**Single Cell RNAseq 5.6K€/sample**	**MethEpic 600€/sample**
Colorectal cancer	FFPE (archived)	30	35		35
	FF (Fresh/frozen)	5			
Lung cancer	FFPE	30	35		35
	FF	5			
Pancreatic cancer	FFPE	30	35	3	35
	FF	5			
	3 purified cell types	40	120		120
			22500	17000	135000
		TOTAL	**~170,000 €**		

COMETH – COmputational METhods in Health

EIT Health call 2020 (600k€)