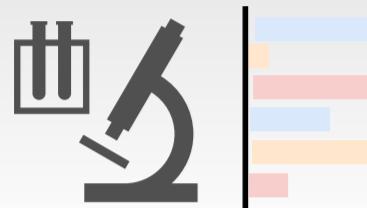


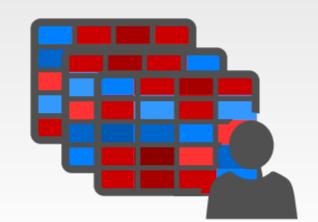


Massive resources,
diverse data types,
varied omics sizes, etc.

- 
- .fastq
 - .maf
 - .bam
 - .txt
 - .csv
 -

Data integration&processing

- Omics recruitment
- Protocol verification
- Cleaning&alignment
-

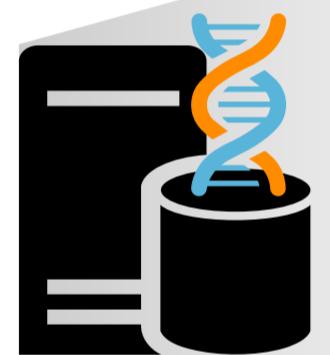
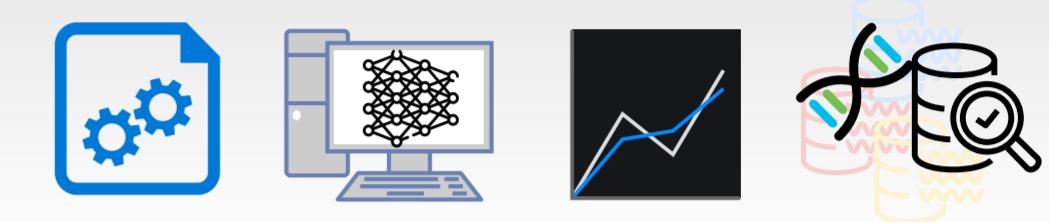


- Table A
- Table B
- Table C



Datasets construction

- Benchmark ML tasks
- Baseline collection
- Metrics and Analysis
- Bio-databases linking

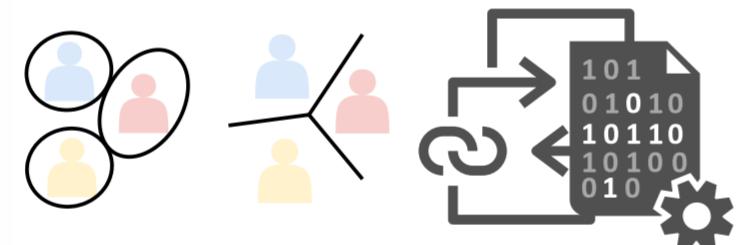


MLomics

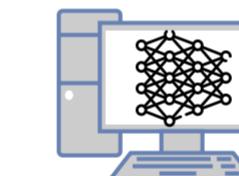
(32 cancer types)



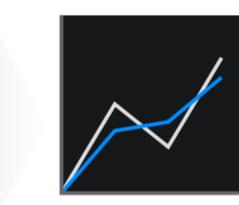
- 20 learning tasks
- 6 classification datasets
 - 9 clustering datasets
 - 5 omics imputation datasets



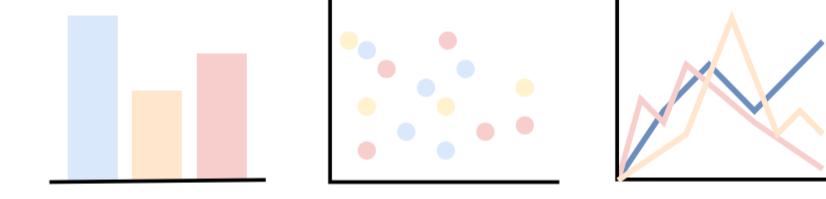
- 10 classification baselines
- 10 clustering baselines
- 7 imputation baselines



Statistical/ML/DL methods



Evaluation, metrics,
and visualization



- Bio-bases mapping
- Clinical records
- Analysis tools
 - Vocano, GO, etc.
-

Bio-knowledge database linking

MLomics: Machine Learning Cancer Multi-Omics Benchmark with Datasets, Tasks, and Baselines



Integrate two resources
and ML models

