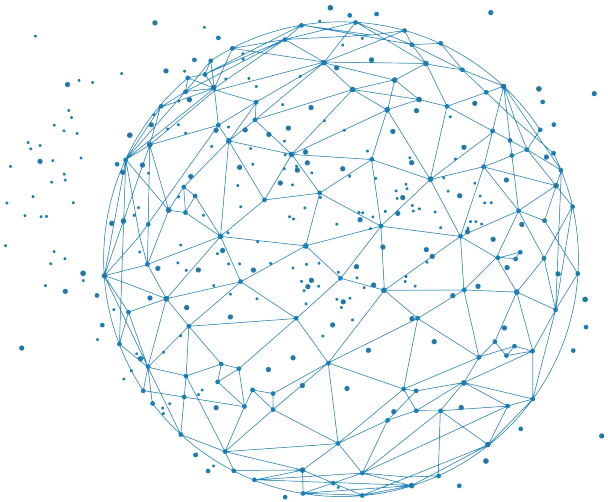
Third edition 2026 in Fréjus

ÉCOLE THÉMATIQUE ETBII BIOINFORMATIQUE INTÉGRATIVE IFB



Omics integration - General aspects

Jimmy Vandel

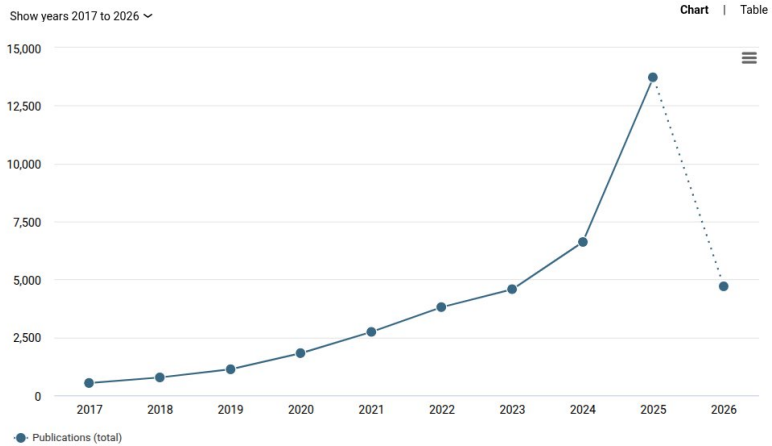
IFB INSTITUT FRANÇAIS DE BIOINFORMATIQUE

elixir FRANCE

cnrs INRAE Inserm cea

Rise in popularity

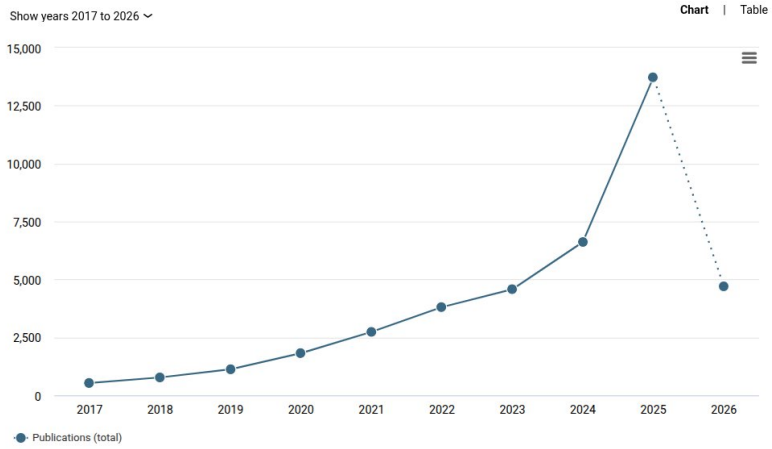
“Multi-omics” citations



https://app.dimensions.ai/discover/publication *(22th Mar. 2026: 163,356,634 referenced publications)*

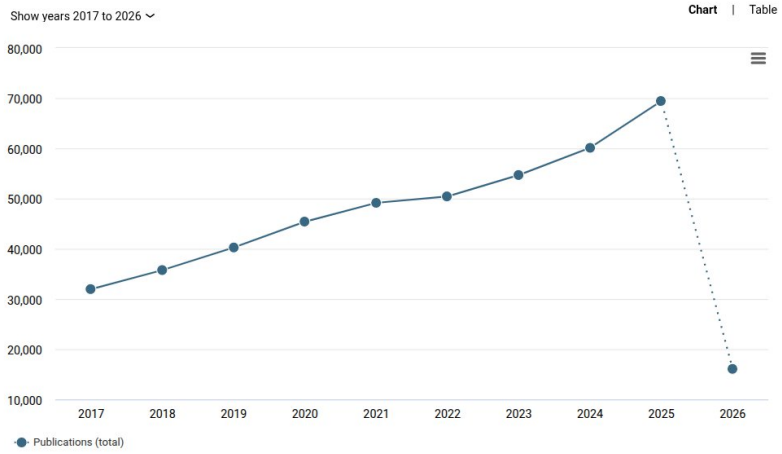
Rise in popularity

“Multi-omics” citations



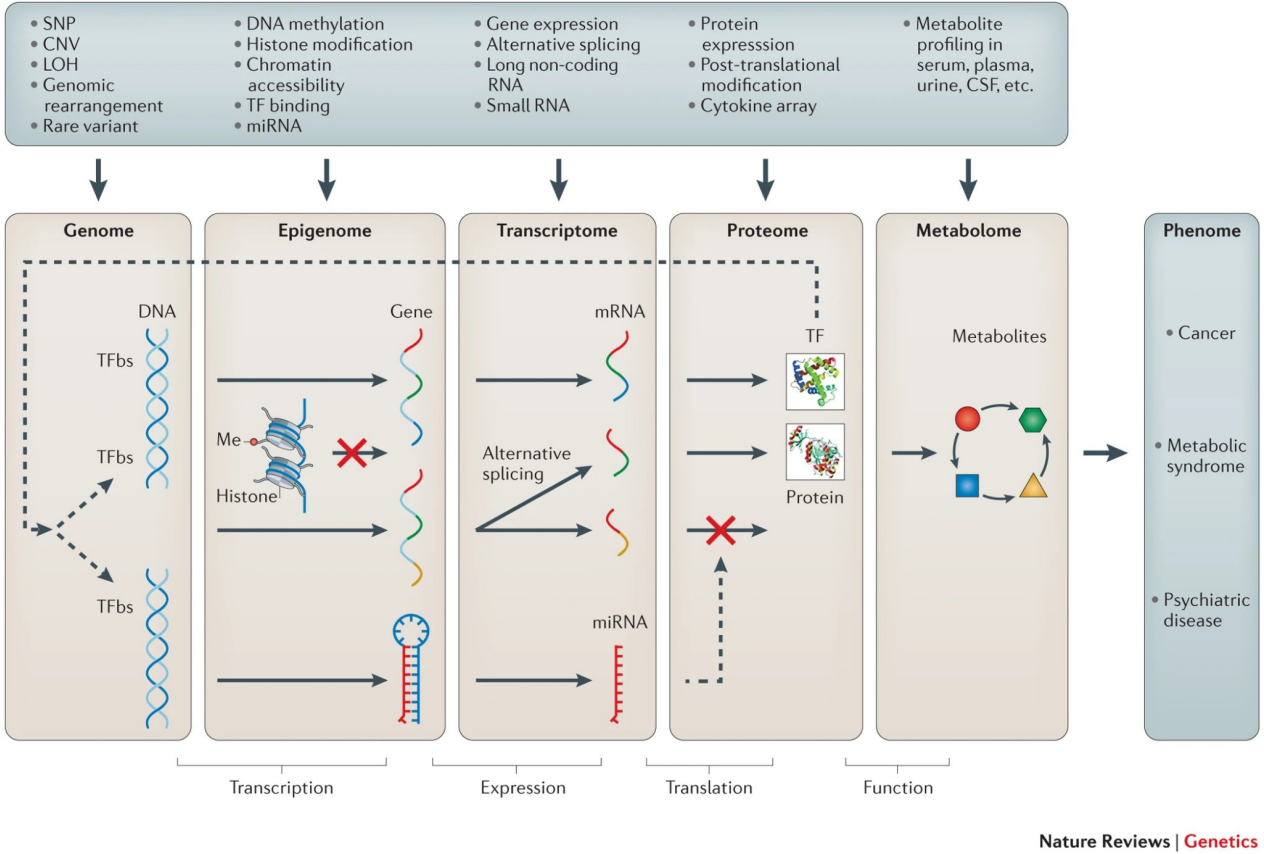
Name Fields of Research (ANZSRC 2020) code	↓Publications	Citations
Biomedical and Clinical Sciences 32	22,930	263,806
Biological Sciences 31	20,195	348,609
Oncology and Carcinogenesis 3211	7,755	73,484
Medical Biochemistry and Metabolomics 3205	6,286	90,509
Genetics 3105	6,141	111,426
Bioinformatics and Computational Biology 3102	5,764	105,554

“Single-cell” citations



Name Fields of Research (ANZSRC 2020) code	↓Publications	Citations
Biomedical and Clinical Sciences 32	515,971	18,438,401
Biological Sciences 31	329,178	14,987,432
Oncology and Carcinogenesis 3211	164,331	4,548,786
Biochemistry and Cell Biology 3101	164,230	7,806,708
Engineering 40	141,413	4,130,923
Clinical Sciences 3202	121,843	3,710,657

Omics… which ones ?

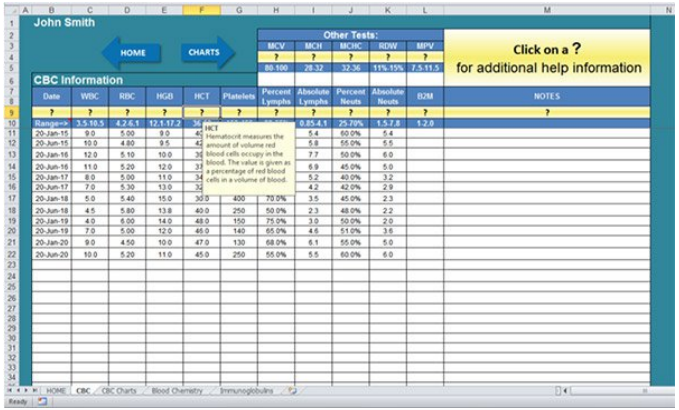


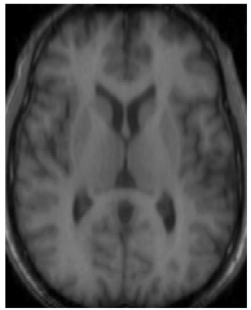
Ritchie, M., Holzinger, E., Li, R. et al. Methods of integrating data to uncover genotype–phenotype interactions. Nat Rev Genet 16, 85–97 (2015).

But also ?

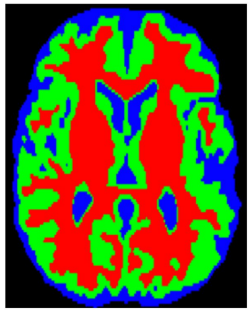
Other related data ?

- clinical data
- imaging data / radiomics / pathomics (full data or extracted characteristics)
- new omics fields : fluxomics, ionomics, microbiomics, glycomics…
- biological knowledge : DNA/protein, protein/protein interactions, DNA recombination
  → a priori in model definition/construction





(a) Axial slice



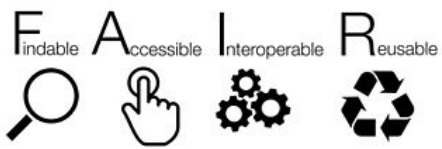
(b) Tissue segmentation

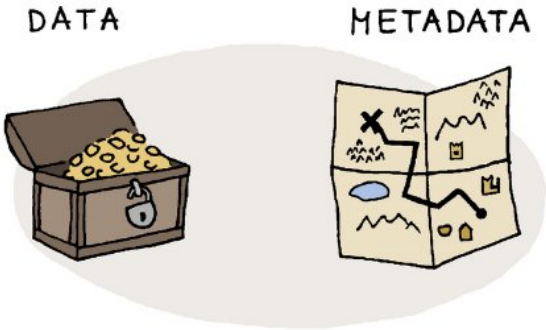
Data format

– **Omics data**: mostly dependent on omics and/or technology (FASTQ, VCF, mzML...)
→ numeric matrices

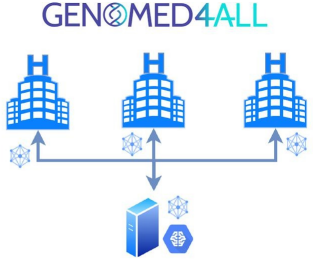
– **Clinical**: very heterogeneous (care/research), coding standards (diagnoses, medication...), electronic health records with interoperable standards such as openEHR or FHIR
→ numeric matrices

– **Imaging data**: Digital Imaging and Communications in Medicine (DICOM) standard (image storage, metadata definition and communication protocols) but also TIFF, NifTI
→ numeric matrices





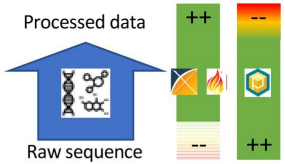
Data accessibility



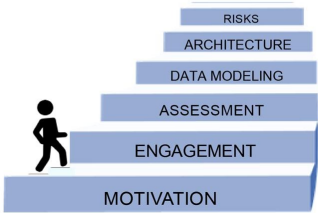
Common data model for federated learning on multi-omics and clinical data?

Challenges

Assess interoperability standards

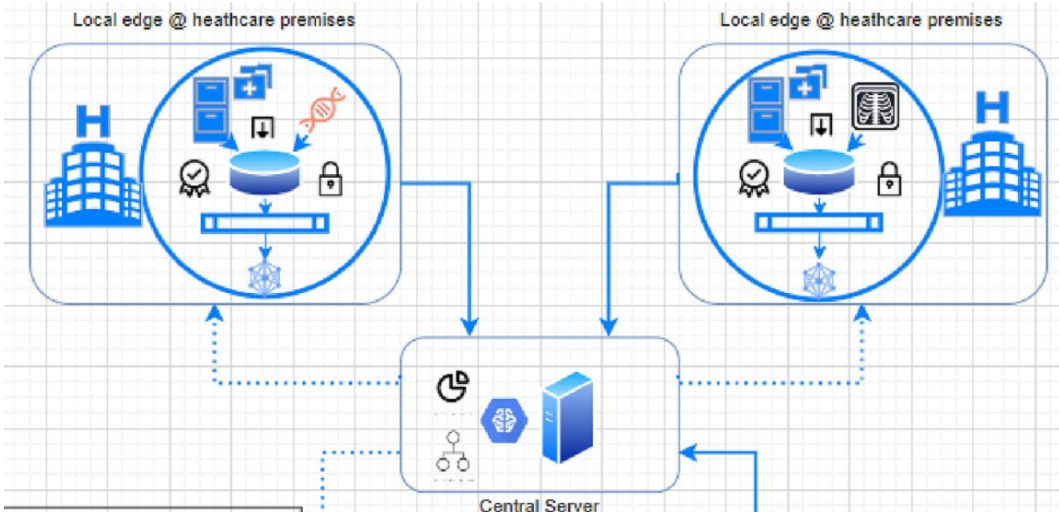


Effective data modeling strategy



Cremonesi F et al.. The need for multimodal health data modeling: A practical approach for a federated-learning healthcare platform. J Biomed Inform. 2023

Federated learning



Wish-list des types de données

Génomique
- WGS
- Génotypage (puces ADN, PMRA)
- Variants génétiques (eQTL, mQTL, GWAS)
- Métagénomique (16S, short/long reads)

Épigénomique
- Méthylation (EPIC, bisulfite sequencing, RRBS, EM-seq)
- Accessibilité de la chromatine (ATAC-seq)
- Modifications d'histones (ChIP-seq, Cut&Run, Cut&Tag)
- Épitranscriptomique (m6A, R-loops)

Transcriptomique
- RNA-seq (bulk, long/short reads, Nanopore, Illumina)
- scRNA-seq/snRNA-seq (10X Genomics, Visium, etc.)
- Nascent RNA-seq
- Métatranscriptomique (microbiote, pathogènes)
- miRNA-seq (puces ou séquençage)
- CITE-seq

Protéomique, métabolomique et lipidomique
- Spectrométrie de masse
- Ribosome profiling (traductome)
- Métaprotéomique (microbiote)
- Protéines associées à la chromatine (Pol II, etc.)

Mais aussi : données spatiales, imagerie, cytométrie, phénotypique, clinique

Vous connaissez/utilisez/aimeriez utiliser

Analyse single-omics :
DESeq2, edgeR, Seurat, Boruta (métabolomique), ChromHMM

Corrélation et clustering

Analyse d'enrichissement :
GO terms

Visualisation :
ShinyR, PCA/MDS (réduction de dimension)

Outils d'intégration :
WNN, WGCNA, SNF, iCluster, PLS-DA, MixOmics, DIABLO, RGCCA, SGCCA, ComDim/CCSWA, MOFA, network diffusion hierarchique, MINI-EX

Vous connaissez/utilisez/aimeriez utiliser

Analyse single-omics :
DESeq2, edgeR, Seurat, Boruta (métabolomique), ChromHMM

Corrélation et clustering

Analyse d'enrichissement :
GO terms

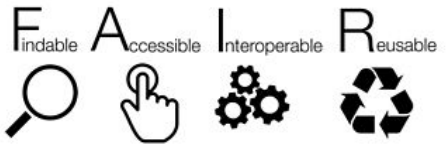
Visualisation :
ShinyR, PCA/MDS (réduction de dimension)

Outils d'intégration :
WNN, WGCNA, SNF, iCluster, PLS-DA, MixOmics, DIABLO, RGCCA, SGCCA, ComDim/CCSWA, MOFA, network diffusion hierarchique, MINI-EX

Integration: motivation and challenges

- Take advantage of the vast amount of available data
  - Data access (local/national regulation, infrastructures…)
  - Data representation (structuration, ontologies…)

  → Need of common representation framework



- Improve our understanding of biological phenomena
  - Data heterogeneity (technology, format, meaning, distribution…)
  - Data complexity (dependances/independances, ad-hoc assumptions…)
  - Amount a data (time/memory consumming)

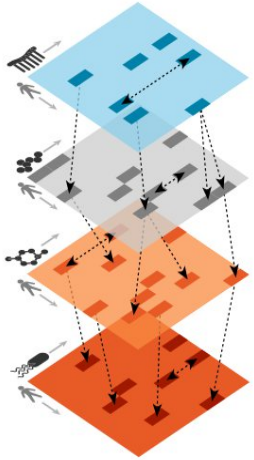
  → Need for methods adapted to these data

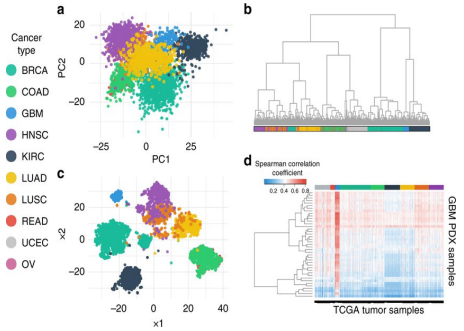
Integration: biological motivation

- Deep insights into biology phenomenon

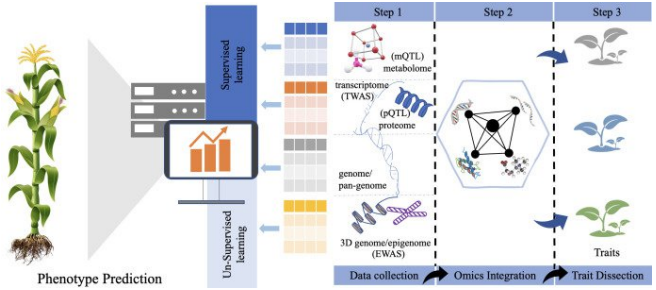
- Subtyping and classification (disease, species, varieties)

- Biomarkers prediction: diagnostic, disease drivers, plant/animal selection…



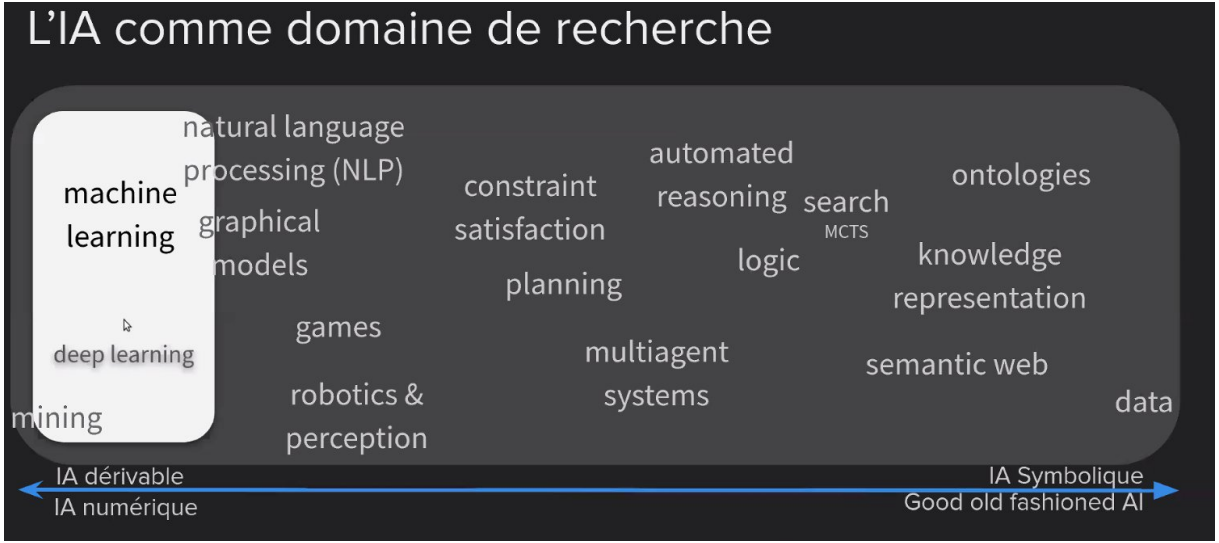


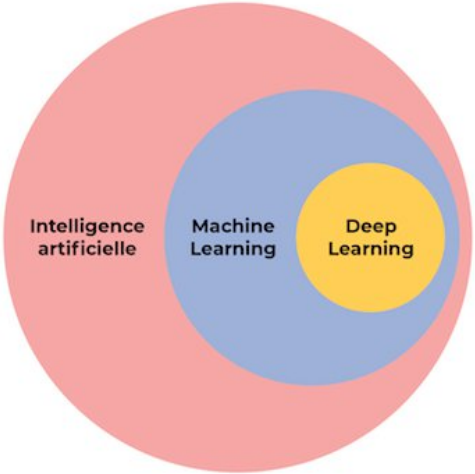
Vasileios et al (2018). Drug and disease signature integration identifies synergistic combinations in glioblastoma. Nature Communications. 9.



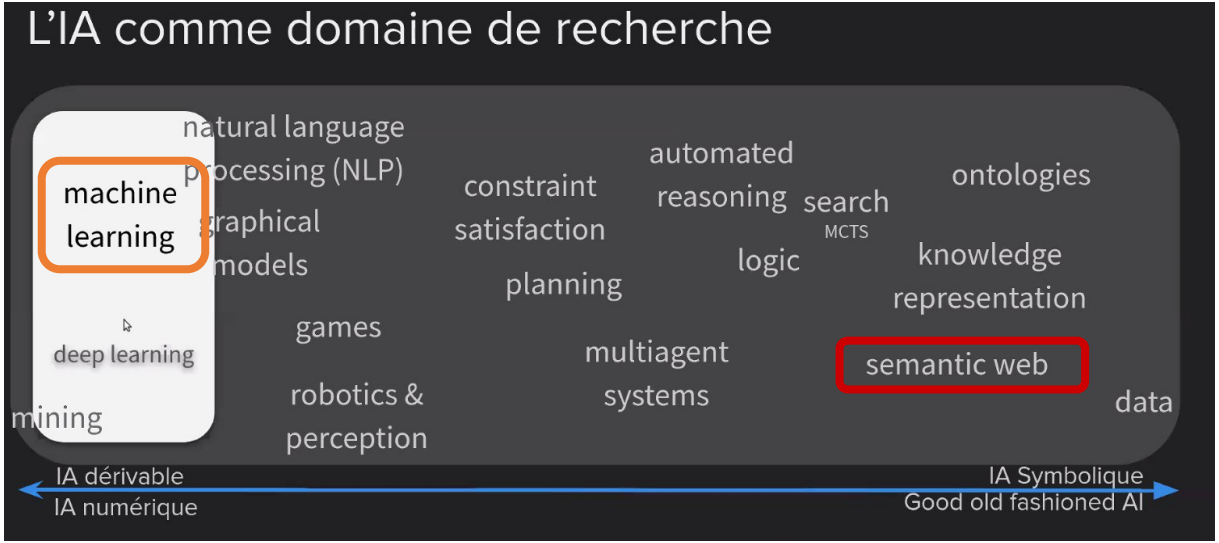
Mahmood et al (2022) Multi-omics revolution to promote plant breeding efficiency. Front Plant Sci Dec 8.

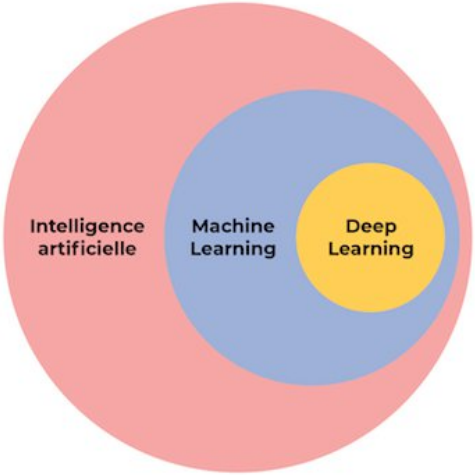
Artificial intelligence of course … and so ?





Artificial intelligence of course … and so ?

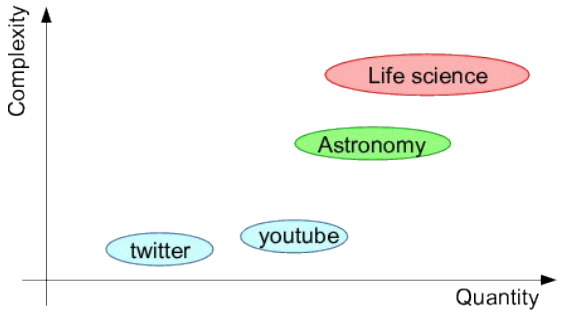




Improve our understanding of biological phenomena

Take advantage of the vast amount of available data

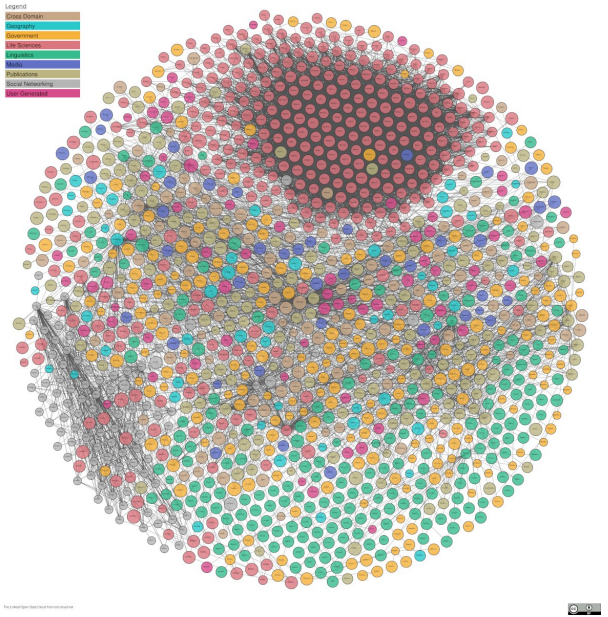
Take advantage of the vast amount of available data



Big Data: Astronomical or Genomical?

Zachary D. Stephens[1], Skylar Y. Lee[1], Faraz Faghri[2], Roy H. Campbell[2], Chengxiang Zhai[3], Miles J. Efron[4], Ravishankar Iyer[1], Michael C. Schatz[5]*, Saurabh Sinha[3]*, Gene E. Robinson[6]*

PLOS Biology | DOI:10.1371/journal.pbio.1002195 July 7, 2015



Life science: 1600+ reference databases

→ integrating heterogeneous data and knowledge is (badly) needed!

Editorial > Nucleic Acids Res. 2022 Jan 7;50(D1):D1-D10. doi: 10.1093/nar/gkab1195.

The 2022 Nucleic Acids Research database issue and the online molecular biology database collection

Daniel J Rigden [1], Xosé M Fernández [2]

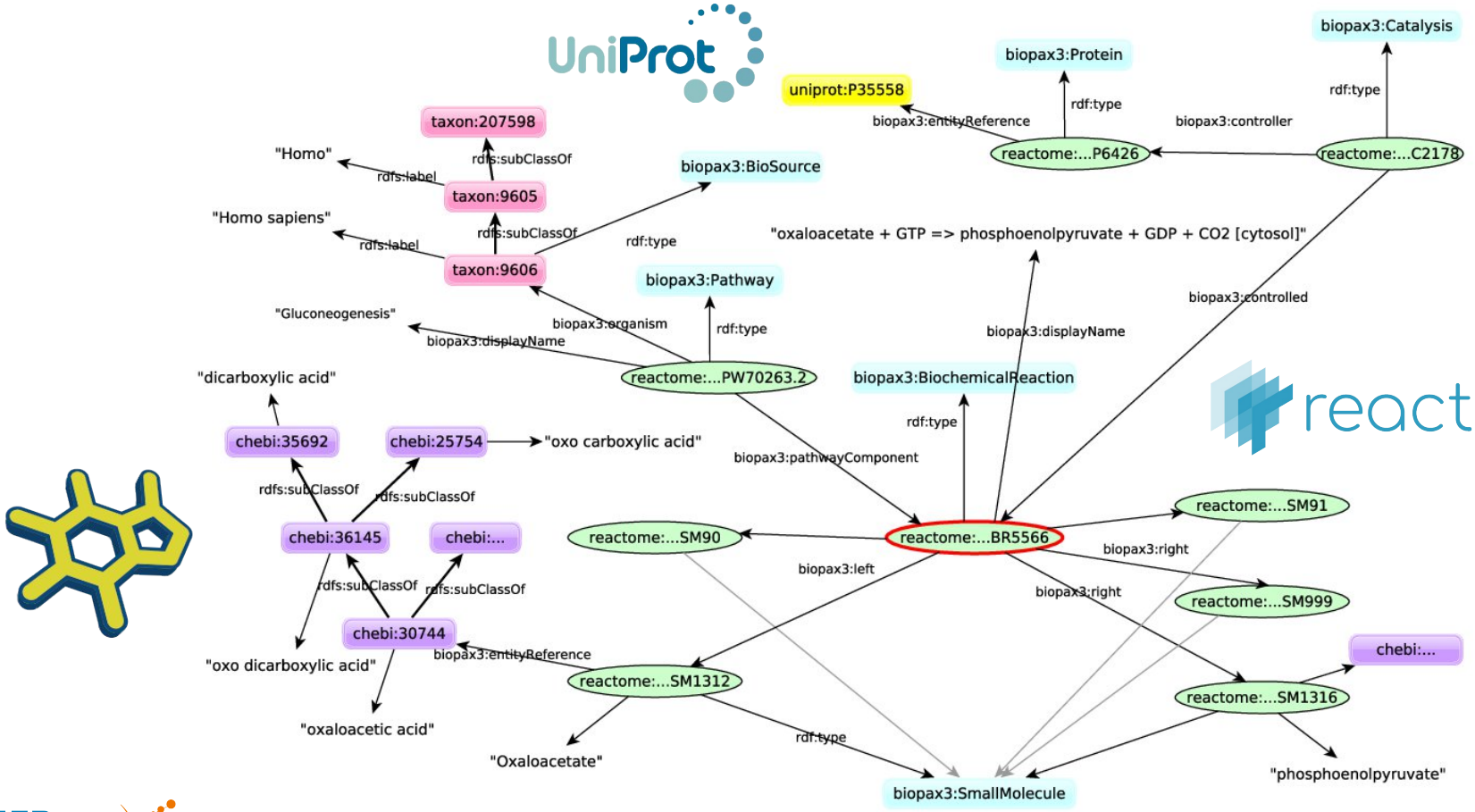
Affiliations + expand

PMID: 34986604 PMCID: PMC8728296 DOI: 10.1093/nar/gkab1195

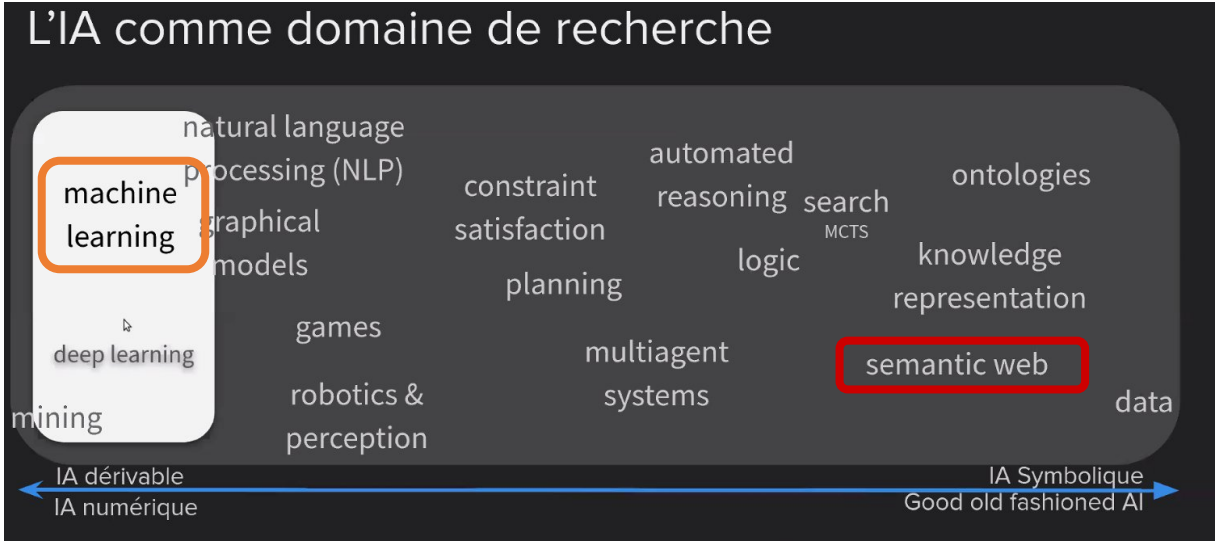
Semantic Web = framework for:

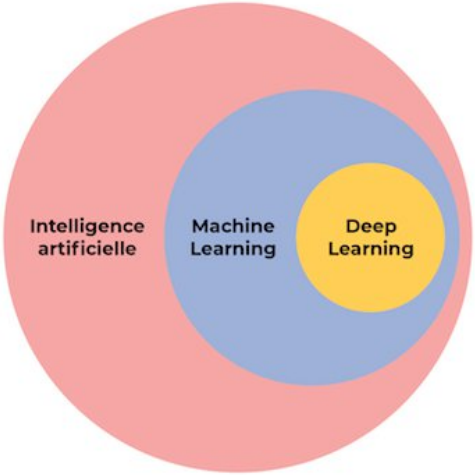
- **integrating** data and knowledge
- **querying**
- **reasoning**

Take advantage of the vast amount of available data



Artificial intelligence of course … and so ?

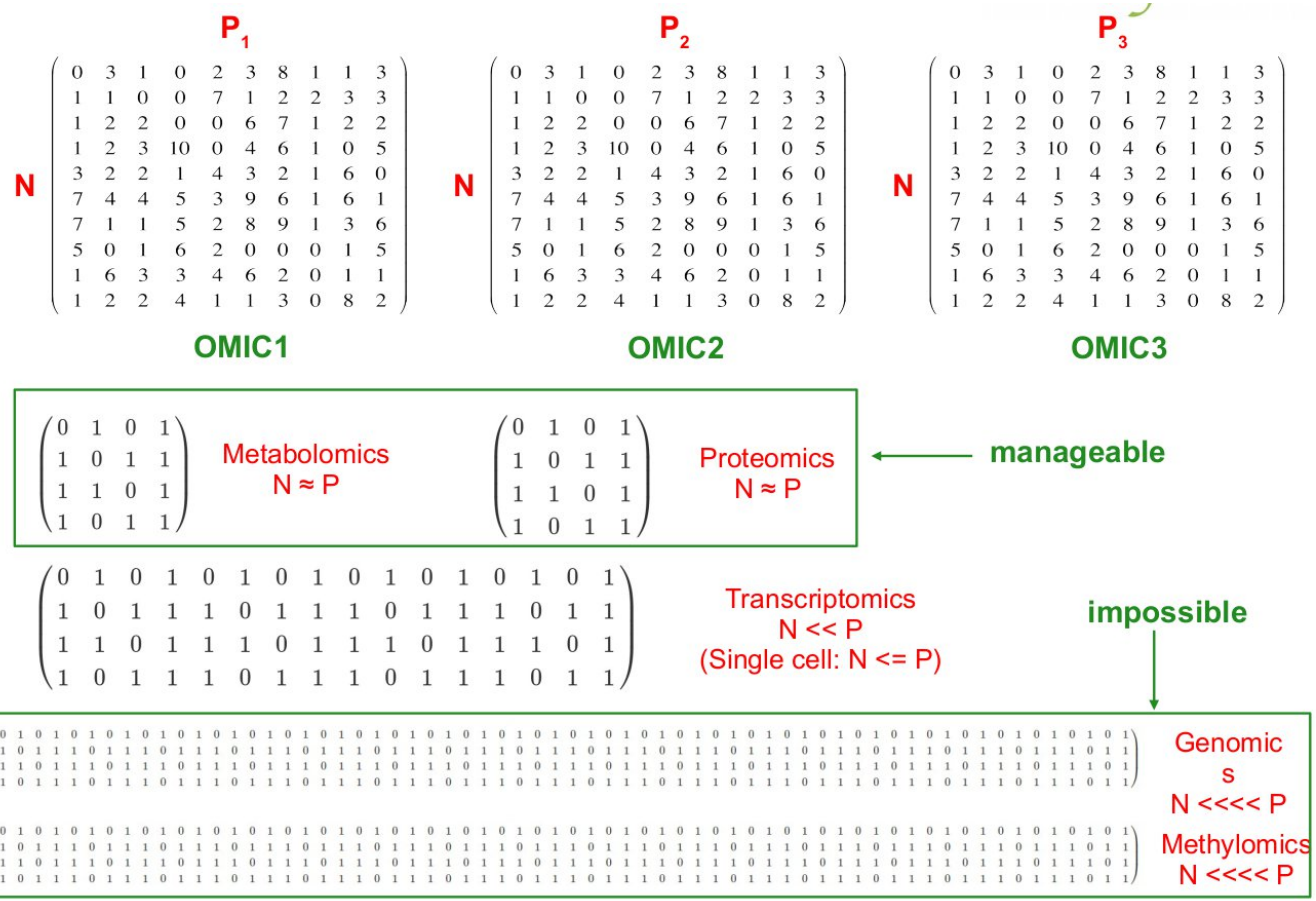




Improve our understanding of biological phenomena

Take advantage of the vast amount of available data

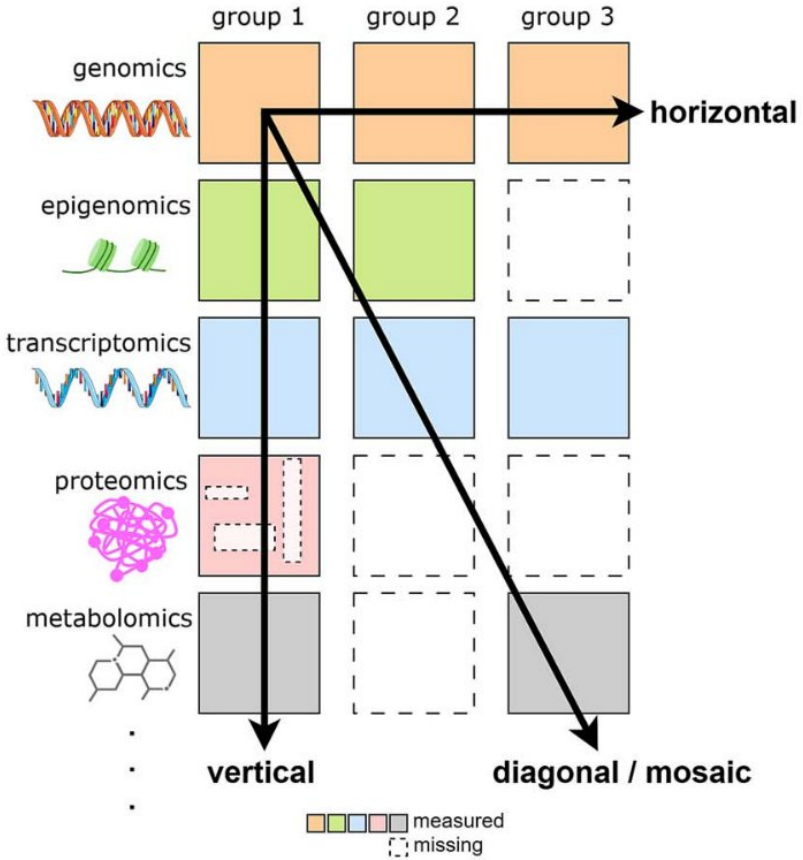
Curse of dimensionality



P : number of features (genes, proteins, genetic variants…)

N : number of observations (samples, cells, nucleotides…)

Integration strategies



- **Vertical integration**: combines different omics modalities within the same group of samples
- **Horizontal integration**: aligns datasets from the same omics layer across different sample groups (e.g. batches, cellular models)
- **Diagonal integration**: combines distinct omics modalities from different sample groups to explore inter-modality relationships across groups
- **Mosaic integration:** leverages overlapping modalities across samples to infer relationships and impute missing modalities

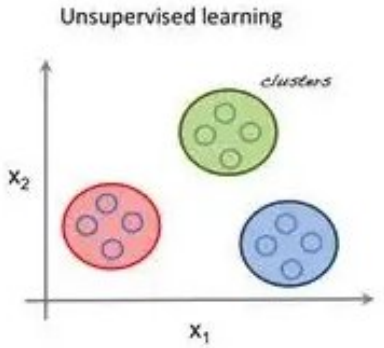
Baião AR. et al. A technical review of multi-omics data integration methods: from classical statistical to deep generative approaches. Brief Bioinform. 2025

- ## Unsupervised learning

find hidden patterns, analyze and organize unlabelled data.

eg: clustering, dimension reduction, density estimation

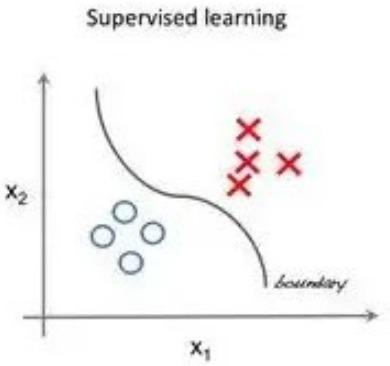
  - Self-supervised learning : optimise a loss function based on a ground truth



- ## Supervised learning

use labeled data to predict new observation outcome (predictive model).

eg: classification task (categorical/numerical), regression (numerical)



- ## Semi-supervised

use labeled and unlabelled data to predict the unlabelled data and/or outcome

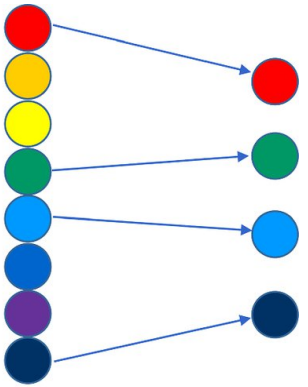
eg: inductive/transductive approaches

And also: transfer learning, reinforcement learning

Feature selection

→ determine a smaller set of features minimizing (relevant) information loss

eg: filtering methods (correlation), recursive elimination, regularization

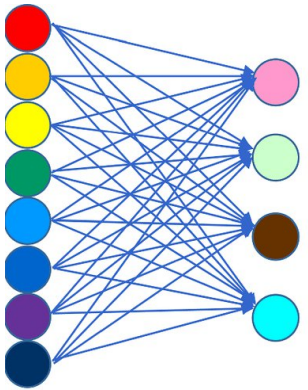


Feature extraction

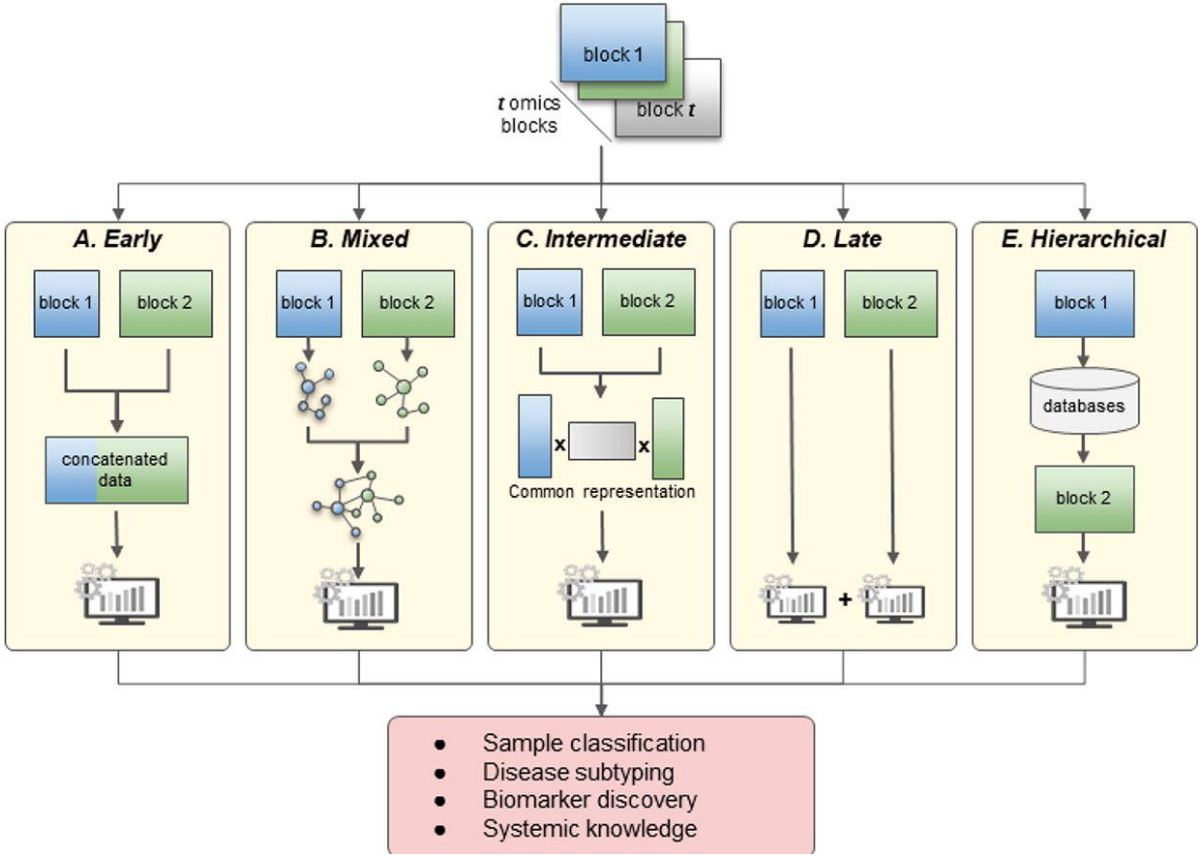
→ combine the input features into another set of variables in a linear or non-linear fashion

eg: PCA, PCoA, ICA…

+ regularization for sparse methods: sPCA, sNMF

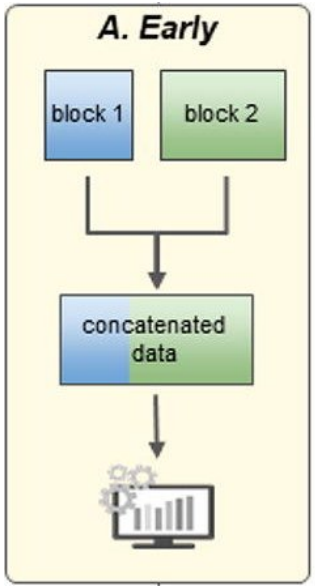


Integration strategies



Picard M. et al. Integration strategies of multi-omics data for machine learning analysis. Comput Struct Biotechnol J. 2021.

Integration strategies



Concatenate every omics datasets into a single large matrix.

Pros :

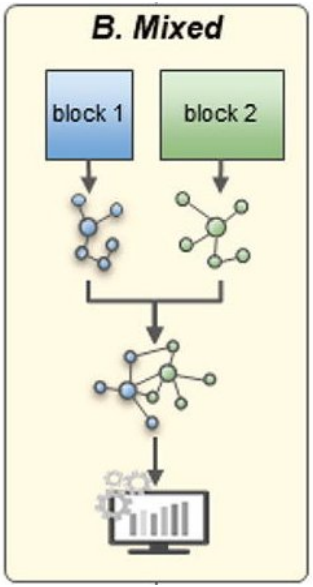
- conceptually simple
- easy to implement
- can reveal interactions between omics

Cons :

- increased noise and dimensionality (concatenated matrix)
- need to deal wih imbalanced omics datasets
- ignores the specific data distribution of each omics
- need a common dimension (rows or columns → samples or features)

Picard M. et al. Integration strategies of multi-omics data for machine learning analysis. Comput Struct Biotechnol J. 2021.

Integration strategies



Transform independently each omics dataset into a simpler representation before integration.

Pros :

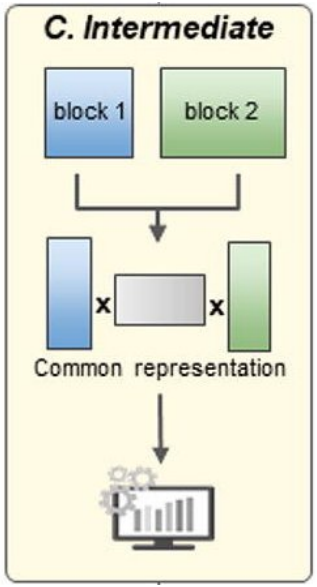
- the new representation is less dimensional and less noisy
- reduce heterogeneity between omics
- classical approaches can be used on the new representation

Cons :

- the choice of the transformation method is not trivial
- risk of information loss during transformation
- need a common dimension (rows or columns → samples or features)

Picard M. et al. Integration strategies of multi-omics data for machine learning analysis. Comput Struct Biotechnol J. 2021.

Integration strategies



Jointly integrate the multi-omics datasets without prior transformation.

Pros :

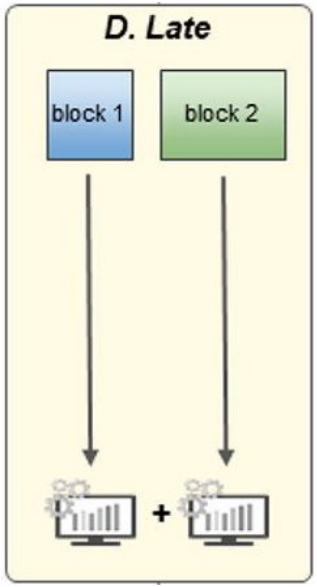
- reduce information loss compare to the mixed strategy
- discover a joint inter-omics structure
- highlight the complementary information between omics

Cons :

- can require robust pre-processing step to reduce heterogeneity
- common latent space assumption

Picard M. et al. Integration strategies of multi-omics data for machine learning analysis. Comput Struct Biotechnol J. 2021.

Integration strategies



Apply machine learning models separately on each omics dataset and then combine results.

Pros :

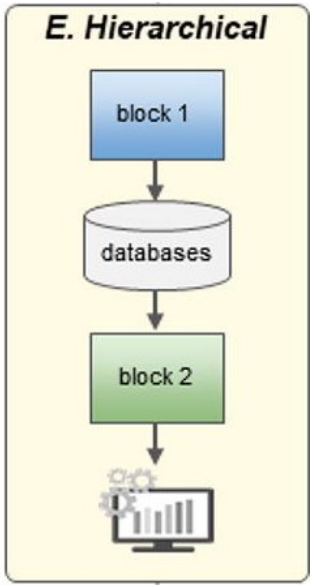
- avoid the (many) challenges associated with direct omics integration
- you can use tools designed specifically for each omics
- classical approaches can be used to combine results

Cons :

- cannot capture direct inter-omics interactions
- complementarity information between omics is not exploited

Picard M. et al. Integration strategies of multi-omics data for machine learning analysis. Comput Struct Biotechnol J. 2021.

Integration strategies



Include prior knowledge of omics relationships.

Pros :

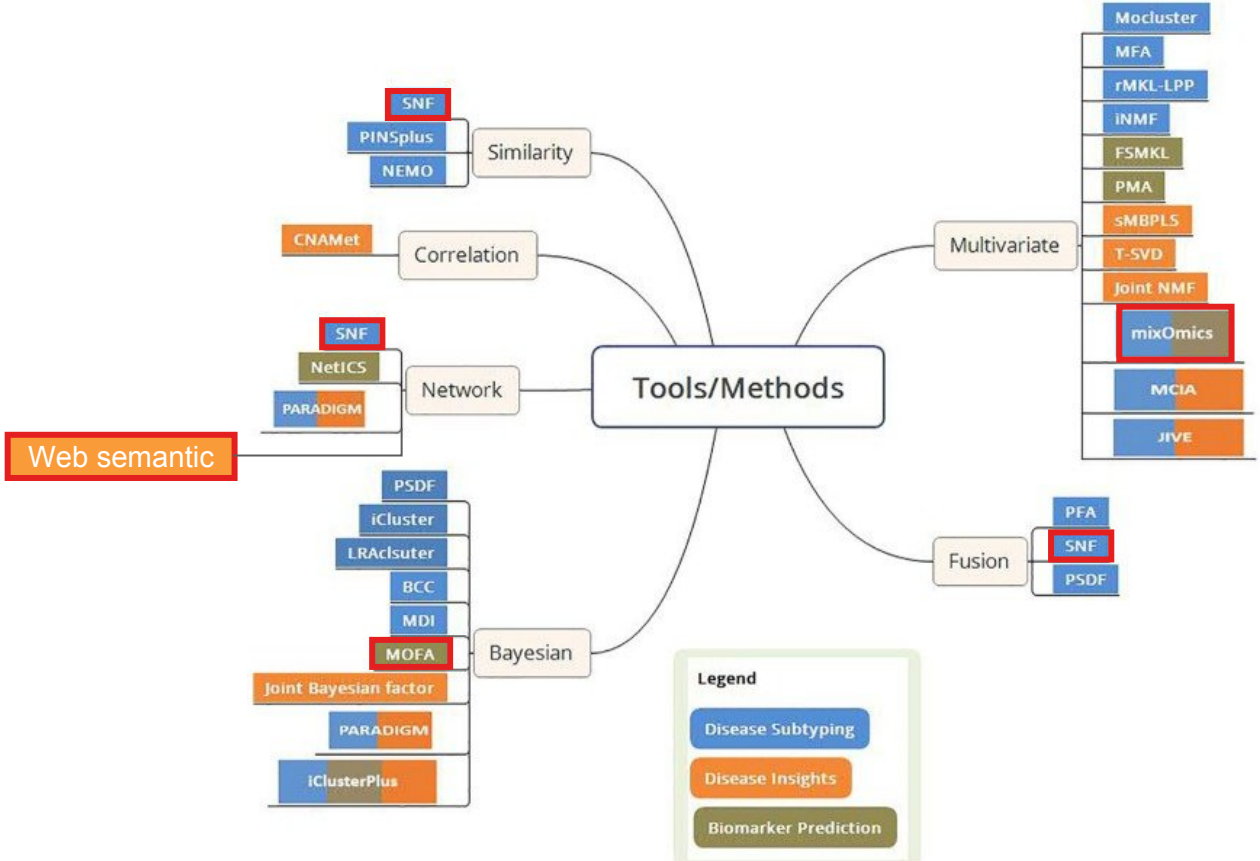
- reduced complexity (sequential integration)
- integrate external knowledge

Cons :

- less generic than previous strategies

Picard M. et al. Integration strategies of multi-omics data for machine learning analysis. Comput Struct Biotechnol J. 2021.

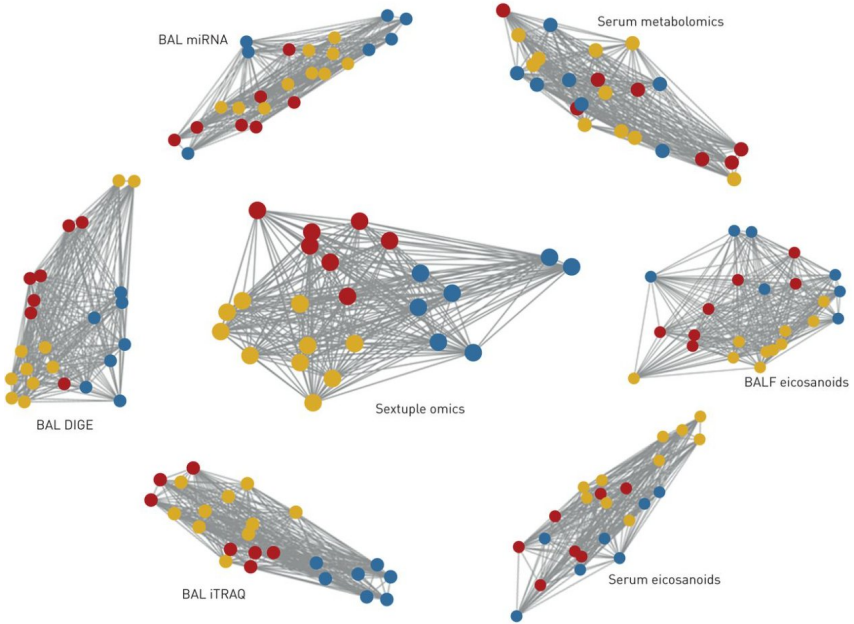
Integration approaches



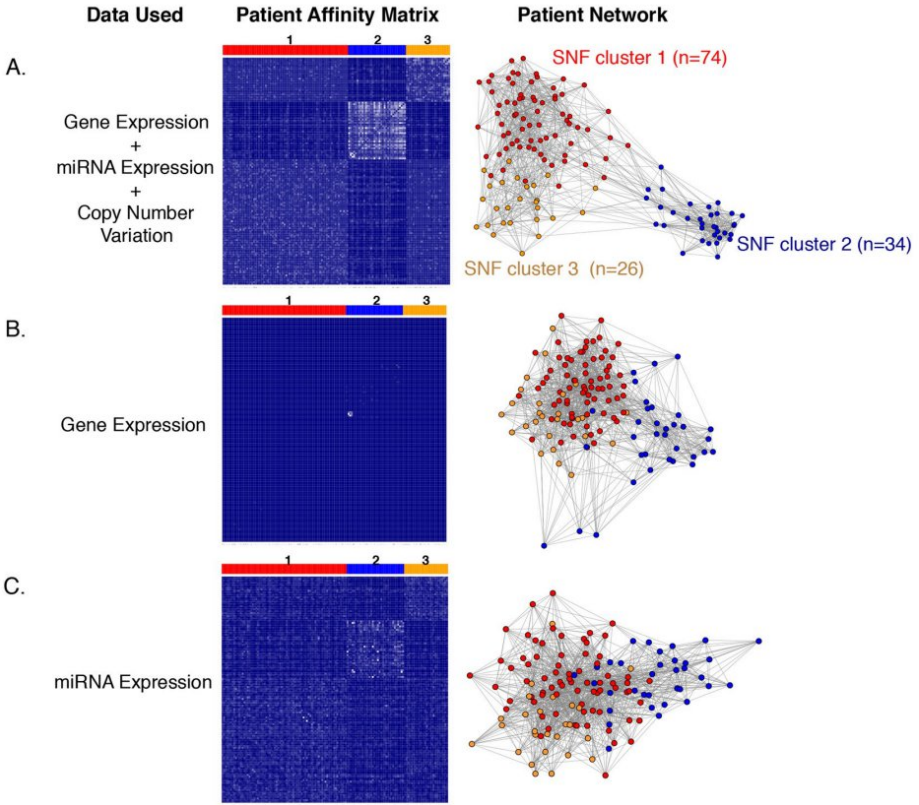
Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics Data Integration, Interpretation, and Its Application. Bioinform Biol Insights. 2020

Network based methods (*eg* SNF)

Capturing shared topological structures and interactions across omics layers through networks

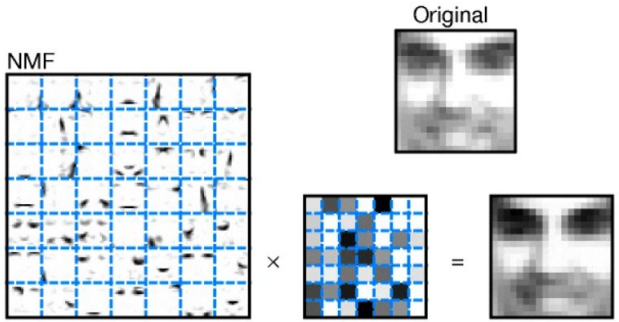


Li C et al. Integration of multi-omics datasets enables molecular classification of COPD. Eur Respir J. 2018

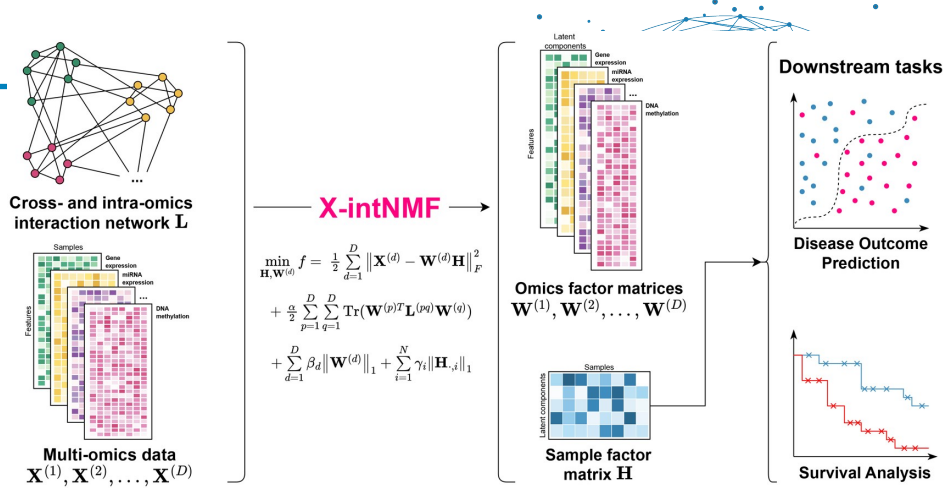


Chiu AM et al. Integrative analysis of the inter-tumoral heterogeneity of triple-negative breast cancer. Sci Rep. 2018

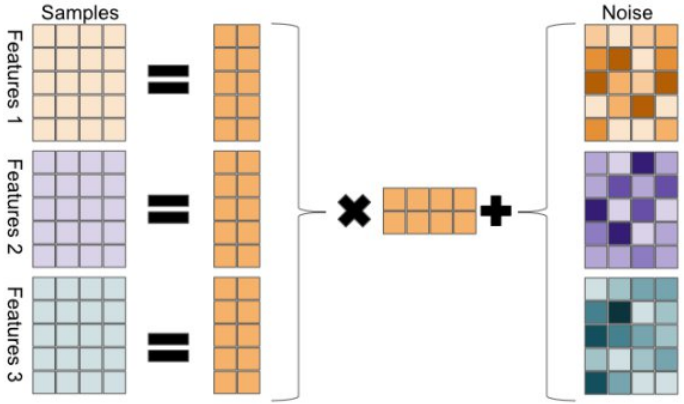
Matrix factorization based methods



Lee, D et al. Learning the parts of objects by non-negative matrix factorization. Nature 1999.



Tien-Thanh Bui, et al. X-intNMF: a cross- and intra-omics regularized NMF framework for multi-omics integration, Bioinformatics 2026



https://compgenomr.github.io/book/matrix-factorization-methods-for-unsupervised-multi-omics-data-integration.html

Probabilistic based methods (*eg* iClusterBayes)

Like PCA, a projection from the initial space onto a low-dimensional *k* subspace

$$y_{ijt} = \mathbf{x}_i \mathbf{\Gamma}_{jt} \boldsymbol{\beta}_{jt} + \varepsilon_{ijt},\ i = 1, \cdots, n,\ j = 1, \cdots, p_t,\ t \in (1, \cdots, m)$$

$\mathbf{x}_i = (1, \mathbf{z}_i) = (1, z_{i1}, \cdots, z_{ik})$ *projection vector in the k-subspace*

$\mathbf{\Gamma}_{jt} = \mathrm{diag}(1, \gamma_{jt}, \cdots, \gamma_{jt})$ *indicative diagonal matrix*

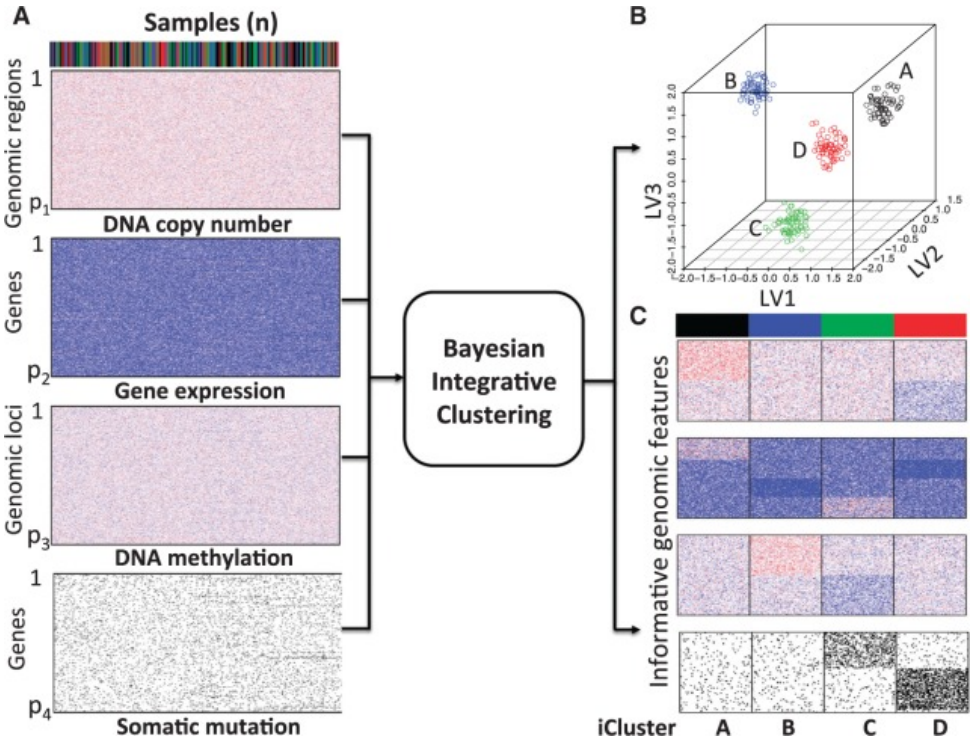
$\boldsymbol{\beta}_{jt} = (\beta_{0jt}, \beta_{1jt}, \cdots, \beta_{kjt})^T$ *coefficients vector*

$\boldsymbol{\beta}_{jt} \sim \mathrm{MVN}(\boldsymbol{\beta}_{0t}, \Sigma_{0t})$ *multivariate normal distribution*

$\sigma_{jt}^2 \sim \mathrm{IG}(\nu_0/2,\ \nu_0\sigma_0^2/2)$ *inverse-gamma distribution*

$\gamma_{jt} \sim \mathrm{Bernoulli}(q_t).$ *Bernoulli distribution*

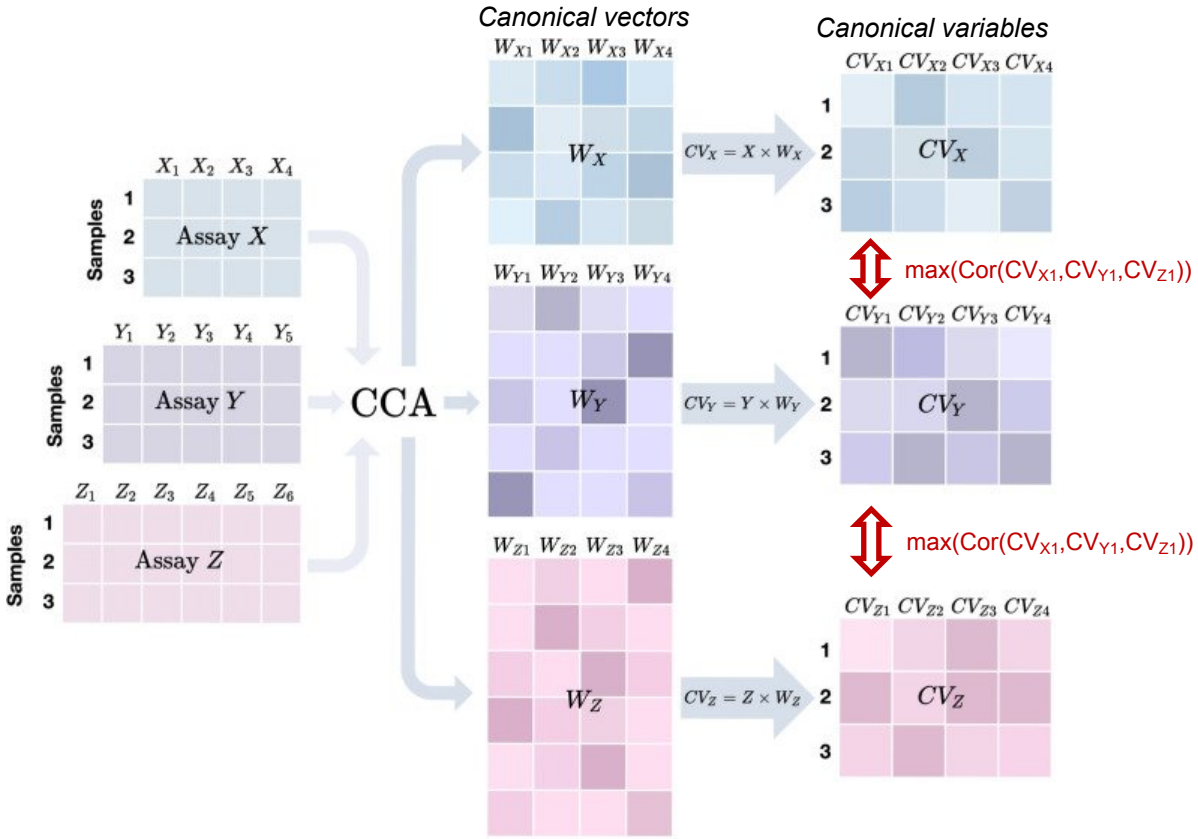
Use the Gibbs and Metropolis–Hasting sampling algorithms to sample for statistical inference.



Mo, Q. et al. A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data. Biostatistics. 2018

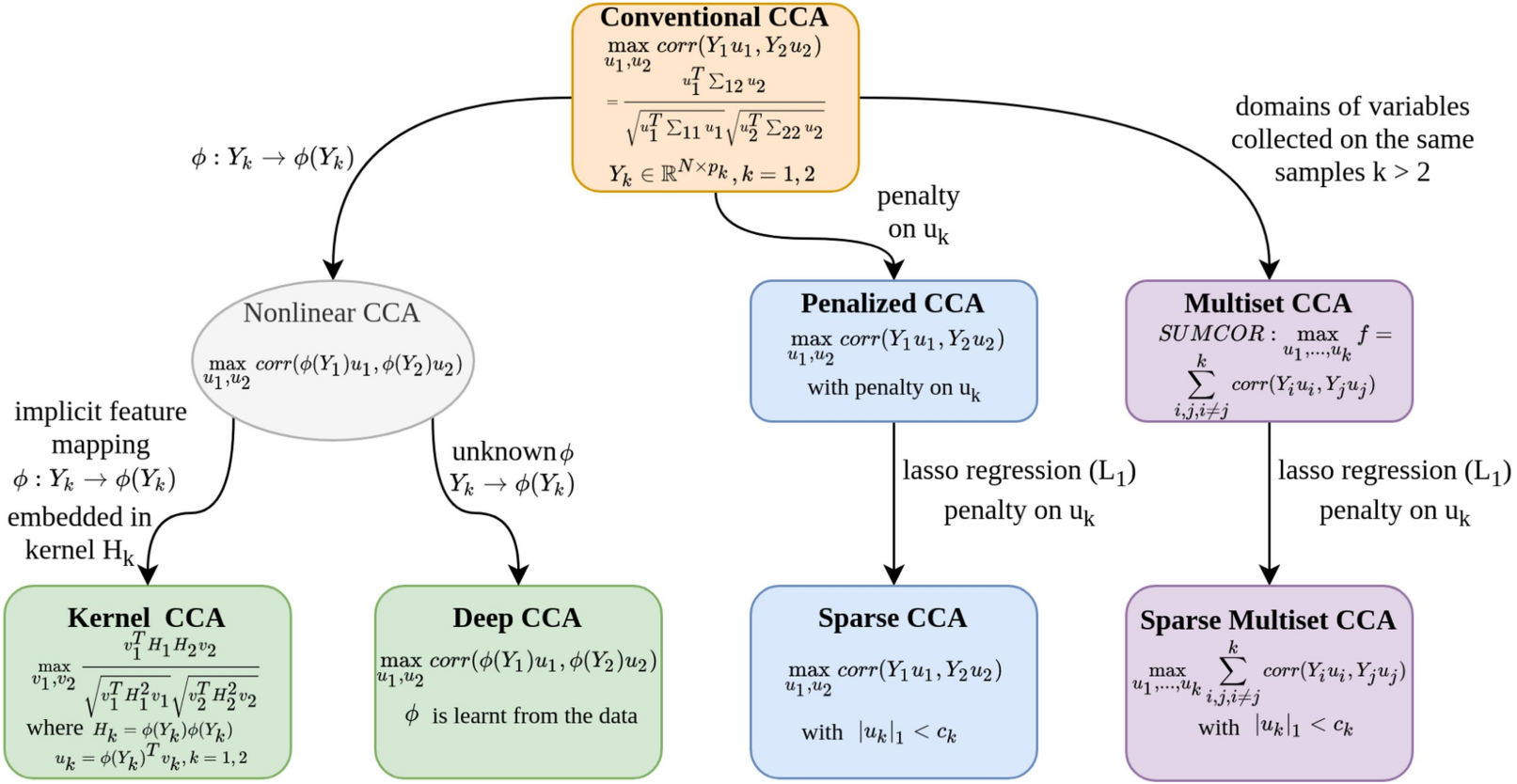
Variance/covariance based methods

Canonical Correlation Analysis (CCA)



Jiang, MZ. et al. Canonical correlation analysis for multi-omics: Application to cross-cohort analysis. PLoS Genet. 2023

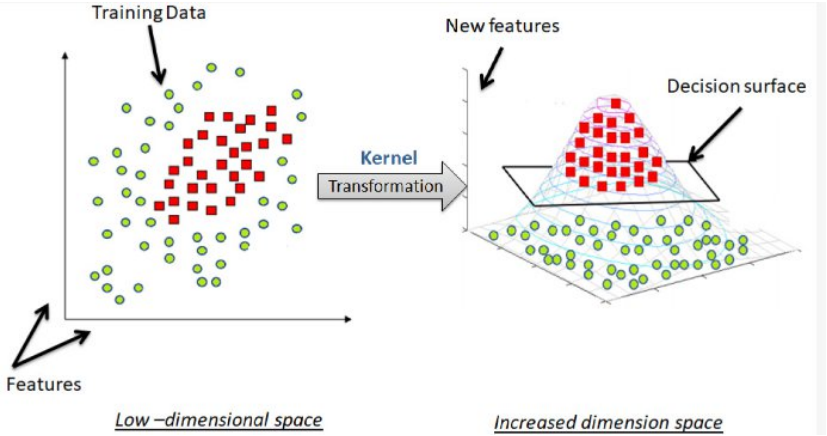
Variance/covariance based methods



Wróbel, S. et al. Data integration through canonical correlation analysis and its application to OMICs research. J Biomed Inform. 2024

Kernel principle

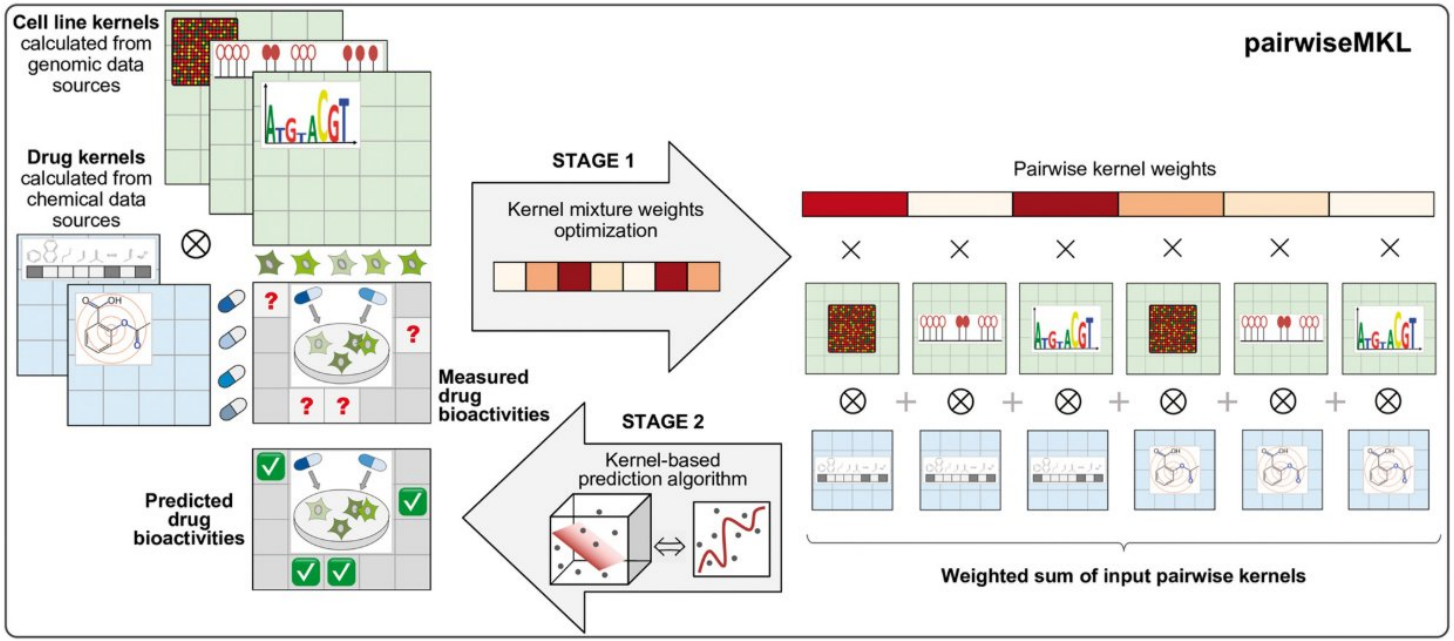
A kernel function maps input features onto a new space, which may be of higher dimension



Learn non-linear models from linear based algorithms using kernel functions $\forall(\boldsymbol{x},\boldsymbol{x}') \in \mathcal{X}^2, \quad K(\boldsymbol{x},\boldsymbol{x}') = \langle \boldsymbol{x},\boldsymbol{x}' \rangle_{\mathcal{X}}$

- Polynomial kernel : $\forall(\boldsymbol{x},\boldsymbol{x}') \in \mathcal{X}^2, \quad K(\boldsymbol{x},\boldsymbol{x}') = (\langle \boldsymbol{x},\boldsymbol{x}' \rangle_{\mathcal{X}})^{\gamma}$
- Gaussian kernel : $\forall(\boldsymbol{x},\boldsymbol{x}') \in \mathcal{X}^2, \quad K(\boldsymbol{x},\boldsymbol{x}') = \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|_2^2}{2\sigma^2}\right)$

Kernel based methods (*eg* pairwiseMKL)



- 10 drugs kernels
- 12 cell-line kernels
- 312 proteins kernels

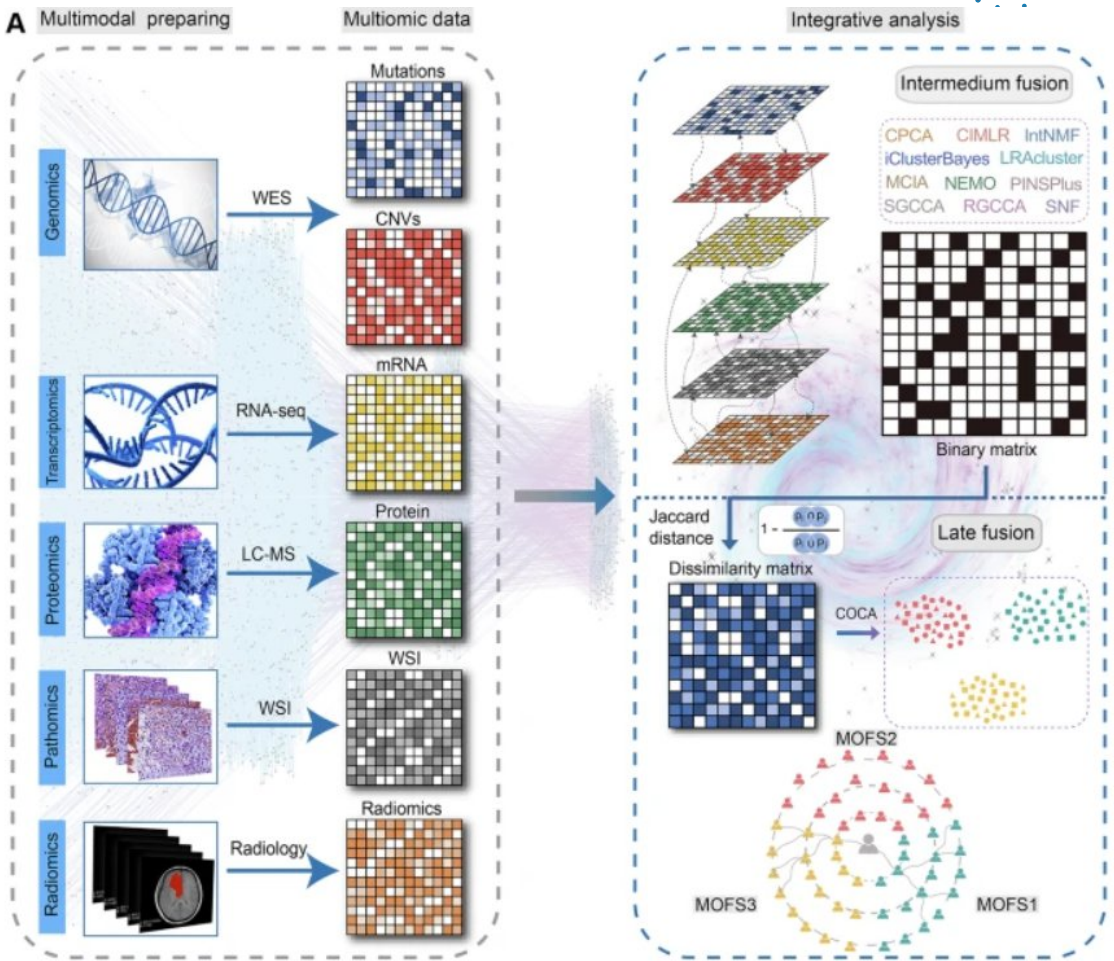
- 120 / 3120 pairwise kernels (Kronecker product)

- regularized model optimization to identify a subset of pairwise kernels weights

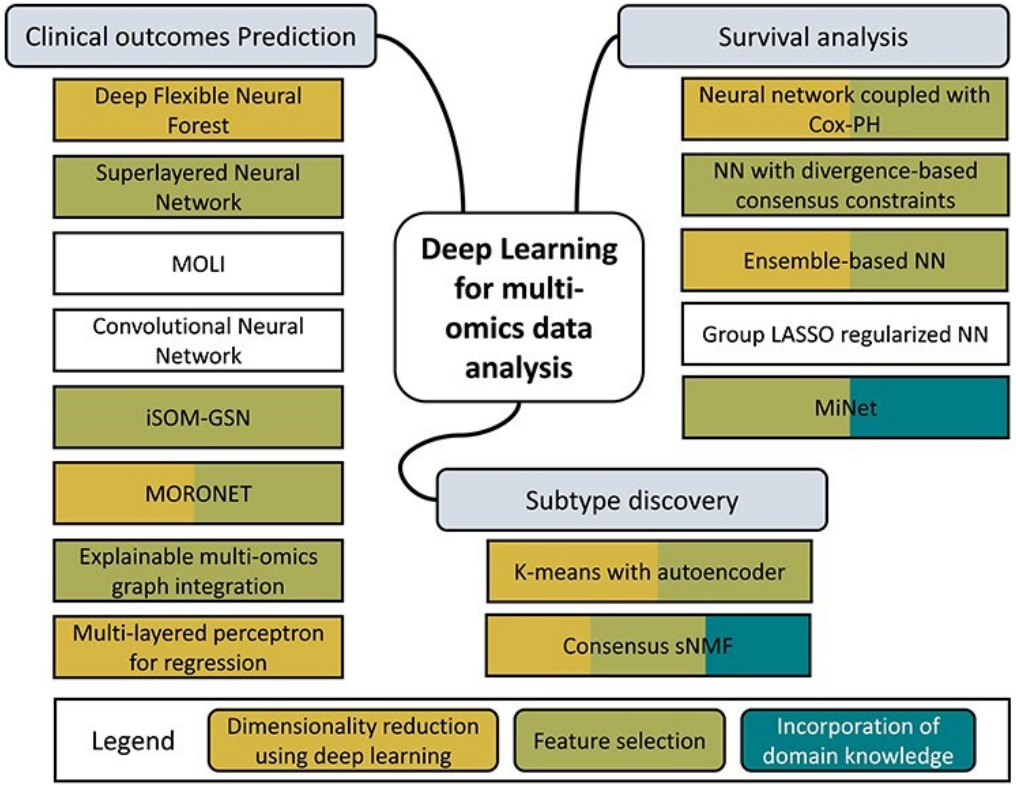
Anna Cichonska, et al. Learning with multiple pairwise kernels for drug bioactivity prediction, Bioinformatics, 2018.

And combine them

- intermediate fusion of multimodal data through 11 algorithms
- each clustering results are converted into a binary matrix
- Jaccard index was calculated for the 11 matrices in order to assess the similarity between the samples
- late fusion of the results obtained from the 11 algorithms in order to derive the final clustering results using the COCA method (Clustering of Cluster Analysis) and arrive at consensus results



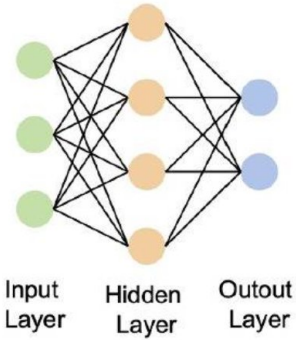
Integration approaches : neural networks



Kang, M., Ko, E., & Mersha, T. B. (2022). A roadmap for multi-omics data integration using deep learning. Briefings in bioinformatics, 23(1).

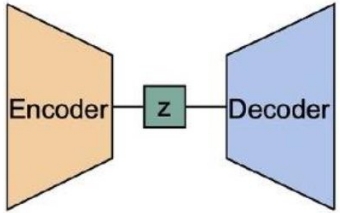
Non-generative methods

Feedforward Neural Network



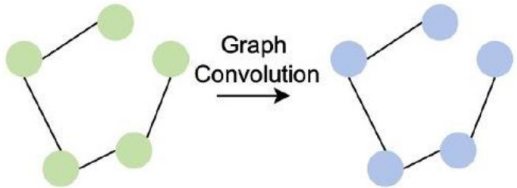
- Inter-modality Interactions
- Biological Interpretability

Autoencoder



- Complementary Learning
- Consensus Learning
- Complementary and Consensus Learning
- Similarity Learning.

Graph Convolutional Neural Network

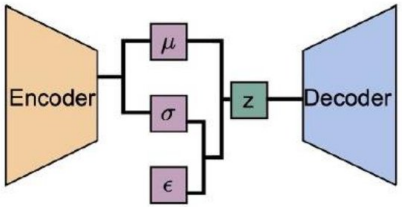


- Knowledge-guided Connectivity
- Data-Driven Connectivity

Ballard JL et al. Deep learning-based approaches for multi-omics data integration and analysis. BioData Min. 2024

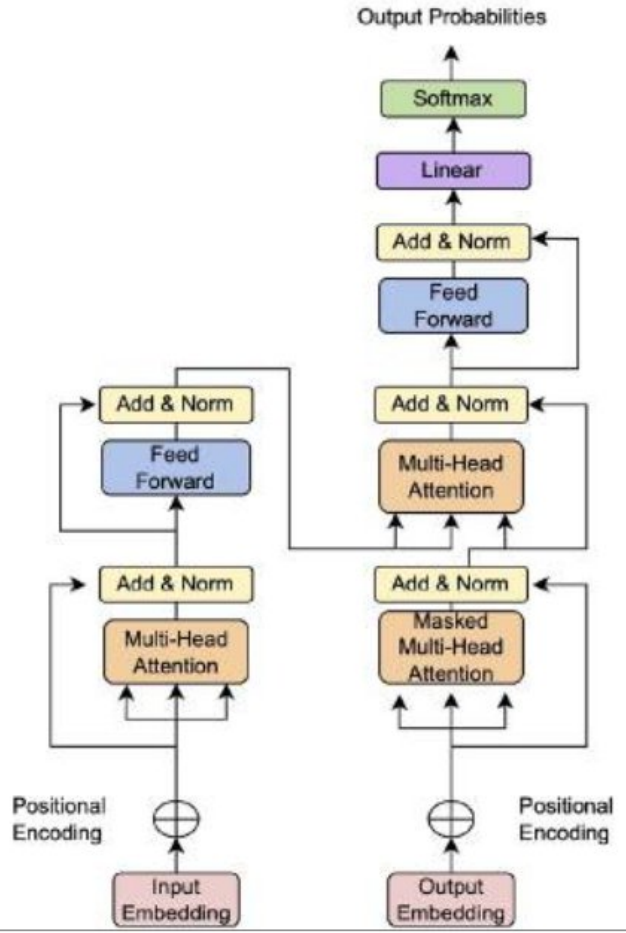
Generative methods

Variational Autoencoder



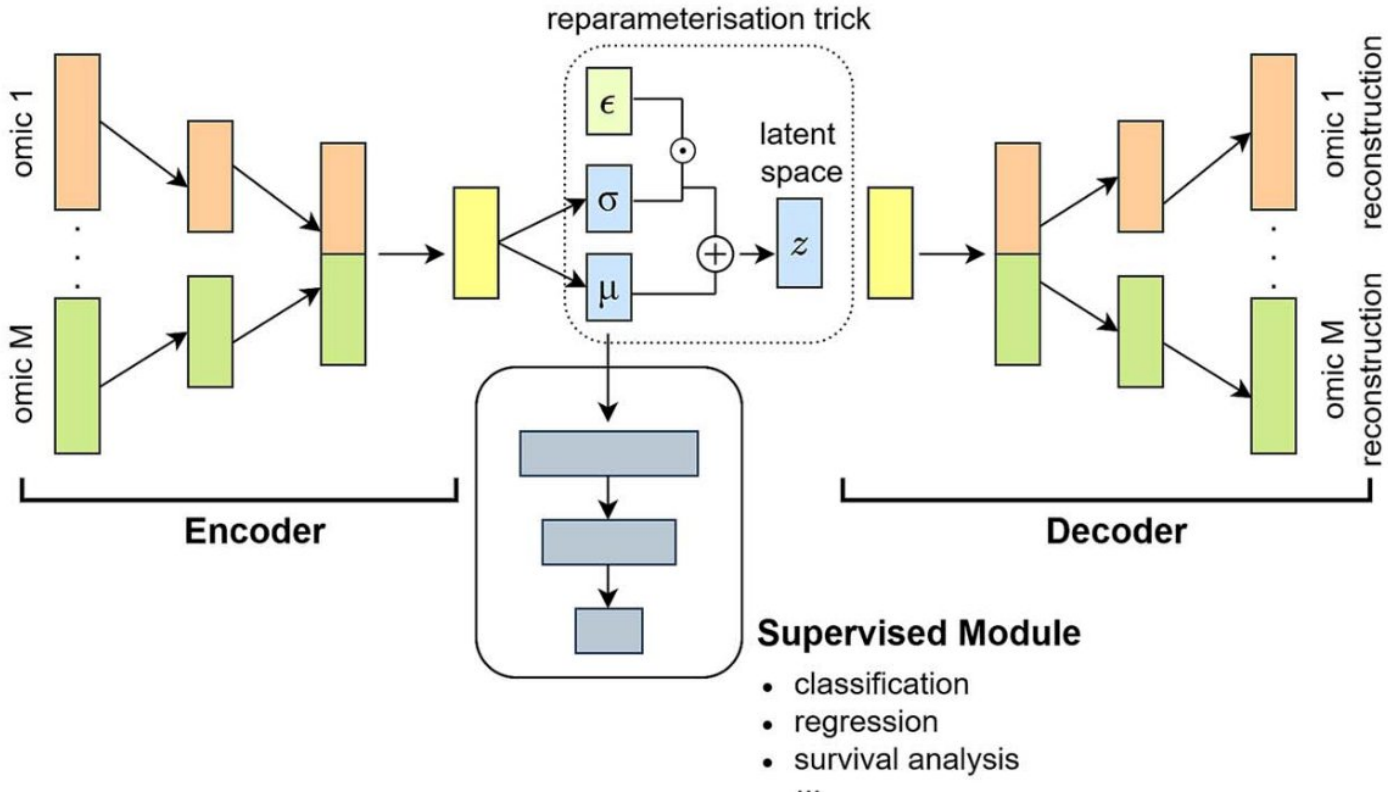
- Unsupervised learning for modeling inter-modality relationships
- Handling Unpaired Data
- Unsupervised learning for dimensionality reduction
- Supervised learning for generating task-relevant embeddings.

Generative Pretrained Transformer



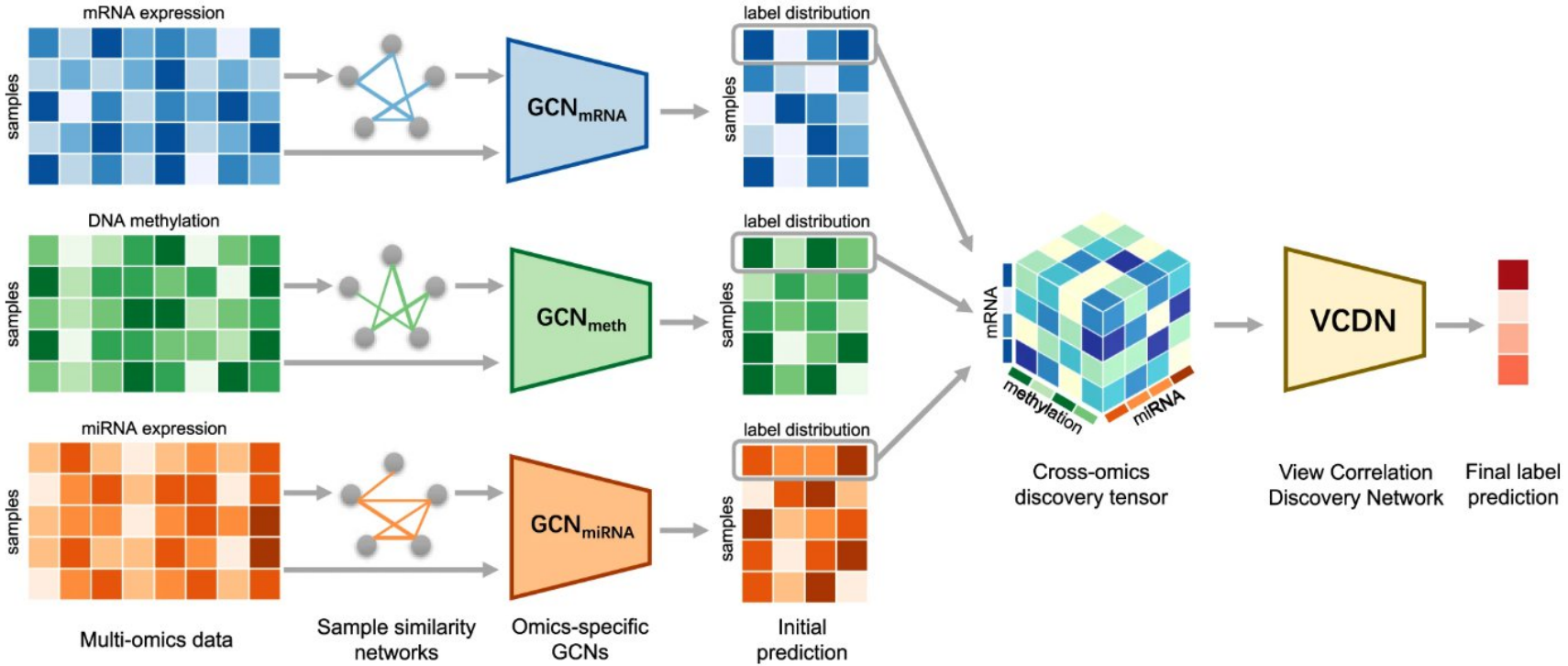
Ballard JL et al. Deep learning-based approaches for multi-omics data integration and analysis. BioData Min. 2024

Variational Auto Encoder (VAE) architecture



Baião AR. et al. A technical review of multi-omics data integration methods: from classical statistical to deep generative approaches. Brief Bioinform. 2025

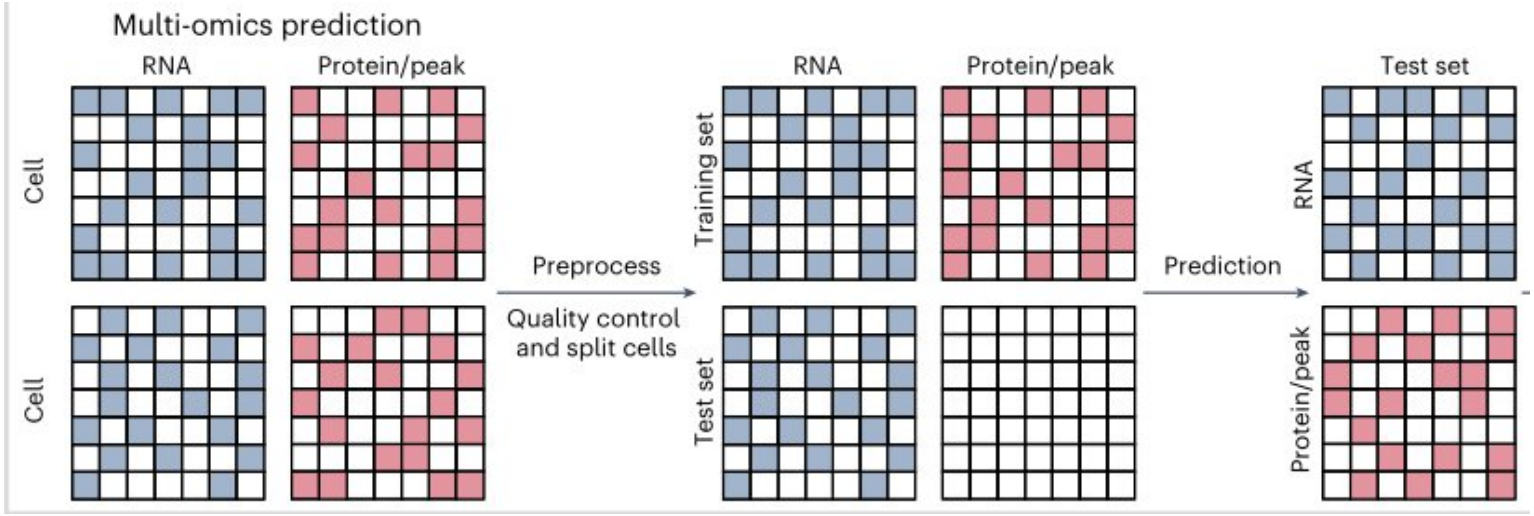
Graph Convolutional Network : MOGONET



Wang, T. et al. MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. Nat Commun 2021

Single-cell integration

- CITE-seq, REAP-seq … : RNA expression and surface proteins abundance for a single cell;
- SHARE-seq, SNARE-seq … : RNA expression and chromatin accessibility for a single cell;
- DOGMA-seq…. : RNA expression, chromatin accessibility and surface proteins abundance;
- … (see Lim, J. et al. Advances in single-cell omics and multiomics for high-resolution molecular profiling. Exp Mol Med. 2024)



Hu, Y.,*et al.* Benchmarking algorithms for single-cell multi-omics prediction and integration. *Nat Methods* **21**, 2024.

Single-cell integration

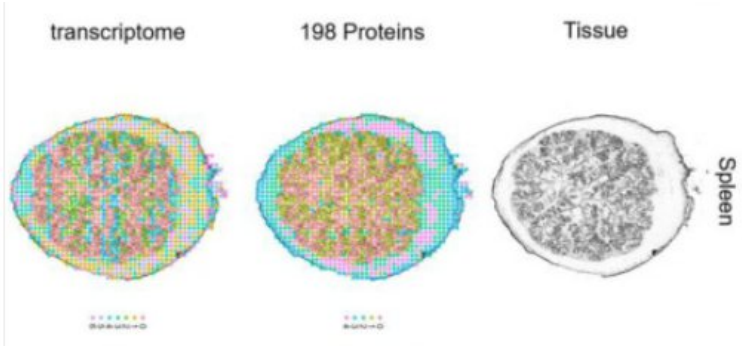
- From same single cell (matched)
  - Deep Learning: DCCA, DeepMAPS, scMVAE, totalVI
  - Probabilistic: MOFA+, BREM-SC, MIRA, MultiVelo
  - Graph / Metric-based: SCHEMA, Seurat v4, citeFUSE, FigR

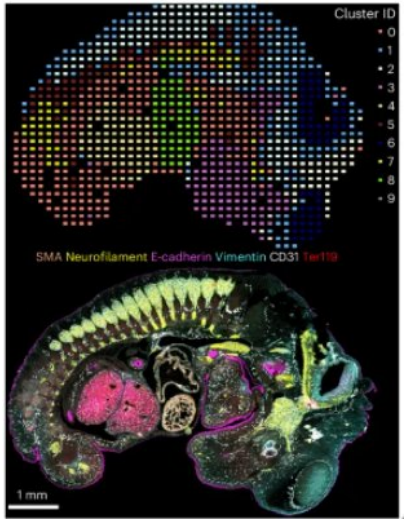
- From different single cells (unmatched)
  - Deep Learning: GLUE, Cobolt
  - Probabilistic: MultiVI, LIGER
  - Graph: Spectrum, Seurat v5
  - Manifold Alignment: MMD-MA, UnionCom, Pamona
  - CCA: BindSC

Baysoy, A. et al. The technological landscape and applications of single-cell multi-omics. Nat Rev Mol Cell Biol 2023

- Spatial (only) transcriptomic: MERFISH, Stereo-seq
  - Prediction of non-spatial multi-omic data (e.g. proteins or open chromatin) by inferring the distribution of 'missing' omic data in space
  - Assume a common distribution between spatial and non-spatial omics data

- Spatial multi-omics technology: spatial-CITE-seq and DBiT-seq

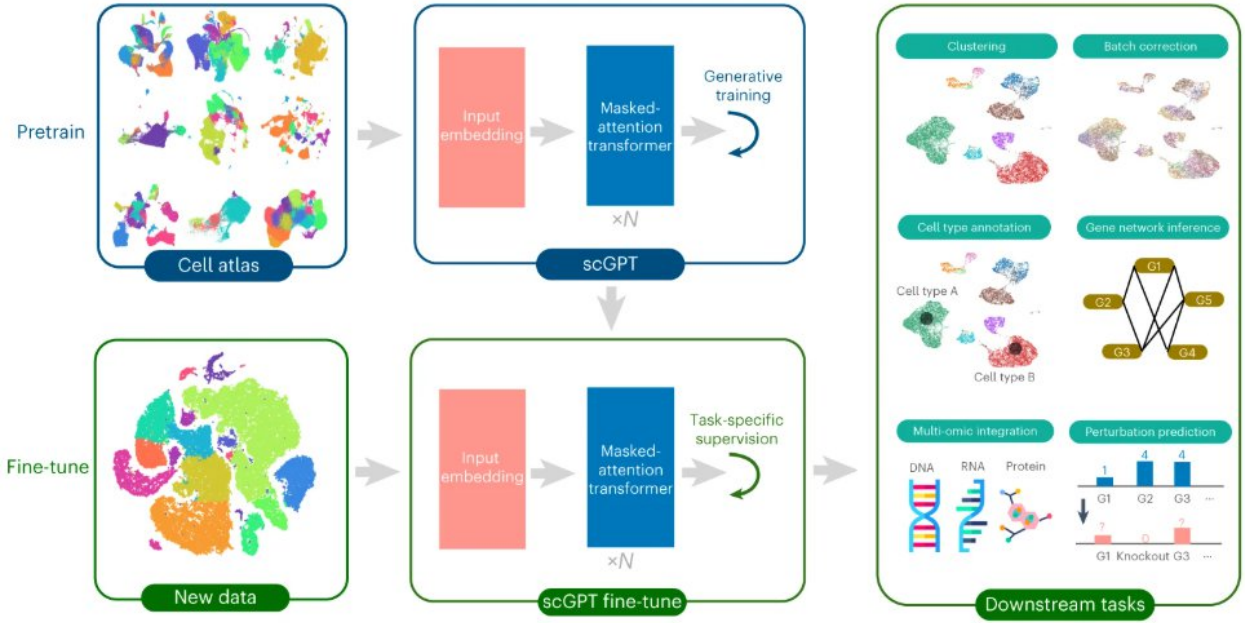




Liu, Y. et al. Spatial-CITE-seq: spatially resolved high-plex protein and whole transcriptome co-mapping. Res Sq 2022.
Enninful, A. et al. Integration of imaging-based and sequencing-based spatial omics mapping on the same tissue section via DBiTplus. Nat Methods 2026.

- Impute missing modalities or integrate multiple modalities

  eg: *Geneformer*, *scGPT* and *scFoundation*



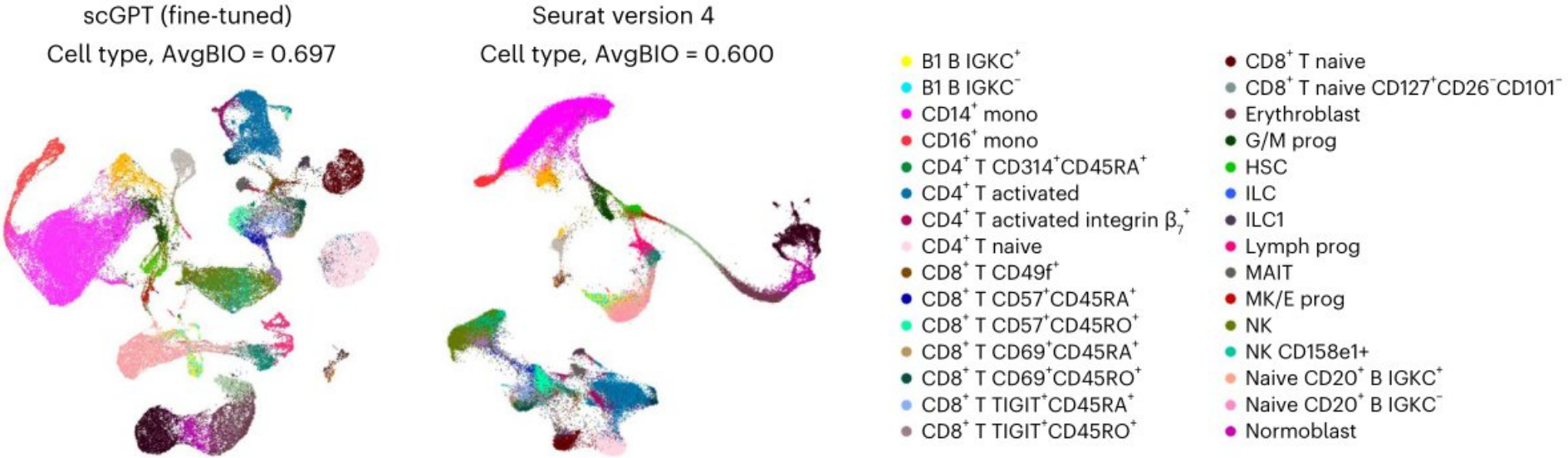
<u>Two training steps:</u>

1. initial pretraining on large cell atlases (33 million non-disease human cells from 51 organs in the CELLxGENE collection)

2. follow-up fine-tuning on smaller datasets for specific applications

Cui, H. et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nat Methods 21, 2024.

scGPT for cell-type annotation with multi-omics data



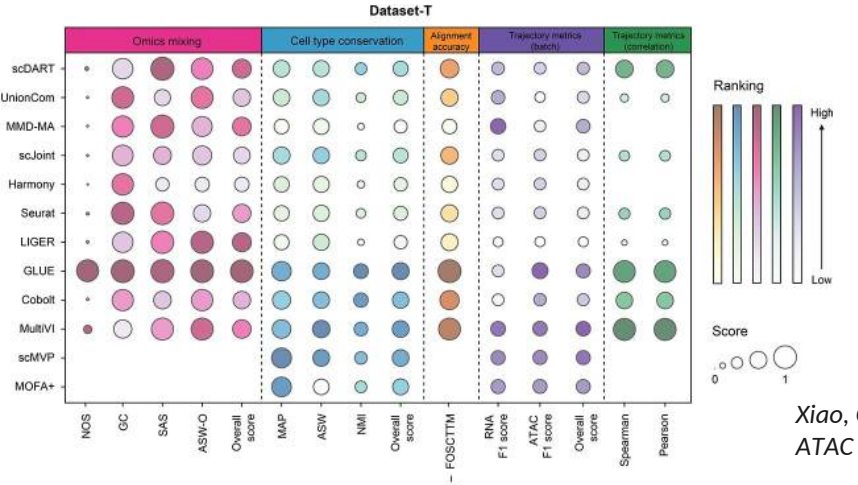
Fine-tuned scGPT model VS Seurat (v.4) on the CITE-seq BMMC dataset (paired RNA and protein data)

- 12 healthy human donors consisting of 12 batches (total of 90,261 cells)
- 13,953 genes and 134 surface proteins

Integration approaches : the good one ?

Integration approaches are not unic

- comparisons exist... for a given application)
- parametrization need expertise
- make your own comparisons
- keep an eye open



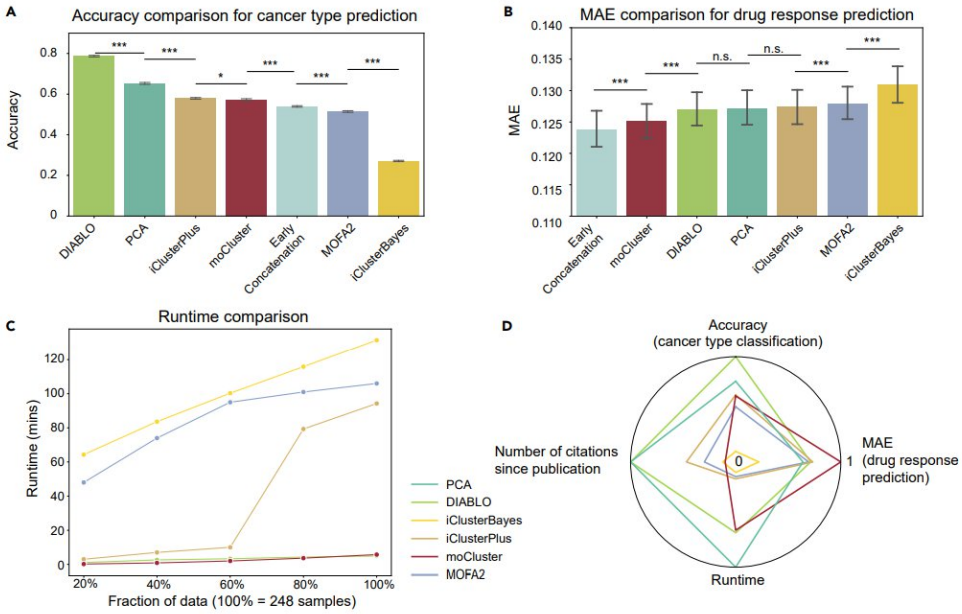


Figure 5. Benchmarking of machine learning-based integration tools using the CCLE multi-omics data

Cai, Z. et al. Machine learning for multi-omics data integration in cancer. iScience. 2022

Xiao, C. et al. Benchmarking multi-omics integration algorithms across single-cell RNA and ATAC data. Brief Bioinform. 2024

Integration approaches are not magic!

You will still need to:

- carefully check design and confounding factors
- perform specific data pre-processing for each omic (at least primary analysis)
- impute missing values* (different meaning → different strategy)
- choose your integration strategy based on your objective and your data (ex. matching between omics) → still no standard pipelines
- the best isn't necessarily the latest
- some omics generate more noise than answers

Multi-omics data portal

Table 1. Key portals for accessing publicly available multi-omics datasets

Name	URL	Omic and other data types	Notes
TCGA (Campbell et al., 2020)	https://portal.gdc.cancer.gov/	• Genomics • Epigenomics • Transcriptomics	• Tumor data • Large coverage of tumors
ICGC (Campbell et al., 2020)	https://dcc.icgc.org/	• Genomics • Transcriptomics	• Tumor data • Powerful online analytics tools
CPTAC	https://cptac-data-portal.georgetown.edu/cptacPublic/	• Proteomics	• Tumor data • The largest proteomic data portal
COSMIC Cell Lines (Iorio et al., 2016)	https://cancer.sanger.ac.uk/cell_lines	• Genomics • Epigenomics • Transcriptomics • Drug response • CRISPR-Cas9 screen	• Cancer cell line data • Manually curated • Large coverage of cell lines
DepMap (Broad, 2020)	https://depmap.org/portal/	• Genomics • Epigenomics • Transcriptomics • Proteomics • Drug response • CRISPR-Cas9 screen	• Cancer cell line data • Large coverage of omic types • Powerful online tools
COSMIC (Tate et al., 2019)	https://cancer.sanger.ac.uk/cosmic	• Genomics • Epigenomics • Transcriptomics	• Tumor data • Manually curated • Focus on genomics • Overlap with other portals

Cai Z, Poulos RC, Liu J, Zhong Q. Machine learning for multi-omics data integration in cancer. iScience. 2022

Multi-omics data portal

Table 1. Key portals for accessing publicly available multi-omics datasets

Name	URL		
TCGA (Campbell et al., 2020)	https://portal.gdc.cancer.gov/		
ICGC (Campbell et al., 2020)	https://dcc.icgc.org/		
CPTAC	https://cptac-data-portal.georgetown.edu/cptacPublic/		
COSMIC Cell Lines (Iorio et al., 2016)	https://cancer.sanger.ac.uk/cell_lines	• Transcriptomics • Drug response • CRISPR-Cas9 screen	• Large coverage of cell lines
DepMap (Broad, 2020)	https://depmap.org/portal/	• Genomics • Epigenomics • Transcriptomics • Proteomics • Drug response • CRISPR-Cas9 screen	
COSMIC (Tate et al., 2019)	https://cancer.sanger.ac.uk/cosmic	• Genomics • Epigenomics • Transcriptomics	• Focus on genomics • Overlap with other portals

Resource | Open access | Published: 13 March 2025

Human BioMolecular Atlas Program (HuBMAP): 3D Human Reference Atlas construction and usage

Katy Börner, Philip D. Blood, Jonathan C. Silverstein, Matthew Ruffalo, Rahul Satija, Sarah A. Teichmann, Gloria J. Pryhuber, Ravi S. Misra, Jeffrey M. Purkerson, Jean Fan, John W. Hickey, Gesmira Molla, Chuan Xu, Yun Zhang, Griffin M. Weber, Yashvardhan Jain, Danial Qaurooni, Yongxin Kong, HRA Team, Andreas Bueckle & Bruce W. Herr II

Nature Methods **22**, 845–860 (2025) | Cite this article

JOURNAL ARTICLE

scMMO-atlas: a single cell multimodal omics atlas and portal for exploring fine cell heterogeneity and cell dynamics

Wenwen Cheng, Changhui Yin, Shiya Yu, Xi Chen, Ni Hong, Wenfei Jin

Nucleic Acids Research, Volume 53, Issue D1, 6 January 2025, Pages D1186–D1194, https://doi.org/10.1093/nar/gkae821

Published: 24 September 2024 **Article history**

Cai Z, Poulos RC, Liu J, Zhong Q. Machine learning for multi-omics data integration in cancer. iScience. 2022

(Many) Web-applications

Profiler *Zirem, Y. et al. Profiler: an open web platform for multi-omics analysis, Bioinformatics 2026*

iSODA *Olivier-Jimenez, D. et al. iSODA: A Comprehensive Tool for Integrative Omics Data Analysis in Single- and Multi-Omics Experiments. Anal Chem. 2025*

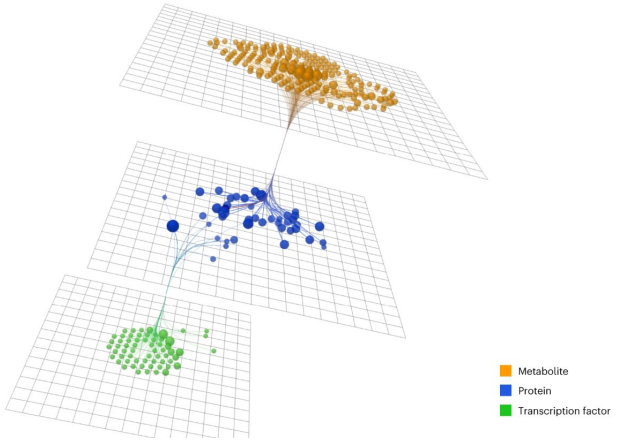
Analyst suite *Ewald, J.D. et al. Web-based multi-omics integration using the Analyst software suite. Nat Protoc. 2024*

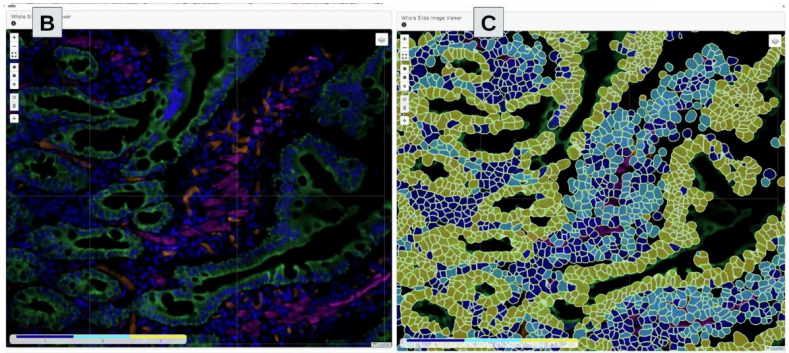
FUSION Border, SP. et al. FUSION: a web-based application for in-depth exploration of multi-omics data with brightfield histology. Nat Commun. 2025

NOODAI *Totu, T. et al. NOODAI: a webserver for network-oriented multi-omics data analysis and integration pipeline, Bioinformatics 2025*

OmNI *Potter, G. Et al. OmNI: a modular open-source framework for interactive multi-omics data integration and visualization. NAR Genom Bioinform. 2026*

Galaxy-P project *(Galaxy-P Project. galaxyp.org.)*





References

Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics Data Integration, Interpretation, and Its Application. Bioinform Biol Insights, 2020.

Picard M. et al. Integration strategies of multi-omics data for machine learning analysis. Comput Struct Biotechnol J. 2021.

Baião AR. et al. A technical review of multi-omics data integration methods: from classical statistical to deep generative approaches. Brief Bioinform. 2025

Wróbel, S. et al. Data integration through canonical correlation analysis and its application to OMICs research. J Biomed Inform. 2024

Kang, M., Ko, E., & Mersha, T. B. A roadmap for multi-omics data integration using deep learning. Briefings in bioinformatics, 2022.

Xiao, C. et al. Benchmarking multi-omics integration algorithms across single-cell RNA and ATAC data. Brief Bioinform. 2024