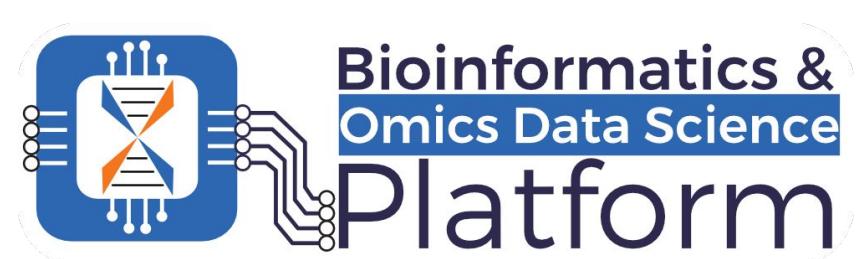




Multi-modal data integration using deep learning and applications in precision oncology

Dr. Bora Uyar

Computational Genomics Workshop
10-16 March 2025



Plan

- Lectures + hands-on + HW + Quizzes
- Follow Google Classroom for announcements
- Helping each other:

https://github.com/BIMSBbioinfo/compgen_course_2025_module3/discussions

The global cancer burden



10 Million

cancer-related deaths in 2020

\$80 Billion

direct medical costs per year just in the US

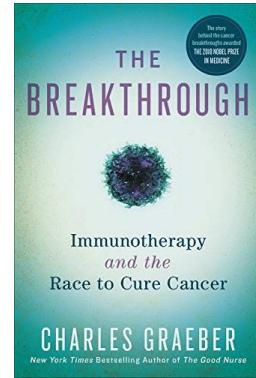
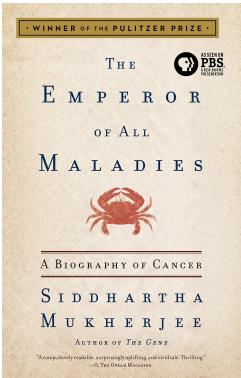
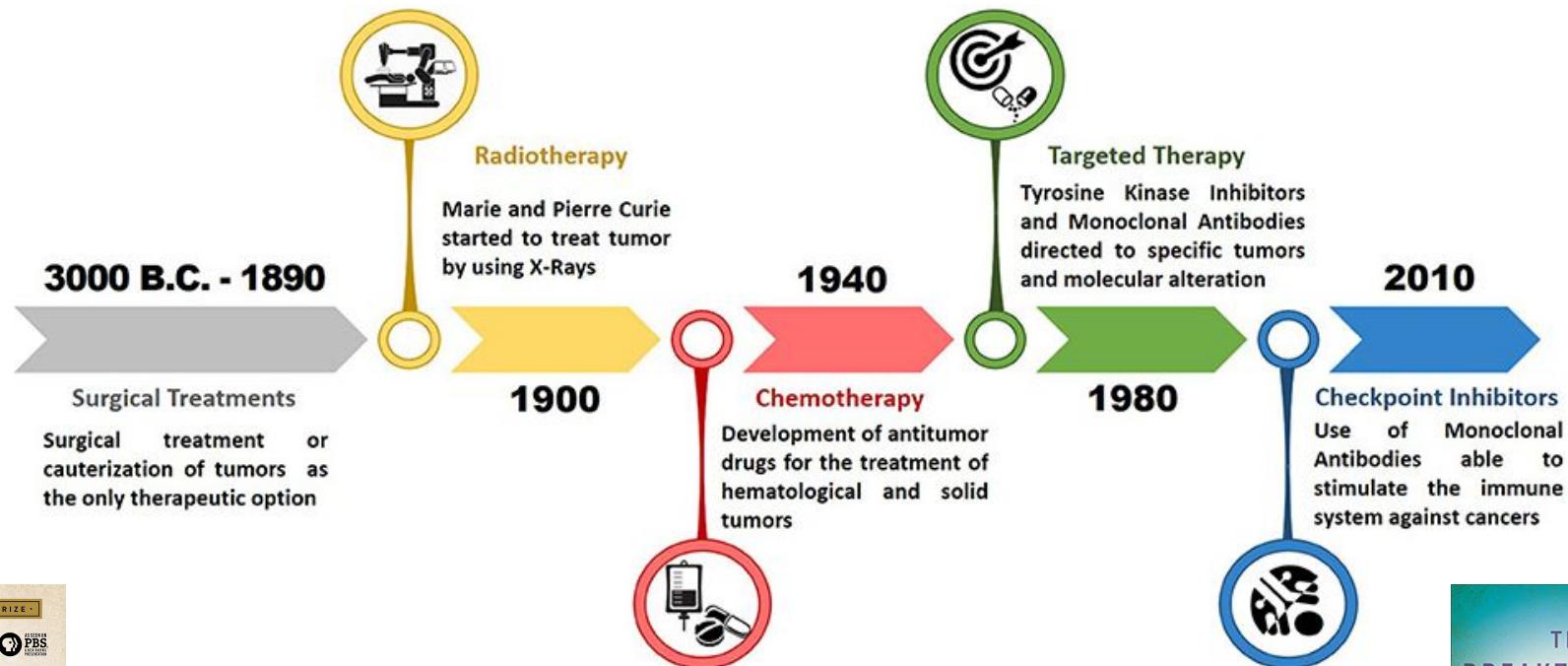
\$2.7 Billion

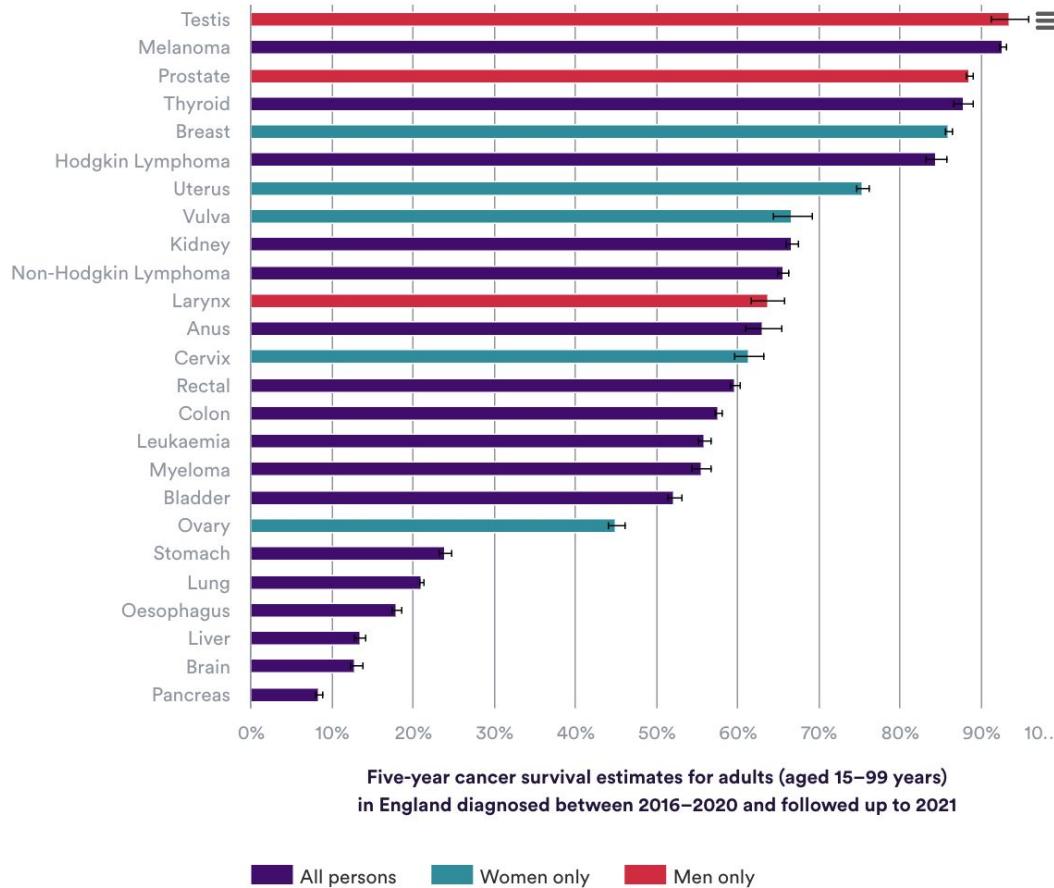
Cost of developing a single cancer drug

Too many deaths, too expensive

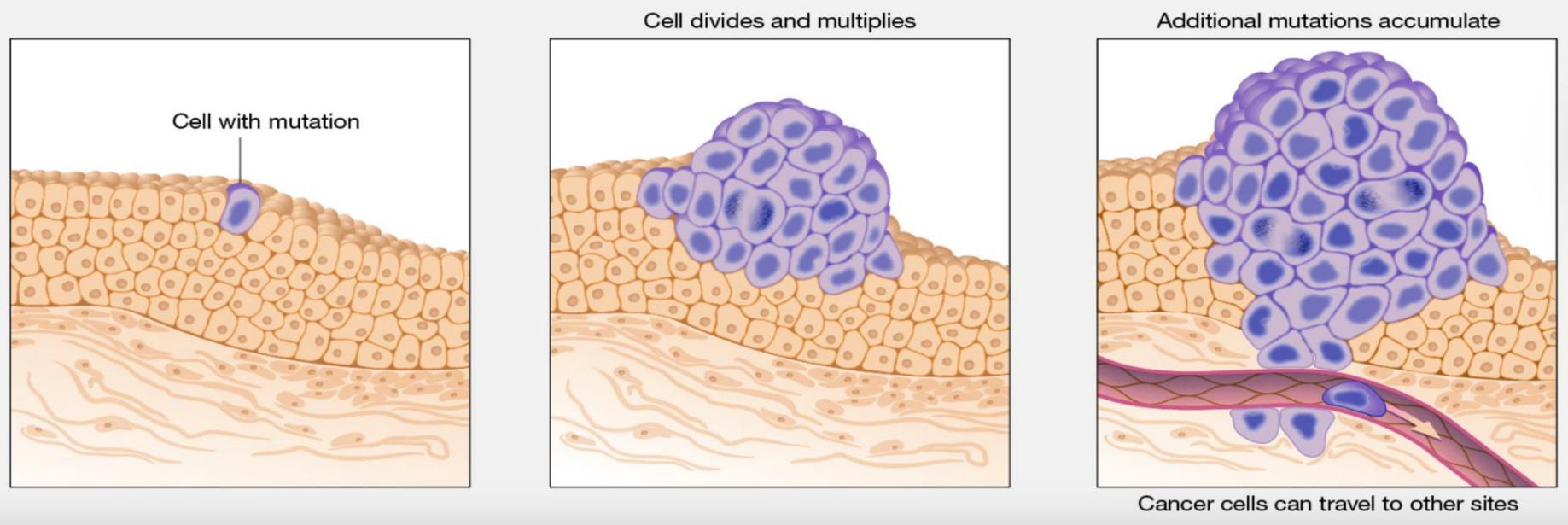
<https://www.cancer.gov/about-cancer/understanding/statistics>

<https://www.cancer.org/cancer/cancer-basics/economic-impact-of-cancer.html>



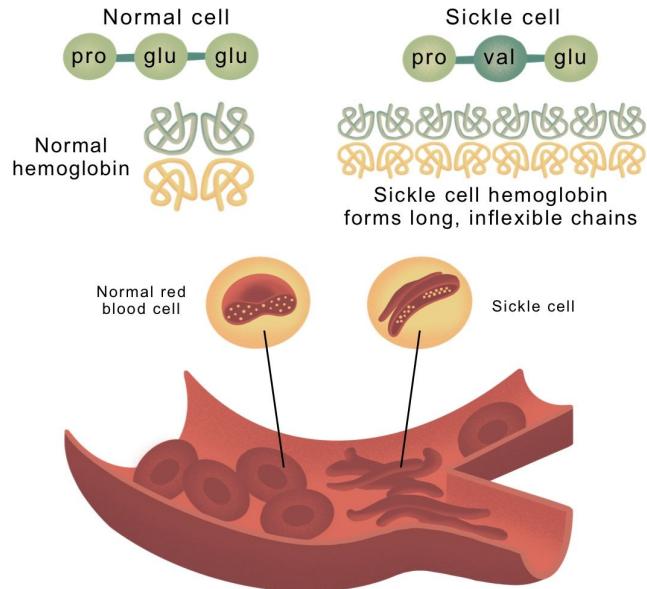


Cancer: uncontrolled cell growth/proliferation



Mendelian Diseases vs Cancer

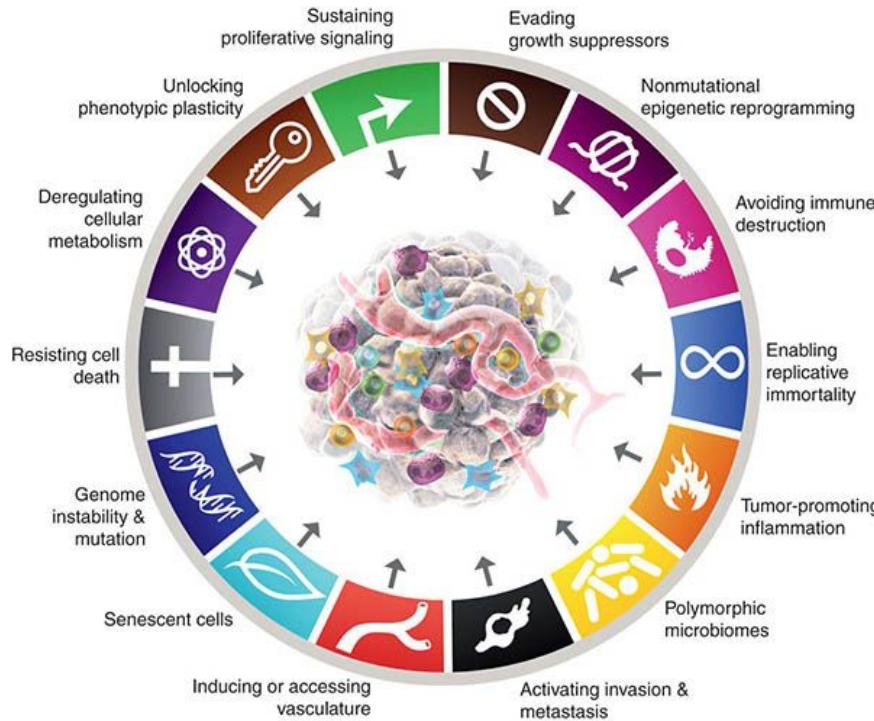
> Usually can be explained by a few genetic defects



Sickle-Cell.com

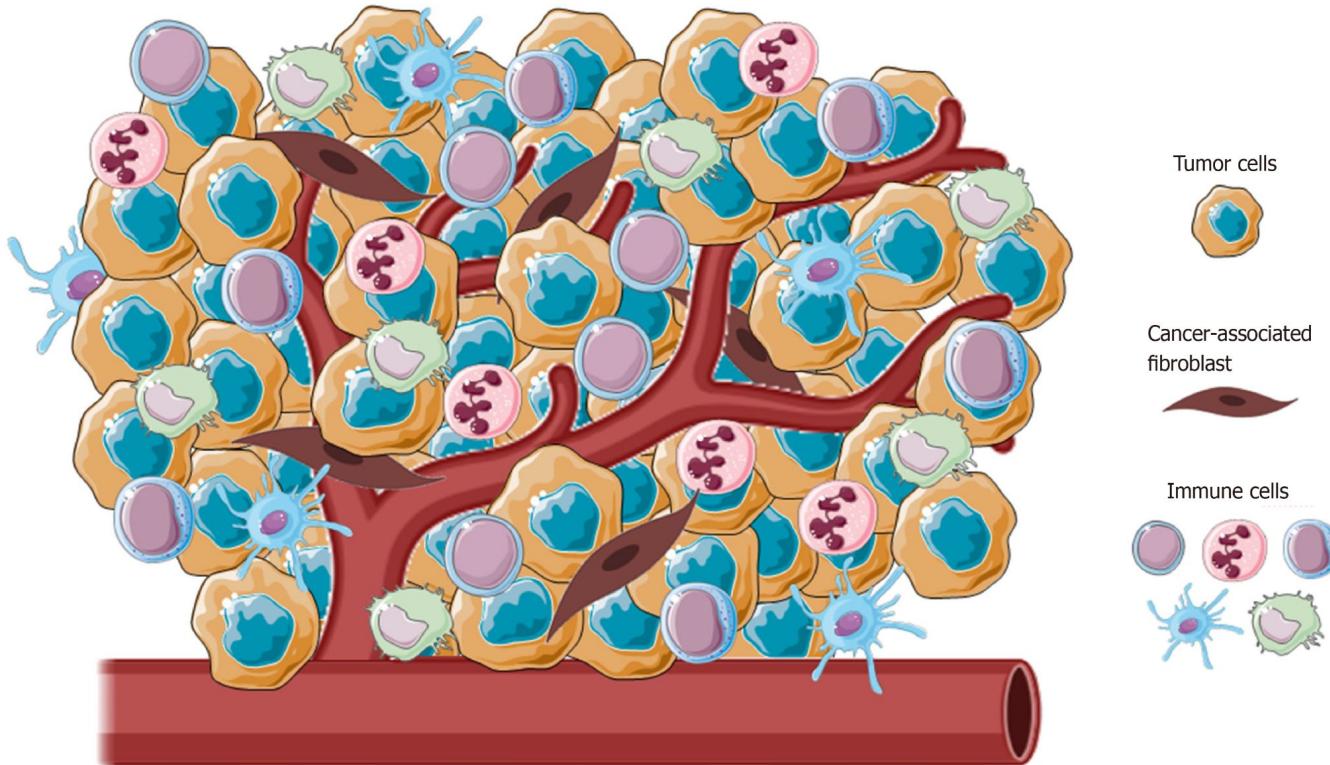
> 7000 rare diseases are known

Hallmarks of Cancer

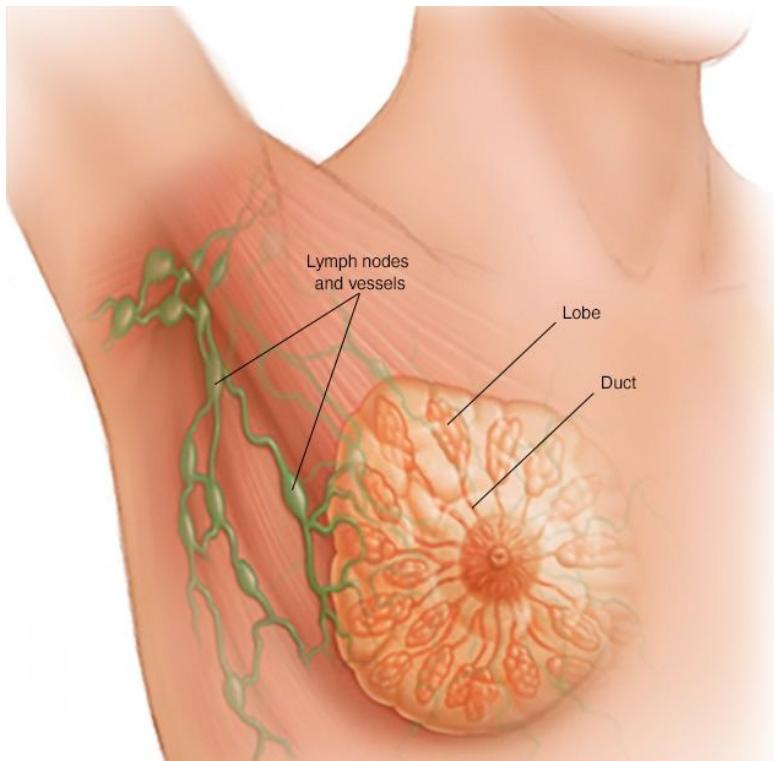


[New Dimensions in Cancer Biology: Updated Hallmarks of Cancer Published - American Association for Cancer Research \(AACR\)](#)

Tumor Microenvironment



Current State of Cancer Diagnosis/Treatment



© MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH. ALL RIGHTS RESERVED.

mayoclinic.org

Principal clinical features to consider:

- Age
- Gender
- Menopause status
- Tumor stage

Principal genetic features to consider:

- HER2, Estrogen receptor, Progesterone receptor (Hormone Therapy)
- BRCA1/2 gene mutations (PARP inhibitors)
- PD-L1 (Immunotherapy)
- Multi-gene panels (Oncotype Dx, MammaPrint) => 50-70 genes (chemotherapy benefit)

Current State of Cancer Diagnosis/Treatment

Gene Panels to guide clinical decisions



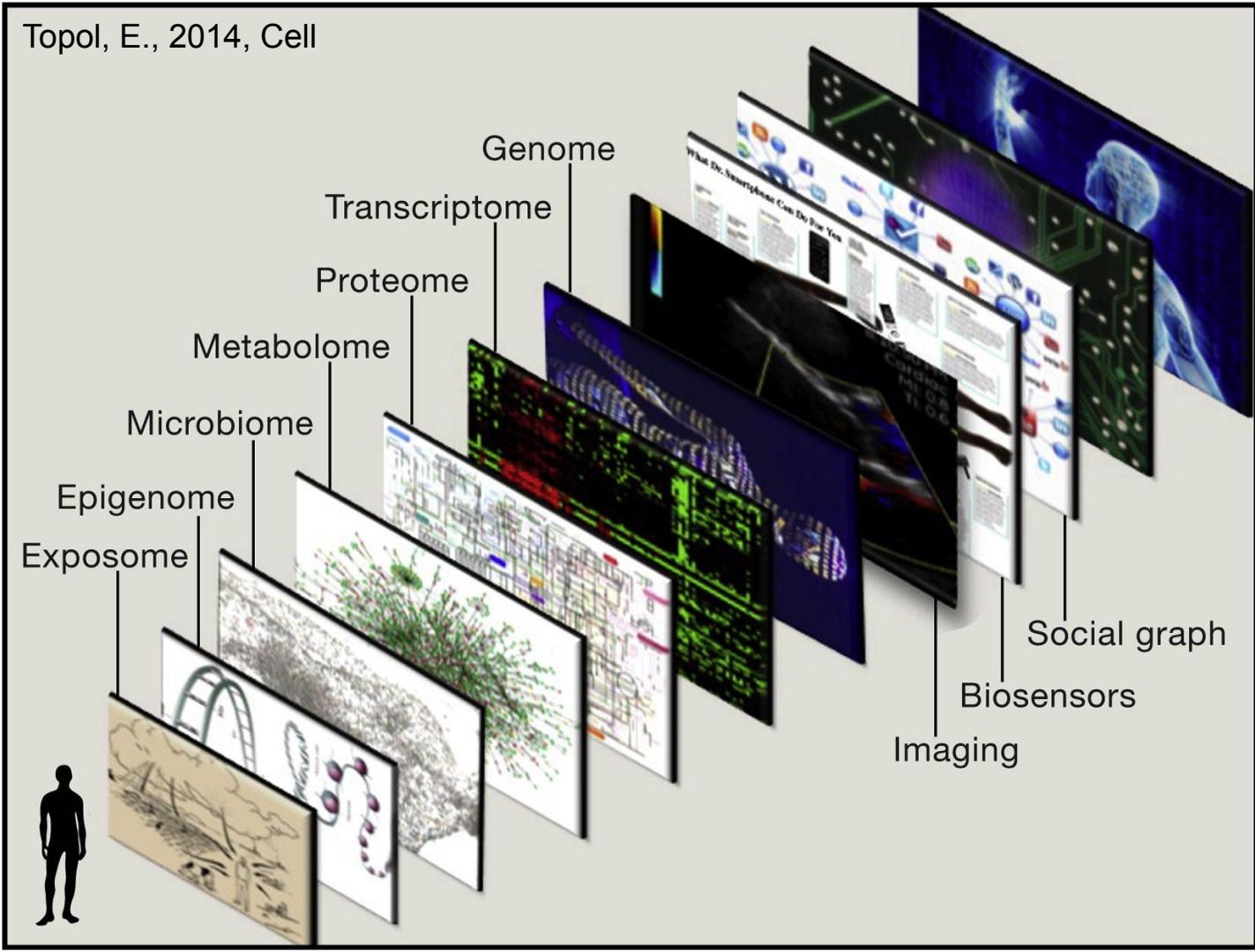
> 350 genes

- Mutations
- Copy Number Alterations
- Derivative features
 - Microsatellite Instability
 - Tumor Mutation Burden

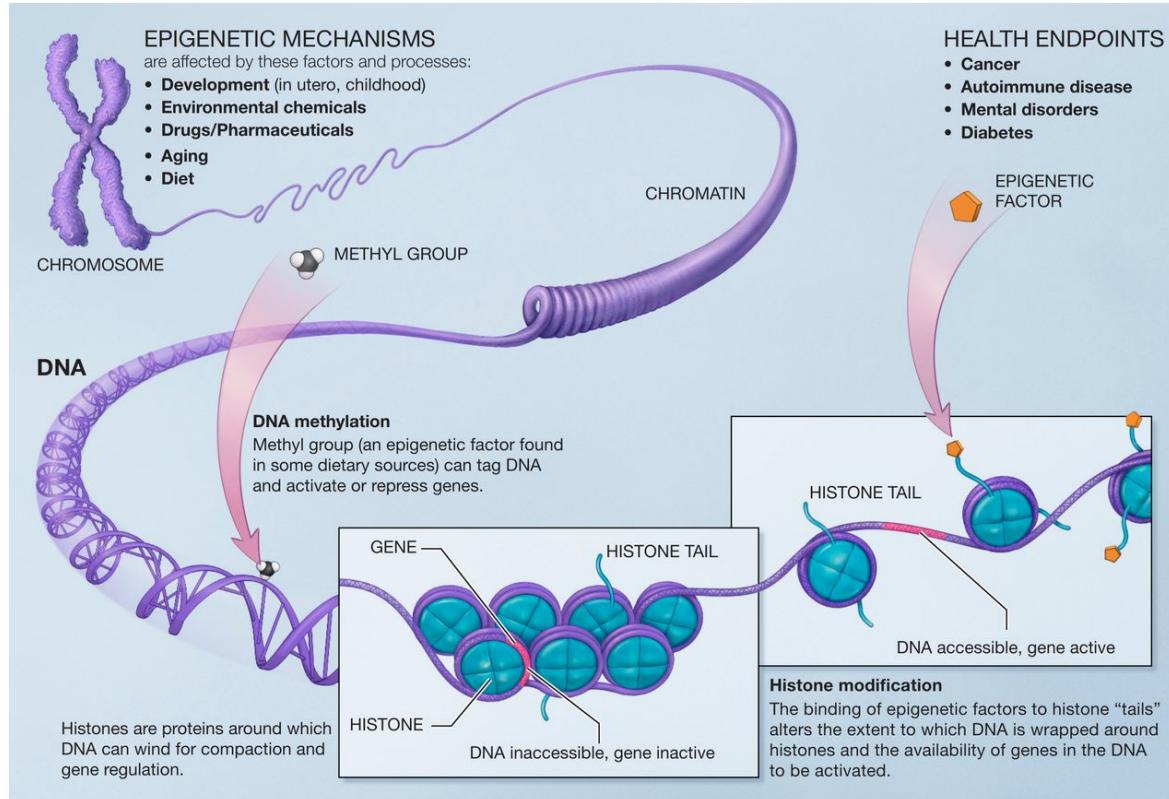


~ 700 genes

Why do we need multi-omics integration?



Epigenome



Genome: structural rearrangements

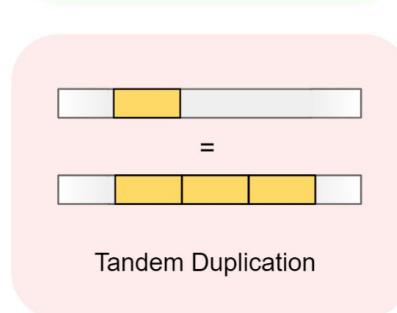
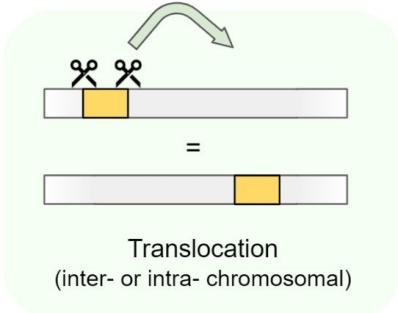
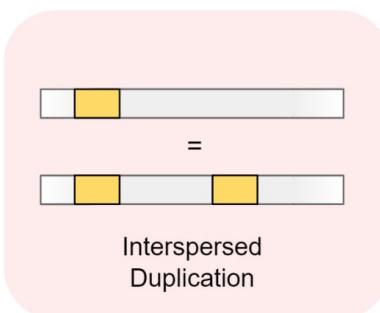
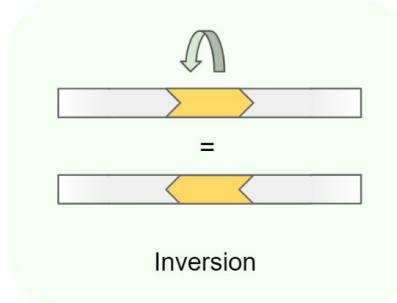
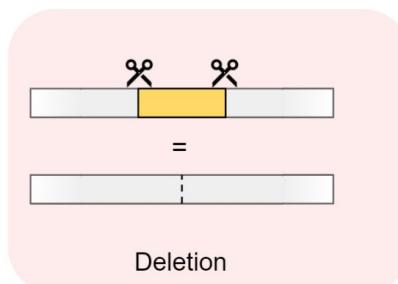
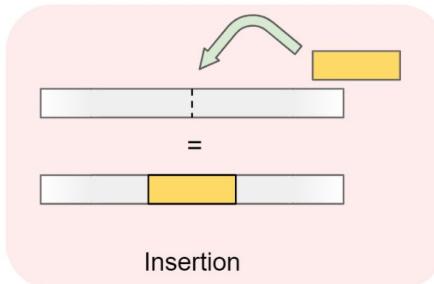
SV Types



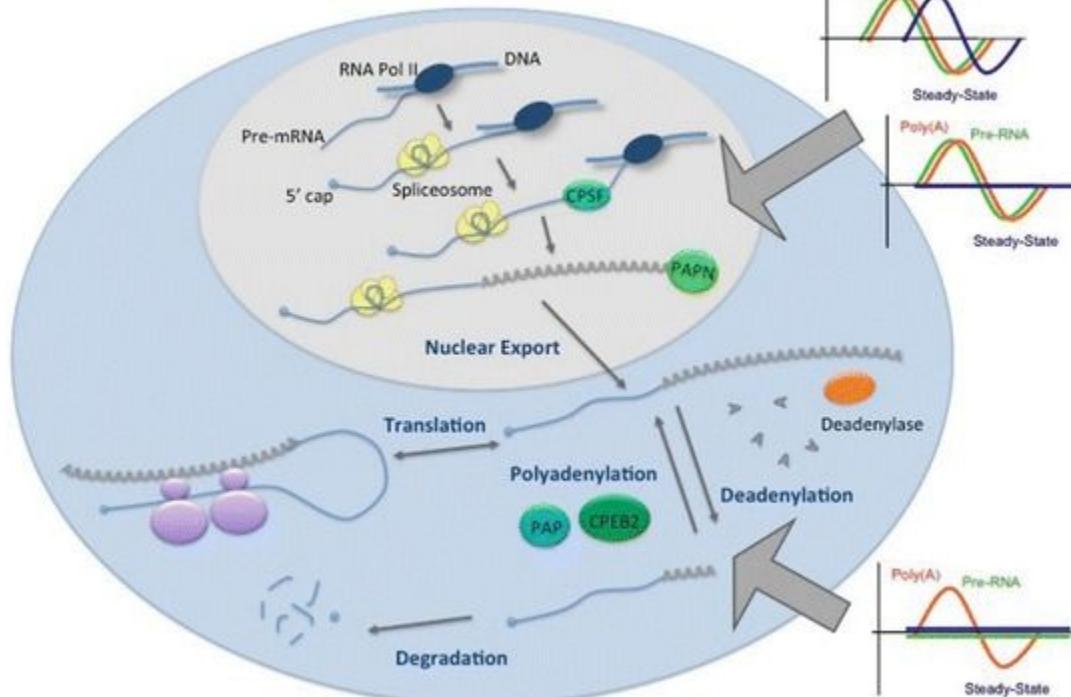
Destructive (non-balanced)



Non-destructive (balanced)

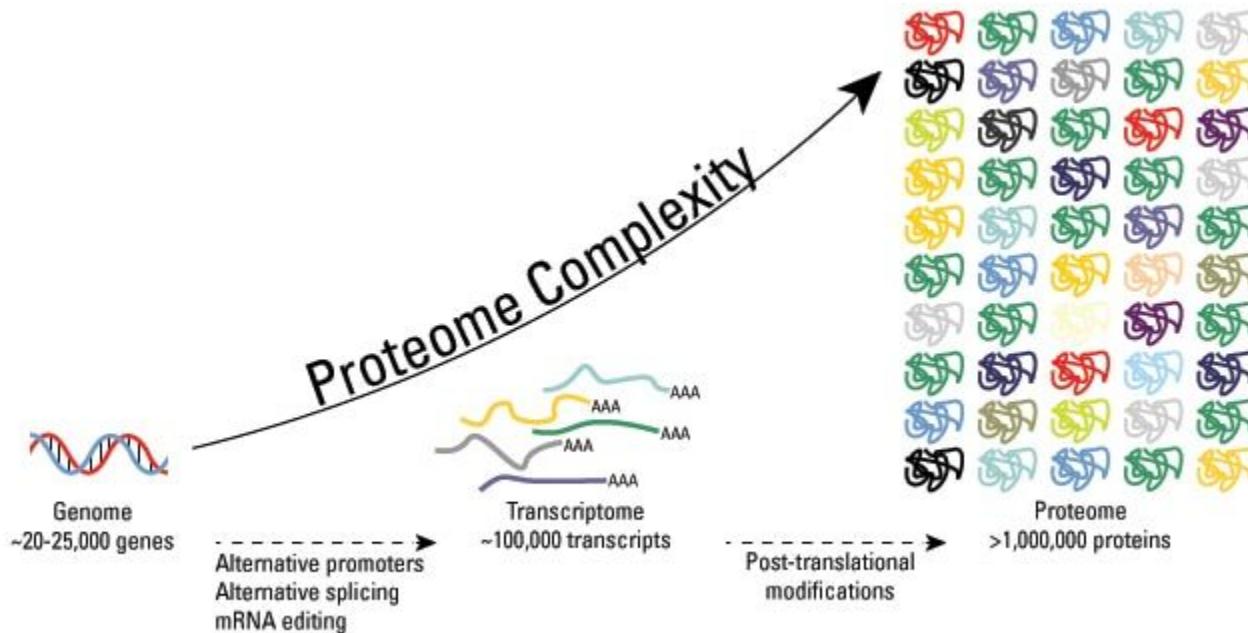


Transcriptome

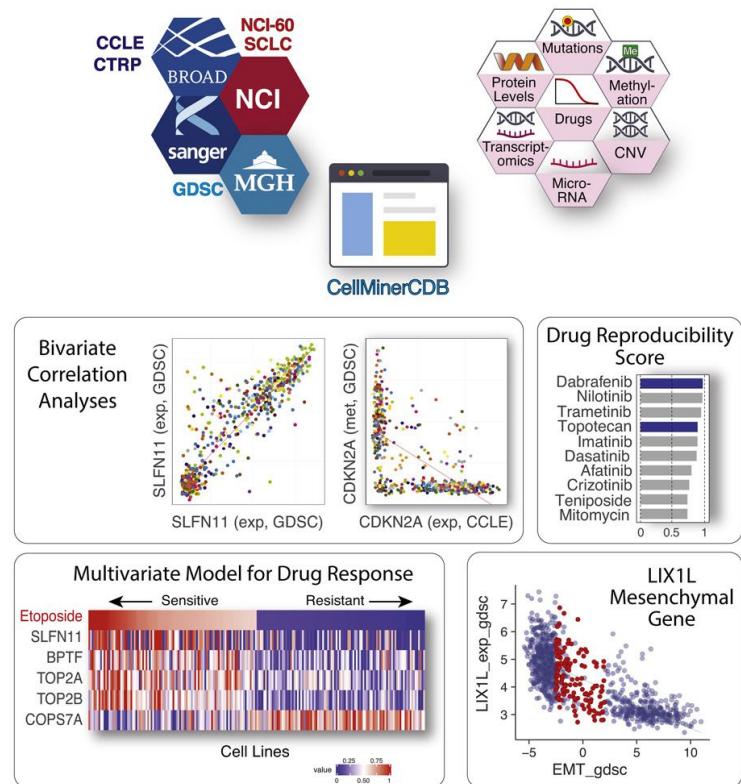
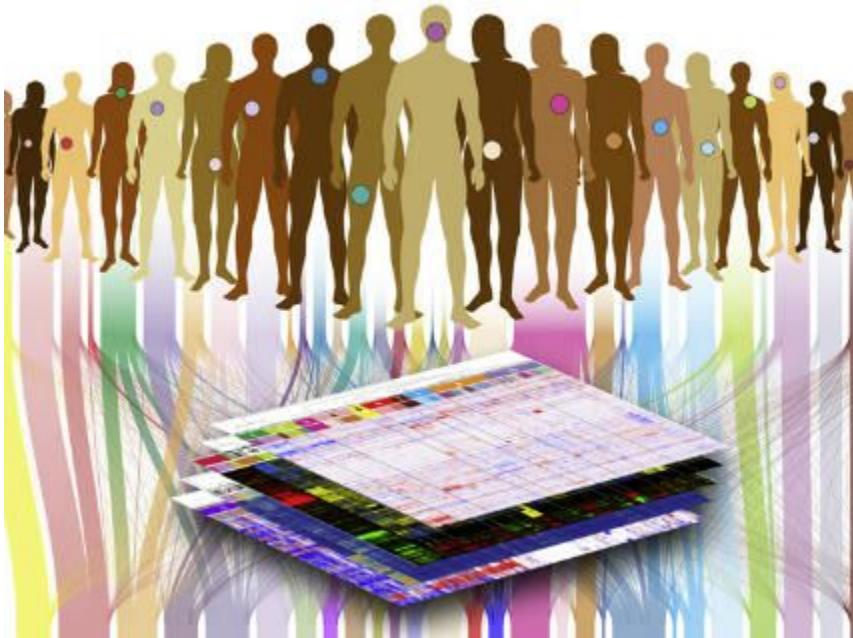


<https://www.ncbi.nlm.nih.gov/books/NBK453179/figure/ch5.Fig1/>

Proteome



Where can I find multi-omics data for cancer?



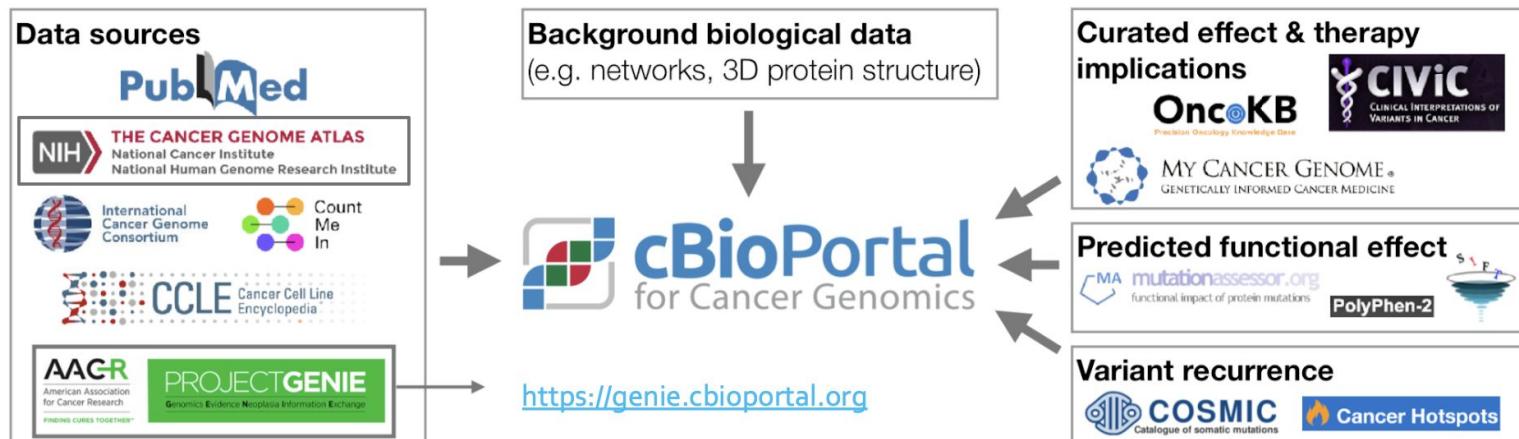
TCGA: [The Cancer Genome Atlas](#)

33 cancer types, >11000 tumor samples, 7 omics types

Drug responses in cancer cell lines
(CCLE/GDSC/NCI)

Where can I find multi-omics data for cancer?

What data is in cBioPortal?



Clinical data:

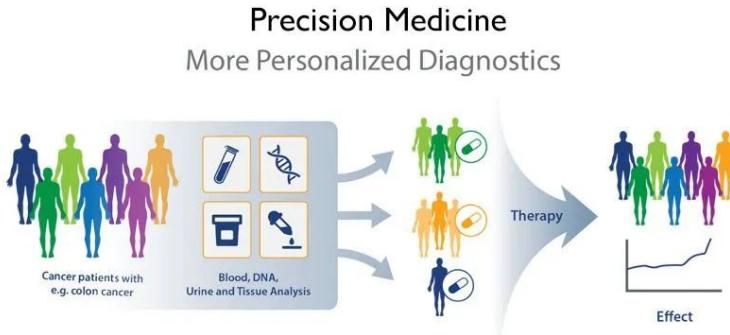
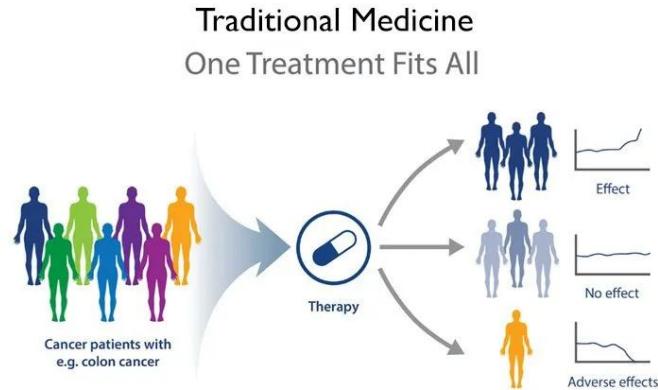
- Treatments
- Survival
- etc

-omic data:

- Mutations
- Fusions
- Copy number
- mRNA expression
- Protein levels
- DNA methylation*

<https://zifornd.com/blogs/cbiportal/>

Precision Medicine



The Manhattan Center for Gynecology offers
Precision Medicine

Patient stratification

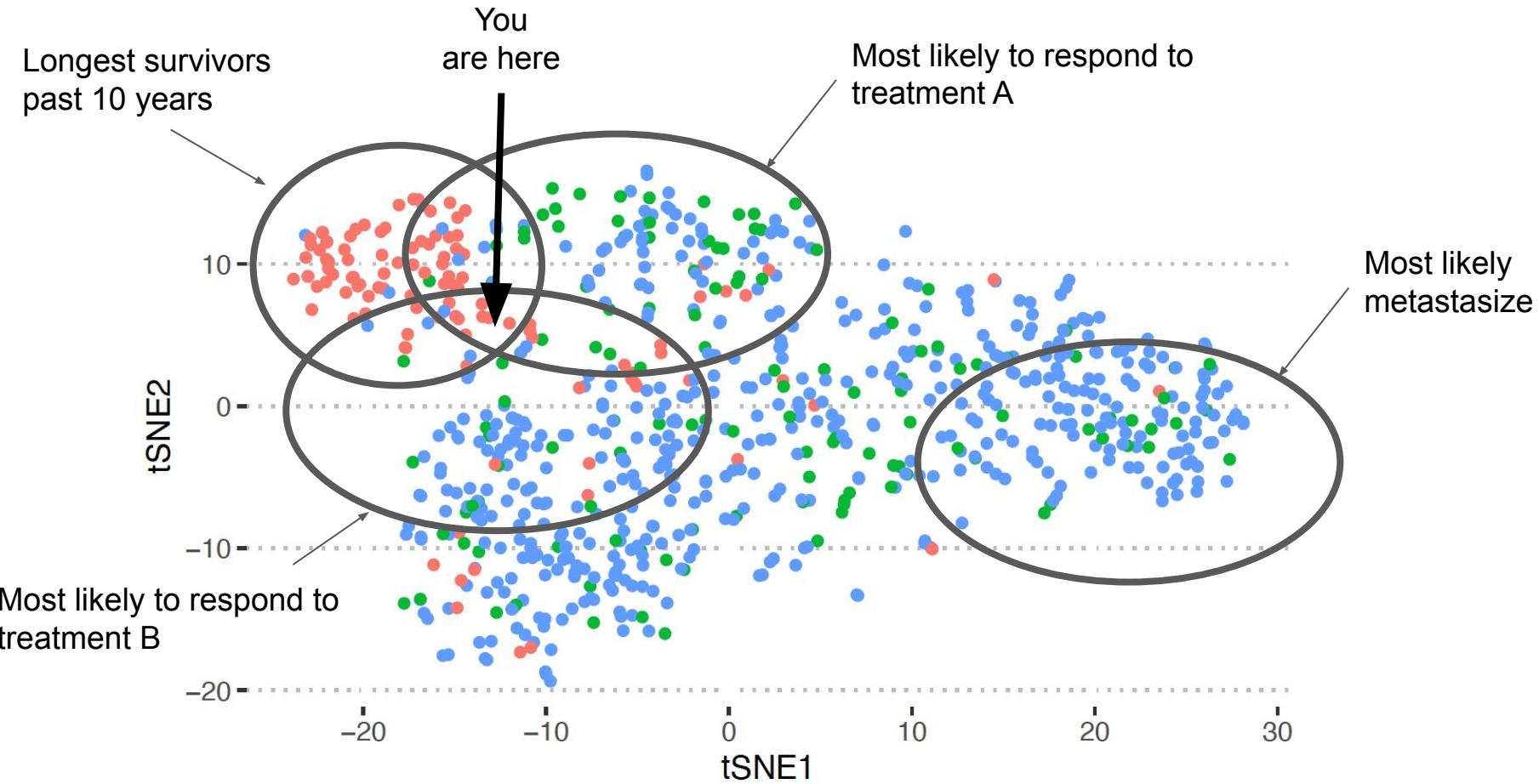
Drug response: does the patient have the molecular profile that will most benefit from this drug?

Survival outcomes: how likely is this patient to have a disease recurrence?

Subtyping: what is the potential underlying molecular mechanism that led to this disease for this patient?

E.g. high mutation burden, structural aberrations, high expression of hormone receptors

Metastasis: how likely is a given tumor to metastasize to other organs?



Multi-omics integration methods

Article | [Open Access](#) | Published: 05 January 2021

Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer

[Laura Cantini](#) , [Pooya Zakeri](#), [Celine Hernandez](#), [Aurelien Naldi](#), [Denis Thieffry](#), [Elisabeth Remy](#) & [Anaïs](#)

[Baudot](#) 

[Nature Communications](#) **12**, Article number: 124 (2021) | [Cite this article](#)

13k Accesses | **14** Citations | **61** Altmetric | [Metrics](#)

PLOS COMPUTATIONAL BIOLOGY

 OPEN ACCESS  PEER-REVIEWED

RESEARCH ARTICLE

Evaluation and comparison of multi-omics data integration methods for cancer subtyping

Ran Duan, Lin Gao , Yong Gao, Yuxuan Hu, Han Xu, Mingfeng Huang, Kuo Song, Hongda Wang, Yongqiang Dong, Chaoqun Jiang, Chenxing Zhang, Songwei Jia

Multi-omics integration methods (some examples)

Dimension Reduction

- MCIA
- MOFA
- JIVE
- RGCCA
- MAUI

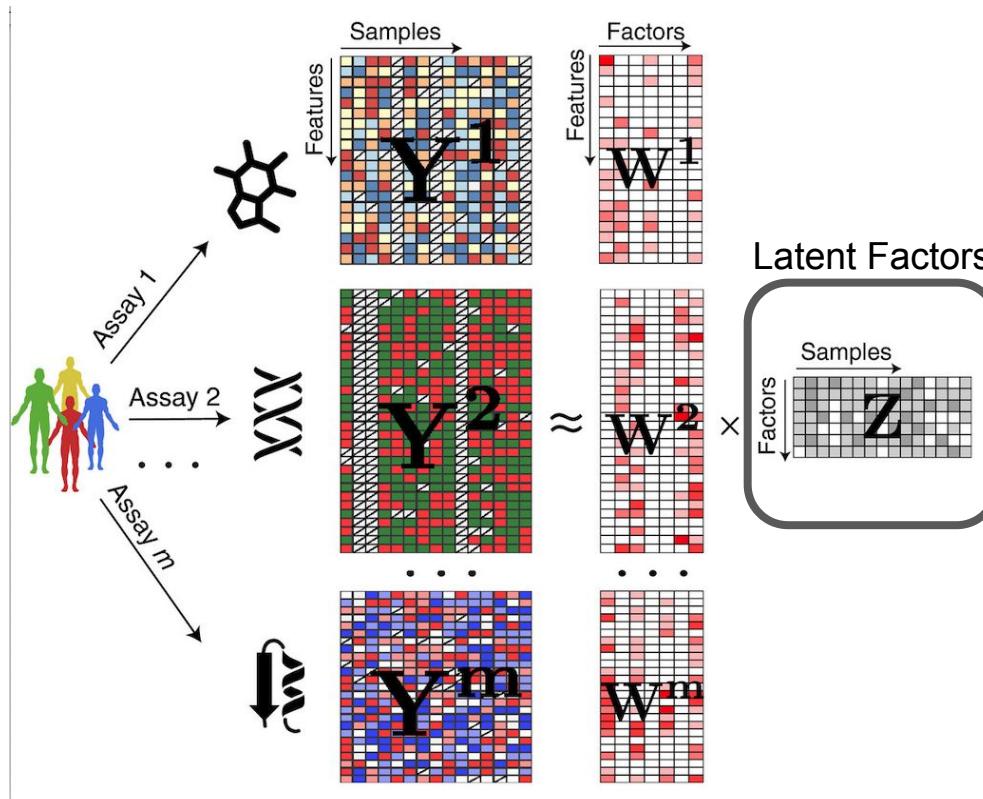
Dimension Reduction with Clustering

- intNMF
- iclusterPlus

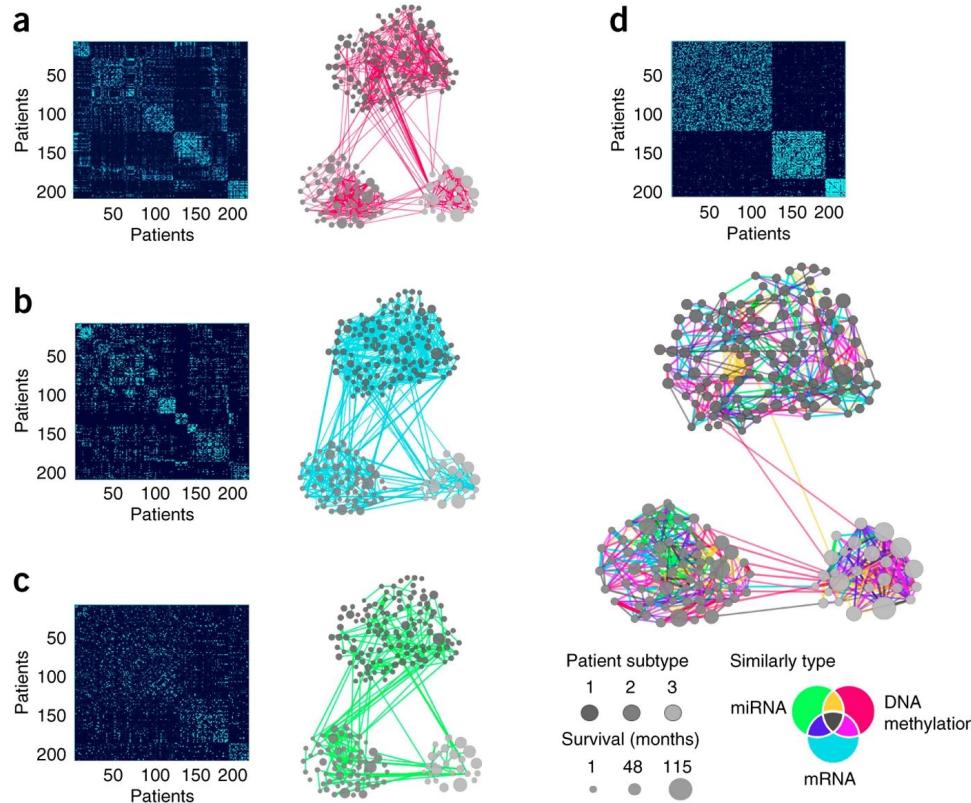
Clustering

- SNF
- COCA

Joint Dimension Reduction and Integration

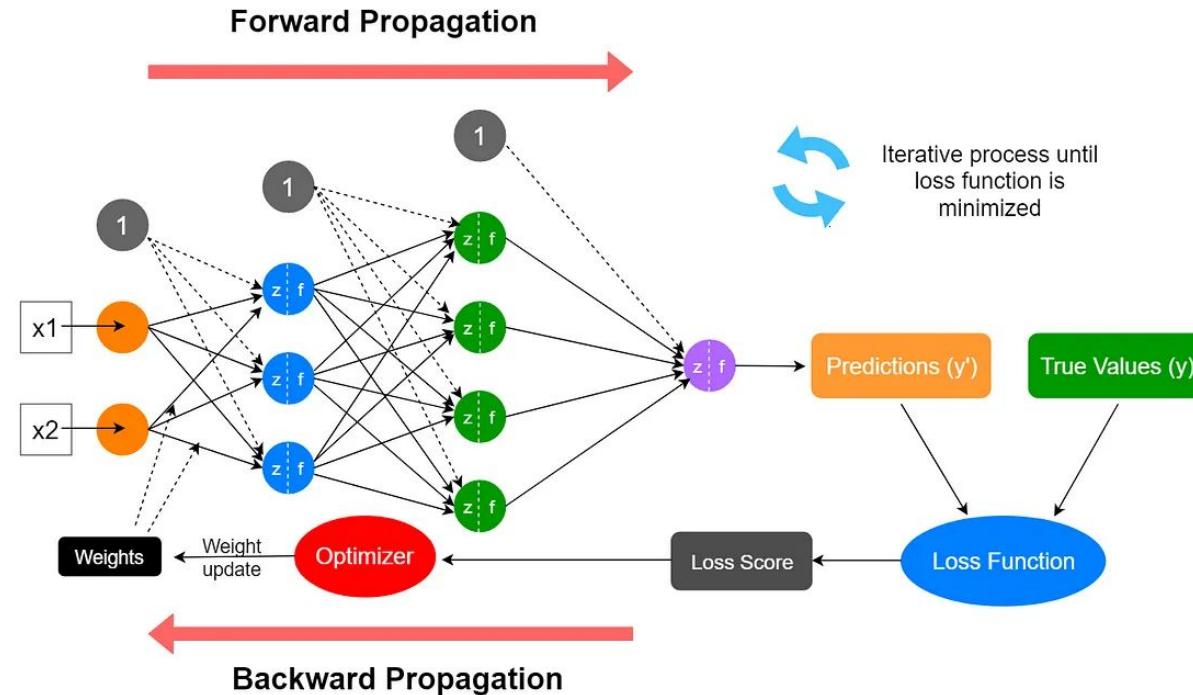


Similarity Network Fusion - Integrative Clustering



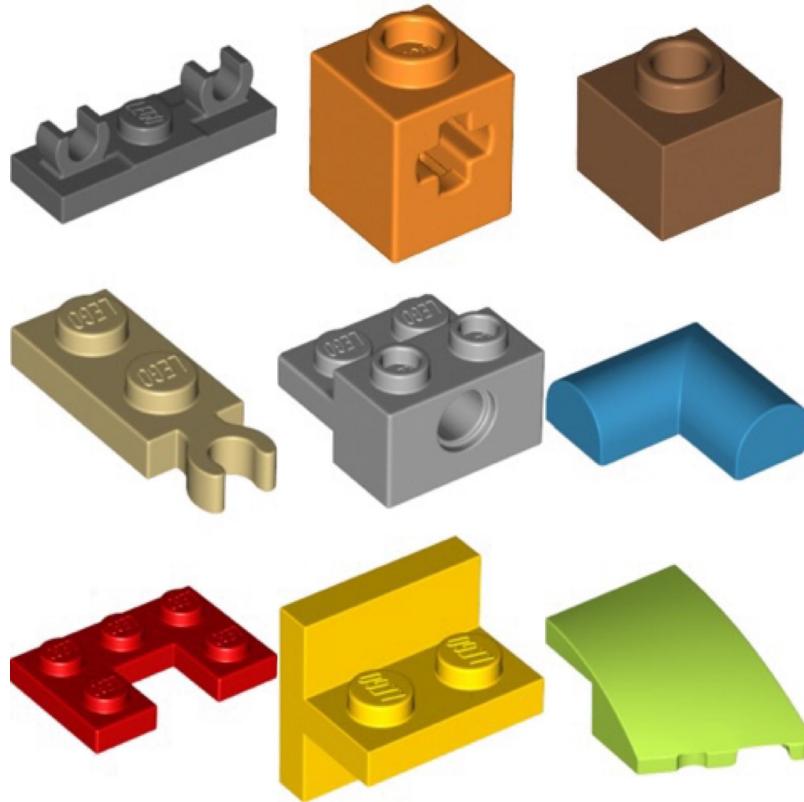
Deep Learning Approaches

Neural Networks



[Decoding Backpropagation and Its Role in Neural Network Learning |](#)
[ml-articles – Weights & Biases](#)

Neural Network Layers can be combined like Lego pieces



Different approaches for integrating omics layers using deep learning

[Tony Hauptmann and Stefan Kramer, Arxiv, 2022](#)

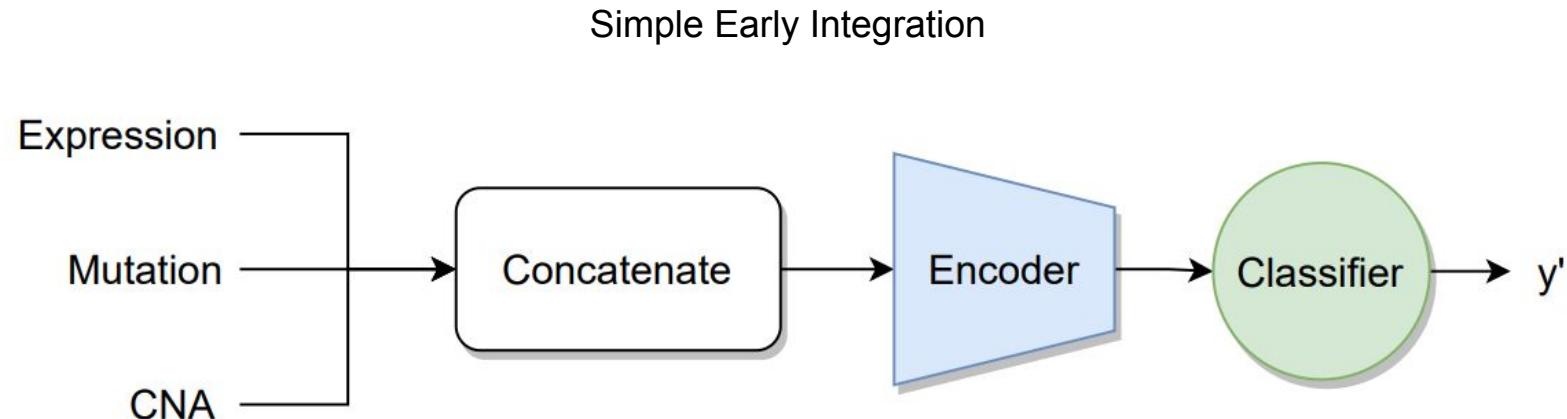


Figure 1: Schematic architecture of Early Integration with three input omics.

Different approaches for integrating omics layers using deep learning

[Tony Hauptmann and Stefan Kramer, Arxiv, 2022](#)

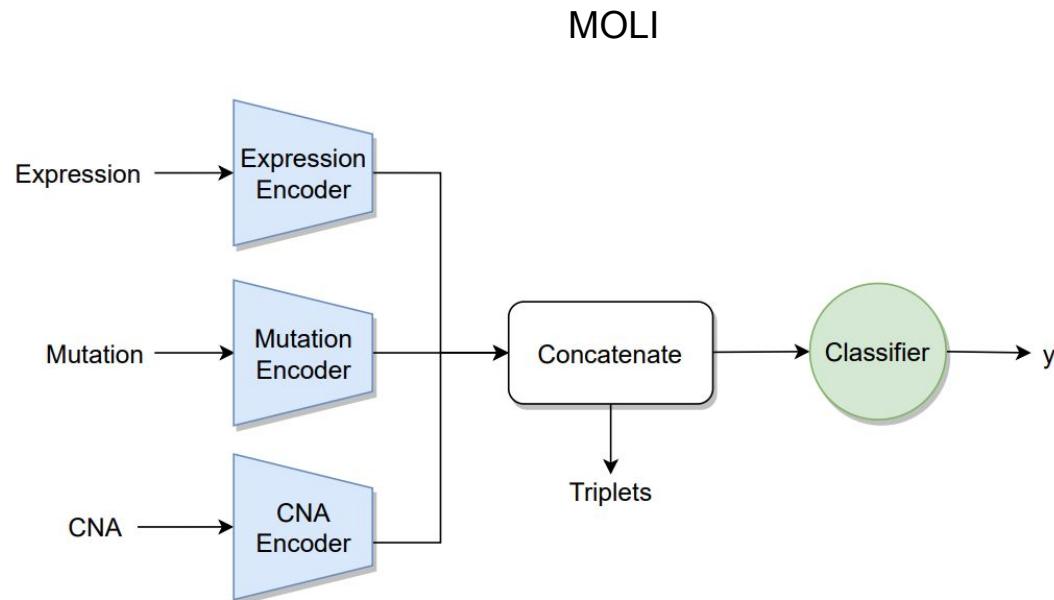


Figure 2: Schematic architecture of MOLI with three input omics.

Different approaches for integrating omics layers using deep learning

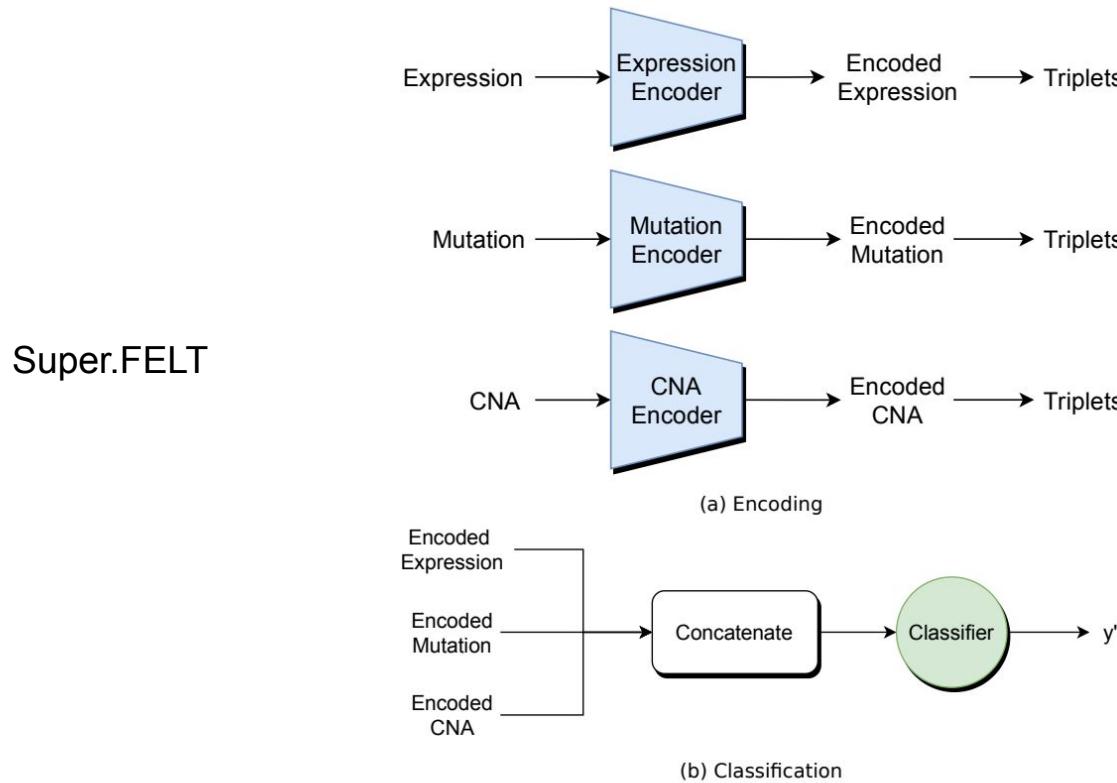


Figure 3: Schematic architecture of Super.FELT with three input omics.

Different approaches for integrating omics layers using deep learning

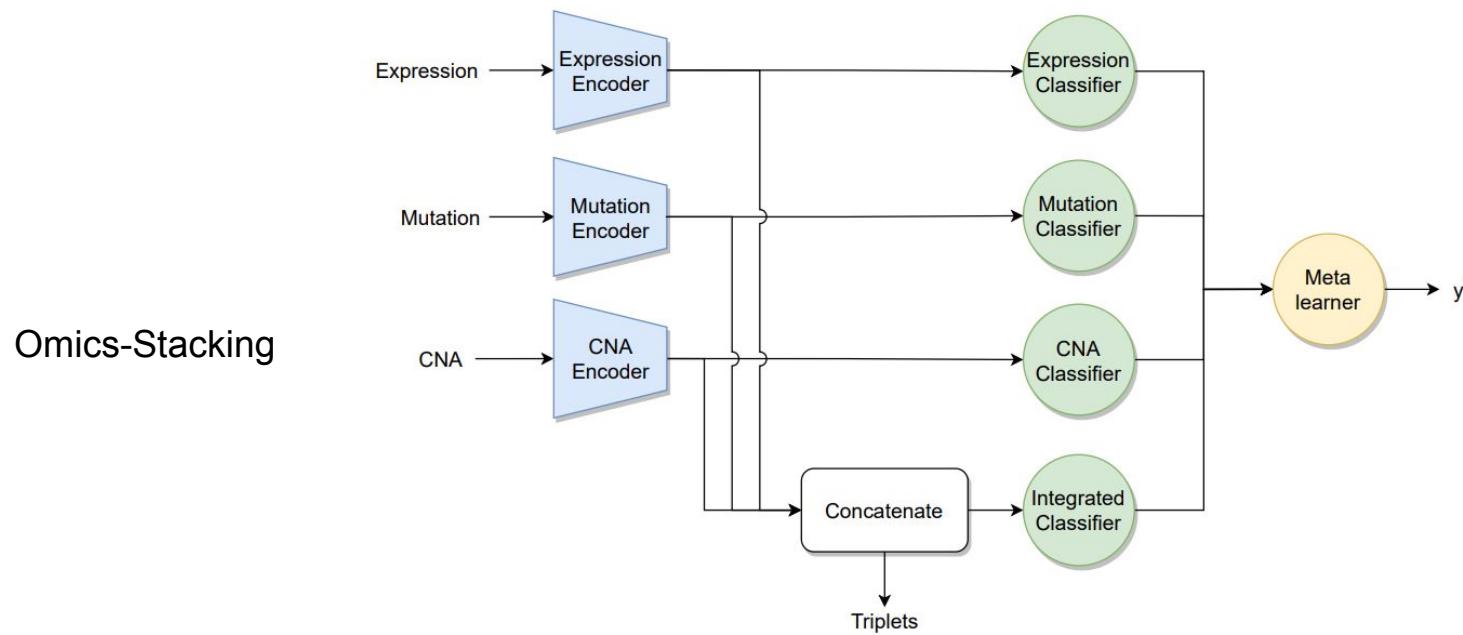


Figure 4: Schematic architecture of Omics-Stacking with three input omics.

Different approaches for integrating omics layers using deep learning

MOMA

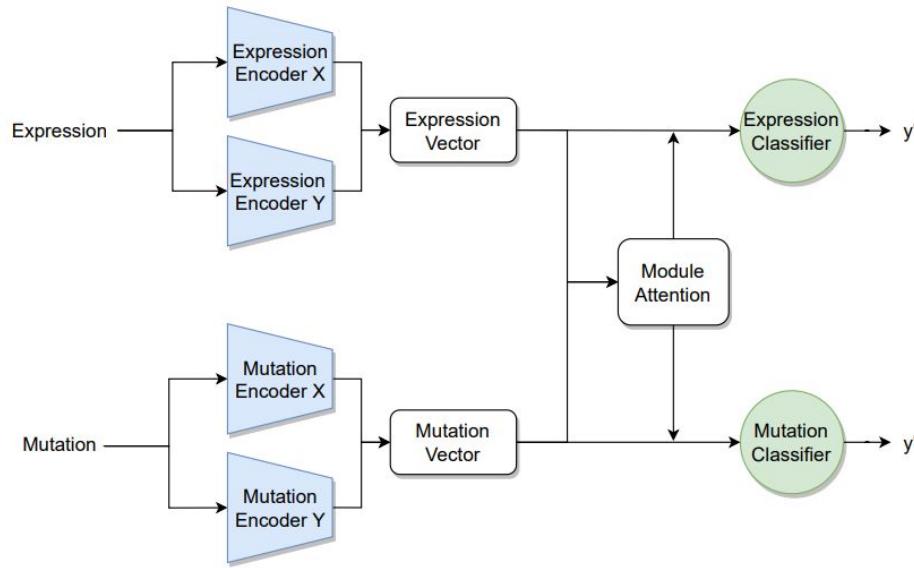


Figure 5: Schematic architecture of the network used in MOMA for two input omics.

Different approaches for integrating omics layers using deep learning

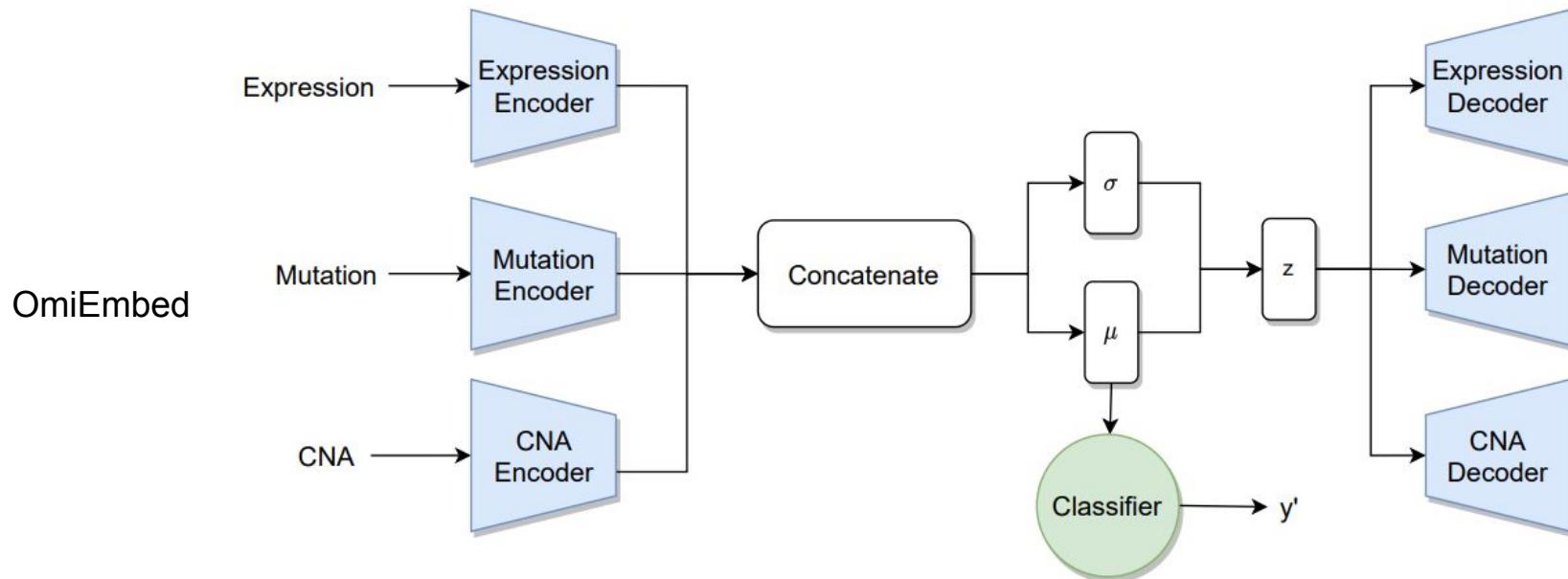


Figure 6: Schematic architecture of OmiEmbed for three input omics.

Conclusions from Some Benchmarking Papers

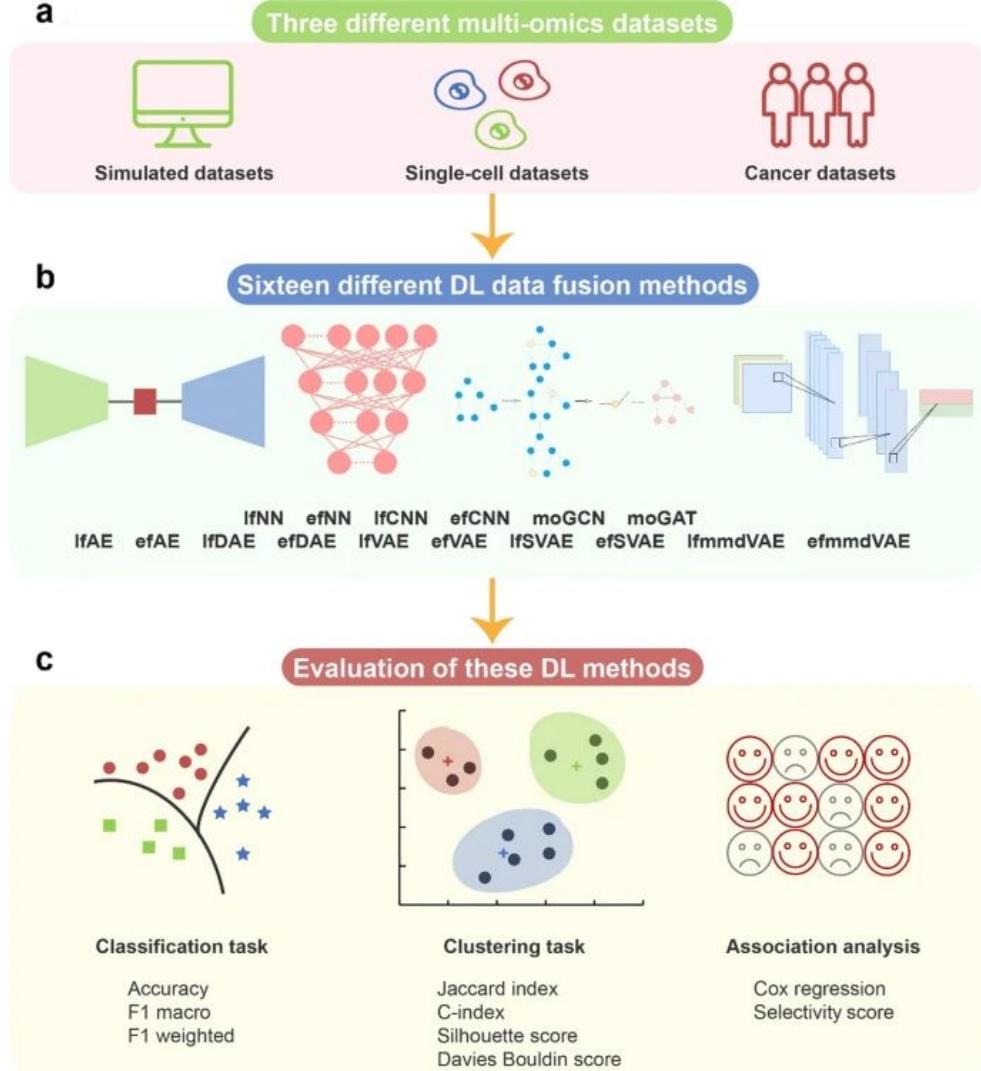
Benchmark by [Tony Hauptmann and Stefan Kramer, Arxiv, 2022](#)

Compared **6 Tools**: Simple Early Integration Classifier, Super.FELT, MOLI, Omics-Stacking, OmiEmbed, MOMA

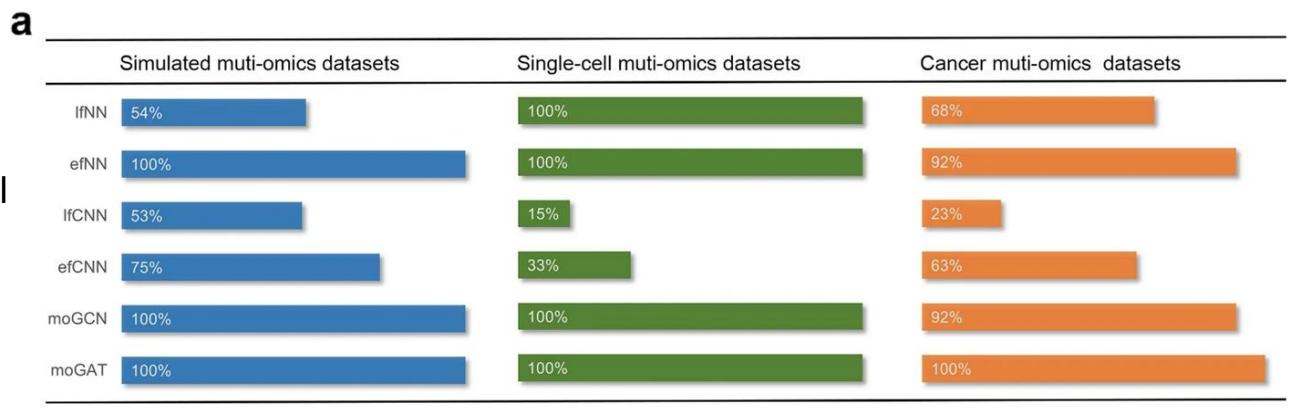
- **Early integration performed significantly worse**
- **No method clearly outperforms** others
- Tools using **triplet loss** perform better than those that don't
- Omiembed (supervised VAE) performed worst on **holdout data**

They test:

- Integration sequence (early/late fusion)
- Encoding
 - FCN
 - Autoencoders
 - Basic
 - Variational
- Loss function (MMD vs KL for autoencoders)
- Convolutions
 - Graph attention network
 - Graph convolution network
 - 1D CNN

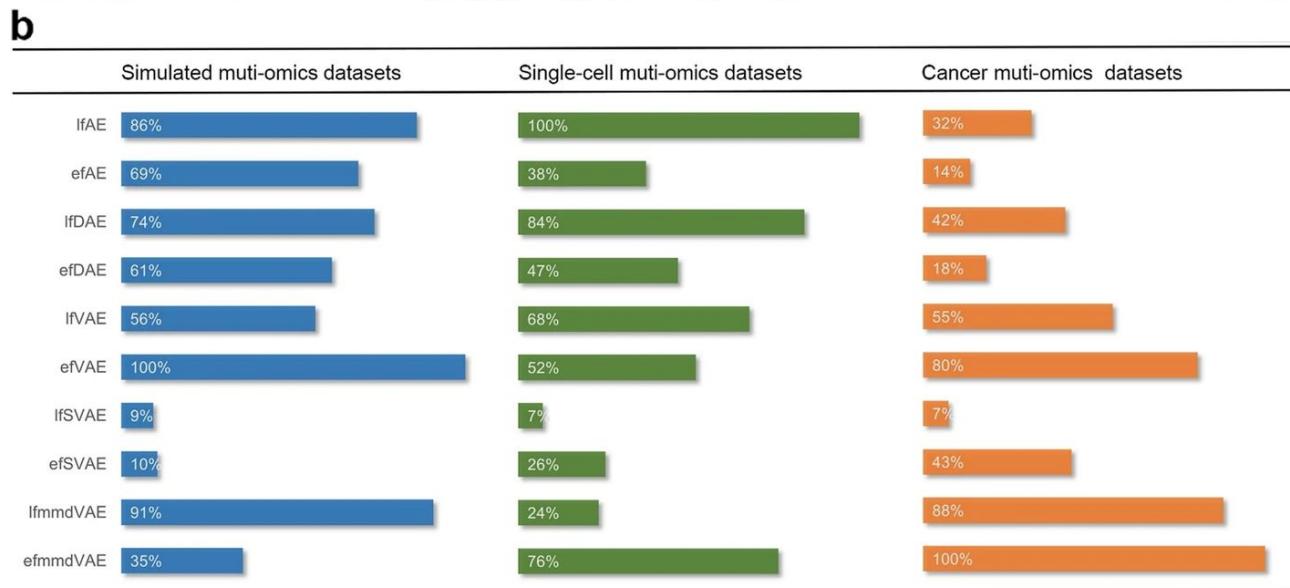


[A benchmark study of deep learning-based multi-omics data fusion methods for cancer | Genome Biology | Full Text](#)



Summary:

- No approach is the best in all scenarios
- Overall Graph Attention Network performs better for classification
- CNNs don't work well



[A benchmark study of deep learning-based multi-omics data fusion methods for cancer | Genome Biology | Full Text](#)

Fusion strategy	Taxonomy Subcategory 1	Taxonomy Subcategory 2	Papers
	Approach	Architecture	
Early fusion	Direct modeling	Fully connected	[17–19]
		Convolutional	[20–23]
		Recurrent	[20, 24]
	Autoencoder	Regular	[25–34]
		Denoising	[33, 35–37]
		Stacked	[37–40]
		Variational	[33, 40–42]
	Branch	Representation	
Intermediate fusion	Homogeneous design	Marginal	[43–49]
		Joint	[21, 28, 38, 41, 50–63]
	Heterogeneous designs	Marginal	[64–68]
		Joint	[69–81]
	Aggregation	Model contribution	
Late fusion	Averaging	Equal	[82–84]
		Weighted	[85–87]
	Meta-learning	Weighted	[83, 88]

JOURNAL ARTICLE

Multimodal deep learning for biomedical data fusion: a review

Sören Richard Stahlschmidt , Benjamin Ulfenborg, Jane Synnergren

Briefings in Bioinformatics, Volume 23, Issue 2, March 2022, bbab569,

<https://doi.org/10.1093/bib/bbab569>

Published: 28 January 2022 Article history ▾

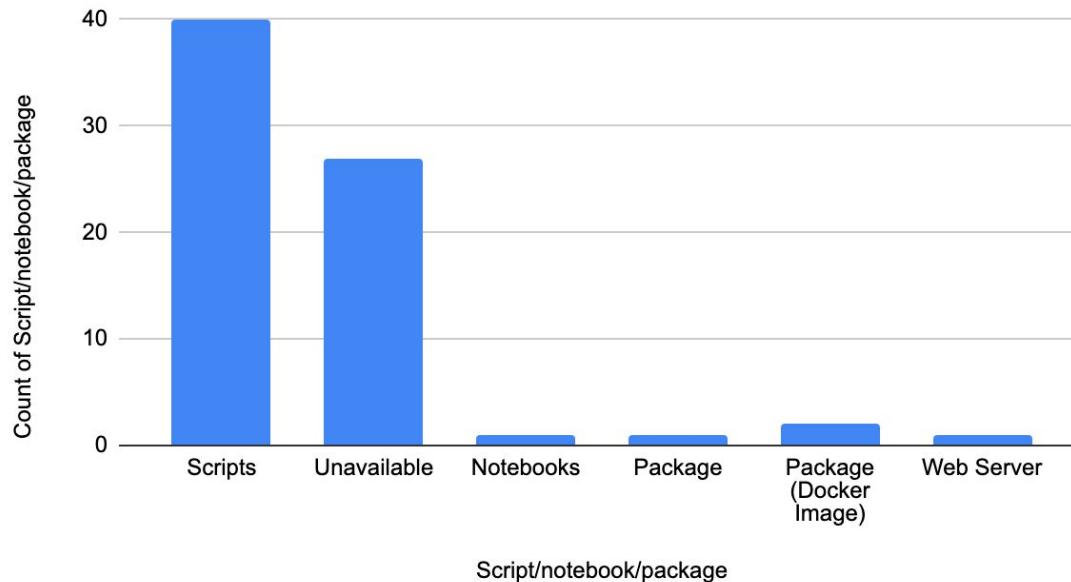
<https://academic.oup.com/bib/article/23/2/bbab569/6516346?login=true>

> 70 Deep Learning Multi-Omics Papers!

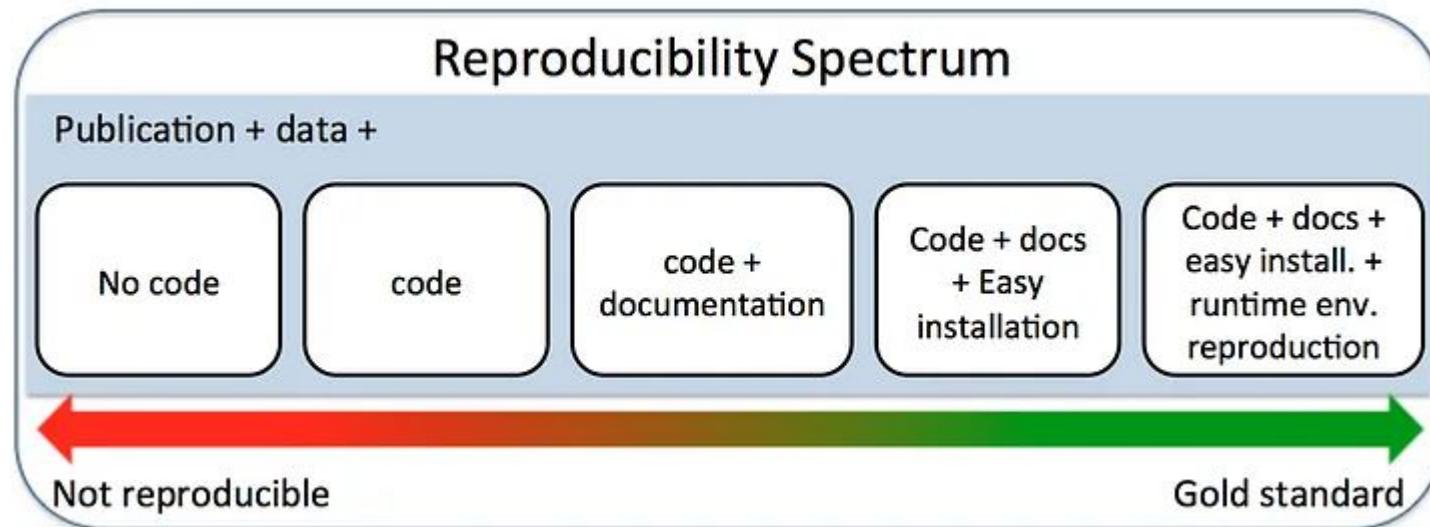
Literature Search Summary

MultiOmics Integration Tools (Deep Learning Based)

Count of Script/notebook/package



Reproducibility Spectrum



Altuna Akalin @AltunaAkalin · Mar 2

Reproducibility is warfare and a journey far from home

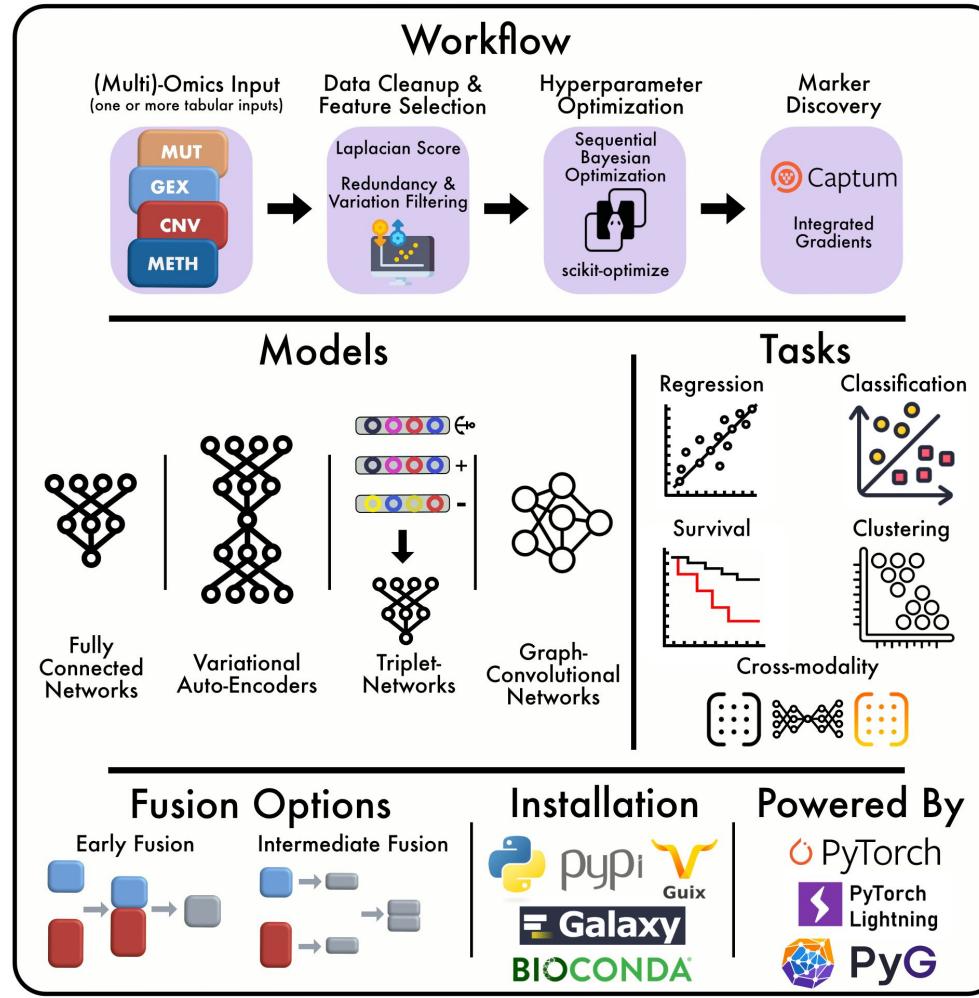
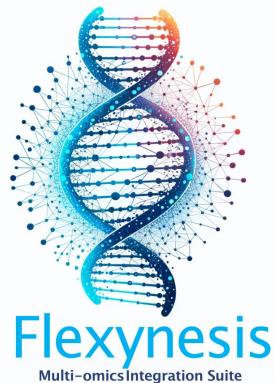


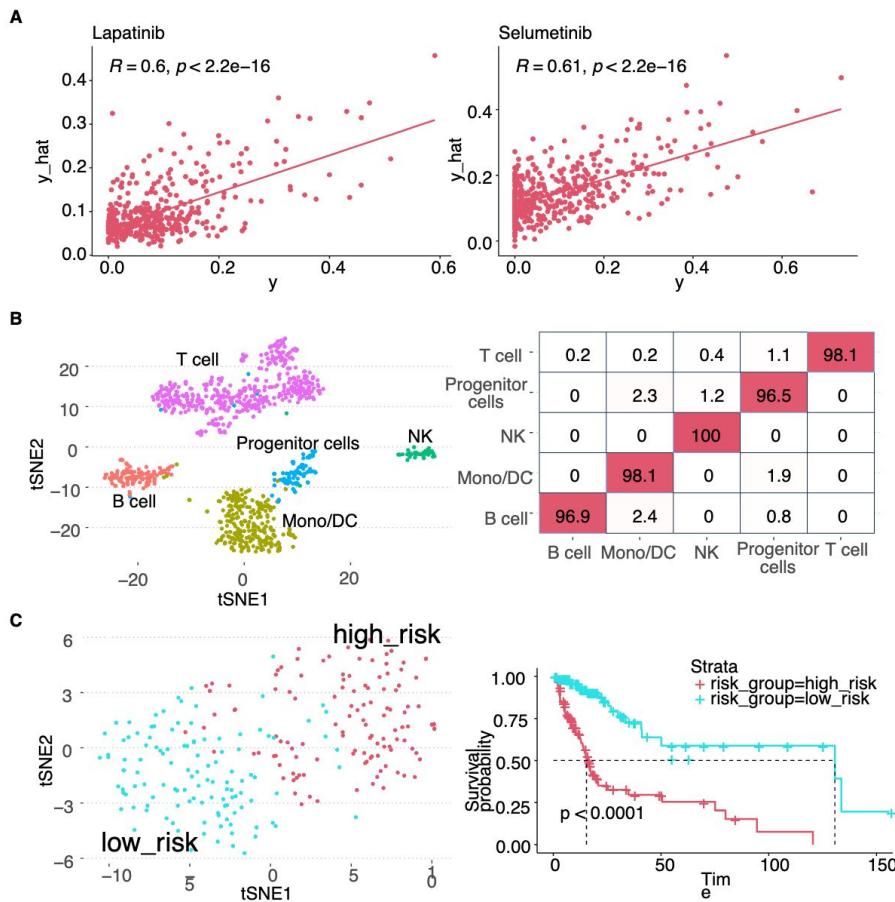
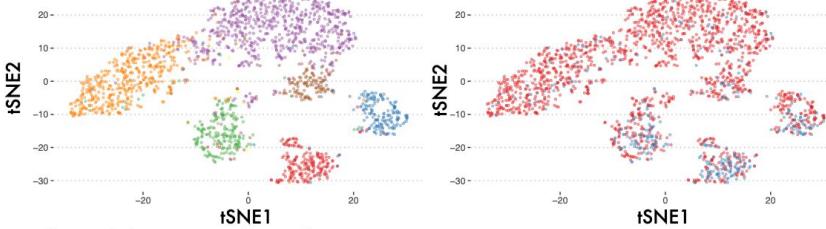
Figure 1

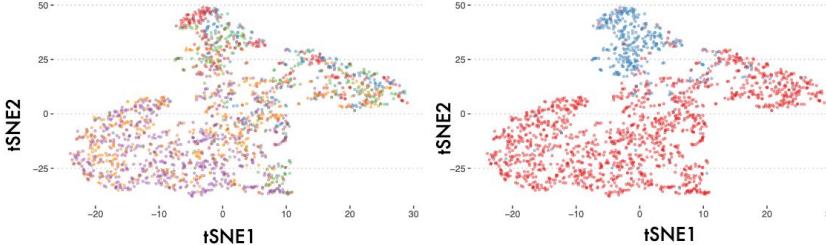
Figure 2



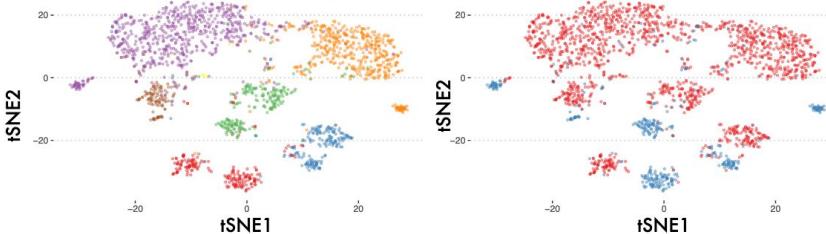
A) single-task learning: subtype

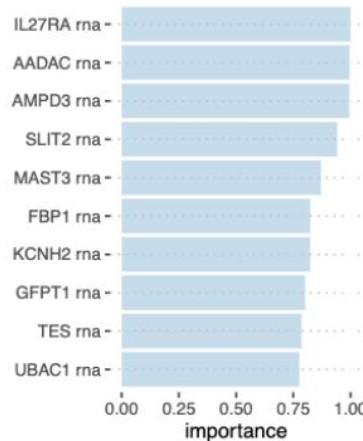
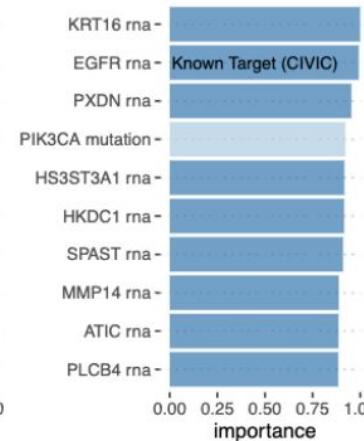
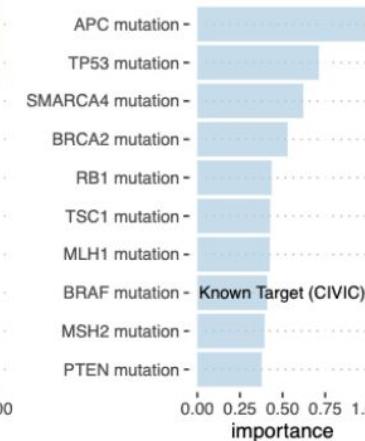
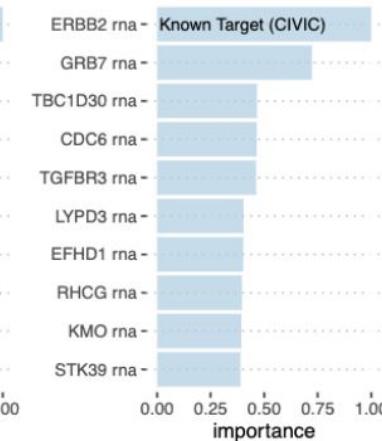
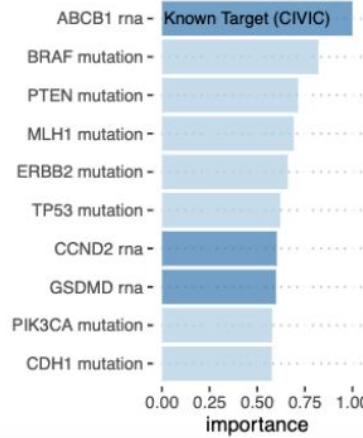
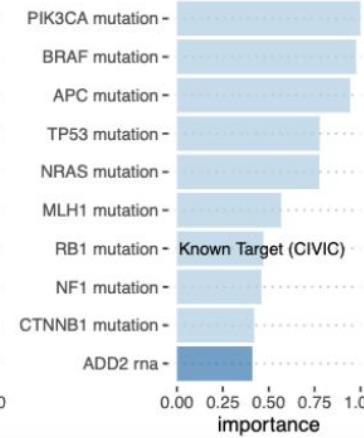
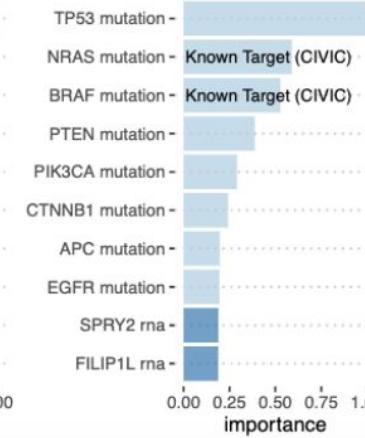
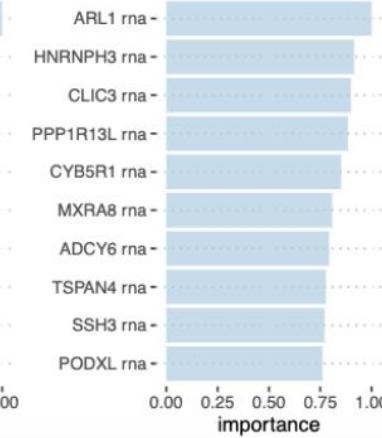


B) single-task learning: chemotherapy response



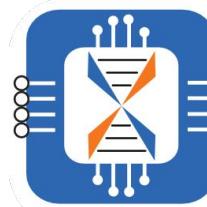
C) multi-task learning: both subtype and chemotherapy response



B**Crizotinib****Erlotinib****Irinotecan****Lapatinib****Paclitaxel****Palbociclib****Selumetinib****Sorafenib**

Acknowledgements

Akalin Lab Members



Bioinformatics &
Omics Data Science
Platform

MAX
DELBRÜCK
CENTER

Day 1 - Homework

1. System Setup

Rolv.io / Docker / Mamba

See

https://github.com/BIMSBbioinfo/compgen_course_2025_module3/blob/main/README.rst#compute-environment

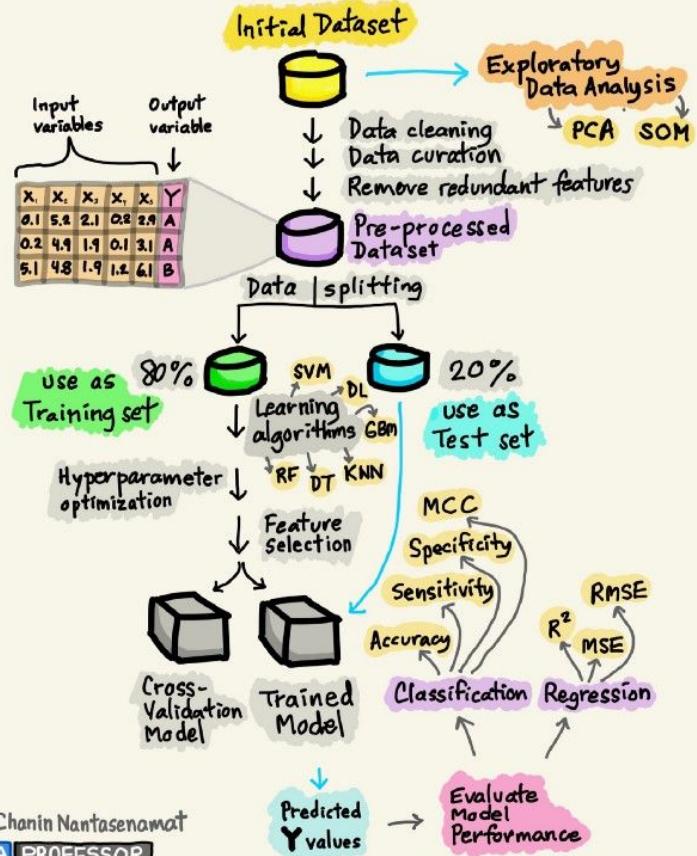
2. See

https://github.com/BIMSBbioinfo/compgen_course_2025_module3/tree/main/homeworks/hw1

Getting Help

https://github.com/BIMSBbioinfo/compgen_course_2025_module3/discussions

BUILDING THE MACHINE LEARNING MODEL



By: Chanin Nantasenamat

DATA PROFESSOR



<http://youtube.com/dataprofessor>

January 1, 2020

How to Build a Machine Learning Model | by
Chanin Nantasenamat | Towards Data Science