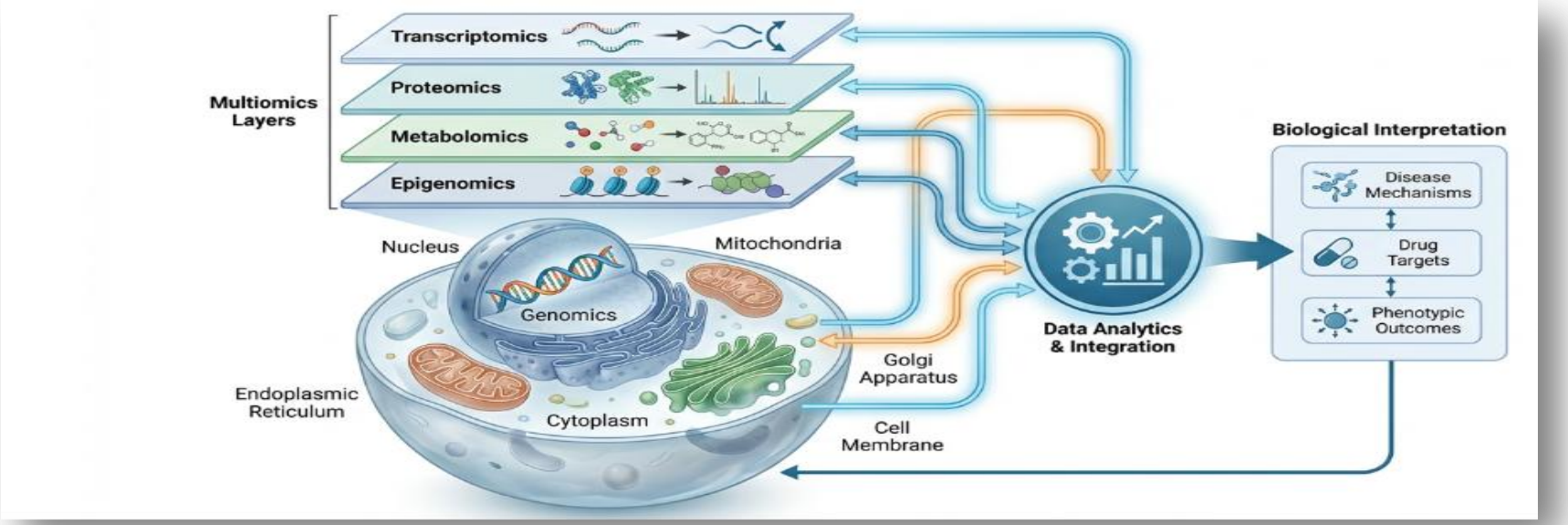
INTERNATIONAL CONFERENCE 2026 April 15 – 16, 2026

THE BIOLOGICAL KNOWLEDGE DISCOVERY (BKD)

Discovery in Biology for application in health care and Agriculture.

Connecting the omics dots: Data analytics for integrated biological insights



Dr. Sajitha Lulu S

Associate Professor Senior

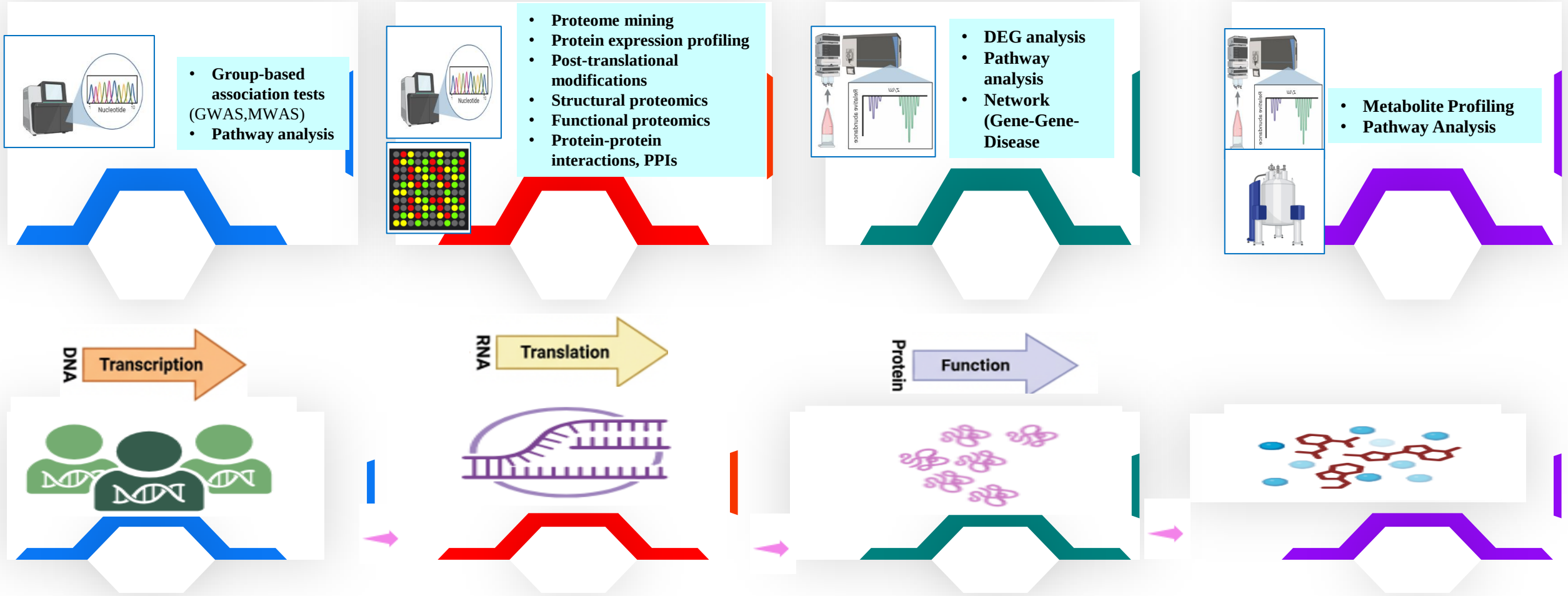
Integrative Multi-omics Laboratory

School of Bio Sciences and Technology

VIT, Vellore

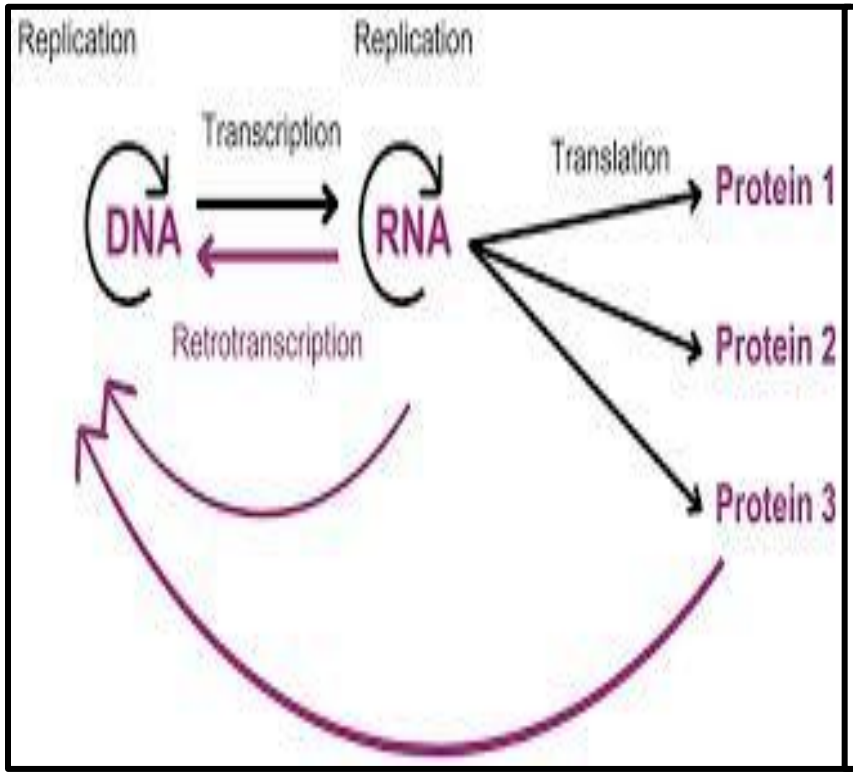
632014

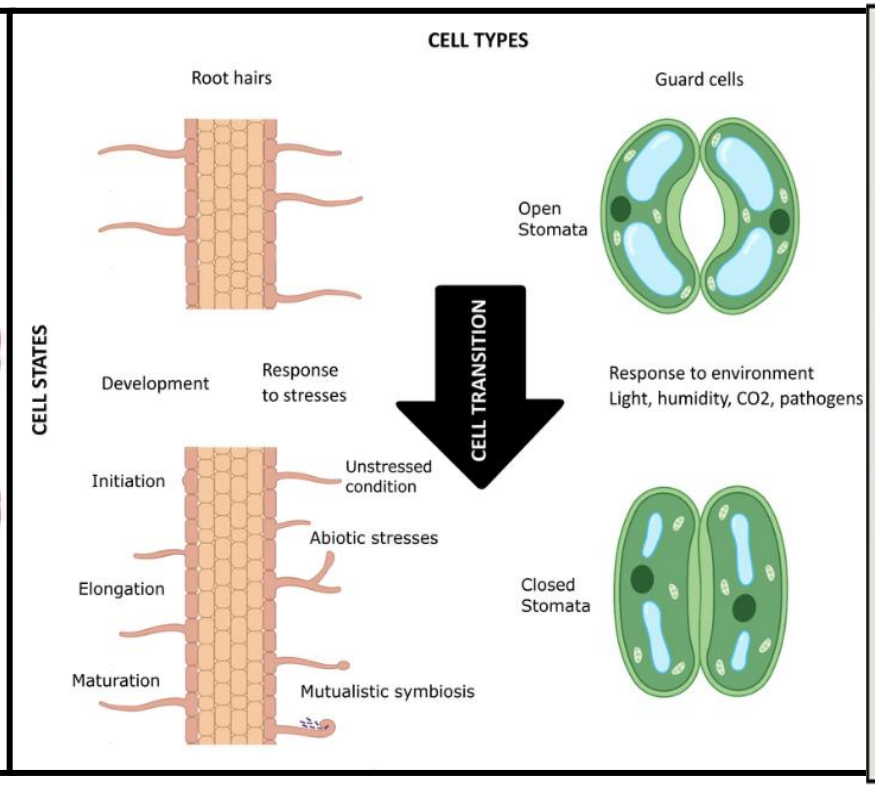
Introduction

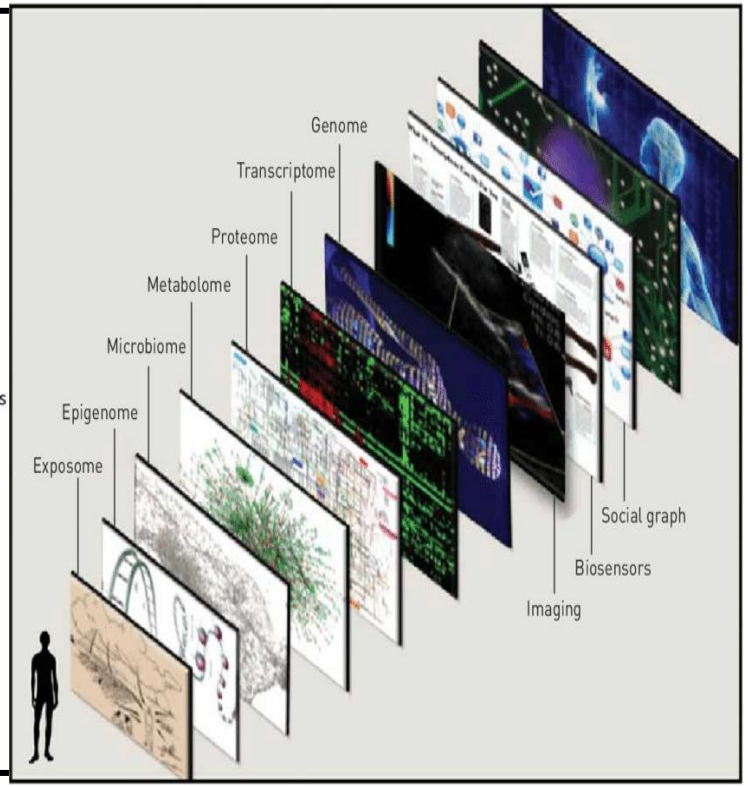


Each omics layer provides a partial view of biology, but only together do they reveal the complete functional state of a system

Biological Complexity



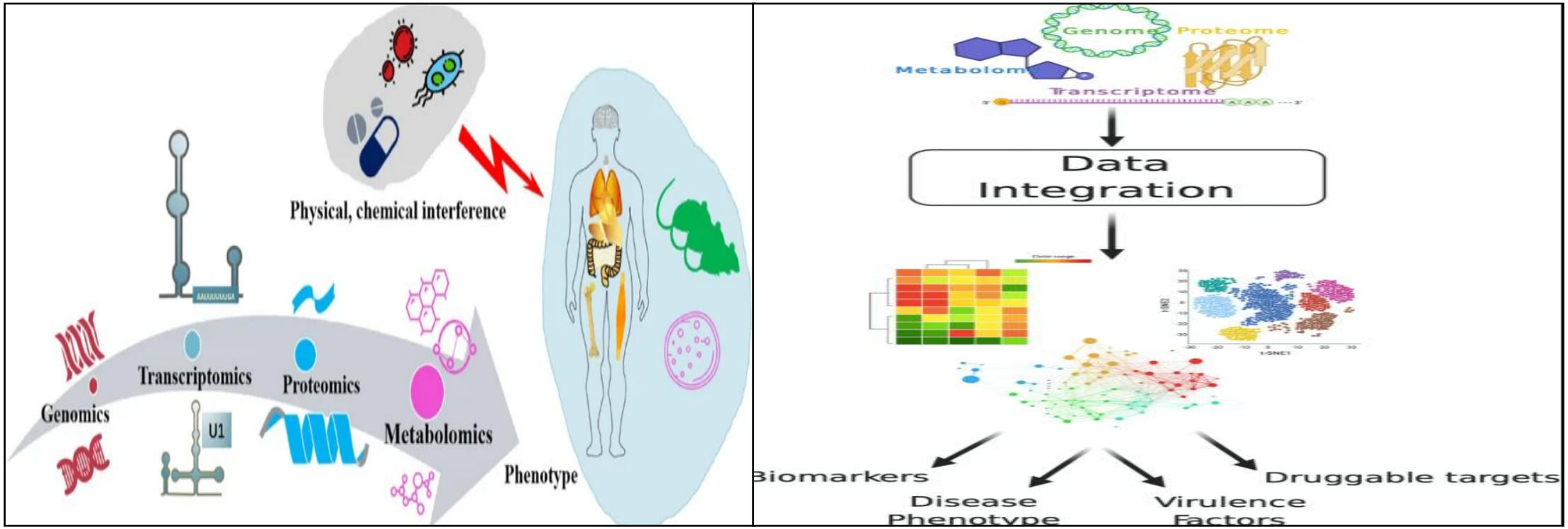




Hasin et al., 2017, Genome Biology

Biology is not linear , it is dynamic, multi-layered, and interconnected

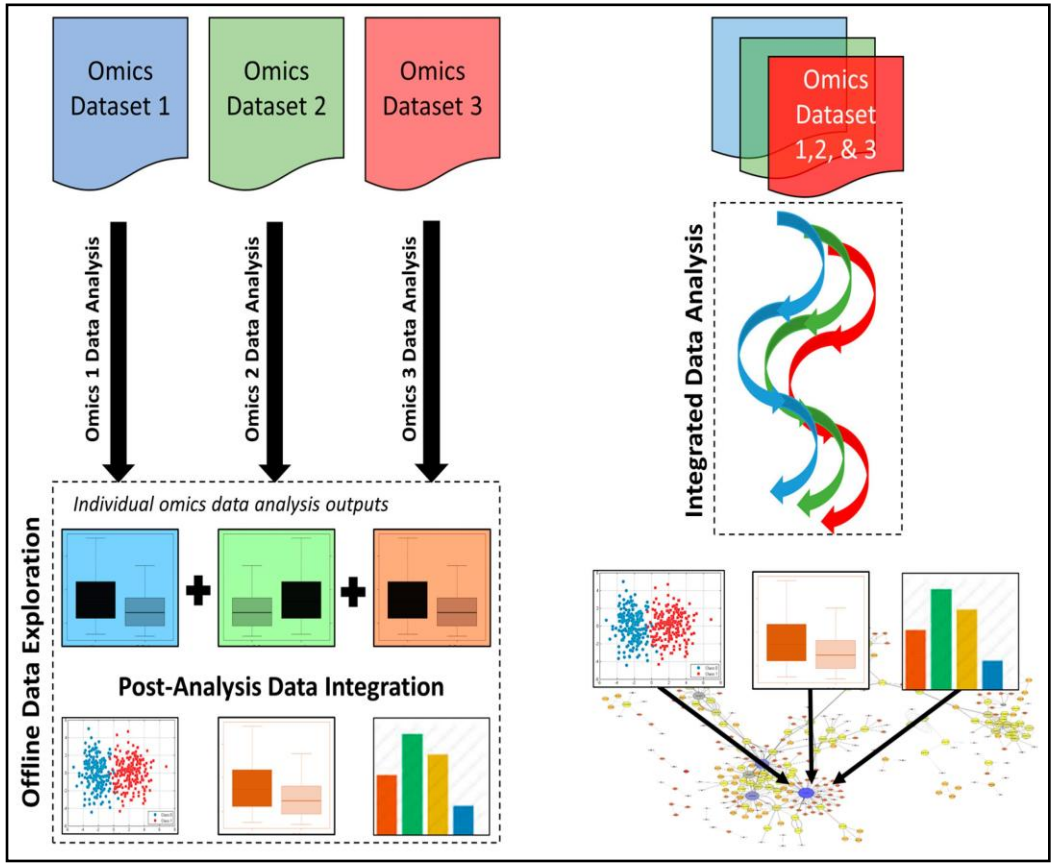
Omics Layers : Multi-omics integration connects molecular data to real-world disease outcomes and clinical decisions.



Karczewski & Snyder, 2018, Nature Reviews Genetics

Single-omics fails to capture biological complexity

Data Integration in Multi-Omics: Horizontal Data Integration



Gene expression (RNA-seq)

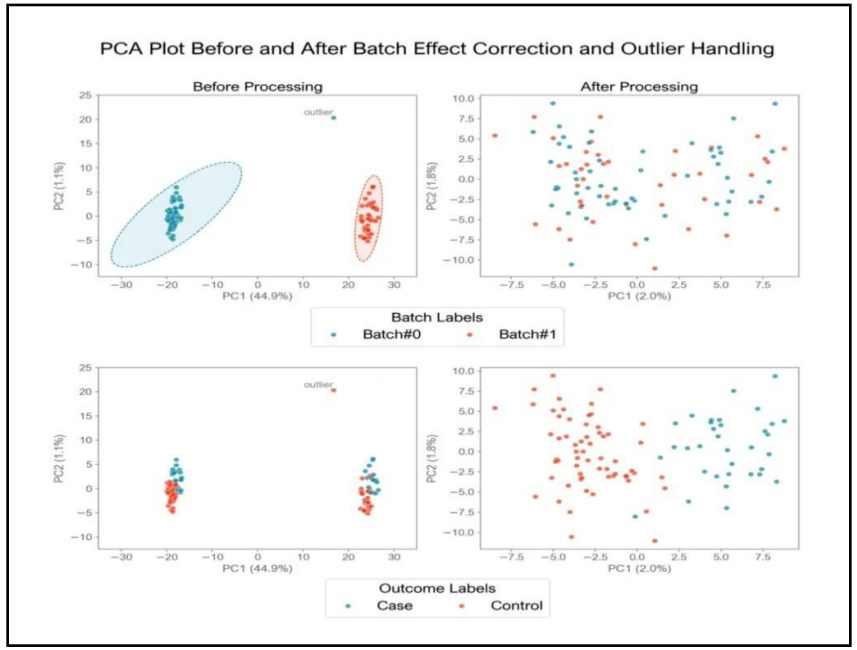
- TCGA
- GEO
- Clinical cohorts

Benifits

- Increase sample size
- Improve statistical power
- Reduce bias

Challenges

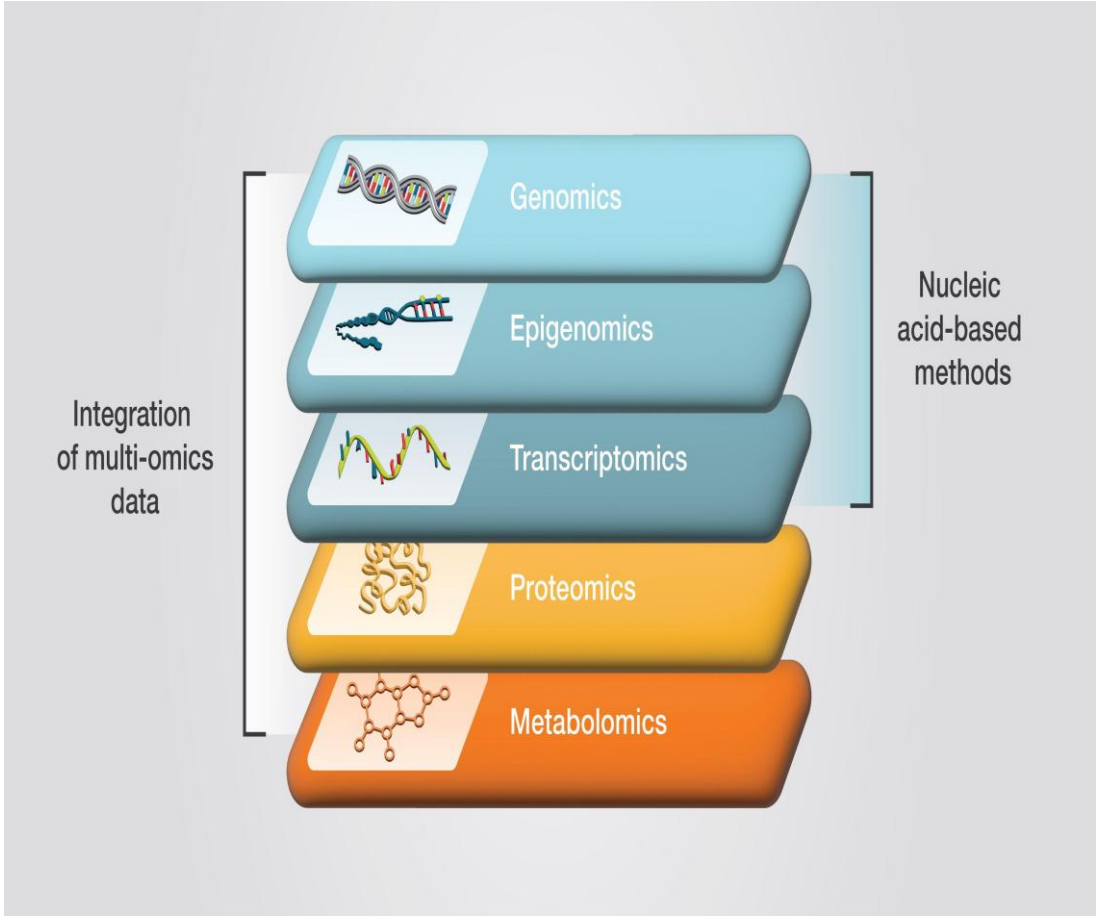
- Batch effects
- Different platforms
- Data normalization



Batch effects can completely mask biological signals, and proper correction reveals the true structure of the data.

Vertical Data Integration

Combining different omics layers for the same samples



Benefits

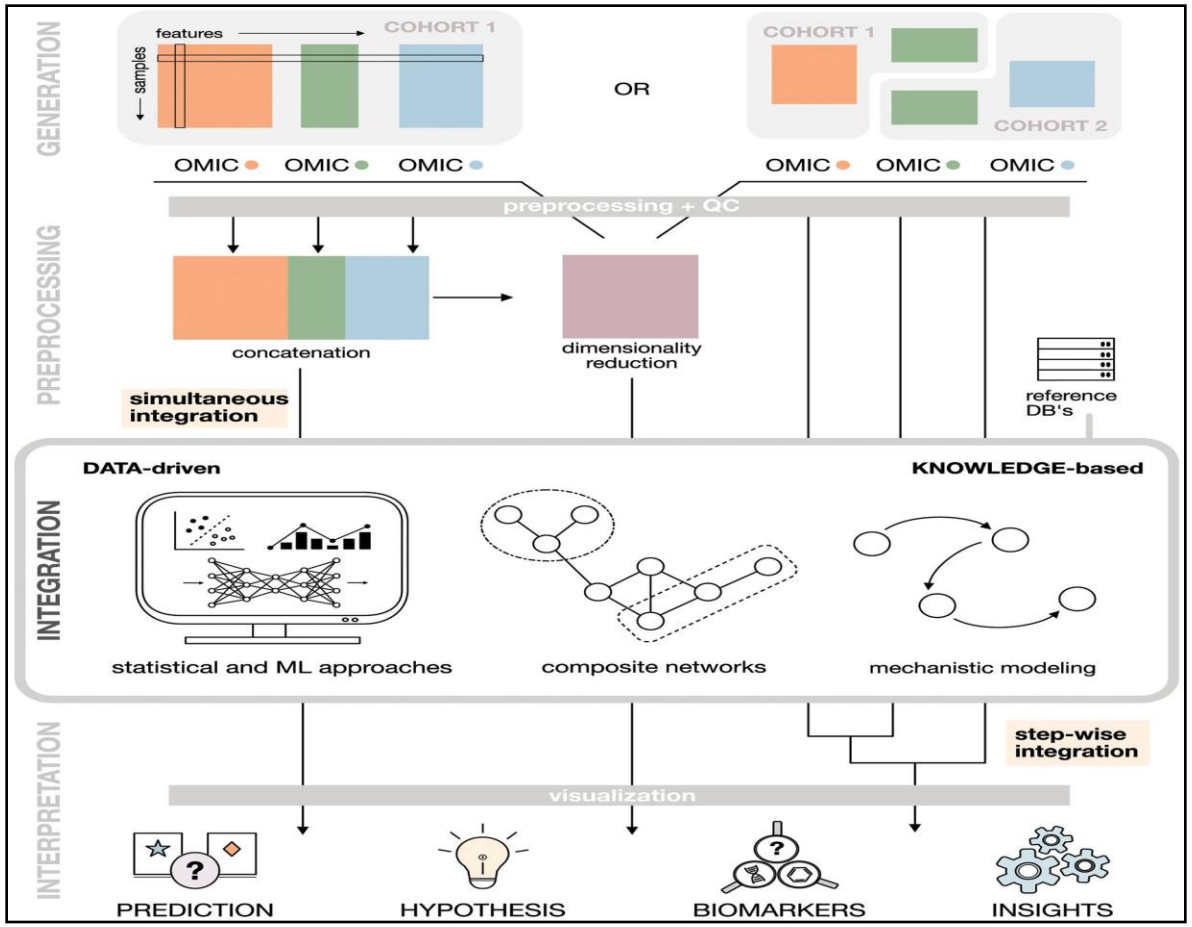
- Understand biological systems
- Capture interactions across layers
- Reveal mechanisms

Challenges

- Data heterogeneity
- Missing data
- Integration complexity

Horizontal integration strengthens evidence; vertical integration reveals biology.

Horizontal Integration - Tools for Integration



Batch Correction / Harmonization

ComBat

- Most widely used
- Empirical Bayes method

Limma

- Strong for differential expression + correction

Harmony

- Excellent for single-cell datasets

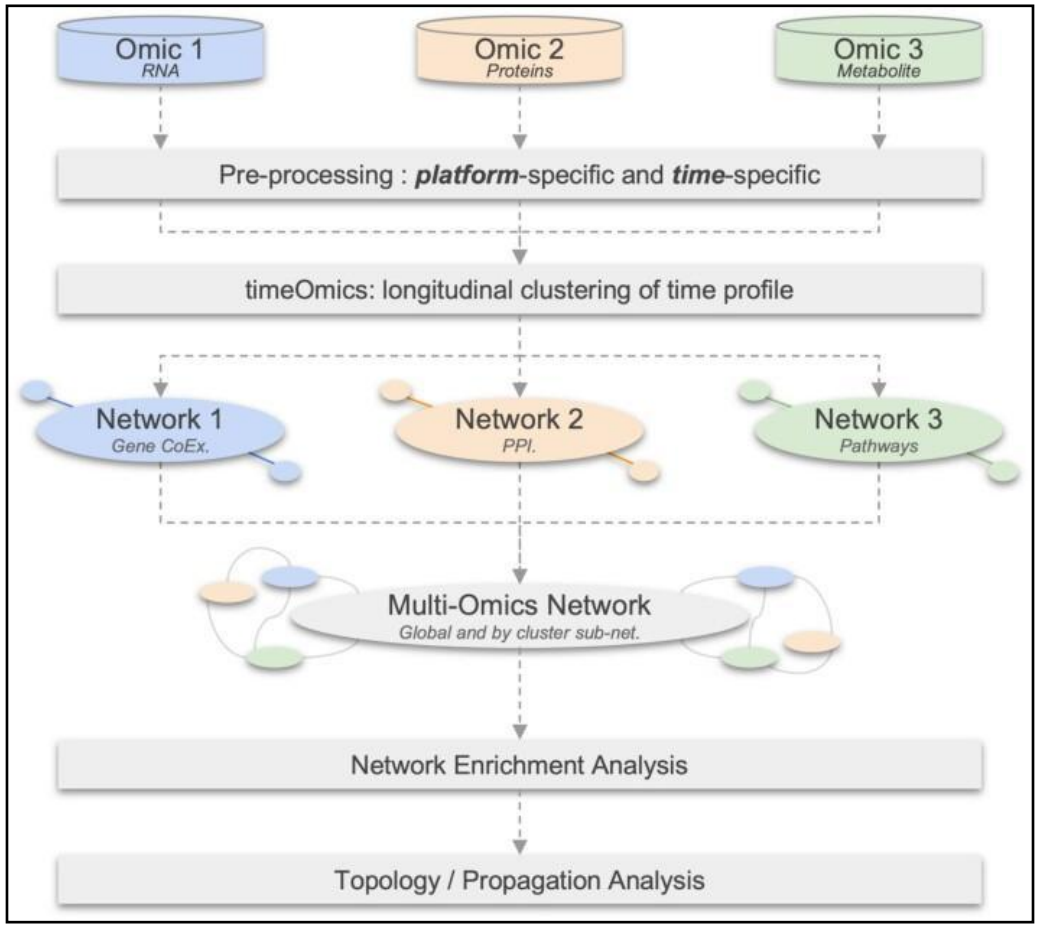
MNN Correct

- Aligns datasets using nearest neighbors

Meta-analysis Tools

- MetaDE
- RankProd

Vertical Integration - Tools for Integration



Integrate different omics layers

- MOFA
- iCluster
- mixOmics
- SNFtool

Analysis & Visualization:

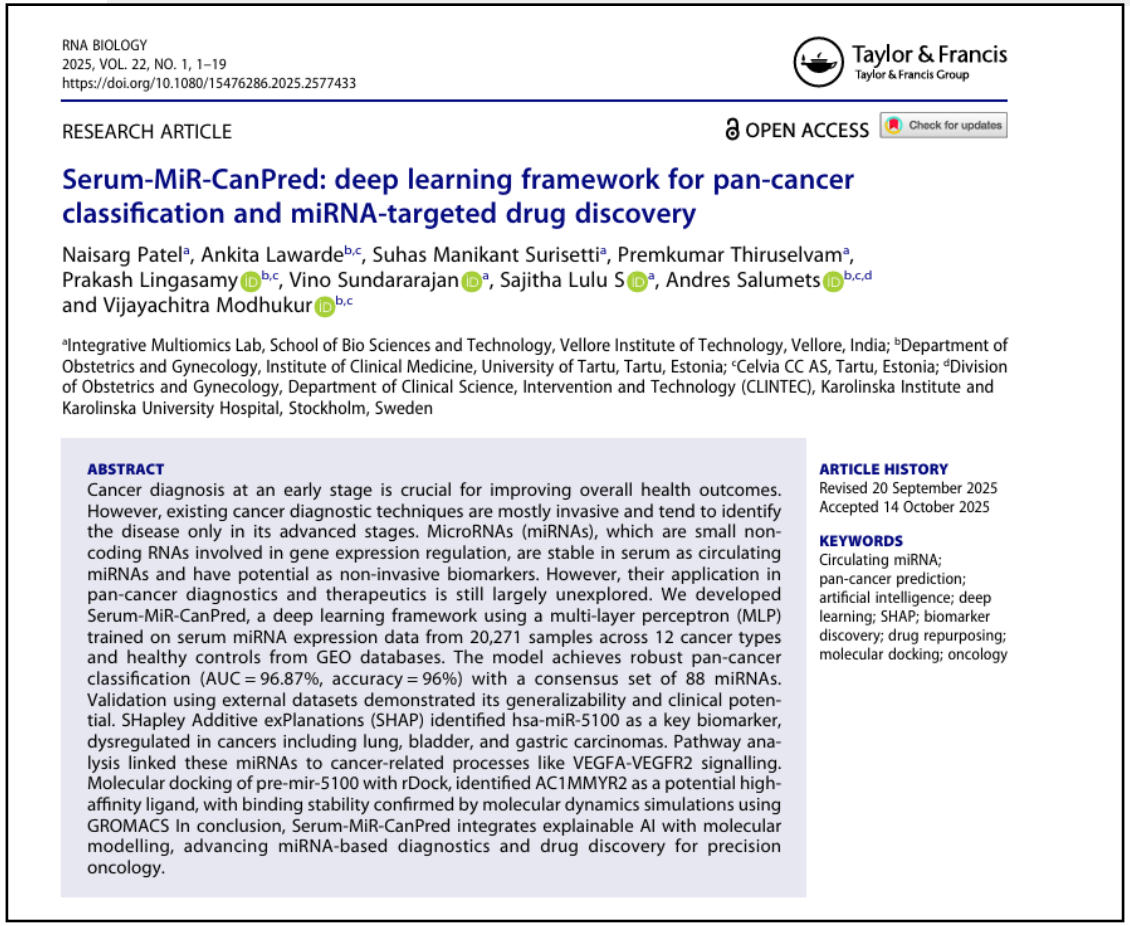
Extract biological meaning

- Seurat
- Scanpy
- Cytoscape
- Heatmaps
- PCA / UMAP
- Networks

Biological Insights

- Biomarkers
- Disease subtypes
- Drug targets
- Pathway insights

Case Study 1: Serum-MiR-CanPred: deep learning framework for pan-cancer classification and miRNA-targeted drug discovery [RNA Biology]



RNA BIOLOGY
2025, VOL. 22, NO. 1, 1–19
https://doi.org/10.1080/15476286.2025.2577433

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Serum-MiR-CanPred: deep learning framework for pan-cancer classification and miRNA-targeted drug discovery

Naisarg Patel[a], Ankita Lawarde[b,c], Suhas Manikant Surisetti[a], Premkumar Thiruselvam[a], Prakash Lingasamy[b,c], Vino Sundararajan[a], Sajitha Lulu S[a], Andres Salumets[b,c,d] and Vijayachitra Modhukur[b,c]

[a]Integrative Multiomics Lab, School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore, India; [b]Department of Obstetrics and Gynecology, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia; [c]Celvia CC AS, Tartu, Estonia; [d]Division of Obstetrics and Gynecology, Department of Clinical Science, Intervention and Technology (CLINTEC), Karolinska Institute and Karolinska University Hospital, Stockholm, Sweden

ABSTRACT

Cancer diagnosis at an early stage is crucial for improving overall health outcomes. However, existing cancer diagnostic techniques are mostly invasive and tend to identify the disease only in its advanced stages. MicroRNAs (miRNAs), which are small non-coding RNAs involved in gene expression regulation, are stable in serum as circulating miRNAs and have potential as non-invasive biomarkers. However, their application in pan-cancer diagnostics and therapeutics is still largely unexplored. We developed Serum-MiR-CanPred, a deep learning framework using a multi-layer perceptron (MLP) trained on serum miRNA expression data from 20,271 samples across 12 cancer types and healthy controls from GEO databases. The model achieves robust pan-cancer classification (AUC = 96.87%, accuracy = 96%) with a consensus set of 88 miRNAs. Validation using external datasets demonstrated its generalizability and clinical potential. SHapley Additive exPlanations (SHAP) identified hsa-miR-5100 as a key biomarker, dysregulated in cancers including lung, bladder, and gastric carcinomas. Pathway analysis linked these miRNAs to cancer-related processes like VEGFA-VEGFR2 signalling. Molecular docking of pre-mir-5100 with rDock, identified AC1MMYR2 as a potential high-affinity ligand, with binding stability confirmed by molecular dynamics simulations using GROMACS In conclusion, Serum-MiR-CanPred integrates explainable AI with molecular modelling, advancing miRNA-based diagnostics and drug discovery for precision oncology.

ARTICLE HISTORY
Revised 20 September 2025
Accepted 14 October 2025

KEYWORDS
Circulating miRNA; pan-cancer prediction; artificial intelligence; deep learning; SHAP; biomarker discovery; drug repurposing; molecular docking; oncology

Serum-MiR-CanPred

- Non-invasive pan-cancer detection
- Uses circulating miRNA (liquid biopsy)
- Combines:
  - Deep Learning
  - Explainable AI
  - Drug discovery
- From diagnosis → mechanism → therapy

Datasets Used

GEO datasets:

- GSE113740
- GSE164174
- GSE211692
- GSE212211

Scale

- 20,271 samples
- 12 cancer types + healthy controls

Data Type

Serum miRNA expression (microarray)

Data Integration Strategy

Horizontal integration

- Preprocessing
- Batch effect correction → PyComBat
- Class imbalance → SMOTE
- Started: 1630 miRNAs
- Selected: 88 miRNAs (Consensus Feature Set)

Machine Learning & Deep Learning Models

ML: kNN ,SVM & Random Forest

DL: MLP & CNN

Model Performance

- Accuracy: **96%**
- AUC: **96.87%**
- External validation AUC: **~99%**

Cross-validation

- Stable across folds (~95%)

Interpretation:

- Model generalizes across datasets → clinically promising

Explainable AI (SHAP)

- Makes DL interpretable
- Identifies top miRNA biomarkers
- Cancer-specific feature importance
- hsa-miR-5100 → top biomarker
- Moves from "black box" → "biological insight

Drug Discovery Integration

Target: miR-5100

Methods:

•Molecular docking → rDock

•Molecular dynamics → GROMACS

Key Result:

Drug candidate: AC1MMYR2

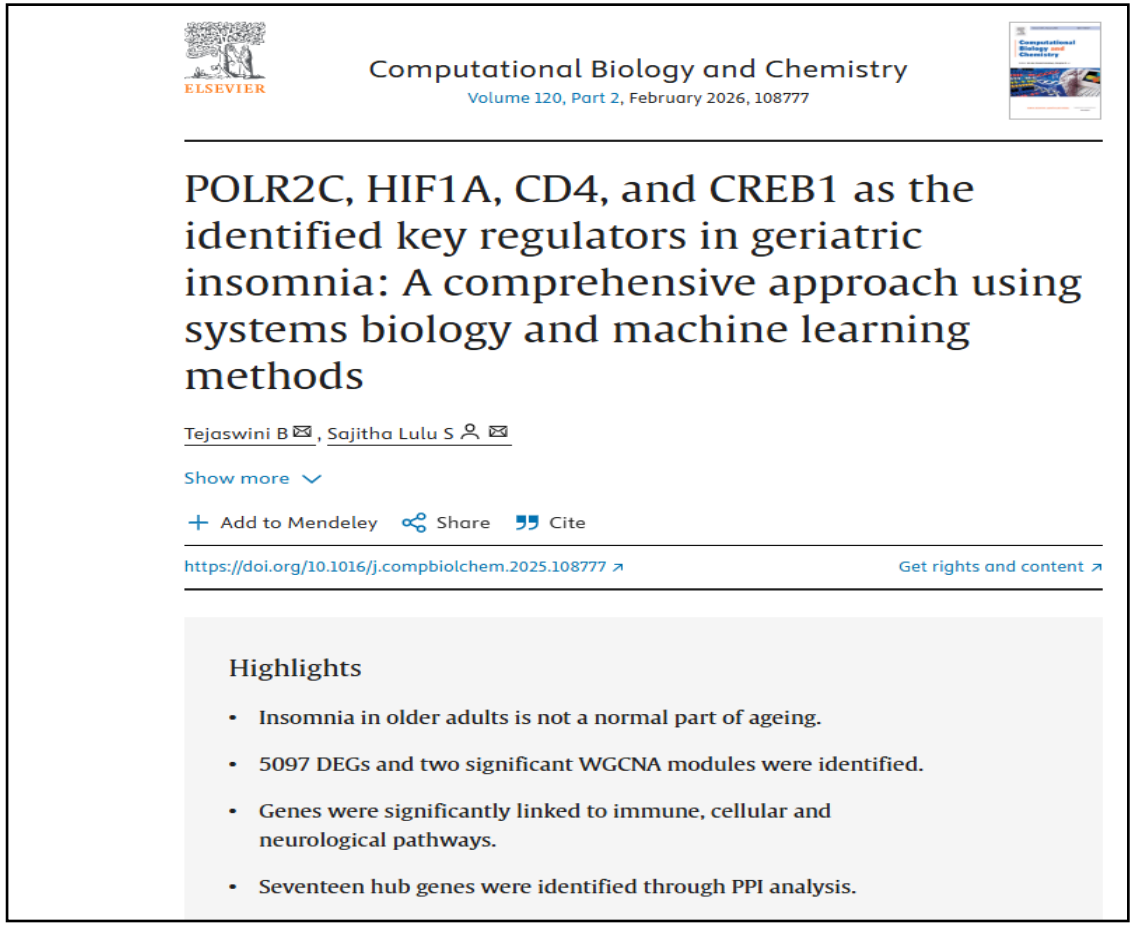
- Strong binding
- Stable interaction

AI → biomarker → drug → full pipeline

Biological Outputs

- **Functional Analysis**
- **Key Pathways:**
- **Visualization:**
- Heatmaps → expression patterns
- PCA / t-SNE → cancer vs healthy separation
- Cancer types share signatures → explains overlap in clustering

Case Study 2



ELSEVIER

Computational Biology and Chemistry

Volume 120, Part 2, February 2026, 108777

POLR2C, HIF1A, CD4, and CREB1 as the identified key regulators in geriatric insomnia: A comprehensive approach using systems biology and machine learning methods

Tejaswini B ✉, Sajitha Lulu S ✉

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.compbiolchem.2025.108777 ↗ Get rights and content ↗

Highlights

- Insomnia in older adults is not a normal part of ageing.
- 5097 DEGs and two significant WGCNA modules were identified.
- Genes were significantly linked to immune, cellular and neurological pathways.
- Seventeen hub genes were identified through PPI analysis.

- Input dataset: GEO-GSE208668
- Parallel analysis branches: DEGs, WGCNA, GSEA
- Functional interpretation: GO and KEGG
- PPI network for hub gene identification
- ML block with LASSO, SVM-RFE, and Random Forest
- Final molecular signature
- Validation using immune infiltration, GO/KEGG, regulatory pathways, and ageing database comparison

Case Study 3:

Heliyon 11 (2025) e41342

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Mapping integral cell-type-specific interferon-induced gene regulatory networks (GRNs) involved in systemic lupus erythematosus using systems and computational analysis

Blessy Kiruba [a,1], Akshayata Naidu [b,1], Vino Sundararajan [c], Sajitha Lulu S [b,*]

[a] Department of Biosciences, School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore, 632 014, Tamil Nadu, India
[b] Department of Biotechnology, School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore, 632 014, Tamil Nadu, India
[c] Department of Biosciences, School of Bio Sciences and Technology, Vellore Institute of Technology, Vellore, 632 014, Tamil Nadu, India

ARTICLE INFO

Keywords:
Systemic lupus erythematosus
Gene regulatory networks
Non-coding RNAs
Interferons
Targets

ABSTRACT

Systemic lupus erythematosus (SLE) is a systemic autoimmune disorder characterized by the production of autoantibodies, resulting in inflammation and organ damage. Although extensive research has been conducted on SLE pathogenesis, a comprehensive understanding of its molecular landscape in different cell types has not been achieved. This study uncovers the molecular mechanisms of the disease by thoroughly examining gene regulatory networks within neutro-

Bulk RNA-seq from GEO [Cell Specific Datasets]
The study integrates multiple cell-type-specific datasets

Cell type	GEO ID	Samples
Neutrophils	GSE153781	6 SLE + 6 controls
Dendritic cells	GSE136731	17 SLE + 11 controls
T cells	GSE97263	13 SLE + 14 controls
B cells	GSE118254	48 SLE + 37 controls

Validation dataset

- Single-cell RNA-seq (scRNA-seq) from GEO

Multi-layer biological integration [Vertical]

- DEG + ncRNA identification
- Functional enrichment (GO + KEGG)
- PPI network construction (STRING + Cytoscape)
- Hub gene detection (CytoHubba – MCC algorithm)
- TF + miRNA regulatory network (TRRUST + NetworkAnalyst)

Outcome

Therapeutic targets

•Explained cell-type-specific disease mechanisms

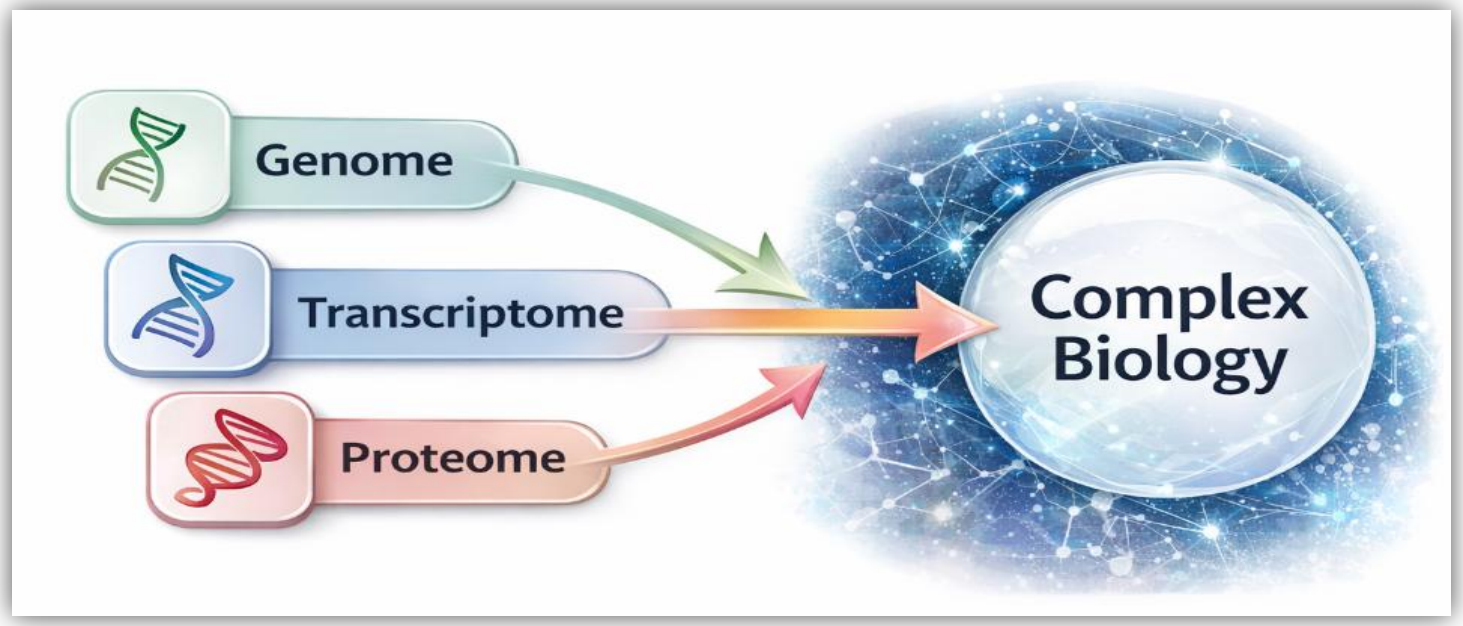
•Linked gene expression → immune pathways → disease phenotype

Multi-omics data and Biological Complexity

Biology is multi-layered, dynamic, and interconnected

- Single-omics methods oversimplify biology
- Disease is driven by networks of interactions

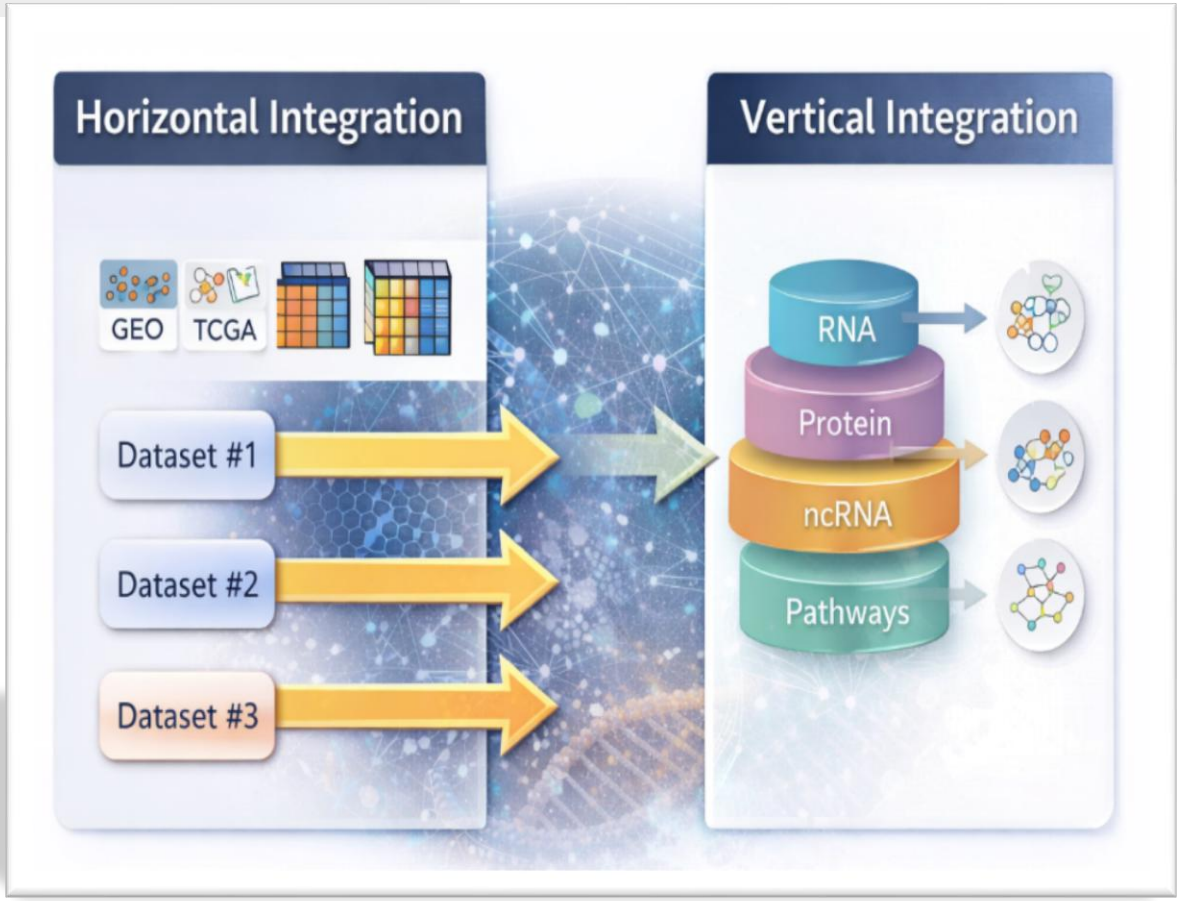
Single-omics gives an incomplete picture of the biological system



Data Integration – The Turning Point

- **Horizontal Integration**
- Combines datasets
- Larger sample size
- Robustness
- **Vertical Integration**
- Combines omics layers
- Pathways
- Mechanisms

Integration strengthens deeper biological insights



AI + Multi-Omics = Transformational Power

Deep learning models deliver high accuracy

- SHAP turns AI into a discovery engine, not just a black box

Pipeline:

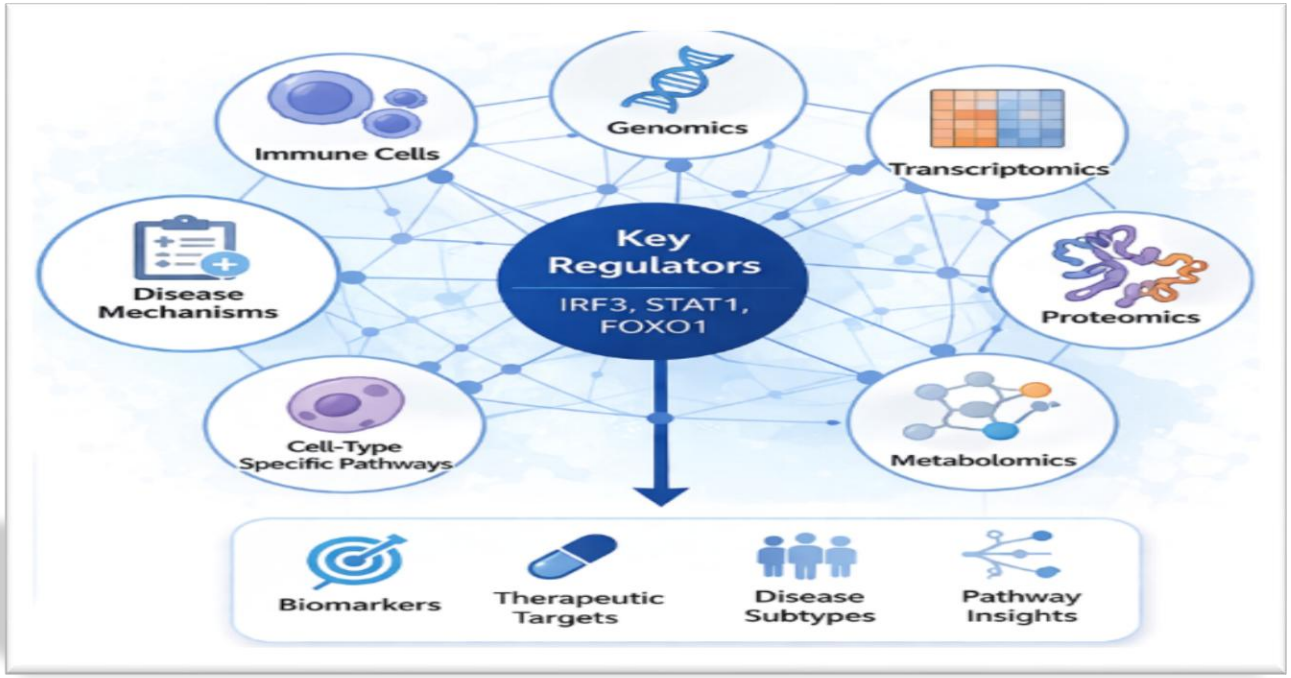
AI → Biomarkers → Mechanisms → Drug Discovery

To truly understand disease, we must move from isolated data → integrated biological systems

Systems Biology Reveals True Disease Mechanisms

- Integration of omics + networks identifies key regulators
- Reveals cell-type-specific pathways driving disease
- Explains immune regulation and therapeutic targets

Diseases are network-driven and only systems-level integration can decode them



ssajithalulu@vit.ac.in