Lecture 12:

Multi-Modal Data Integration

Integrative biology

Systems medicine

Holistic approaches

Introduction to Biomedical Data Science

Lecture Contents

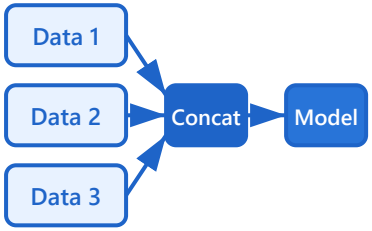
Part 1/3:

Integration Methods

- Mathematical frameworks
- Computational approaches
- Evaluation metrics

Early vs Late Fusion

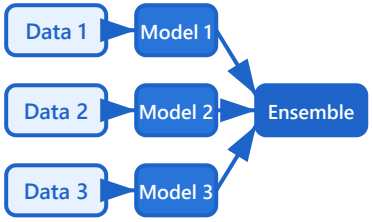
Early Fusion



Feature-level concatenation

- ✓ Captures feature interactions
- ✓ Joint representation learning
- ✗ Computationally expensive

Late Fusion



Decision-level combination

- ✓ Flexible and modular
- ✓ Independent training
- ✗ Misses feature interactions

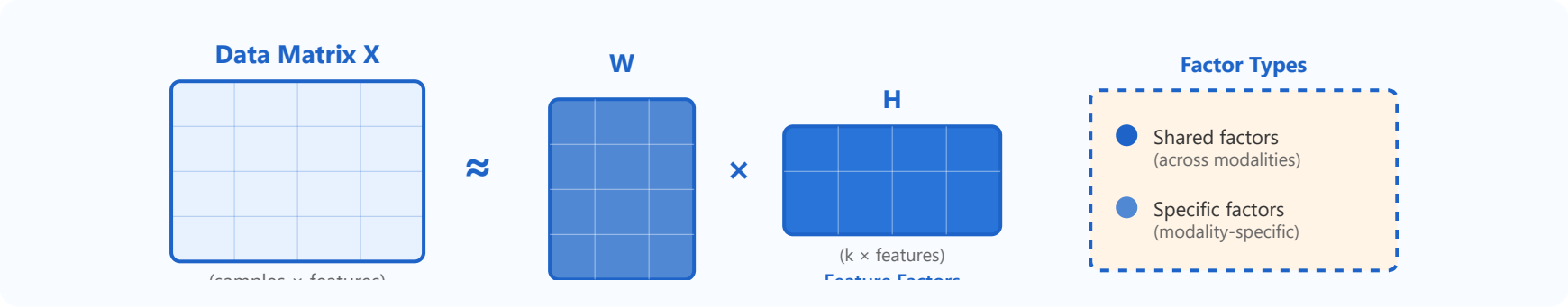
Intermediate Fusion

Combines features at intermediate layers, balancing both approaches

Trade-offs

Early: captures interactions but computationally expensive | Late: flexible but misses feature interactions

Matrix Factorization for Multi-Omics Integration



NMF Methods

Non-negative Matrix Factorization for parts-based representation

Joint NMF

Simultaneous factorization of multiple data matrices

iCluster

Integrative clustering of multiple cancer genomic data types

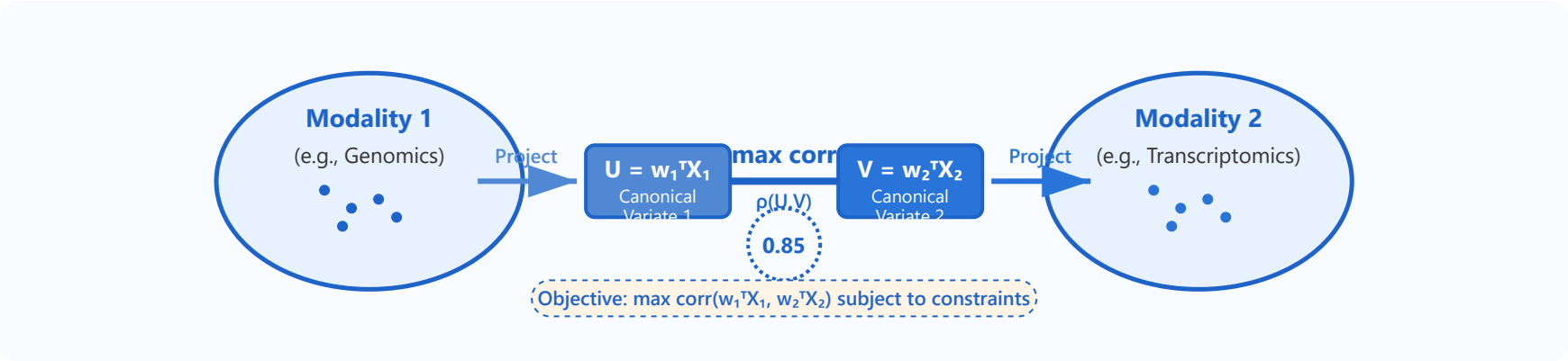
Integrative NMF

Shared and data-specific factors

Interpretation

Biological meaning of latent factors

Canonical Correlation Analysis (CCA)



CCA Principles

Finding linear combinations with maximum correlation

Sparse CCA

Feature selection through sparsity constraints

Kernel CCA

Nonlinear relationships via kernel methods

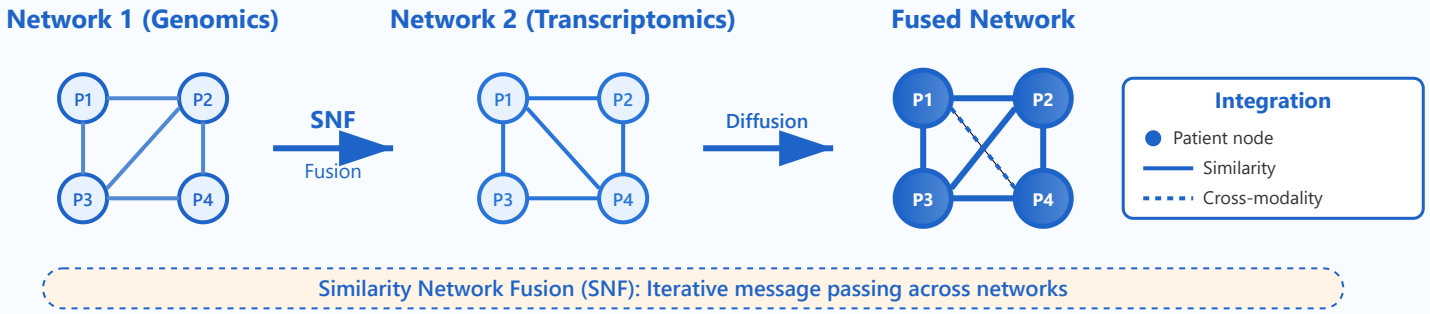
Deep CCA

Neural network-based correlation learning

Multi-view CCA

Extension to more than two data modalities

Graph-based Integration



Similarity Networks

Patient or feature similarity graphs

Network Fusion

SNF: fusing multiple similarity networks

Random Walk

Diffusion-based integration on networks

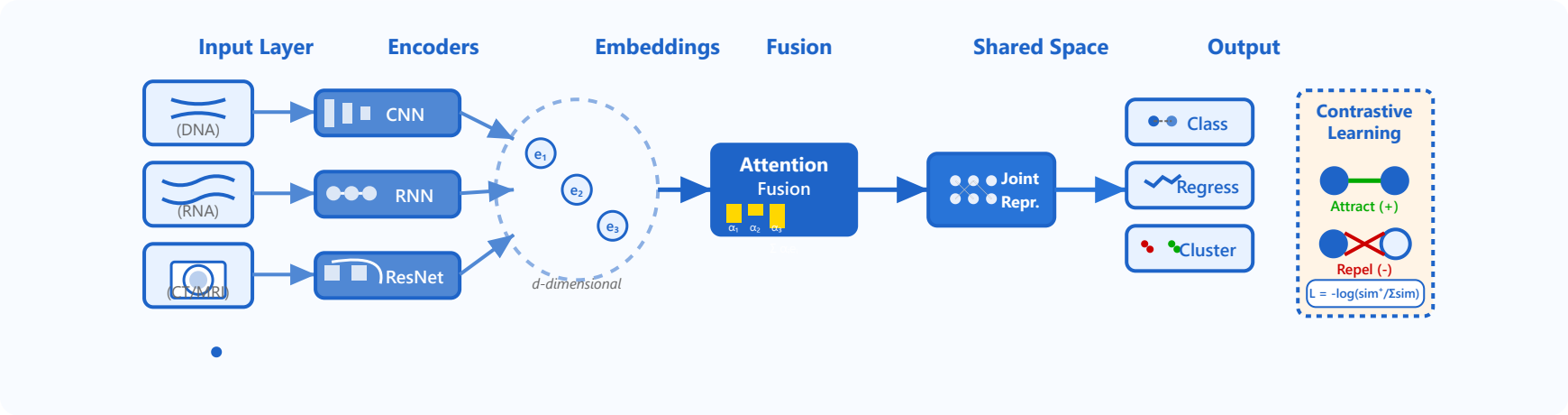
Graph Neural Networks

Deep learning on graph-structured data

Multiplex Networks

Multi-layer network representations

Deep Learning Fusion Strategies



Multi-modal Architectures

Parallel networks for different modalities

Shared Representations

Common latent space across modalities

Cross-modal Attention

Attending to relevant features across data types

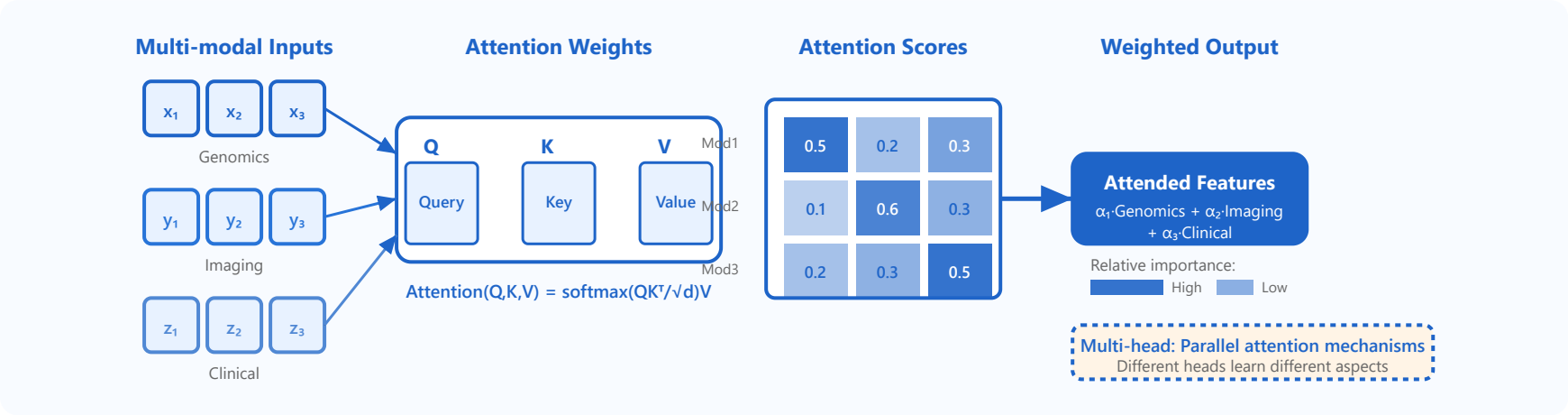
Contrastive Learning

Learning by contrasting positive and negative pairs

Autoencoder Fusion

Reconstruction-based integration

Attention-based Integration



Self-attention Fusion

Learning importance weights within modalities

Cross-attention

Attention between different data modalities

Multi-head Attention

Multiple attention perspectives simultaneously

Hierarchical Attention

Feature and sample-level attention

Interpretability

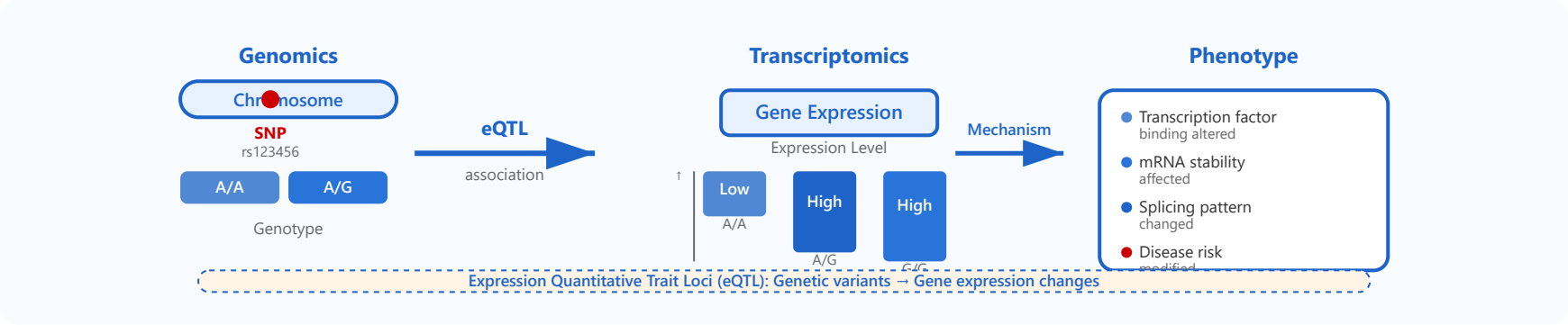
Attention weights as biological insights

Part 2/3:

Multi-Omics Applications

- Biological integration
- Technical challenges
- Analysis workflows

Genomics + Transcriptomics Integration



eQTL Analysis

Expression quantitative trait loci mapping

ASE Detection

Allele-specific expression patterns

Splicing QTLs

Genetic variants affecting RNA splicing

Regulatory Variants

Non-coding variants and gene expression

Allele-specific Binding

Transcription factor binding affected by SNPs

Proteogenomics

Variant Peptides

Protein sequences from genomic variants

Novel ORFs

Discovering new protein-coding regions

PTM Sites

Post-translational modification mapping

Protein Isoforms

Alternative splicing products in proteomics

Neo-antigens

Tumor-specific antigens for immunotherapy

Imaging-genomics (Radiogenomics)

Radiogenomics

Linking imaging phenotypes to genotypes

Imaging Features

Quantitative features from medical images

Genetic Associations

GWAS-style analysis with imaging

Outcome Prediction

Combining imaging and genomics for prognosis

Treatment Response

Predicting therapy efficacy

Clinical + Molecular Data Integration

EHR Integration

Electronic health records with omics data

Lab Values

Clinical laboratory measurements

Imaging Reports

Radiology and pathology findings

Molecular Profiles

Genomic, transcriptomic, proteomic data

Temporal Alignment

Synchronizing time-series clinical and molecular data

Temporal Integration

Longitudinal Designs

Repeated measurements over time

Time Series Alignment

Synchronizing different measurement schedules

Dynamic Modeling

Capturing temporal dynamics

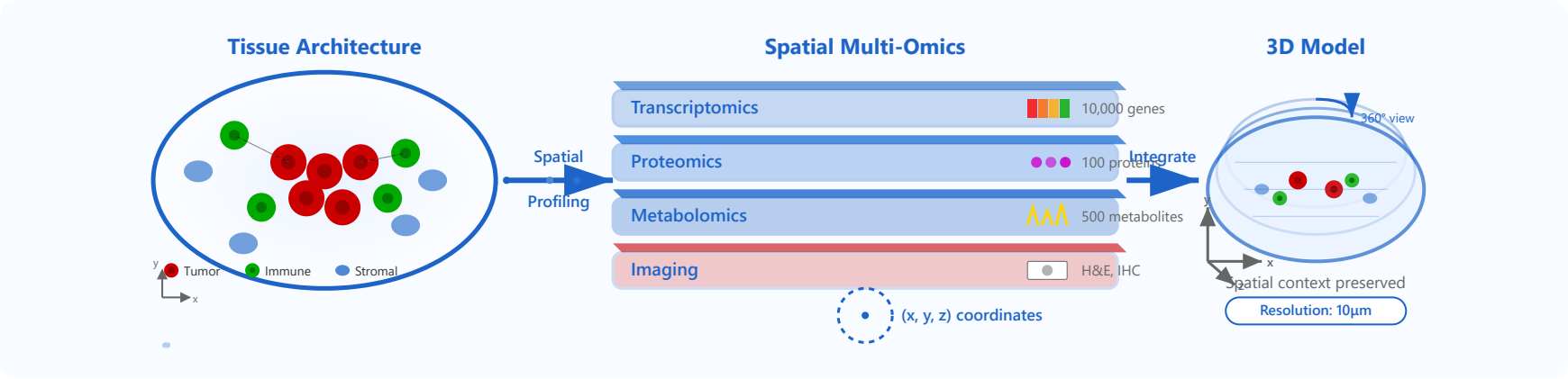
State Transitions

Disease progression and treatment response

Trajectory Inference

Reconstructing continuous processes

Spatial Multi-Omics Integration



Spatial Omics

Spatial transcriptomics, proteomics, metabolomics

Image Registration

Aligning multi-modal spatial data

Cellular Neighborhoods

Microenvironment characterization

Tissue Architecture

Structural organization patterns

3D Reconstruction

Building 3D tissue models

Part 3/3:

Clinical Applications and Future Directions

- Disease understanding
- Clinical translation
- Future directions

Disease Subtyping

Molecular Subtypes

Identifying disease subtypes from multi-omics

Clinical Correlates

Linking subtypes to outcomes

Consensus Clustering

Robust subtype identification

Stability Analysis

Assessing subtype reproducibility

Validation Cohorts

Independent validation of subtypes

Prognosis Prediction

Multi-modal Signatures

Prognostic signatures from integrated data

Risk Stratification

Identifying high-risk patients

Survival Models

Cox regression and deep survival models

Time-dependent ROC

Evaluating time-to-event predictions

Clinical Utility

Decision curve analysis

Drug Response Prediction

Sensitivity Prediction

Predicting drug effectiveness

Resistance Markers

Identifying resistance mechanisms

Combination Effects

Drug synergy and antagonism

Pharmacogenomics

Genetic variants affecting drug response

Clinical Trials

Integration in precision medicine trials

Biomarker Panels

Multi-analyte Tests

Combining multiple biomarkers

Optimal Combinations

Feature selection for panels

Performance Metrics

Sensitivity, specificity, PPV, NPV

Cost-benefit

Clinical and economic considerations

Regulatory Approval

FDA/EMA approval pathways

Systems Medicine

Network Medicine

Disease as network perturbations

Disease Modules

Interconnected disease components

Comorbidities

Shared molecular mechanisms

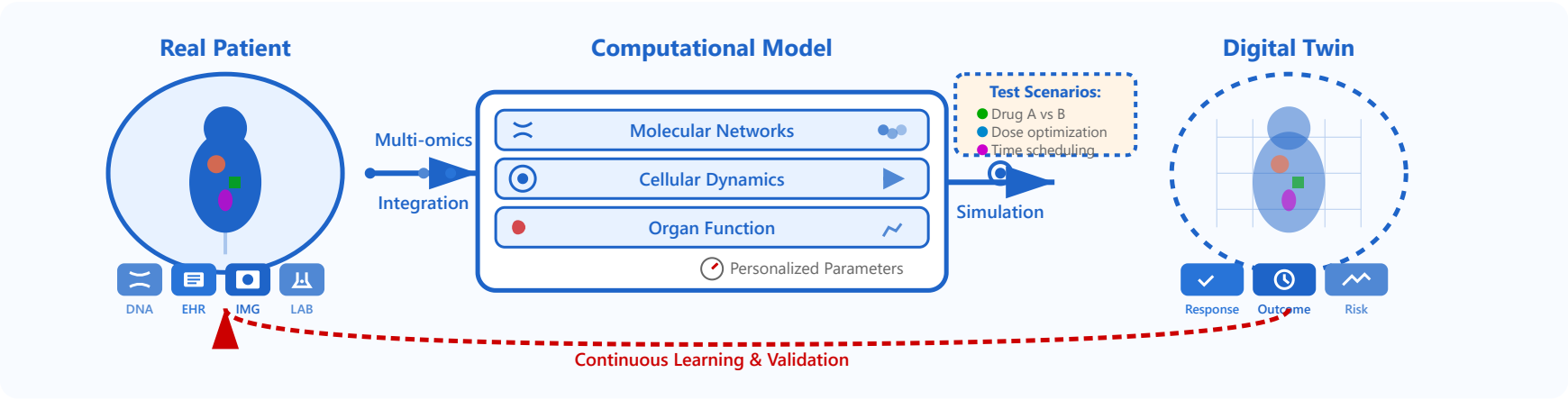
Drug Repurposing

Network-based drug discovery

Personalized Networks

Patient-specific network models

Digital Twins in Medicine



Patient Models

Computational patient representations

Simulation Frameworks

In silico clinical trials

Parameter Estimation

Personalizing model parameters

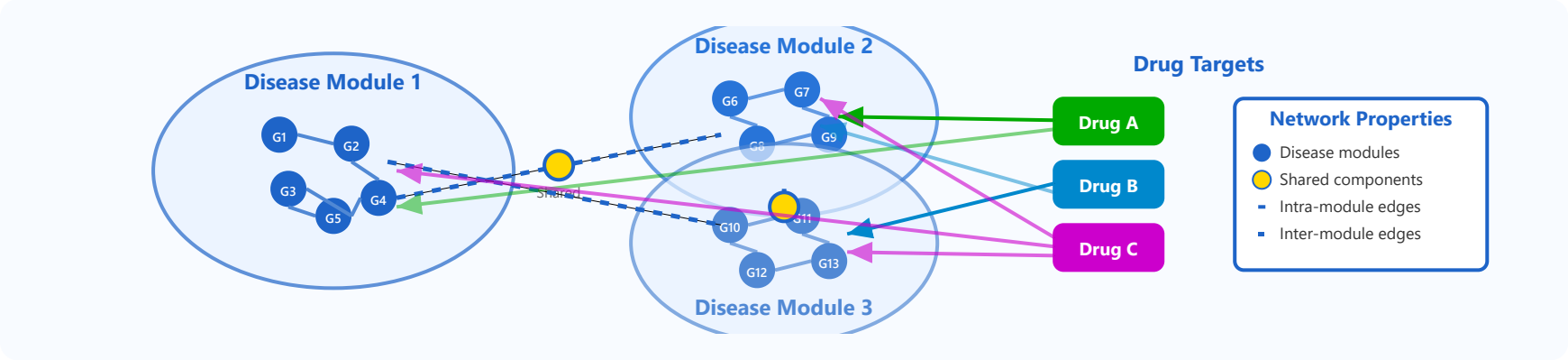
Treatment Optimization

Simulating treatment strategies

Validation Approaches

Comparing predictions to reality

Network Medicine



Disease Networks

Molecular interaction networks in disease

Interactome

Protein-protein interaction networks

Disease-disease Relationships

Shared pathways and comorbidities

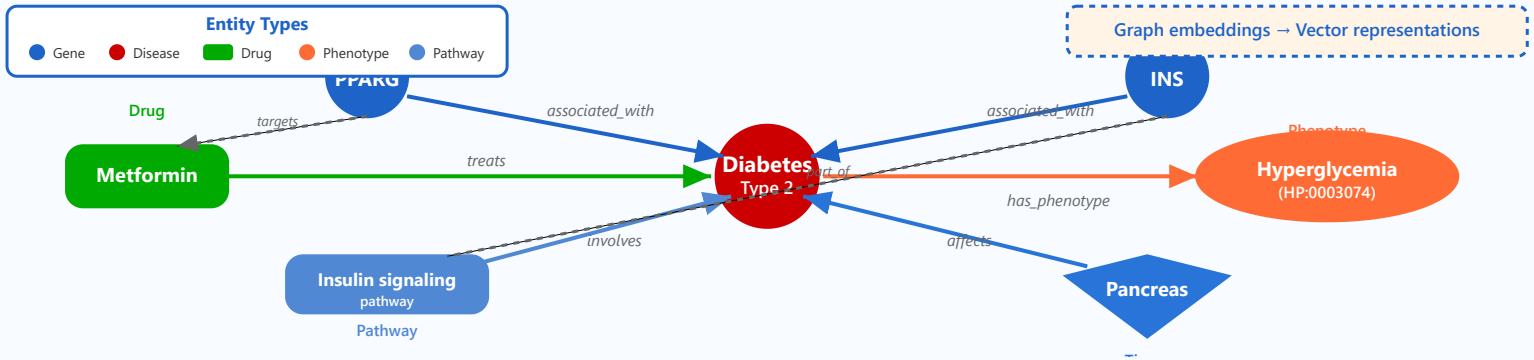
Drug-target Networks

Polypharmacology and off-targets

Network Pharmacology

Systems-level drug discovery

Biomedical Knowledge Graphs



Biomedical Ontologies

Structured vocabularies (GO, HPO, etc.)

Entity Relationships

Genes, diseases, drugs, phenotypes

Graph Embeddings

Learning representations from graphs

Link Prediction

Discovering new relationships

Query Systems

Biological question answering

Case Studies

TCGA Pan-cancer

The Cancer Genome Atlas multi-omics integration

METABRIC

Molecular taxonomy of breast cancer

LINCS

Library of Integrated Network-based Cellular Signatures

HuBMAP

Human BioMolecular Atlas Program

Clinical Examples

Real-world clinical integration

Challenges in Multi-Modal Integration

Missing Data

Incomplete measurements across modalities

Batch Effects

Technical variation across platforms

Scale Differences

Different measurement scales and distributions

Interpretability

Understanding integrated models

Validation

Reproducibility and generalization

Hands-on: MOFA (Multi-Omics Factor Analysis)

Data Preparation

Formatting multi-omics datasets

Model Training

Running MOFA analysis

Factor Interpretation

Understanding learned factors

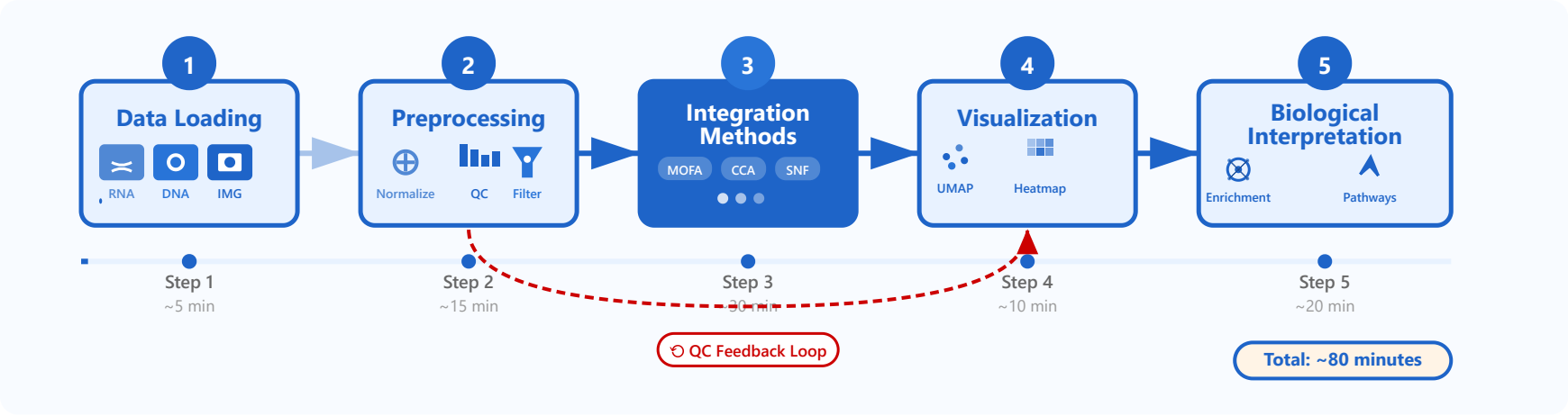
Variance Decomposition

Attributing variance to factors

Downstream Analysis

Using factors for prediction

Hands-on: Integration Workflow



Data Loading

Reading multi-modal datasets

Preprocessing

Normalization and quality control

Integration Methods

Applying different integration approaches

Visualization

UMAP, t-SNE, heatmaps

Biological Interpretation

Functional enrichment analysis

Thank you

Emerging methods in multi-modal integration

Clinical impact and translational opportunities

Research opportunities in systems medicine

Introduction to Biomedical Data Science