

Integrative Analysis Extra Content: PARADIGM

Lab 4, Pathway and Network Analysis 2026

Duration: 1 hour

Format: Reading

Prerequisites: PNAS Modules 1-4

What is PARADIGM?

PARADIGM (PATHway Recognition Algorithm using Data Integration on Genomic Models, Vaske et al., 2010) is a probabilistic graphical model framework for integrating multiple types of genomic data with pathway knowledge to infer patient-specific pathway activities. It became the standard multi-omics integration method in the TCGA pipeline and was used across all 33 cancer types in the Pan-Cancer Atlas.

The Problem PARADIGM Solves

Consider a gene like *MYC* in a tumor sample:

- It might be amplified (CNA data shows gain)
- But not highly expressed (RNA-seq shows moderate levels)
- And its downstream targets are activated (target gene expression is high)

What is *MYC*'s functional activity in this tumor? Traditional analysis gives us three separate, potentially conflicting signals. PARADIGM integrates them with pathway structure to infer a single, coherent Inferred Pathway Level (IPL) that represents *MYC*'s functional state.

How PARADIGM Works: The Factor Graph

PARADIGM uses a factor graph, a type of probabilistic graphical model, that encodes three types of information:

1. Pathway Structure
From curated databases (NCI-PID, Reactome, KEGG) describing
 - Which genes regulate which
 - Activation vs. inhibition relationships
 - Protein complexes and their components
2. Multi-Omic Observations
For each patient, we might have
 - Copy number (is the gene amplified/deleted?)
 - Expression (is it transcribed?)
 - Or other molecular profiles such as methylation, protein levels, *etc*
3. Probabilistic Relationships
Paradigm encodes conditional probabilities representing how data at one node predicts data at another node, such as
 - DNA copy number → expected mRNA level
 - Regulator activity → target gene expression
 - Complex member states → complex activity

PARADIGM uses Belief Propagation to infer the most likely activity state of every pathway entity (genes, complexes, abstract processes) given all observed data and pathway constraints.

PARADIGM vs. Standard Enrichment Methods

Feature	PARADIGM	Standard GSEA/OR
Input	Multiple data types (CNA + expression + ...)	Single ranked list or gene set
Output	Per-patient, per-pathway-entity activity scores (IPLs)	Pathway enrichment p-values
Unit of analysis	Individual patient	Cohort comparison
Pathway structure	Explicitly modeled (directed edges, activation/inhibition)	Ignored (gene sets are unordered)
Conflicting signals	Resolved probabilistically	Not addressed

The Output: Inferred Pathway Levels (IPLs)

PARADIGM produces an IPL matrix:

- Rows: ~13,000 pathway features (genes, complexes, processes)
- Columns: patients
- Values: continuous activity scores (typically centered around 0)

Positive values indicate activation relative to normal; negative values indicate repression.

This matrix can be used for:

- Unsupervised clustering to find patient subtypes
- Survival analysis to find prognostic pathway signatures
- Differential pathway activity between tumor subtypes
- Correlation with clinical features (stage, grade, mutations)

Example: Resolving Conflicting Signals

Imagine a tumor sample where:

- *CDKN2A* is deleted (CNA = -2)
- *RB1* expression is normal
- But E2F targets are highly expressed

PARADIGM reasons through the pathway:

- *CDKN2A* deletion → loss of p16 protein
- Loss of p16 → CDK4/6 uninhibited
- Active CDK4/6 → *RB1* phosphorylated (inactivated)
- Inactive *RB1* → E2F released
- Active E2F → target genes transcribed

Even though *RB1* expression is normal, PARADIGM infers *RB1* activity is low because the upstream regulator is deleted and downstream targets are activated. This is consistent with *RB1* being functionally inactivated via phosphorylation — reasoning that is impossible with expression-only analysis.

PARADIGM in TCGA

PARADIGM was incorporated into the standard TCGA analysis pipeline and run on all cancer types. Key findings enabled by PARADIGM include:

- Glioblastoma (2010): Identified patient subtypes with distinct pathway activities that correlated with survival
- Breast cancer (2012): Revealed FOXM1 transcription factor network as a key driver of proliferation
- Pan-Cancer (2018): Computed IPLs for 3,531 samples across 12 cancer types, enabling cross-cancer pathway comparisons

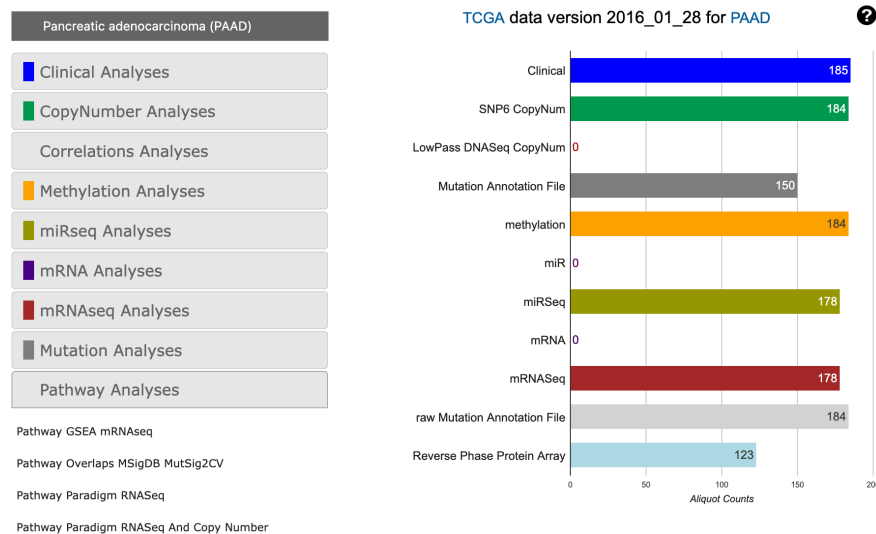
Limitations

- Computational complexity: Requires significant compute time and expertise to run
- Software availability: The original code is difficult to install; no web interface or R package
- Pathway dependence: Results are only as good as the pathway models; poorly annotated pathways give poor inferences
- Interpretation challenges: 13,000 pathway features × hundreds of patients = difficult to interpret without dimensionality reduction

Accessing PARADIGM Results

Since running PARADIGM yourself is challenging, you can access pre-computed results:

- Firebrowse (Broad GDAC): <https://gdac.broadinstitute.org> — has PARADIGM IPLs for all TCGA cancers



- TCGA publications: Supplementary data often includes IPL matrices
- cBioPortal: Some studies include PARADIGM-derived pathway scores

Alternatives and Extensions

Tool	Relationship to PARADIGM
MPAC (2024)	Modernized PARADIGM with improved state discretization and accessibility
Pathifier	Similar goal (patient-specific pathway scores) but different method (PCA-based)
GSVA/ssGSEA	Single-sample enrichment; simpler but doesn't integrate multi-omics or pathway structure
PRECISE	Network-based patient-specific pathway scores from proteomics data

Summary: A Different Philosophy

PARADIGM represents a different philosophy from enrichment-based pathway analysis:

Enrichment (GSEA/ORa)	PARADIGM
"Which pathways are different between groups?"	"What is each pathway's activity in each patient?"
Cohort-level inference	Patient-specific inference
Expression only	Multi-omics integration
Pathways as gene sets	Pathways as causal networks

In this course, you have learned multiple approaches to pathway analysis

Module	Method	Question Answered
DESeq2	Differential expression	Which genes change between conditions?
ORA	Over-representation analysis	Which pathways contain my DE genes?
GSEA	Ranked enrichment	Which pathways show coordinated shifts?
ReactomeFI	Network clustering	Which modules are disrupted by mutations?
GeneMANIA	Network integration	What genes connect my multi-omic results?
PARADIGM	Probabilistic integration	What is each pathway's activity per patient?

Each method answers a different question. While you won't run PARADIGM in this workshop (due to software constraints), understanding its approach helps you appreciate both the power and the challenges of multi-omics pathway integration.