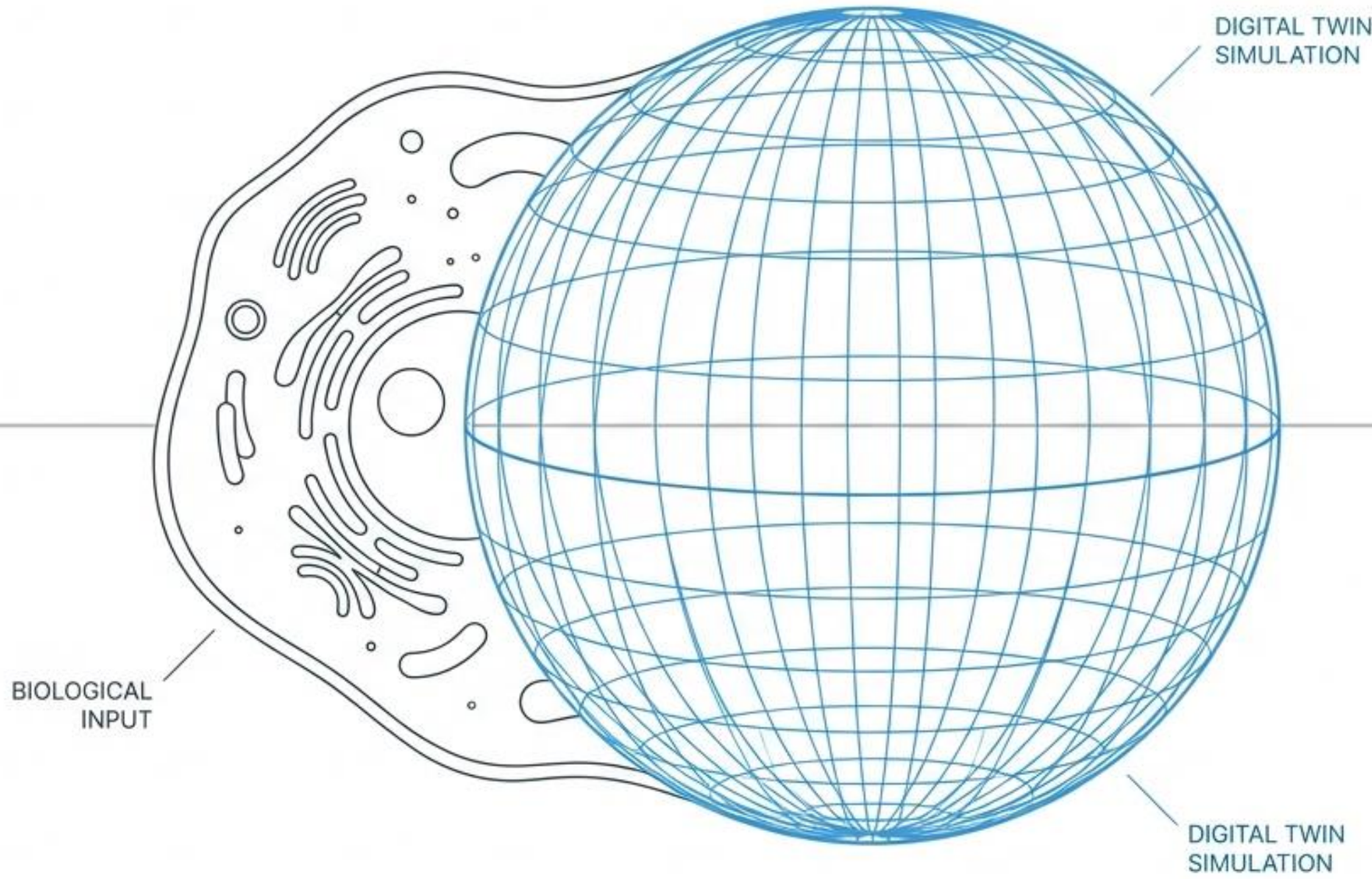
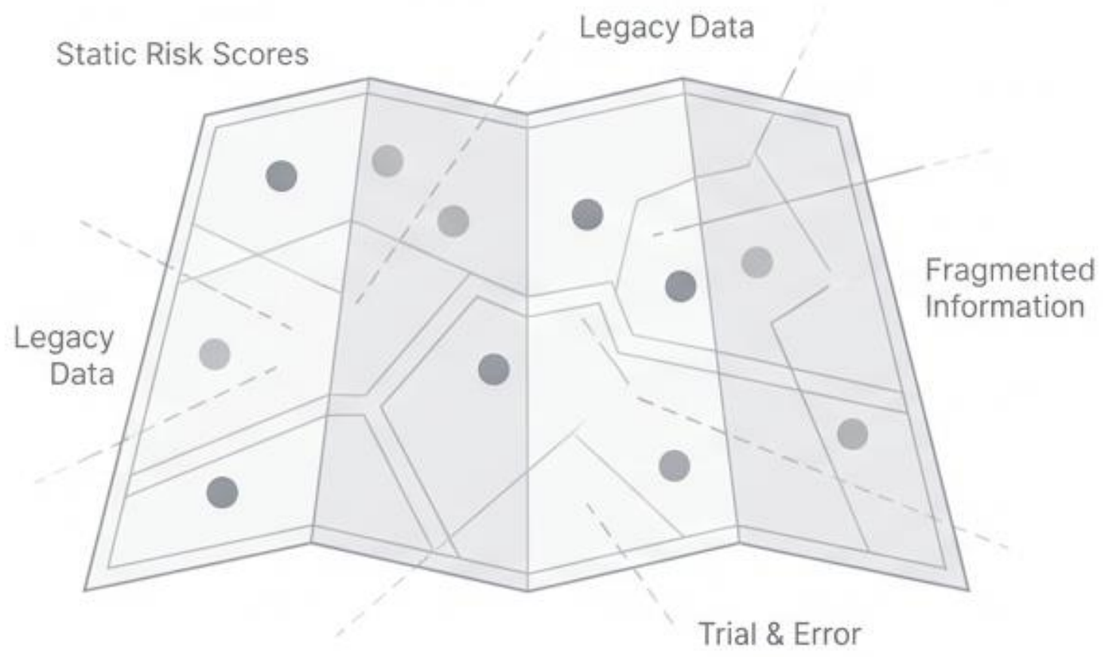
Constructing the Digital Twin

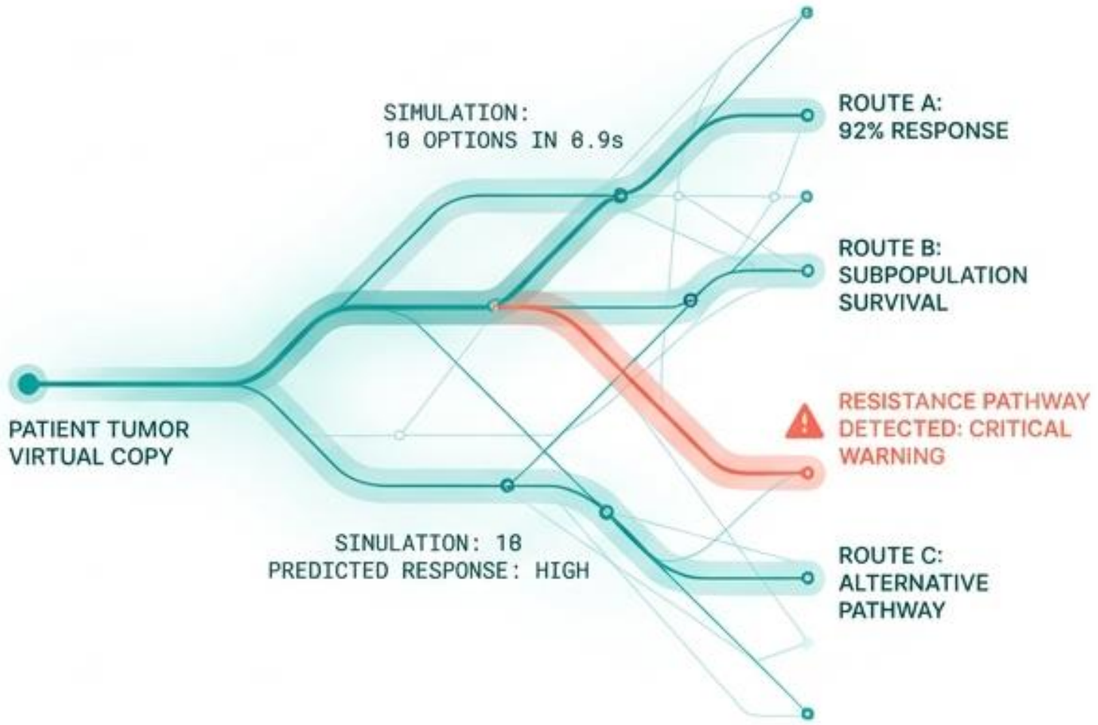
From High-Dimensional Biology to Inspectable Simulation



The shift from statistical guessing to individualized simulation.



Current Standard of Care: Trial and error based on static risk scores and statistical guessing.



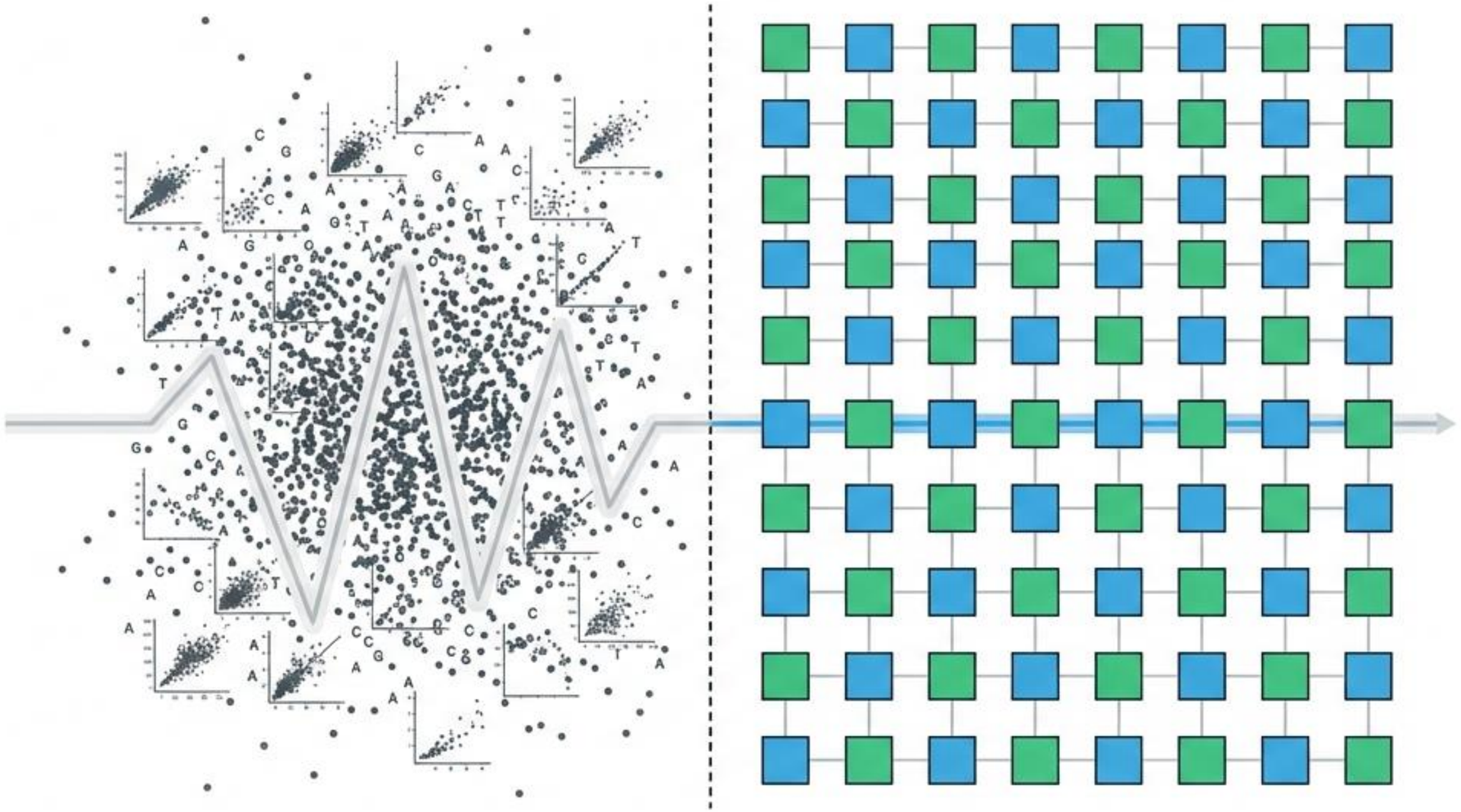
The DNAI Reality: Running treatments on a virtual copy of the patient's tumor to watch responses, resistance, and subpopulation survival.
Testing ten distinct options in under a second.

The Limits of Human Synthesis

Every cancer diagnosis represents a unique, evolving system. A single patient's tumor generates roughly 6,000 individual measurements across four molecular data types.

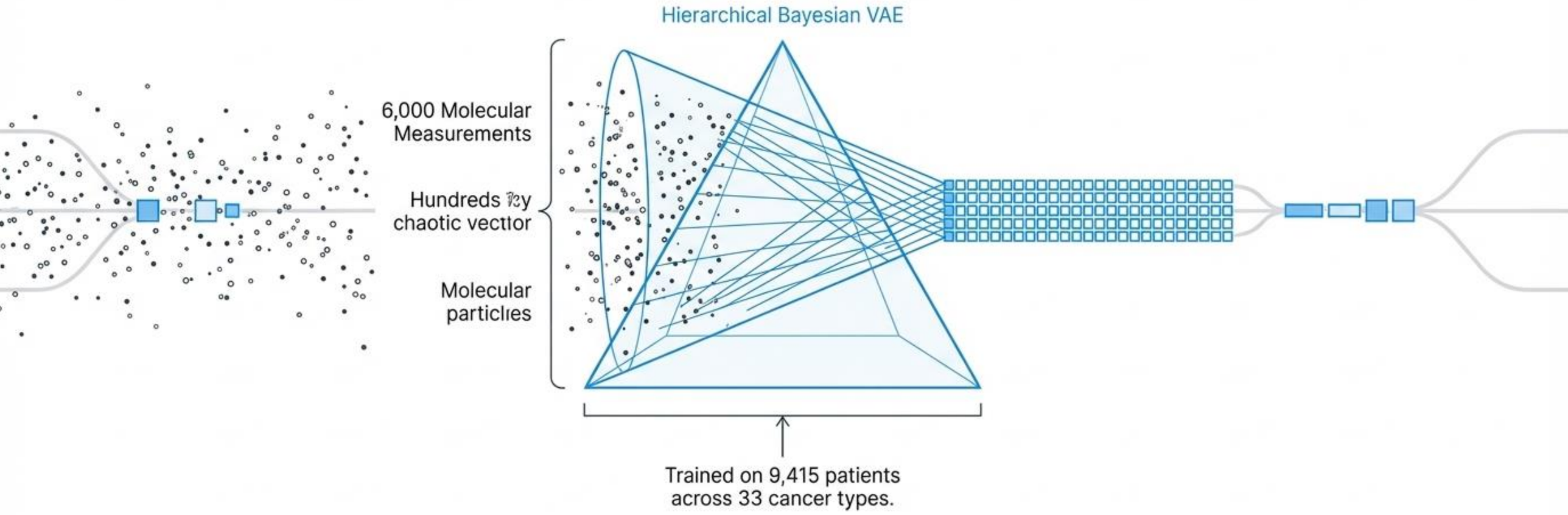
- Gene Expression (activity levels)
- DNA Mutations (broken pathways)
- Copy Number Variation (duplications/deletions)
- Methylation (chemical silencing)

This multidimensional noise is mathematically incompatible with standard clinical decision-making.



Compressing Biology into a Structured Fingerprint

The foundation model distills 6,000 disparate molecular measurements into exactly 328 latent dimensions. Through a Product-of-Experts fusion architecture, it handles missing data gracefully, ensuring similar tumors occupy neighboring mathematical spaces—even if superficial mutation profiles differ completely.



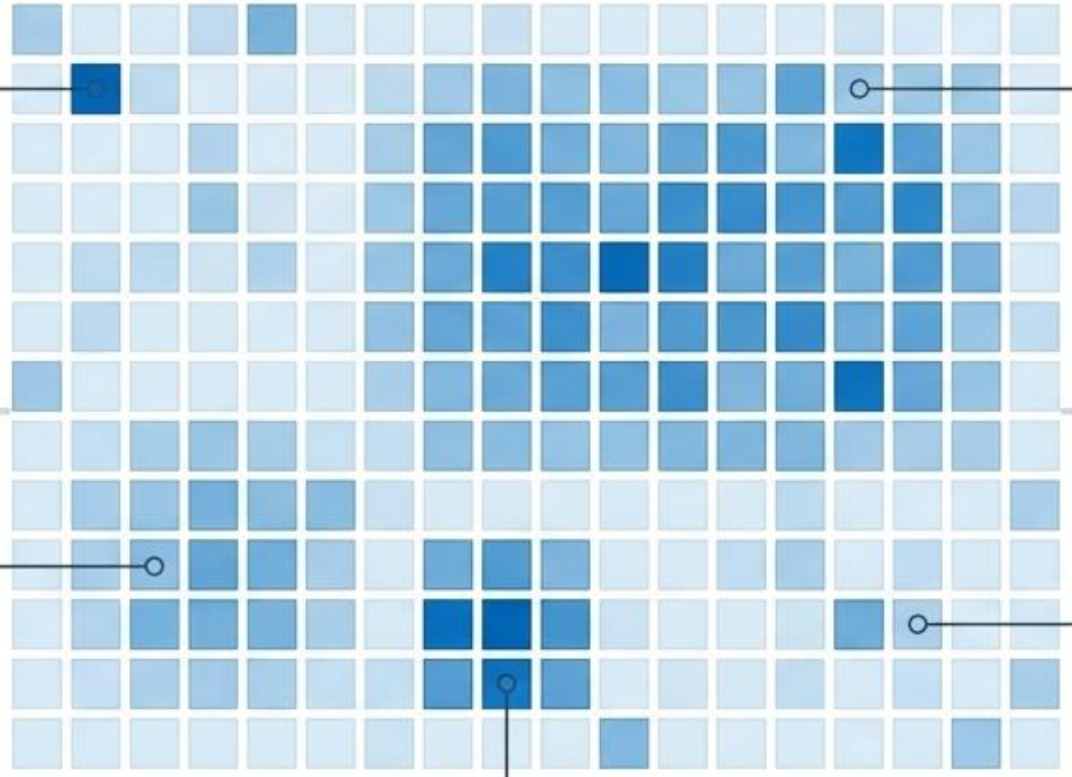
The Anatomy of the 328-Dimensional Fingerprint

Our pipeline transforms chaotic biology into an inspectable simulation. Every prediction chain of biologically named computations—from raw expression to time-resolved trajectories with neatty.

1 Dimension:
Tumor Growth Speed (0.96 correlation with standard proliferation markers).

200 Dimensions:
50 Known Cancer Pathways (e.g., inflammation, DNA repair, metabolic reprogramming—4 dimensions per pathway).

Pipeline Track



48 Dimensions:
Epigenetic Patterns.

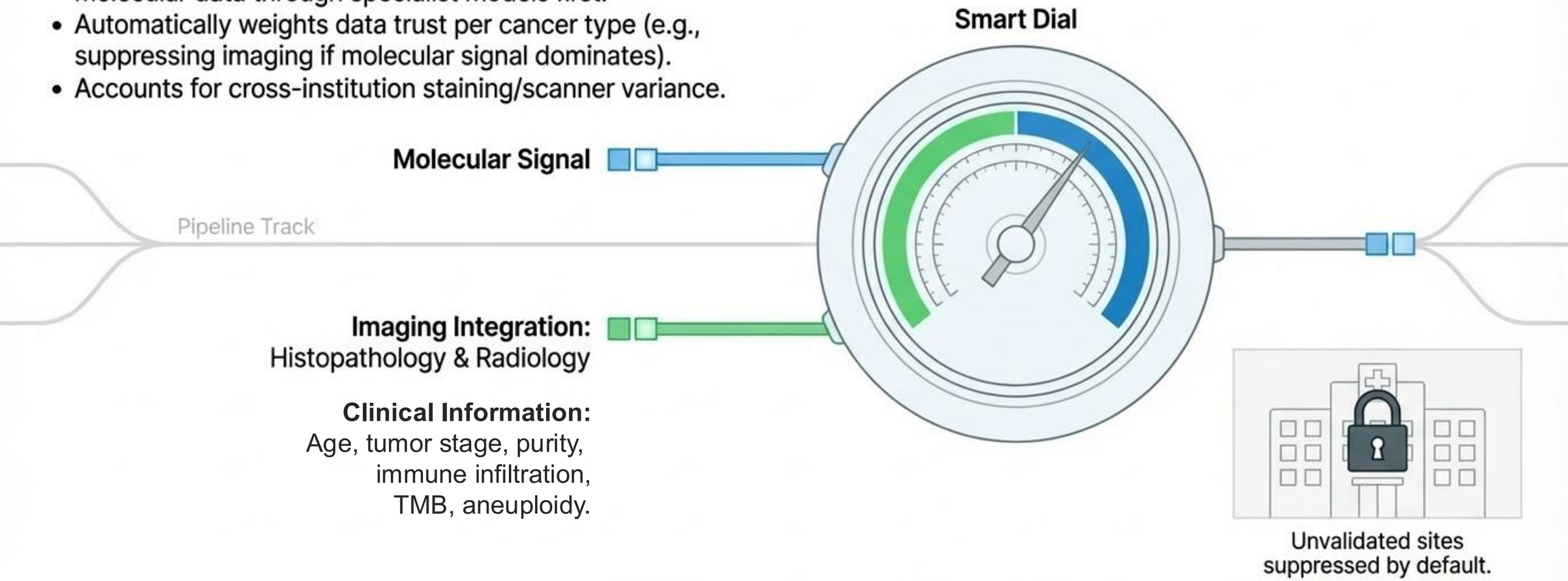
Residual Dimensions:
Broader biological context and unclassified signals.

32 Dimensions:
Chromosomal Instability.

Late Gated Fusion: Adding Eyes to the Tumor

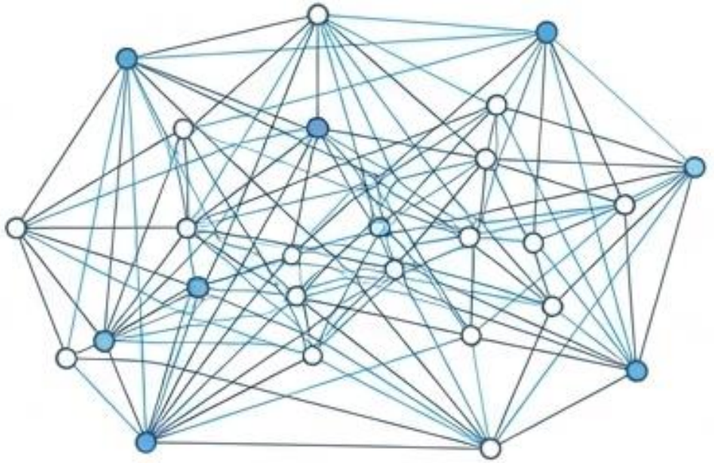
Sequencing cannot reveal physical architecture, immune infiltration, or spatial arrangement.
We integrate whole-slide images and scans using a specialized learned gate.

- Maintains signal integrity by processing imaging and molecular data through specialist models first.
- Automatically weights data trust per cancer type (e.g., suppressing imaging if molecular signal dominates).
- Accounts for cross-institution staining/scanner variance.



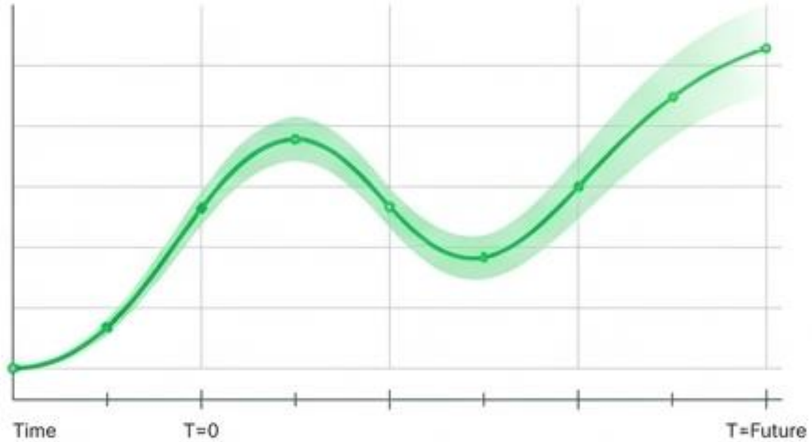
Two Parallel Analysis Paradigms

The Static Path:
What is driving this cancer right now?



Pipeline Track

The Dynamic Path:
How will this tumor evolve over time?



The Mechanistic Evidence Engine: A Fully Traceable Pipeline to Precision Oncology

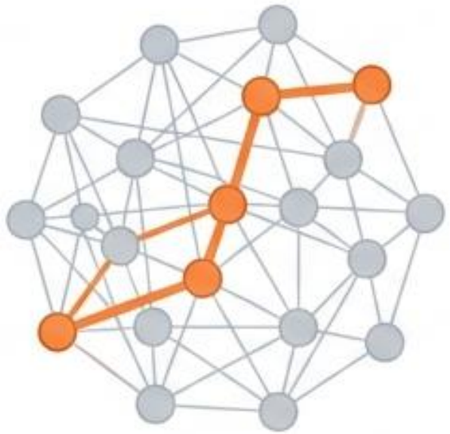
Every recommendation maps directly to specific genes, causal pathway edges, and clinical evidence. The antithesis of a black box.

Phase 1: Driver Identification

[IntOGen: 633 Genes] [COSMIC: 95 Genes]

The Static Path. Patient mutations are matched against established population-level cancer drivers to identify potential targets.

Phase 2: Pathway Activation



[MSigDB Hallmark] [KEGG]

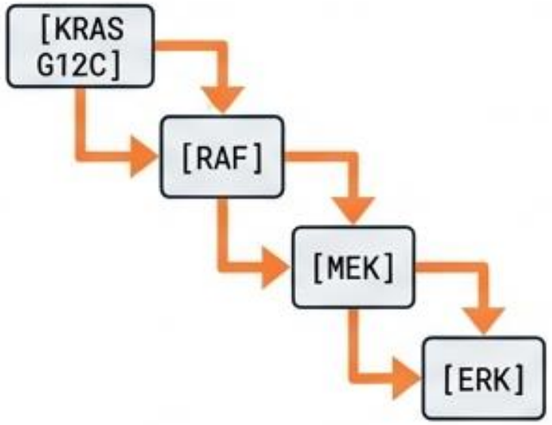
[Reactome] (Total: 169 Pathways)

The Dynamic State. A mutation alone is insufficient. The engine analyzes 169 pathways to verify if the biological program is actively signaling in this specific patient.

Validation: KRAS signaling is significantly higher in KRAS-mutated patients ($p = 8.5 \times 10^{-29}$).

Phase 3: Downstream Tracing

[SIGNOR: 1,743 directed causal edges]

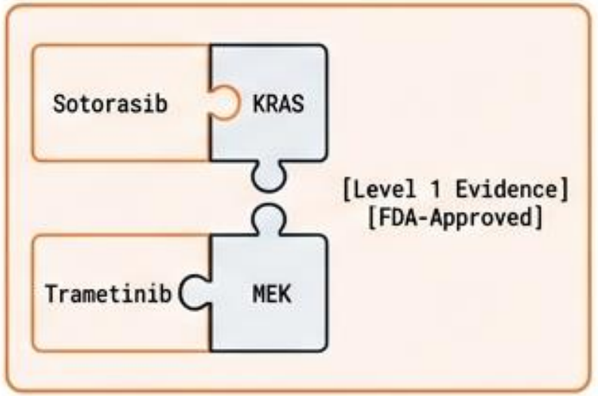


The Causal Network. Active signaling is traced downstream through verified causal edges to identify druggable targets at every subsequent node.

Phase 4: Precision Drug Matching

[130 Curated Variant-Drug Associations]

[114 Drugs] [75 Genes]



Structured Abstention

Known Resistance: Resistance mutations automatically override sensitivity predictions.

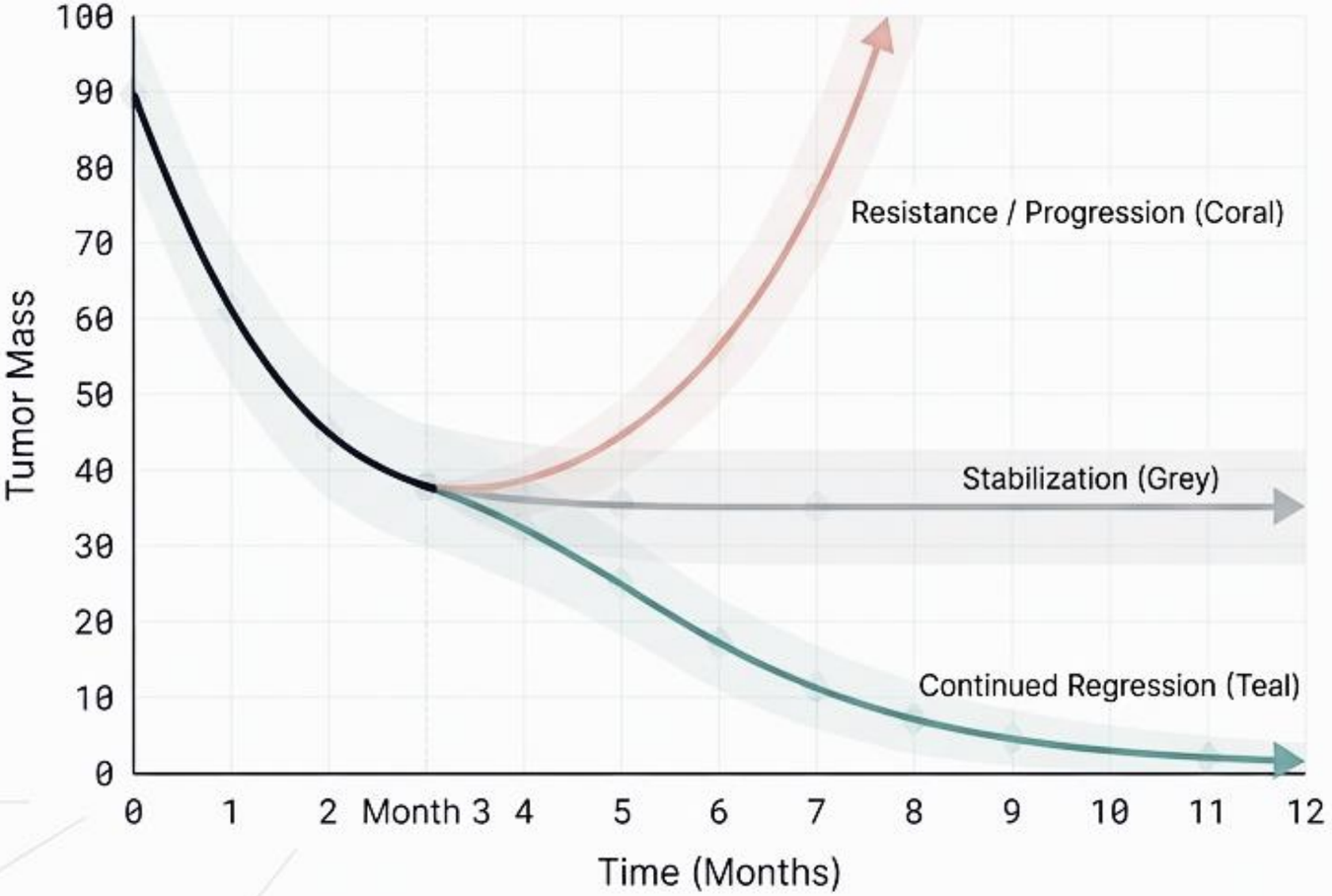
System Integrity: The system refuses to guess. If evidence is insufficient, it abstains and outputs the exact missing data required.

Simulating Time-Resolved Trajectories

Physics-constrained neural ODEs project personalized treatment outcomes

1 **The digital twin becomes real.**

A physics-constrained neural ODE takes personalized parameters (growth rate, drug sensitivity, immune response) and projects them forward. This models the patient's unique biological trajectory.



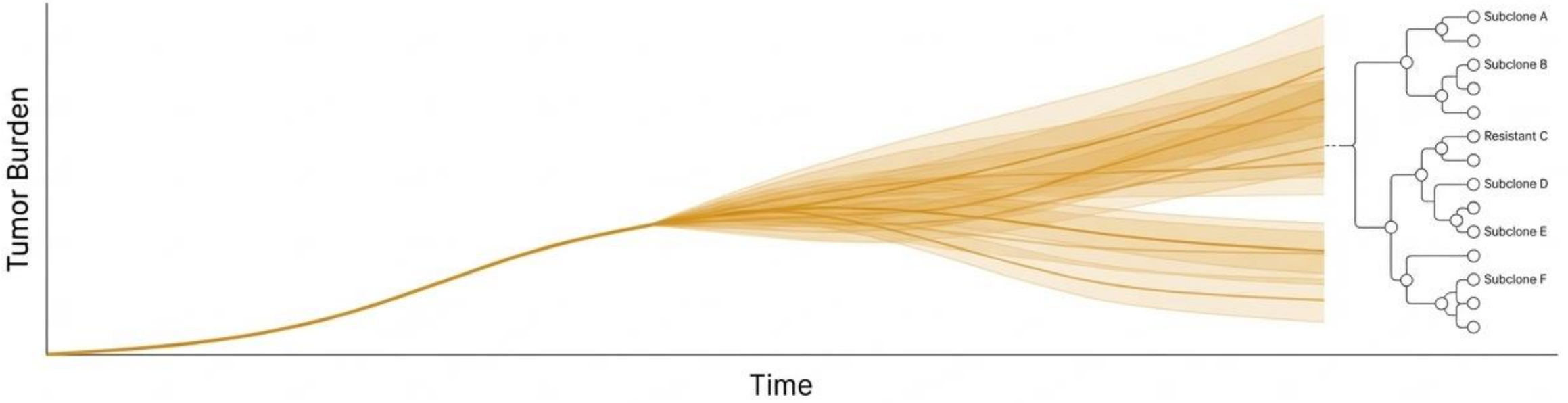
2 **Biological Boundaries:**

The math mirrors ecological predator-prey dynamics. Growth rates cannot be negative; kill rates stay positive. The AI is bounded by physical law, ensuring realistic biological outputs.

3 **Real-Time Emulation:**

A fast emulator runs each branching treatment scenario in 5 milliseconds, enabling real-time exploration of thousands of sequential treatment strategies to optimize the clinical path.

Stochastic simulation accounts for biological randomness and emergent resistance.



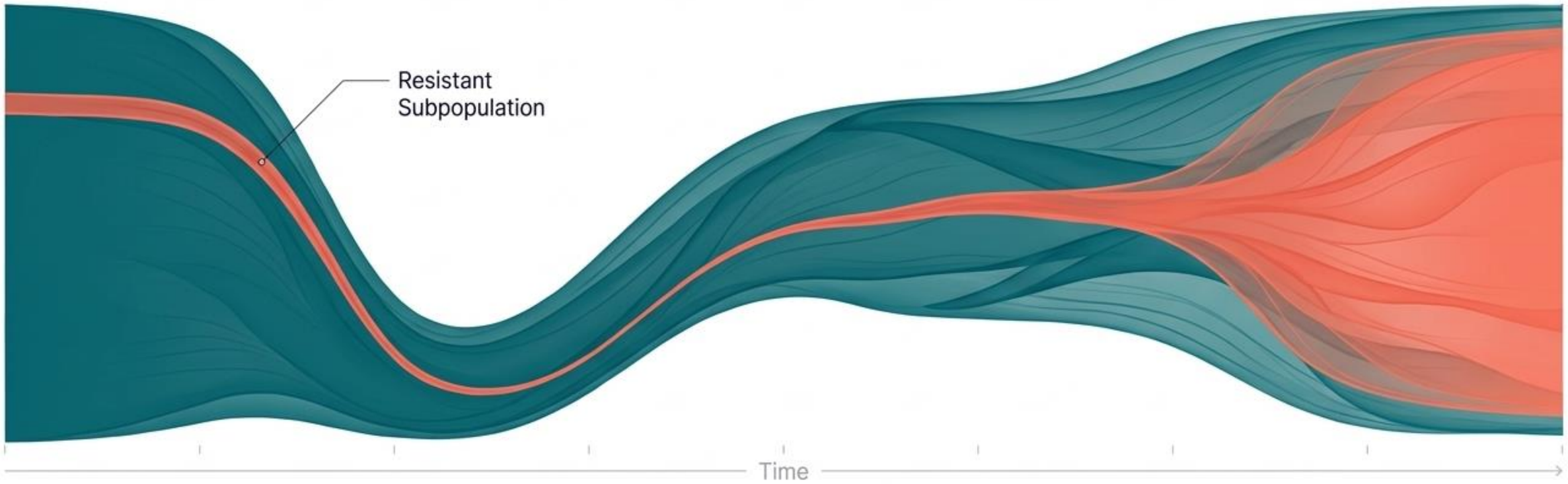
Adding Chaos: On top of the deterministic physics, the engine simulates unpredictable mutation events and the chance dynamics of cancer.

Mapping the Future: Rather than a single brittle prediction, the system outputs a range of possible outcomes with calibrated confidence intervals.

Phylogenetic Tracking: Anticipates the emergence of new resistant subclones, showing exactly how different tumor populations relate and compete over time.

The Resistance Sentinel and Clonal Evolution

Kill 95% of a tumor, and a pre-existing 3% resistant clone will expand to take over. DNAI reconstructs this evolutionary family tree.



Integrating Clinical Evidence

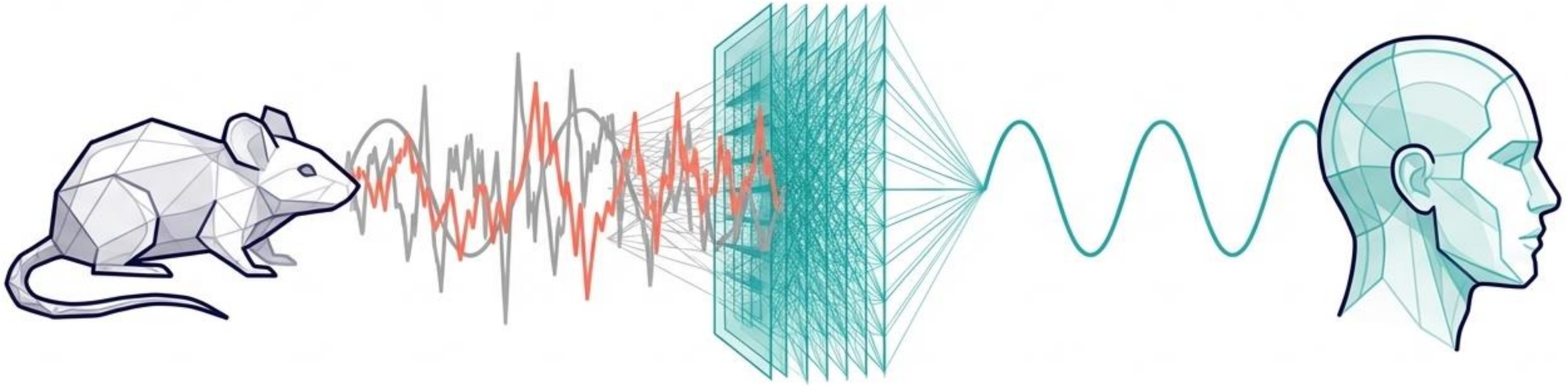
Curated knowledge bases (OncoKB, CIViC) directly inform the simulation. If a clone carries EGFR T790M, the system natively knows its resistance profile.

Stochastic Ensembles

A 100-member ensemble produces a distribution of possible futures. For near-extinction clones, an exact stochastic algorithm (Gillespie SSA) takes over to model individual cell birth and death events.

Filtering the Preclinical Signal

Mouse models generate massive amounts of drug response data, but species-specific biology contaminates the signal.



Domain Separation Network

Strips out mouse artifacts and isolated variables, retaining only the universal tumor biology that seamlessly transfers to human predictions.

Filling the Gaps

Accurately imputes missing data types (like methylation and copy number) that are rarely measured in preclinical models.

Translating Simulation into Actionable Treatment Design

Four specialized output modules turn simulations into definitive clinical strategy:

1. Treatment Optimization

Pareto optimization across efficacy, toxicity, and resistance. Sequentially plans three lines of treatment.

2. Synthetic Lethality Scoring

Maps 28 validated gene pairs to available drugs, finding vulnerabilities created by the tumor's own lost genes.

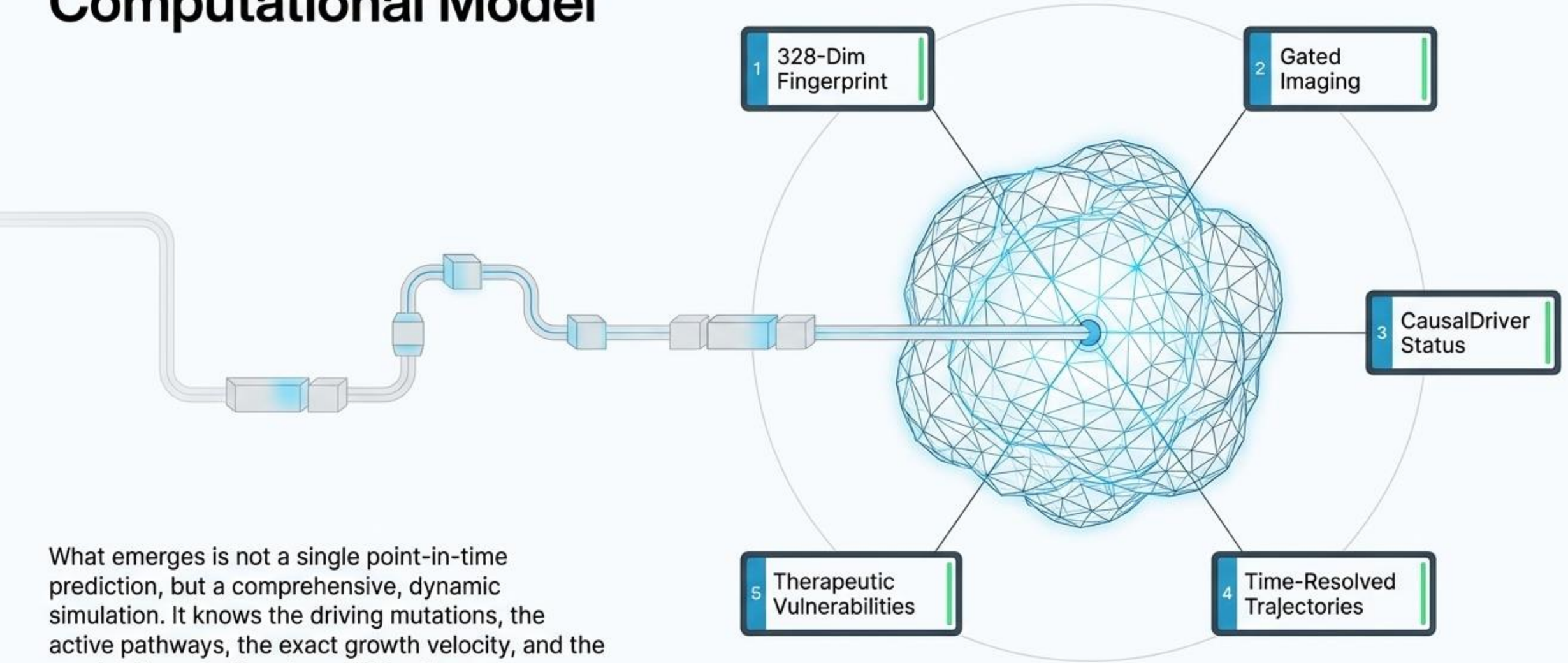
3. Immunogenic Variant Prioritization

Reads the microenvironment to classify tumors as hot, warm, or cold. Critically, it proactively abstains from recommending immunotherapy for cold tumors.

4. Methylation Decoding

Captures 76% of epigenetic variation from just 48 dimensions, identifying chemically silenced (and reversible) tumor suppressors.

The Result: A Living Computational Model



What emerges is not a single point-in-time prediction, but a comprehensive, dynamic simulation. It knows the driving mutations, the active pathways, the exact growth velocity, and the precise therapeutic vulnerabilities. Every recommendation maps directly back to an inspectable biological mechanism.