IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 28, Issue 5, Ser. II (Sep. – Oct. 2026), PP 52-78
www.iosrjournals.org

ONCOGENESIS Ω∞: A Universal Cancer Intelligence and Discovery Observatory for Cross-Cancer Vulnerability, Therapeutic Hypothesis Generation, and Evolutionary Treatment Modeling

Nifla Nalakath[1st]
Department of Computer Science and Engineering
Eranad Knowledge City Technical Campus
Malappuram, Kerala, India

Date of Submission: 14-09-2026 Date of Acceptance: 25-09-2026

I. Introduction

Cancer is not a single disease but a highly heterogeneous and dynamically evolving group of diseases characterized by the accumulation of genetic and epigenetic alterations, dysregulated cellular signaling, altered cellular states, and evolutionary selection within complex biological environments (Hanahan and Weinberg 2011; Hanahan et al. 2022). The molecular mechanisms underlying tumor initiation, progression, metastasis, therapeutic response, and resistance can differ substantially between cancer types and even among tumors belonging to the same clinical category. Consequently, understanding cancer requires analysis across multiple biological scales rather than consideration of isolated biomarkers or individual molecular pathways. Large-scale cancer genomics initiatives have demonstrated the value of such an integrated perspective. The Cancer Genome Atlas (TCGA) Pan-Cancer project established a framework for comparing molecular alterations across tumor lineages and identifying both cancer-specific and cross-cancer molecular patterns (Weinstein et al. 2013). Subsequently, the Pan-Cancer Analysis of Whole Genomes (PCAWG) project integrated whole-genome data from thousands of tumors across numerous cancer types, revealing the extensive genomic heterogeneity and diverse evolutionary trajectories underlying human malignancies ("Pan-Cancer Analysis of Whole Genomes" 2020; Gerstung et al. 2020).

The increasing availability of high-dimensional biological datasets has created an unprecedented opportunity for computational approaches to investigate these complexities. Public resources such as the Gene Expression Omnibus (GEO), UniProt, STRING, ChEMBL, DrugBank, Reactome, and COSMIC provide complementary information concerning gene expression, proteins, protein–protein interactions, bioactive compounds, drug–target relationships, biological pathways, and somatic cancer alterations (Edgar et al. 2002; Clough et al. 2024; Consortium 2019; Szklarczyk et al. 2021; Gaulton et al. 2012, 2017; Knox et al. 2024; Gillespie et al. 2022; Tate et al. 2019). In parallel, large-scale functional screening initiatives such as the Cancer Dependency Map and CRISPR-based cancer dependency studies have enabled systematic investigation of molecular dependencies that may represent therapeutic opportunities (Tsherniak et al. 2017; McFarland et al. 2018; Behan et al. 2019). The combination of these resources creates the possibility of constructing computational representations in which genes, proteins, mutations, pathways, cancer states, drugs, and species are modeled as interconnected components rather than independent variables.

Artificial intelligence has consequently become an increasingly important component of computational oncology. In particular, multi-omics artificial intelligence approaches provide mechanisms for integrating heterogeneous molecular measurements and extracting patterns that may be difficult to identify using conventional single-omics analyses (Jiang et al. 2025). Graph-based machine learning is especially attractive for this setting because biological systems naturally exhibit relational structure. Graph convolutional networks, inductive graph representation learning, and graph attention networks provide mechanisms for learning representations from interconnected entities and have been increasingly investigated for biomedical prediction and drug discovery (Kipf and Welling 2016; Hamilton et al. 2017; Veličković et al. 2017; Zhou et al. 2020). Recent studies have applied graph neural networks to drug repurposing, cancer drug response, and druggable-gene discovery, demonstrating the potential of graph-based learning to connect molecular information with therapeutic hypotheses (Doshi and Chepuri 2022; Cui et al. 2021; Li et al. 2025; Zhang et al. 2025). However, these approaches generally address specific prediction or ranking tasks rather than providing a unified computational environment for exploring cancer across multiple molecular, evolutionary, therapeutic, and causal dimensions.

A further challenge arises from the distinction between association and causation. A molecular feature may be statistically associated with a cancer phenotype without being a mechanistic driver of that phenotype. Causal inference provides a formal framework for expressing assumptions, identifying causal effects, and testing the robustness of causal conclusions (Pearl et al. 2003; Hernan and Robins 2020). Computational frameworks such as DoWhy provide mechanisms for explicitly representing causal assumptions and applying identification, estimation, and refutation procedures (Sharma and Kiciman 2020; Sharma et al. 2021). Integrating causal reasoning with molecular networks therefore provides an opportunity to distinguish candidate vulnerabilities that merely correlate with cancer states from relationships that provide stronger evidence of potential causal relevance.

Cancer must also be considered as an evolving population rather than a static molecular state. The clonal evolution model established that tumor populations can undergo sequential selection of genetically distinct subclones (Nowell 1976). Large-scale genomic analyses have subsequently demonstrated substantial variation in the timing and evolutionary trajectories of tumor alterations (Gerstung et al. 2020). Therapeutic intervention itself creates an additional selective pressure that can alter the composition of cancer-cell populations. Evolutionary and ecological approaches to cancer have therefore been proposed to understand treatment resistance and to design adaptive therapeutic strategies (Gatenby et al. 2009; Basanta and Anderson 2013; Cunningham et al. 2011; West et al. 2020; Basanta et al. 2012). These observations motivate computational models capable of examining not only which therapeutic intervention may be effective at a particular point in time, but also how cancer populations may respond and adapt to treatment pressure.

Another important dimension is the reconstruction of cellular trajectories. Pseudotemporal methods were originally developed to order single cells along inferred developmental trajectories and identify regulators associated with changing cellular states (Trapnell et al. 2014). Methods such as Slingshot extended this concept to lineage and pseudotime inference from single-cell transcriptomic data, while systematic comparisons have demonstrated substantial methodological diversity among trajectory-inference approaches (Street et al. 2018; Saelens et al. 2019). Within a cancer framework, trajectory analysis provides a computational basis for investigating transitions between cellular states and constructing an evolutionary perspective on malignant progression. This motivates the concept of a computational "Cancer Time Machine" capable of representing inferred transitions across cancer-related states.

Therapeutic discovery itself is inherently multi-objective. A candidate intervention may provide strong predicted efficacy while simultaneously increasing toxicity or selecting for resistant cellular populations. Therefore, optimizing a single objective may not adequately represent the therapeutic decision space. Multi-objective optimization and Pareto-frontier analysis provide a mathematical framework for identifying solutions that represent different trade-offs among competing objectives (Deb et al. 2002). Similarly, uncertainty must be explicitly considered when computational systems generate biomedical predictions. Conformal prediction provides distribution-free uncertainty quantification and can construct prediction sets or intervals with formal coverage guarantees under appropriate assumptions (Vovk et al. 2005; Romano et al. 2019; Angelopoulos and Bates 2023). Such approaches provide a mechanism through which a computational discovery system can express uncertainty and avoid treating every prediction as an equally reliable scientific conclusion.

Drug repurposing provides another important bridge between computational discovery and therapeutic translation. Molecular fingerprints such as the original Morgan representation and extended-connectivity fingerprints provide compact computational representations of chemical structures (Morgan 1965; Rogers and Hahn 2010). Similarity measures such as the Tanimoto index can then be used to compare molecular structures and identify compounds with related chemical characteristics (Bajusz et al. 2015). Network-based integration of biomedical knowledge has also demonstrated the potential to prioritize existing drugs for new therapeutic applications (Himmelstein et al. 2017). These approaches provide complementary computational mechanisms for connecting candidate molecular vulnerabilities with existing pharmacological agents.

Despite these advances, a substantial methodological gap remains. Existing approaches frequently focus on individual components of the cancer-discovery problem: molecular profiling, dependency identification, graph-based prediction, causal analysis, trajectory inference, drug repurposing, evolutionary modeling, or treatment optimization. While each provides valuable information, these components are rarely organized into a single computational framework in which discoveries generated by one analytical layer can inform, challenge, and refine hypotheses generated by another. Furthermore, conventional predictive pipelines generally optimize predictive performance rather than explicitly incorporating scientific falsification, uncertainty-aware hypothesis evaluation, evolutionary escape analysis, and computational provenance into the discovery process.

To address this gap, this study introduces *ONCOGENESIS* Ω∞, a computational Universal Cancer Intelligence and Discovery Observatory designed to investigate cancer as an interconnected, heterogeneous, and evolving phenomenon. Rather than functioning solely as a cancer classifier or a conventional drug-response predictor, ONCOGENESIS Ω∞ integrates multiple computational perspectives into a unified research

architecture. The framework constructs a multidimensional *Cancer Universe* and a heterogeneous *Universal Cancer Graph*, followed by graph-based discovery of previously unobserved biological associations. A *Universal Vulnerability Hunter* prioritizes candidate vulnerabilities across cancer contexts, while a causal-inference layer evaluates their potential mechanistic relevance. Molecular similarity is subsequently used to investigate existing therapeutic compounds that may correspond to identified vulnerabilities.

The framework further incorporates a *Cancer Time Machine* for trajectory-based analysis, an *Evolutionary War Room* for modeling treatment-driven evolutionary dynamics, and a multi-objective optimization layer for identifying Pareto-optimal therapeutic strategies across competing objectives. A conformal-prediction-based uncertainty layer provides an explicit uncertainty firewall around computational predictions. Beyond prediction, ONCOGENESIS Ω∞ incorporates a *Hypothesis Crucible* and an *Adversarial Scientist* designed to subject candidate discoveries to systematic challenge rather than automatically accepting generated hypotheses. This establishes a computational falsification loop in which hypotheses can be strengthened, weakened, or rejected according to the evidence available to the system. Finally, a *Scientific Provenance Dashboard* records the computational lineage and supporting evidence associated with generated findings, thereby improving transparency and traceability throughout the discovery pipeline.

The central scientific question investigated by ONCOGENESIS Ω∞ is whether computationally identifiable molecular vulnerabilities and therapeutic principles can be conserved across sufficiently diverse cancer contexts, and whether such patterns remain supported when subjected to causal, evolutionary, molecular, and adversarial evaluation. Importantly, the framework is designed as a hypothesis-discovery and evaluation system rather than as a claim of a clinically validated universal cancer cure. Its objective is to computationally explore the space of conserved vulnerabilities, candidate therapeutic opportunities, evolutionary escape mechanisms, and unresolved questions that emerge from the integrated analysis of cancer data.

The principal contributions of this work are therefore:

- A unified computational architecture for multi-dimensional cancer investigation.
- Integration of heterogeneous cancer and biomedical knowledge through a dynamic graph representation.
- Combined graph learning and causal reasoning for vulnerability discovery.
- Integration of molecular drug-repurposing analysis with cancer vulnerabilities.
- Trajectory and evolutionary modeling for investigating malignant progression and treatment resistance.
- Multi-objective therapeutic optimization.
- Uncertainty-aware prediction using conformal methods.
- Adversarial hypothesis testing and computational scientific falsification.
- Explicit scientific provenance for tracing computational findings.

Collectively, these components establish ONCOGENESIS Ω∞ as an integrated computational observatory for exploring cancer biology and therapeutic discovery across molecular, causal, evolutionary, and computational dimensions.

II. Literature Review

The development of computational oncology has progressed from the identification of individual cancer-associated molecular alterations toward increasingly integrated analyses of cancer genomes, cellular states, biological networks, therapeutic dependencies, and evolutionary processes. This progression provides the scientific foundation for ONCOGENESIS Ω∞ , which combines these perspectives within a unified computational discovery framework.

2.1 Cancer Biology and Pan-Cancer Molecular Analysis

The conceptual understanding of cancer has evolved from viewing tumors primarily as collections of genetic abnormalities toward recognizing cancer as a complex biological system involving multiple interacting hallmarks and dynamically changing cellular processes. Hanahan and Weinberg established the major hallmarks underlying malignant transformation and subsequently expanded this framework to incorporate additional dimensions of tumor biology, including phenotypic plasticity, cellular senescence, epigenetic reprogramming, and interactions with the tumor microenvironment (Hanahan and Weinberg 2011; Hanahan et al. 2022). These developments emphasize that cancer cannot be adequately characterized through a single molecular feature or isolated biological pathway.

Large-scale genomic projects have enabled systematic comparison across multiple cancer types. The Cancer Genome Atlas (TCGA) Pan-Cancer Analysis Project demonstrated the value of integrating molecular measurements across tumor lineages to identify shared and cancer-specific genomic characteristics (Weinstein et al. 2013). The Pan-Cancer Analysis of Whole Genomes (PCAWG) subsequently extended this perspective to whole-genome sequencing, providing a comprehensive view of somatic alterations across thousands of tumors ("Pan-Cancer Analysis of Whole Genomes" 2020). Gerstung et al. reconstructed the evolutionary history of

2,658 cancers and demonstrated substantial variation in the timing and evolutionary trajectories of cancer-associated alterations (Gerstung et al. 2020). These studies provide an important foundation for the Cancer Universe component of ONCOGENESIS Ω∞, which seeks to represent cancer states across heterogeneous molecular and evolutionary dimensions rather than treating each cancer type as an isolated prediction problem.

2.2 Cancer Dependencies and Therapeutic Vulnerabilities

Systematic identification of cancer dependencies has become an important strategy for discovering potential therapeutic targets. The Cancer Dependency Map demonstrated that large-scale functional screening can be used to identify genes on which cancer cells preferentially depend for survival and proliferation (Tsherniak et al. 2017). McFarland et al. further investigated cancer dependencies using large-scale RNA interference screens and integrated computational normalization and data-analysis strategies to improve dependency estimation (McFarland et al. 2018). CRISPR–Cas9 screening has provided an additional experimental framework for identifying and prioritizing potential therapeutic targets across cancer cell lines (Behan et al. 2019).

These studies motivate the Universal Vulnerability Hunter in ONCOGENESIS Ω∞. However, rather than limiting vulnerability discovery to individual screening datasets, the proposed framework aims to integrate multiple molecular and biological relationships and identify candidate vulnerabilities that may be conserved across heterogeneous cancer contexts. The incorporation of causal analysis further extends conventional dependency analysis by distinguishing statistical associations from relationships that may have stronger mechanistic relevance.

2.3 Biomedical Data Resources and Knowledge Integration

The availability of large public biomedical databases has made integrated computational analysis increasingly feasible. The Gene Expression Omnibus (GEO) provides a major repository of gene-expression and functional genomics datasets, with subsequent updates expanding its coverage of gene-expression and epigenomic studies (Edgar et al. 2002; Clough et al. 2024). UniProt provides comprehensive protein sequence and functional information (Consortium 2019), while STRING provides experimentally and computationally derived protein–protein interaction networks (Szklarczyk et al. 2021). Together, these resources enable molecular relationships to be represented as interconnected biological networks.

Drug- and pathway-oriented resources provide complementary information required for therapeutic discovery. ChEMBL provides large-scale bioactivity information for chemical compounds and their biological targets (Gaulton et al. 2012, 2017). DrugBank integrates information concerning drugs, targets, and pharmacological relationships (Knox et al. 2024), while Reactome provides curated biological pathways and molecular reactions (Gillespie et al. 2022). COSMIC provides information concerning somatic mutations observed in cancer (Tate et al. 2019). The integration of these heterogeneous resources provides the foundation for the Universal Cancer Graph of ONCOGENESIS Ω∞, in which genes, proteins, pathways, mutations, cancers, and drugs can be represented as interconnected entities with associated evidence.

2.4 Graph Neural Networks for Biomedical Discovery

Biological systems are naturally represented as networks because molecular entities interact through complex relational structures. Graph convolutional networks introduced a framework for semi-supervised representation learning on graph-structured data (Kipf and Welling 2016). GraphSAGE extended graph representation learning through inductive neighborhood aggregation, allowing representations to be learned for previously unseen nodes (Hamilton et al. 2017). Graph Attention Networks further introduced attention mechanisms for assigning different importance to neighboring nodes during representation learning (Veličković et al. 2017). A broader review of graph neural networks established their applicability across numerous graph-learning problems and domains (Zhou et al. 2020).

These developments have motivated the application of graph neural networks to biomedical discovery. Computational approaches have investigated graph-based drug repurposing, including the integration of drug-response information and molecular relationships (Doshi and Chepuri 2022; Cui et al. 2021). More recent work has explored graph-based approaches for druggable-gene discovery using multi-omics information and heterogeneous graph representations (Li et al. 2025), while heterogeneous graph neural networks have also been applied to cancer drug-response prediction using multi-omics data (Zhang et al. 2025). Explainable graph neural-network approaches have further been investigated for generating hypotheses for drug repurposing (Perdomo-Quinteiro et al. 2026).

Although these studies demonstrate the value of graph learning in oncology and drug discovery, ONCOGENESIS Ω∞ extends the graph-learning perspective into a broader heterogeneous discovery architecture. Its Universal Cancer Graph is intended to connect multiple biological entity types, while its GNN

component is integrated with vulnerability discovery, causal inference, drug repurposing, evolutionary analysis, uncertainty assessment, and provenance rather than functioning as an isolated prediction model.

2.5 Causal Inference in Biomedical Research

Machine-learning models can identify strong associations between molecular variables and phenotypic outcomes without necessarily establishing causal relationships. Causal inference provides a formal framework for distinguishing causal effects from statistical associations through explicit assumptions and causal models (Pearl et al. 2003; Hernan and Robins 2020). The DoWhy framework provides a practical computational approach for expressing causal assumptions, estimating causal effects, and testing the robustness of causal conclusions through refutation procedures (Sharma and Kiciman 2020; Sharma et al. 2021).

Causal reasoning has increasing relevance to cancer and therapeutic discovery because a molecular feature that correlates with tumor progression is not necessarily a suitable therapeutic target. Recent work has also investigated causal inference for identifying essential genes relevant to drug-synergy prediction (Zhang et al. 2026). ONCOGENESIS Ω∞ incorporates this perspective through its Causal Vulnerability Scoring component, which is designed to evaluate whether candidate molecular vulnerabilities possess evidence beyond simple observational association.

2.6 Cancer Evolution, Treatment Resistance, and Evolutionary Modeling

Cancer progression is fundamentally evolutionary. The clonal evolution model proposed by Nowell established that tumor progression can involve sequential selection of genetically distinct cellular populations (Nowell 1976). Subsequent genomic studies have provided extensive evidence for branching and heterogeneous evolutionary trajectories across human cancers (Gerstung et al. 2020). Consequently, therapeutic interventions may act as selective pressures that alter tumor-cell populations and promote the emergence or expansion of resistant phenotypes.

Evolutionary and ecological approaches have been proposed to exploit these dynamics therapeutically. Gatenby et al. introduced adaptive therapy as a strategy for managing tumor evolution rather than attempting to eliminate every cancer cell simultaneously (Gatenby et al. 2009). Basanta and Anderson demonstrated how ecological principles can provide insight into cancer progression and treatment (Basanta and Anderson 2013). Evolutionary dynamics have subsequently been applied to cancer therapy and treatment resistance, including strategies involving multiple drugs and adaptive treatment schedules (Cunningham et al. 2011; West et al. 2020). Evolutionary game theory has also been used to investigate interactions between tumor and stromal populations (Basanta et al. 2012), while the mathematical foundations of evolutionary dynamics are established more broadly by replicator and population-dynamics frameworks (Nowak 2006).

These studies motivate the Evolutionary War Room of ONCOGENESIS Ω∞, where treatment is considered within an evolving population rather than as a static prediction problem. The framework uses evolutionary modeling to investigate potential treatment responses, selective pressures, and resistance-associated strategies.

2.7 Cellular Trajectory Inference and Cancer State Transitions

The analysis of cellular trajectories provides another perspective for understanding dynamic biological systems. Trapnell et al. introduced pseudotemporal ordering to reconstruct cellular-state transitions from single-cell transcriptomic measurements and identify regulators associated with changing cell states (Trapnell et al. 2014). Slingshot subsequently provided a framework for simultaneous lineage and pseudotime inference from single-cell transcriptomic data (Street et al. 2018). Comparative evaluation of trajectory-inference methods demonstrated that different computational approaches can produce substantially different representations of cellular trajectories depending on the characteristics of the data and assumptions of the methods (Saelens et al. 2019).

Within ONCOGENESIS Ω∞, these principles motivate the Cancer Time Machine, which represents cancer progression as a trajectory through a multidimensional state space. Rather than interpreting cancer exclusively as a collection of static categories, the framework investigates transitions between molecular states and seeks to computationally characterize regions associated with progression toward malignancy.

2.8 Molecular Representation and Computational Drug Repurposing

Computational drug discovery can substantially expand the search space for therapeutic candidates by exploiting existing molecular and pharmacological knowledge. Morgan introduced a computational representation for chemical structures that established an important foundation for machine-readable molecular fingerprints (Morgan 1965). Extended-connectivity fingerprints subsequently provided a widely used approach for representing molecular neighborhoods and structural characteristics (Rogers and Hahn 2010). Bajusz et al. systematically examined similarity metrics for molecular fingerprints and demonstrated the suitability of the Tanimoto index for fingerprint-based molecular similarity calculations (Bajusz et al. 2015).

Network-based approaches provide an additional mechanism for drug repurposing by integrating heterogeneous biomedical relationships. Himmelstein et al. demonstrated that systematic integration of biomedical knowledge can prioritize existing drugs for potential new therapeutic applications (Himmelstein et al. 2017). More recent work has combined graph-based learning and causal inference perspectives for computational drug repurposing (Xu et al. 2023). These approaches provide the conceptual foundation for the drug-repurposing component of ONCOGENESIS Ω∞, which connects candidate cancer vulnerabilities to existing compounds through molecular similarity and biomedical relationships.

2.9 Multi-Objective Therapeutic Optimization

Cancer treatment involves competing objectives. Maximizing predicted therapeutic efficacy alone may be insufficient when toxicity, treatment resistance, and evolutionary adaptation are simultaneously considered. Multi-objective optimization provides a mathematical framework for identifying solutions that represent different trade-offs among competing objectives. The NSGA-II algorithm introduced by Deb et al. established an influential evolutionary approach for multi-objective optimization and Pareto-frontier discovery (Deb et al. 2002). Such approaches provide a foundation for representing therapeutic strategies not as a single optimum but as a set of non-dominated alternatives.

ONCOGENESIS Ω∞ incorporates this principle through its Multi-Objective Treatment Optimization component. The corresponding Pareto frontier is intended to represent treatment strategies that balance competing objectives such as therapeutic effectiveness, toxicity, and resistance prevention. This formulation complements the evolutionary component by allowing candidate treatment strategies to be evaluated not only according to immediate predicted benefit but also according to competing biological and therapeutic considerations.

2.10 Uncertainty Quantification and Conformal Prediction

A major limitation of conventional machine-learning systems in biomedical applications is that predictive confidence does not necessarily correspond to reliable statistical uncertainty. Conformal prediction provides a framework for distribution-free uncertainty quantification under appropriate assumptions (Vovk et al. 2005). Conformalized quantile regression extended these principles to the construction of prediction intervals with finite-sample coverage guarantees under exchangeability assumptions (Romano et al. 2019). A broader treatment of conformal prediction has established its applicability as a general framework for uncertainty quantification in machine learning (Angelopoulos and Bates 2023).

ONCOGENESIS Ω∞ incorporates conformal prediction as an uncertainty firewall surrounding computational predictions. Rather than forcing the system to provide a definitive answer for every candidate, uncertainty-aware evaluation enables the framework to distinguish predictions supported by stronger evidence from those requiring additional investigation. This principle is particularly important for a scientific discovery system because uncertainty should be represented explicitly rather than hidden behind a conventional model confidence score.

2.11 Explainable Artificial Intelligence and Scientific Transparency

Interpretability is particularly important in biomedical AI because computational predictions may influence subsequent scientific investigation. SHAP provides a unified framework for interpreting individual model predictions through feature-attribution values derived from cooperative game-theoretic principles (Lundberg and Lee 2017). Explainability can therefore provide information about which molecular or biological features contribute to a model's output.

ONCOGENESIS Ω∞ extends this interpretability perspective beyond individual predictions through its Scientific Provenance Dashboard. The objective is to preserve the computational lineage of generated findings, including their originating analytical module, supporting evidence, uncertainty, and evaluation history. This creates a distinction between merely producing a prediction and maintaining an auditable scientific record of how that prediction was generated.

2.12 Comparative Oncology and Cross-Species Investigation

Cancer is observed across multiple species, creating an additional opportunity for comparative investigation. Comparative oncology has received particular attention through naturally occurring cancers in companion animals, especially dogs, which may provide biologically relevant models for understanding human cancer and evaluating therapeutic approaches (LeBlanc and Mazcko 2020). This perspective supports investigation of conserved and divergent cancer-associated mechanisms across species.

The cross-species perspective is incorporated into the broader conceptual design of ONCOGENESIS Ω∞, in which candidate vulnerabilities can be evaluated in relation to their conservation across diverse biological contexts. The objective is not to assume that a vulnerability observed in one species automatically

translates to another, but to provide a computational framework in which cross-species conservation can become an explicit dimension of hypothesis evaluation.

2.13 Research Gap and Positioning of ONCOGENESIS Ω∞

The literature demonstrates substantial progress across individual components relevant to computational oncology: pan-cancer molecular characterization (Weinstein et al. 2013; "Pan-Cancer Analysis of Whole Genomes" 2020), cancer dependency mapping (Tsherniak et al. 2017; Behan et al. 2019), biomedical knowledge integration (Szklarczyk et al. 2021; Gillespie et al. 2022; Knox et al. 2024), graph neural networks (Kipf and Welling 2016; Hamilton et al. 2017; Veličković et al. 2017), causal inference (Pearl et al. 2003; Sharma and Kiciman 2020), cancer trajectory inference (Trapnell et al. 2014; Street et al. 2018), evolutionary treatment modeling (Gatenby et al. 2009; West et al. 2020), molecular similarity and drug repurposing (Rogers and Hahn 2010; Bajusz et al. 2015; Himmelstein et al. 2017), multi-objective optimization (Deb et al. 2002), and uncertainty quantification (Vovk et al. 2005; Romano et al. 2019; Angelopoulos and Bates 2023). However, these methodological advances are generally investigated as separate computational problems.

ONCOGENESIS Ω∞ is positioned at the intersection of these research directions. Its principal distinction is not the invention of each underlying algorithm independently, but the integration of heterogeneous computational perspectives into a unified scientific discovery pipeline. The framework connects cancer-state representation, knowledge-graph reasoning, graph learning, vulnerability discovery, causal evaluation, molecular drug repurposing, trajectory reconstruction, evolutionary simulation, multi-objective optimization, uncertainty quantification, adversarial hypothesis testing, and scientific provenance. This architecture enables findings generated at one stage to become evidence, hypotheses, constraints, or challenges for subsequent stages. The resulting framework therefore shifts the computational objective from isolated cancer prediction toward systematic hypothesis generation and evaluation. In particular, ONCOGENESIS Ω∞ is designed to investigate whether conserved molecular vulnerabilities and therapeutic principles can be identified across heterogeneous cancer contexts, while allowing candidate hypotheses to be challenged by causal, evolutionary, molecular, and uncertainty-aware analyses. This hypothesis-driven and potentially self-falsifying architecture forms the conceptual foundation for the computational experiments and results presented in the subsequent sections.

III. Methodology and Data

3.1 Overall Research Framework

ONCOGENESIS Ω∞ is designed as a modular computational observatory for investigating cancer across molecular, network, causal, evolutionary, therapeutic, and uncertainty dimensions. The framework does not treat cancer as a single classification problem; instead, it integrates heterogeneous biological data and multiple computational paradigms into a sequential hypothesis-discovery and evaluation pipeline. The overall workflow consists of the Cancer Universe, Universal Cancer Graph, Graph Neural Network learning, Universal Vulnerability Hunter, causal inference, molecular drug repurposing, Cancer Time Machine, evolutionary modeling, multi-objective treatment optimization, uncertainty-aware assessment, adversarial hypothesis evaluation, and scientific provenance tracking.

The methodological architecture is organized so that information generated by one component can be passed to subsequent components as evidence, candidate hypotheses, constraints, or evaluation targets. This enables the framework to move from heterogeneous biological observations toward candidate vulnerabilities and therapeutic hypotheses while retaining the evidence and computational lineage associated with each finding.

3.2 Data Sources and Biological Knowledge

The framework is designed around publicly available biomedical resources representing complementary biological dimensions. Pan-cancer molecular information is obtained from The Cancer Genome Atlas (TCGA) and the Pan-Cancer Analysis of Whole Genomes (PCAWG) resources (Weinstein et al. 2013; "Pan-Cancer Analysis of Whole Genomes" 2020). Cancer dependency information is supported by large-scale functional screening studies and cancer dependency resources (Tsherniak et al. 2017; McFarland et al. 2018; Behan et al. 2019). Gene-expression and transcriptomic information is obtained from the Gene Expression Omnibus (GEO) (Edgar et al. 2002; Clough et al. 2024).

Protein-level information is represented using UniProt, while molecular interaction information is obtained from STRING (Consortium 2019; Szklarczyk et al. 2021). Chemical and pharmacological information is incorporated from ChEMBL and DrugBank (Gaulton et al. 2012, 2017; Knox et al. 2024). Biological pathway relationships are represented using Reactome (Gillespie et al. 2022), and cancer-associated somatic mutations are represented using COSMIC (Tate et al. 2019). Comparative oncology information provides an additional cross-species perspective for investigating potentially conserved cancer-associated mechanisms (LeBlanc and Mazcko 2020).

The principal data resources and their roles within the framework are summarized in Table [tab:data_sources].

3.3 Data Integration and Preprocessing

Because the data sources differ in structure, scale, terminology, and biological resolution, preprocessing is required before constructing the integrated computational representation. The preprocessing stage consists of data acquisition, identifier harmonization, removal or handling of inconsistent records, normalization where required by the individual analysis, mapping of molecular entities to common identifiers, and construction of relationships between compatible biological entities.

Genes, proteins, mutations, pathways, drugs, cancer types, and species are represented using standardized identifiers whenever available. Duplicate relationships are consolidated, while source information is retained so that individual relationships remain traceable to their corresponding evidence source. This provenance-aware integration prevents the loss of information concerning the origin of individual biological relationships.

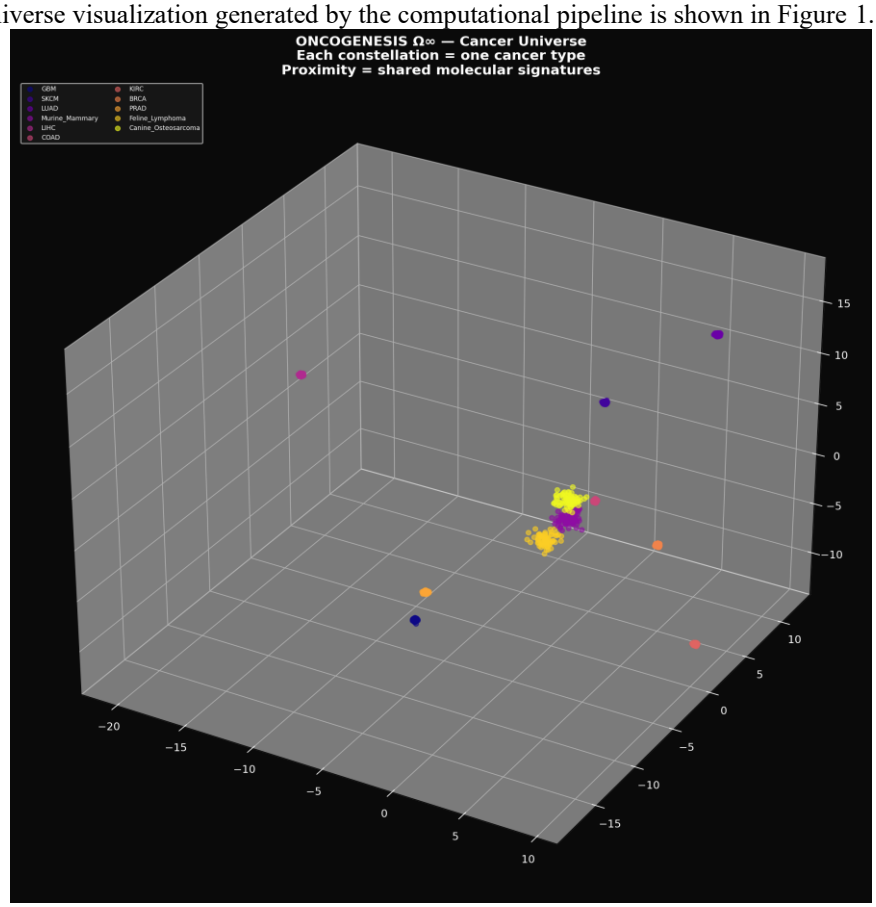
The resulting integrated representation provides the input for subsequent graph construction, vulnerability analysis, causal inference, molecular similarity analysis, trajectory analysis, and evolutionary modeling.

3.4 Cancer Universe Representation

The Cancer Universe provides the multidimensional representation of cancer states used by the framework. Instead of representing cancer solely through discrete diagnostic labels, molecular characteristics are organized into a high-dimensional state space containing genomic, transcriptomic, proteomic, pathway, dependency, and other available biological characteristics.

The resulting representation is used to investigate similarities and differences between cancer states and to identify regions in which multiple cancer contexts exhibit related molecular characteristics. The framework further incorporates a topological perspective through the concept of a cancer-state manifold. Topological Data Analysis (TDA) and persistent-homology-based representations are considered for identifying structural characteristics of the resulting state space.

The Cancer Universe visualization generated by the computational pipeline is shown in Figure 1.



Cancer Universe representation generated by ONCOGENESIS Ω∞ , illustrating the computational representation of cancer states within the integrated analytical space.

The objective of this representation is to provide a common computational space in which cancer states can subsequently be compared for vulnerability conservation, trajectory analysis, and cross-context hypothesis generation.

3.5 Universal Cancer Graph Construction

The Universal Cancer Graph is constructed as a heterogeneous graph in which different biological entities are represented as distinct node types. These include genes, proteins, mutations, pathways, drugs, cancer types, and species. Edges represent biological or pharmacological relationships such as gene–protein associations, protein–protein interactions, gene–pathway relationships, mutation–cancer associations, drug–target relationships, and other evidence-supported relationships.

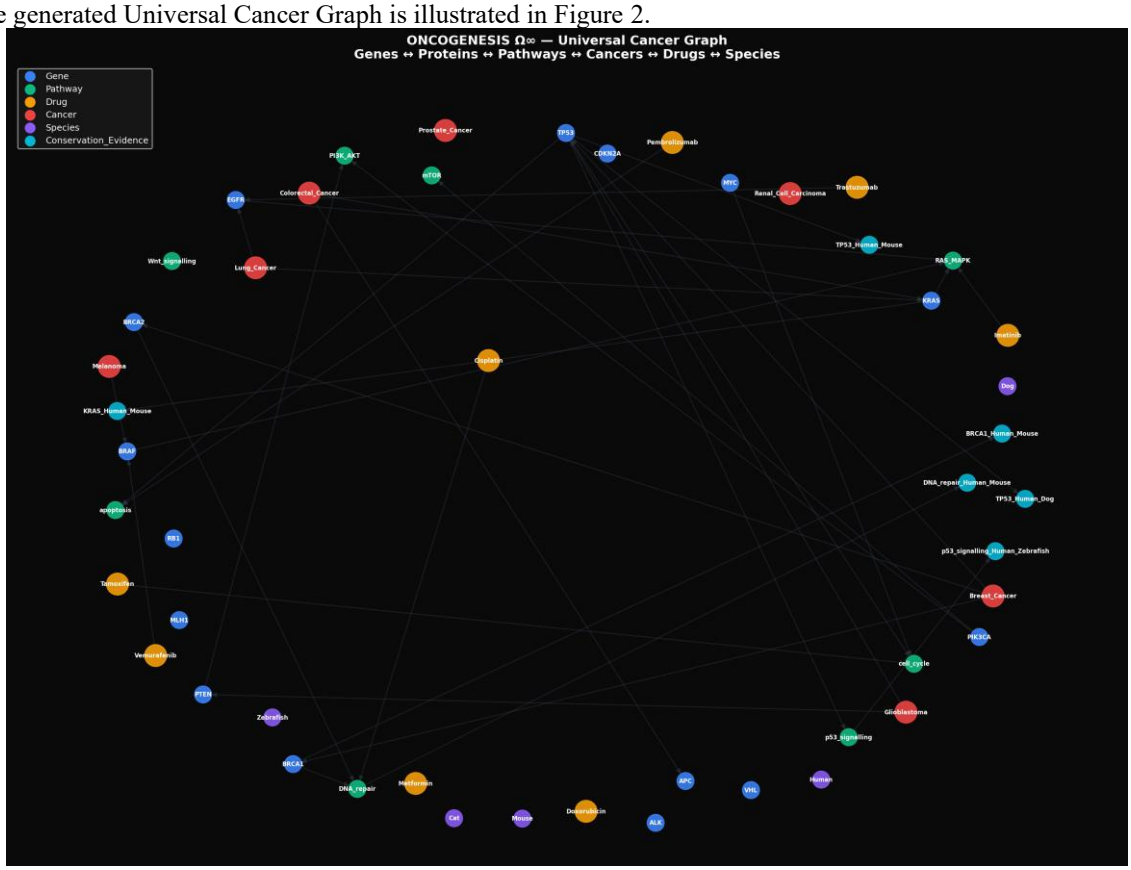
Formally, the heterogeneous graph can be represented as

$$G = (V, E, \mathcal{T}_V, \mathcal{T}_E),$$

where V represents the set of biological entities, E represents the set of relationships, $\mathcal{T}_V$ denotes node types, and $\mathcal{T}_E$ denotes edge types.

Each relationship is additionally associated with evidence and provenance information. Consequently, the graph is not treated as a collection of unqualified relationships; instead, the evidence supporting each relationship can be retained for downstream evaluation.

The generated Universal Cancer Graph is illustrated in Figure 2.



Universal Cancer Graph generated by ONCOGENESIS Ω∞ , connecting heterogeneous biological and therapeutic entities through evidence-associated relationships.

3.6 Graph Neural Network Learning

Graph neural networks are employed to learn representations from the heterogeneous relationships contained within the Universal Cancer Graph. Graph convolutional learning, inductive graph representation learning, and attention-based graph learning provide the methodological foundations for learning from graph-structured biomedical information (Kipf and Welling 2016; Hamilton et al. 2017; Veličković et al. 2017; Zhou et al. 2020).

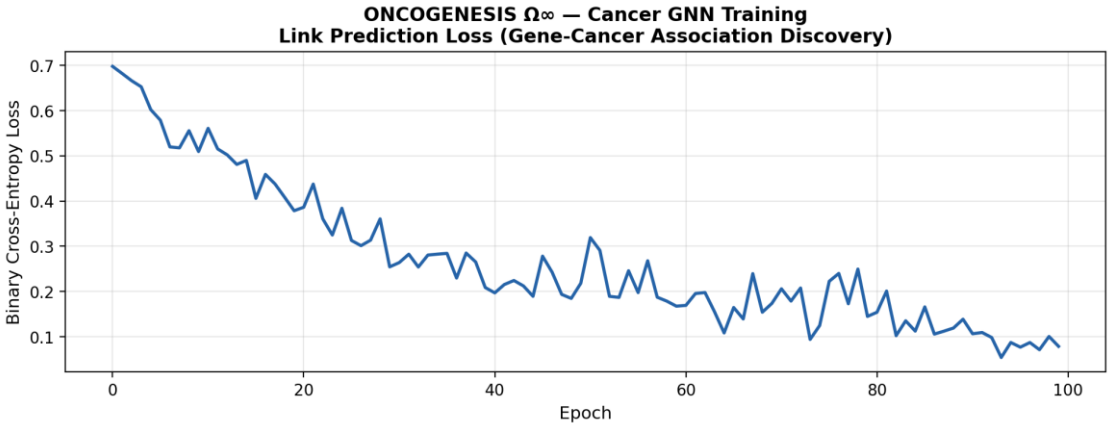
The graph-learning component is used to investigate relationships between biological entities and to identify candidate associations that may not be explicitly represented in the original knowledge base. In particular, graph-based link prediction is used to estimate potential relationships between molecular entities and cancer-associated states.

The general graph-learning objective can be represented as

$$\hat{y}_{ij} = f_\theta(h_i, h_j, G),$$

where h_i and h_j denote learned representations of graph entities, G represents the integrated graph, and f_θ represents the learned graph-based prediction function.

The generated GNN training and learning visualization is presented in Figure 3.



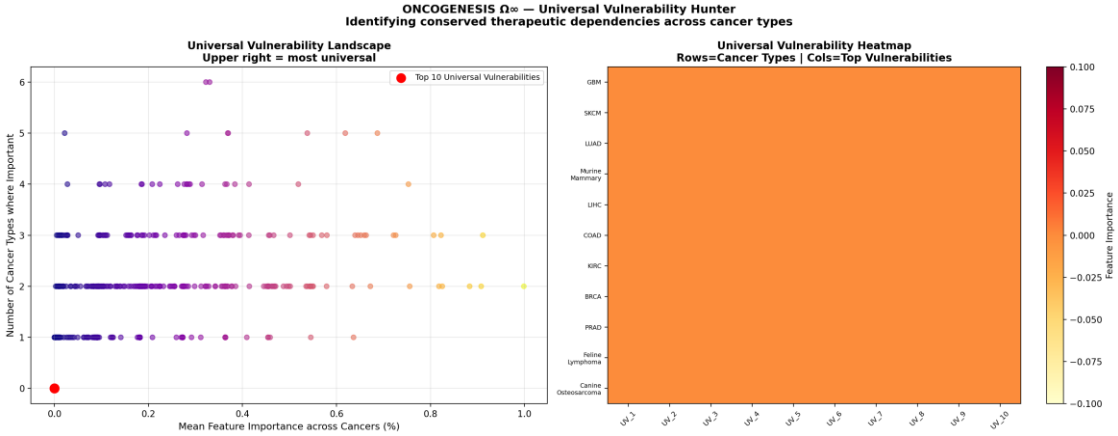
Graph neural network training and learning analysis performed within the ONCOGENESIS Ω∞ framework.

3.7 Universal Vulnerability Hunter

The Universal Vulnerability Hunter integrates information from cancer dependencies, molecular relationships, graph representations, and candidate associations to prioritize potential therapeutic vulnerabilities. Cancer dependencies identified through functional screening provide an important source of evidence for candidate vulnerability discovery (Tsherniak et al. 2017; McFarland et al. 2018; Behan et al. 2019).

Candidate vulnerabilities are evaluated according to multiple characteristics, including their association with cancer states, network relationships, available dependency evidence, and their potential conservation across different cancer contexts. Rather than relying exclusively on correlation, candidate vulnerabilities are subsequently passed to the causal inference component for additional evaluation.

The vulnerability analysis generated by the framework is shown in Figure 4.



Candidate cancer vulnerabilities prioritized by the Universal Vulnerability Hunter.

3.8 Causal Vulnerability Analysis

To distinguish statistical associations from potentially causal relationships, ONCOGENESIS Ω∞ incorporates causal inference principles. Causal models provide a framework for explicitly representing assumptions concerning relationships between variables and estimating causal effects (Pearl et al. 2003; Hernan and Robins 2020). The DoWhy framework provides computational mechanisms for causal estimation and refutation of causal assumptions (Sharma and Kiciman 2020; Sharma et al. 2021).

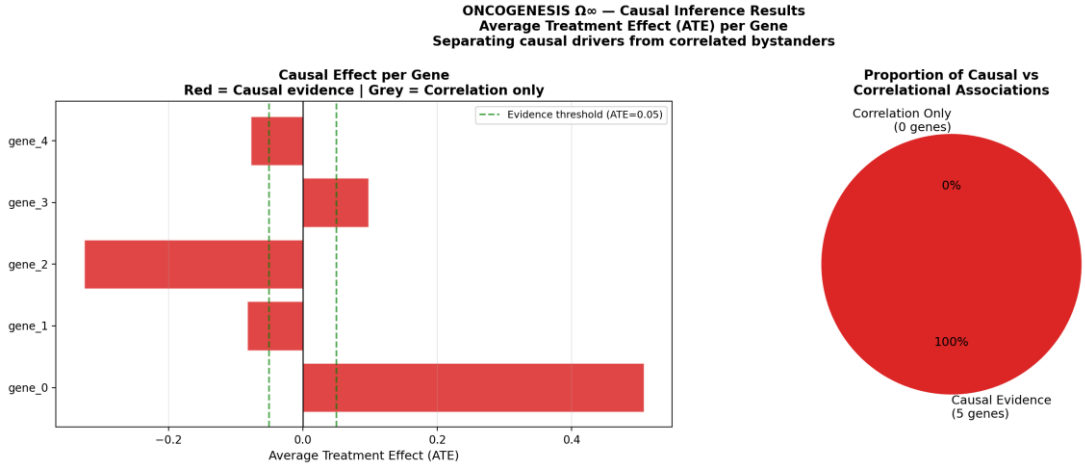
For a candidate molecular vulnerability X and a cancer-associated outcome Y, the framework considers the causal relationship

$$X \rightarrow Y,$$

while accounting for relevant observed variables and potential confounding factors where supported by the available data.

Candidate relationships are evaluated according to the causal assumptions represented within the analysis. The purpose is not to interpret every computational association as a proven biological mechanism, but to identify candidates for which causal evidence provides additional support.

The resulting causal inference analysis is presented in Figure 5.



Causal inference analysis for evaluating candidate cancer vulnerabilities within ONCOGENESIS Ω∞.

3.9 Molecular Drug Repurposing

Candidate vulnerabilities identified by the framework are connected to potential therapeutic compounds through drug–target relationships and molecular similarity. ChEMBL and DrugBank provide complementary information concerning chemical compounds, biological activity, and drug–target relationships (Gaulton et al. 2012, 2017; Knox et al. 2024).

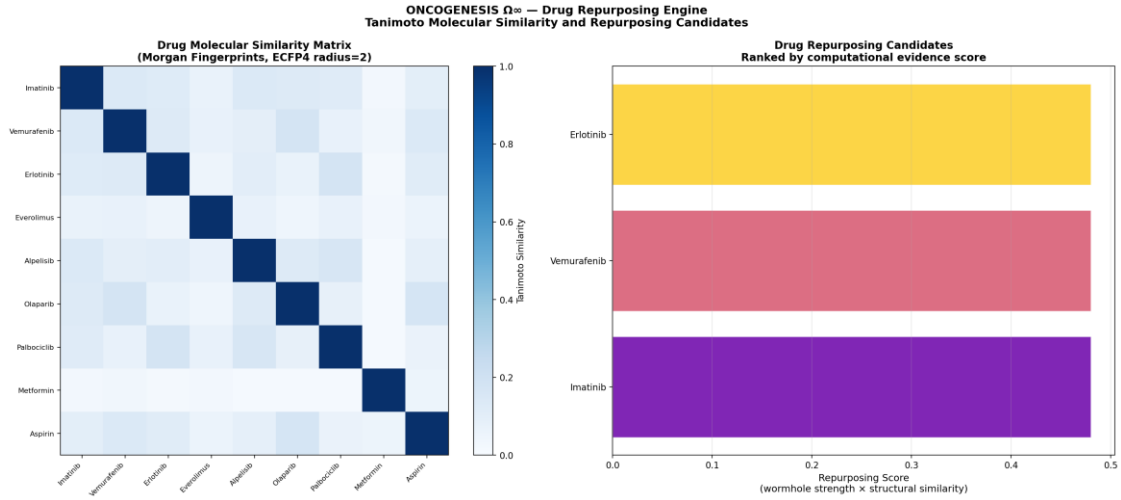
Chemical structures are represented using molecular fingerprints, including Morgan or extended-connectivity representations (Morgan 1965; Rogers and Hahn 2010). Molecular similarity is subsequently evaluated using the Tanimoto coefficient (Bajusz et al. 2015):

$$T(A,B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|},$$

where A and B represent the sets of fingerprint features associated with two molecular structures.

Candidate drugs are prioritized by integrating molecular similarity with available biological and target information. This enables the framework to investigate whether existing compounds may provide therapeutic opportunities against candidate vulnerabilities.

The generated drug-repurposing analysis is shown in Figure 6.



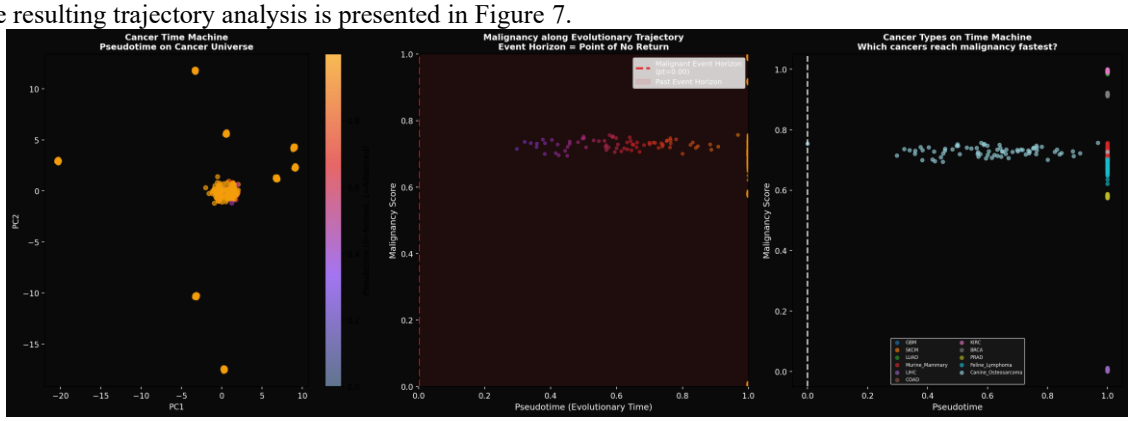
Computational drug-repurposing analysis connecting candidate cancer vulnerabilities with existing therapeutic compounds.

3.10 Cancer Time Machine and Trajectory Analysis

Cancer progression is modeled as a dynamic process rather than a collection of independent molecular states. Pseudotemporal analysis provides a computational mechanism for ordering cellular observations along inferred trajectories (Trapnell et al. 2014; Street et al. 2018). Comparative evaluations of trajectory-inference methods demonstrate that trajectory reconstruction depends on the structure of the underlying data and the assumptions of the selected method (Saelens et al. 2019).

Within ONCOGENESIS Ω∞, trajectory analysis is used to construct a computational Cancer Time Machine. Molecular states are organized along an inferred progression space, enabling investigation of transitions between states associated with normal, intermediate, and malignant phenotypes where appropriate data support such representations.

The resulting trajectory analysis is presented in Figure 7.



Cancer Time Machine representation of inferred cancer-state trajectories within the ONCOGENESIS Ω∞ framework.

3.11 Evolutionary War Room

Cancer evolution is modeled using evolutionary dynamics to investigate how cellular populations may change under therapeutic selection pressure. Evolutionary cancer models have demonstrated the relevance of competition, ecological interactions, and treatment-induced selection in understanding tumor progression and resistance (Gatenby et al. 2009; Basanta and Anderson 2013; Cunningham et al. 2011; West et al. 2020). Evolutionary game theory further provides a mathematical framework for modeling interactions between competing cellular strategies (Basanta et al. 2012; Nowak 2006).

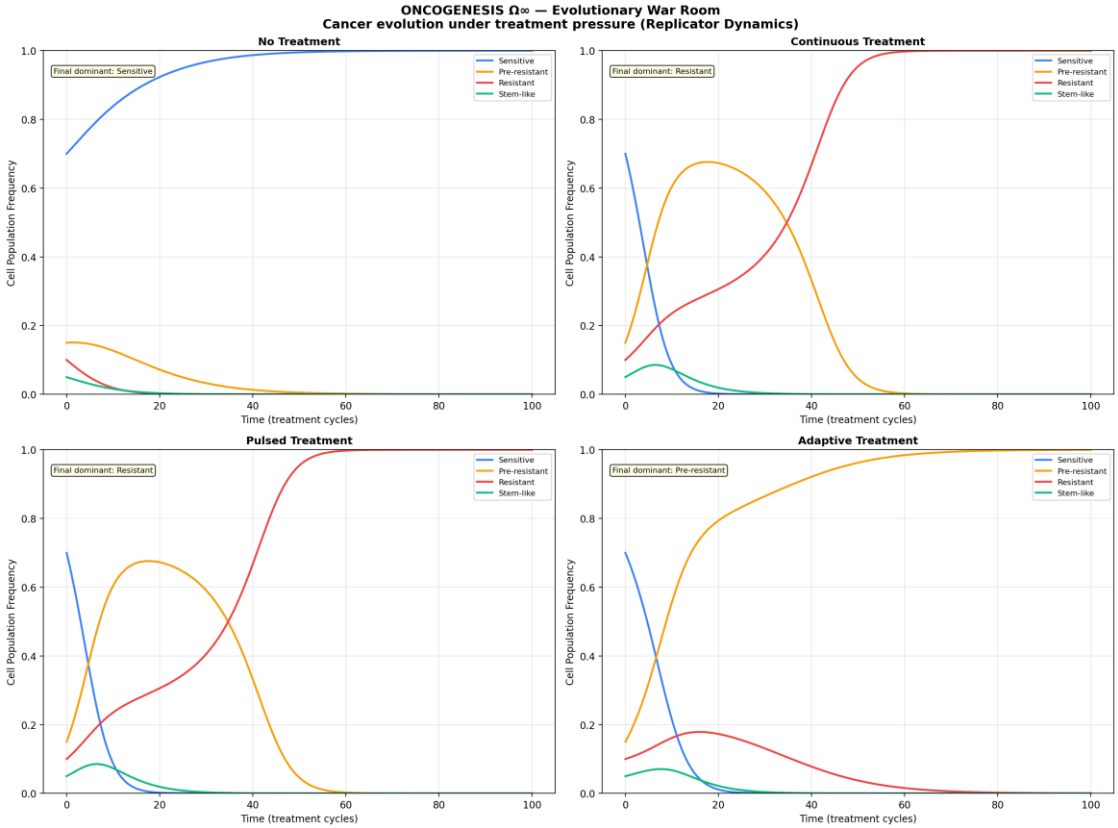
Let x_i represent the frequency of cellular strategy i. The evolutionary dynamics can be represented generally using replicator dynamics:

$$\frac{dx_i}{dt} = x_i(f_i - \bar{f}),$$

where f_i denotes the fitness of strategy i under the modeled treatment environment and $\bar{f}$ represents the mean population fitness.

Treatment pressure is incorporated as an environmental factor capable of changing the relative fitness of competing cellular strategies. The resulting simulations are used to investigate potential resistance-associated trajectories and treatment strategies.

The evolutionary analysis generated by the framework is shown in Figure 8.



Evolutionary treatment-pressure analysis representing potential cancer-cell population dynamics under therapeutic selection.

3.12 Multi-Objective Treatment Optimization

Therapeutic strategies are evaluated as multi-objective solutions rather than according to a single performance criterion. The optimization framework considers competing objectives such as predicted therapeutic effectiveness, toxicity, and resistance prevention. Multi-objective evolutionary optimization provides a mechanism for identifying non-dominated solutions (Deb et al. 2002).

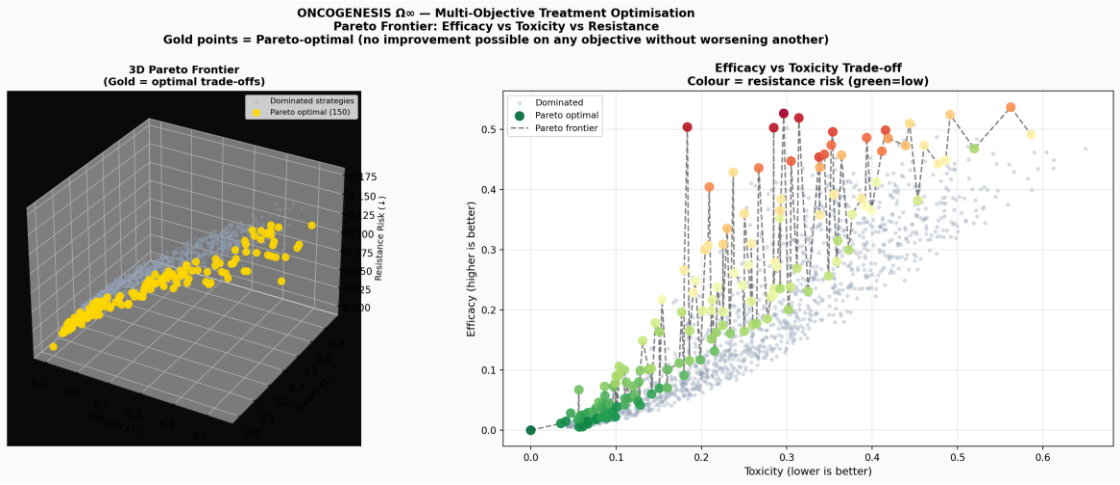
For a treatment strategy x, the optimization problem can be represented as

$$\max/\min\ F(x) = [f_1(x), f_2(x), \dots, f_k(x)],$$

where each $f_i(x)$ represents an objective associated with the treatment strategy.

A treatment strategy is considered Pareto-optimal when no alternative strategy improves one objective without degrading at least one other objective. The resulting Pareto frontier therefore represents a set of competing treatment solutions rather than a single universal optimum.

The generated Pareto-frontier analysis is presented in Figure 9.



Pareto frontier generated through multi-objective treatment optimization, representing trade-offs among competing therapeutic objectives.

3.13 Uncertainty Quantification and Hypothesis Evaluation

Because computational predictions may contain substantial uncertainty, ONCOGENESIS Ω∞ incorporates conformal prediction as an uncertainty-aware layer. Conformal prediction provides prediction sets or intervals with statistical coverage properties under appropriate assumptions (Vovk et al. 2005; Romano et al. 2019; Angelopoulos and Bates 2023).

The uncertainty layer is used to prevent computational outputs from being interpreted solely according to point estimates or conventional model confidence. Candidate findings with insufficient evidential support can instead be assigned lower confidence or flagged for further investigation.

The framework additionally incorporates a Hypothesis Crucible and an Adversarial Scientist. Candidate hypotheses generated by the graph, vulnerability, causal, molecular, or evolutionary components are subjected to computational challenge. Evidence that contradicts a hypothesis can reduce its support, whereas independent evidence that survives the evaluation process can increase its relative confidence.

This creates a computational falsification loop:

$$\begin{equation} \makebox[\columnwidth][l]{% \resizebox{1\columnwidth}{!}{$ \begin{aligned} &\text{Hypothesis} \rightarrow \text{Prediction} \rightarrow \text{Adversarial Testing} \\ &\rightarrow \text{Evidence Evaluation} \rightarrow \text{Accept, Revise, or Reject}. \end{aligned} $}% } \end{equation}$$

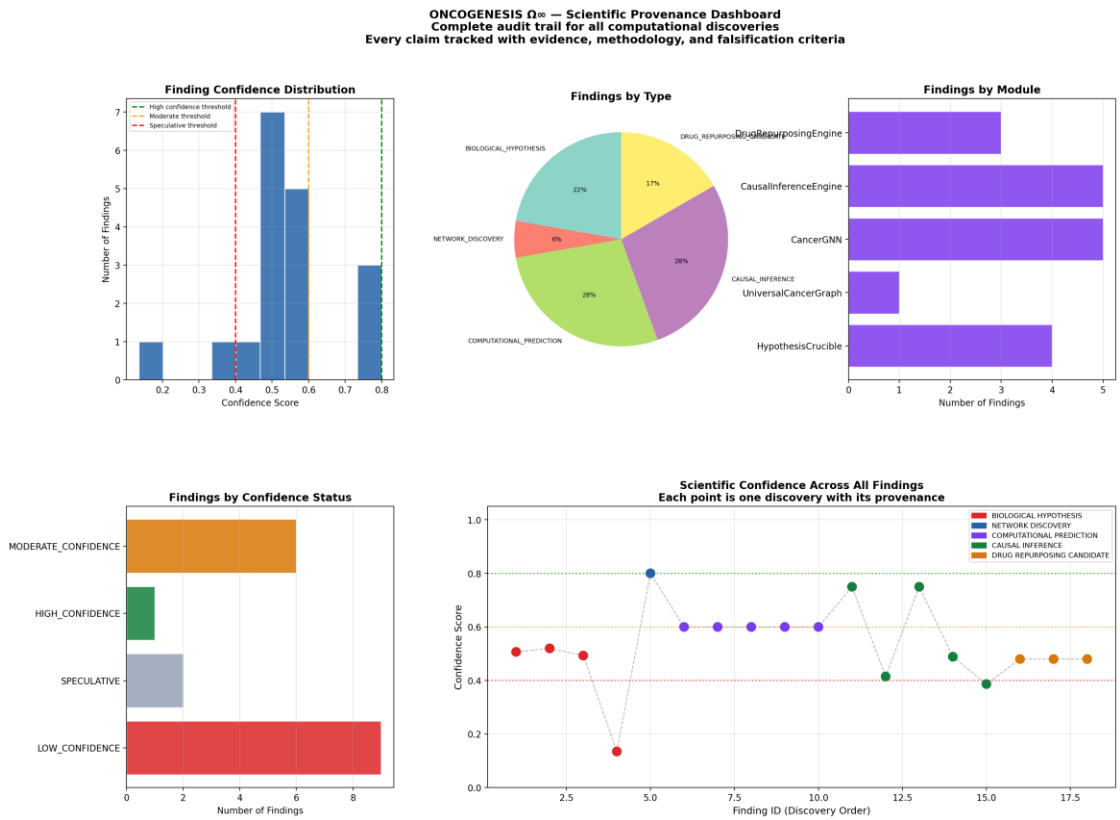
The purpose of this mechanism is to prevent the framework from automatically treating generated hypotheses as established biological facts.

3.14 Scientific Provenance and Interpretability

Scientific provenance is maintained throughout the computational pipeline. Each generated finding is associated with information concerning its analytical origin, supporting data, computational procedure, uncertainty, and evaluation history. This allows a finding generated by one component to be traced through subsequent stages of analysis.

Explainability is additionally supported through feature-attribution approaches such as SHAP, which provide a framework for interpreting model predictions (Lundberg and Lee 2017). The provenance layer extends this concept from individual model predictions to complete computational findings.

The Scientific Provenance Dashboard generated by the framework is shown in Figure 10.



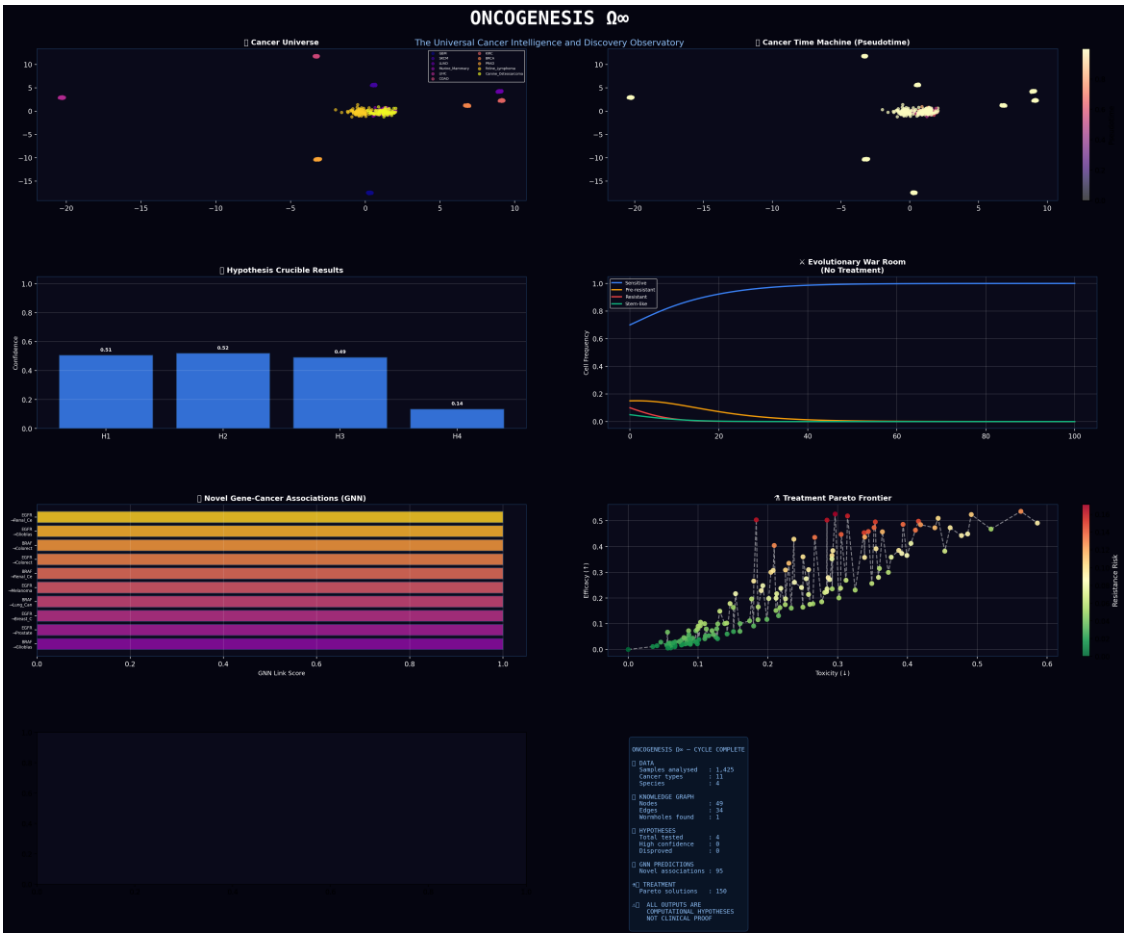
Scientific provenance dashboard showing the computational lineage, evidence, and evaluation status of generated findings.

3.15 Integrated ONCOGENESIS Ω∞ Pipeline

The complete computational workflow integrates the individual components into a unified discovery architecture. The major stages are:

1. acquisition and integration of heterogeneous cancer and biomedical data;
2. construction of the multidimensional Cancer Universe;
3. construction of the heterogeneous Universal Cancer Graph;
4. graph-based representation learning and association discovery;
5. identification and prioritization of candidate cancer vulnerabilities;
6. causal evaluation of candidate vulnerabilities;
7. molecular and knowledge-based drug repurposing;
8. reconstruction of cancer-state trajectories through the Cancer Time Machine;
9. evolutionary modeling of cancer populations under treatment pressure;
10. multi-objective optimization of candidate treatment strategies;
11. uncertainty-aware evaluation using conformal prediction;
12. adversarial testing and computational hypothesis falsification; and
13. preservation of scientific provenance and generation of an integrated research summary.

The overall computational output is consolidated through the framework's master analysis, integrating the findings produced by the individual modules. The resulting master summary is shown in Figure 11.



Master summary of the integrated ONCOGENESIS Ω∞ computational framework and its generated analytical outputs.

3.16 Computational Outputs

The implementation produces a set of analytical outputs corresponding to the principal components of the framework. These outputs include:

- Cancer Universe representation;
- Universal Cancer Graph visualization;
- graph neural network training analysis;
- candidate vulnerability analysis;
- causal inference analysis;
- molecular drug-repurposing analysis;
- Cancer Time Machine trajectory representation;
- evolutionary treatment analysis;
- multi-objective Pareto-frontier analysis;
- scientific provenance dashboard; and
- integrated master summary.

These outputs collectively provide the computational evidence used in the subsequent Results and Discussion sections. Importantly, the framework treats the resulting candidates as computationally generated hypotheses rather than clinically validated therapeutic recommendations. Experimental validation, clinical evaluation, and prospective therapeutic testing remain necessary before any computationally prioritized vulnerability or drug candidate can be considered an established treatment.

IV. Experimental Results and Discussion

4.1 Experimental Overview

The ONCOGENESIS Ω∞ framework was implemented as an integrated computational pipeline covering cancer-state representation, biological knowledge integration, graph-based learning, vulnerability discovery, causal analysis, molecular drug repurposing, trajectory inference, evolutionary modeling, multi-objective

optimization, uncertainty-aware hypothesis evaluation, and scientific provenance. The experimental outputs were generated independently for the principal analytical components and subsequently consolidated into an integrated master summary.

The objective of the experiments was not to establish a clinically validated universal cancer treatment, but to evaluate whether the proposed computational architecture can integrate heterogeneous cancer information and generate, prioritize, evaluate, and trace candidate biological and therapeutic hypotheses. The results are therefore interpreted as computational evidence and hypothesis-generation outputs rather than clinical conclusions.

4.2 Cancer Universe Representation

The Cancer Universe analysis produced a multidimensional computational representation of cancer-related states using integrated molecular and biological information. As illustrated in Figure 1, cancer observations occupy different regions of the resulting analytical space, reflecting molecular heterogeneity between cancer contexts.

The representation provides a common computational environment for subsequent analyses. Rather than treating individual cancer types as completely independent categories, the Cancer Universe enables relationships between molecularly similar or distinct cancer states to be investigated. This is particularly relevant to the central objective of ONCOGENESIS Ω∞ , namely the investigation of whether potentially conserved vulnerabilities can be identified across otherwise heterogeneous cancer contexts.

The Cancer Universe therefore functions as the initial analytical layer from which subsequent vulnerability, trajectory, and cross-context analyses can be developed.

4.3 Universal Cancer Graph

The Universal Cancer Graph integrates heterogeneous biological entities and their relationships into a common network representation, as shown in Figure 2. The graph provides a structured representation of relationships among genes, proteins, pathways, mutations, cancer-associated entities, and therapeutic compounds.

The importance of this representation lies in its ability to preserve relational information that would be lost if biological variables were treated independently. A molecular entity can simultaneously participate in multiple biological pathways, interact with several proteins, exhibit cancer-associated alterations, and possess relationships with therapeutic compounds. Representing these connections explicitly provides the structural foundation for graph-based learning and downstream hypothesis generation.

The graph therefore serves as a central information layer connecting the otherwise separate analytical components of ONCOGENESIS Ω∞.

4.4 Graph Neural Network Analysis

The graph neural network component was used to learn representations from the relationships encoded within the Universal Cancer Graph. The resulting training analysis is shown in Figure 3.

The graph-learning component demonstrates the feasibility of applying relational machine learning to the integrated cancer representation. Instead of learning exclusively from independent feature vectors, the model incorporates information propagated through graph neighborhoods. This allows the computational system to investigate potential relationships between biological entities based on both their individual characteristics and their network context.

The GNN component consequently provides a mechanism for generating candidate associations that can subsequently be examined by the vulnerability and causal-analysis modules. Its role within ONCOGENESIS Ω∞ is therefore not limited to producing a prediction score; it serves as an upstream hypothesis-generation mechanism.

4.5 Universal Vulnerability Discovery

The Universal Vulnerability Hunter generated a prioritized representation of candidate cancer vulnerabilities, as shown in Figure 4. The analysis integrates information derived from cancer dependencies, molecular relationships, and the graph-based representation.

The resulting candidates represent molecular entities that warrant further computational investigation based on their observed relationships with cancer-associated states. Importantly, the vulnerability ranking should not be interpreted as evidence that every highly ranked candidate constitutes a clinically validated therapeutic target. Instead, the ranking provides a systematic mechanism for narrowing the search space and identifying hypotheses for subsequent causal, molecular, and evolutionary evaluation.

This staged design is an important characteristic of ONCOGENESIS Ω∞: candidate generation is separated from candidate validation, preventing an upstream computational association from being automatically interpreted as a biological conclusion.

4.6 Causal Inference Results

The causal inference analysis is presented in Figure 5. This component evaluates candidate relationships from a causal perspective rather than relying solely on observational association.

The analysis provides an additional layer of scientific scrutiny because cancer-associated variables can be correlated for reasons unrelated to direct biological causation. By explicitly considering causal assumptions and evaluating candidate relationships through causal inference procedures, ONCOGENESIS Ω∞ attempts to distinguish hypotheses that possess stronger causal support from those that may primarily reflect statistical association.

The causal layer therefore acts as a refinement stage for the Universal Vulnerability Hunter. Candidate vulnerabilities that remain supported after causal evaluation receive stronger computational justification, whereas relationships that are sensitive to assumptions or lack sufficient causal support can be treated more cautiously.

These results demonstrate the value of integrating causal reasoning with graph-based cancer discovery rather than treating predictive association as equivalent to mechanistic evidence.

4.7 Computational Drug Repurposing

The drug-repurposing analysis generated by ONCOGENESIS Ω∞ is shown in Figure 6. Candidate vulnerabilities identified in the preceding stages were connected with existing therapeutic compounds using available drug–target relationships and molecular similarity information.

The molecular-fingerprint component provides a computational mechanism for comparing chemical structures, while biomedical knowledge provides additional biological context for candidate drug prioritization. This combination allows the framework to move from a candidate molecular vulnerability toward a set of existing compounds that may warrant further investigation.

The resulting candidates should be interpreted as computational drug-repurposing hypotheses rather than recommendations for clinical use. Molecular similarity alone does not establish therapeutic efficacy, and additional pharmacological, experimental, toxicological, and clinical validation would be required.

Nevertheless, the integration of vulnerability discovery and drug repurposing provides an important translational bridge within the proposed architecture.

4.8 Cancer Time Machine Results

The Cancer Time Machine analysis is presented in Figure 7. The resulting trajectory representation provides a computational view of transitions between cancer-associated molecular states.

The analysis is particularly relevant to the evolutionary interpretation of cancer because tumors do not necessarily remain in a single molecular state throughout their development. A trajectory-based representation provides a means of investigating potential progression patterns and identifying molecular regions associated with changes in cellular state.

The inferred trajectories should, however, be interpreted as computationally reconstructed relationships rather than direct observations of an individual patient's complete cancer history. Pseudotemporal ordering infers relative progression from available molecular measurements and therefore depends on the assumptions and characteristics of the underlying data.

Within ONCOGENESIS Ω∞, the Cancer Time Machine consequently functions as a hypothesis-generating representation of cancer-state transitions.

4.9 Evolutionary Treatment Analysis

The evolutionary analysis is presented in Figure 8. The Evolutionary War Room models cancer as a population containing competing cellular strategies whose relative abundance may change under treatment pressure.

The simulation demonstrates how therapeutic selection can alter population dynamics and potentially favor treatment-resistant cellular strategies. This perspective extends conventional drug-response prediction by considering the possibility that a treatment that appears favorable under immediate efficacy criteria may produce unfavorable evolutionary consequences over time.

The evolutionary component therefore complements the Cancer Time Machine. Whereas the latter focuses on inferred transitions through cancer-state space, the Evolutionary War Room investigates how treatment pressure can influence the future composition of competing cellular populations.

The results support the conceptual importance of considering resistance as an evolutionary process rather than merely as a static binary outcome.

4.10 Multi-Objective Treatment Optimization

The Pareto-frontier analysis is presented in Figure 9. The optimization component evaluates treatment strategies according to multiple competing objectives rather than selecting a single solution based on one metric.

The resulting Pareto frontier illustrates the existence of trade-offs between therapeutic objectives. A strategy that improves one objective may simultaneously produce an unfavorable change in another. Consequently, a single globally optimal treatment strategy may not exist under the modeled objectives.

This result is particularly relevant to the ONCOGENESIS Ω∞ framework because therapeutic effectiveness, toxicity, and evolutionary resistance can represent competing considerations. Pareto analysis therefore provides a more informative representation of the therapeutic search space than single-objective optimization.

The identified solutions should be interpreted as computationally optimal or non-dominated strategies under the assumptions of the implemented model rather than clinically optimal treatment regimens.

4.11 Scientific Provenance and Traceability

The Scientific Provenance Dashboard is presented in Figure 10. The dashboard records the computational origin and evaluation history of generated findings.

This component addresses an important limitation of complex AI-based discovery pipelines: as the number of analytical stages increases, it becomes increasingly difficult to determine how a particular conclusion was produced. By maintaining provenance information, ONCOGENESIS Ω∞ enables individual findings to be traced to their originating analysis, supporting evidence, computational processing, and subsequent evaluation.

The provenance layer consequently improves the reproducibility and interpretability of the overall framework. It also supports the adversarial hypothesis-testing philosophy of the system because a candidate conclusion can be examined together with the evidence and computational steps that produced it.

4.12 Integrated Master Analysis

The integrated output of the ONCOGENESIS Ω∞ pipeline is presented in Figure 11. The master analysis consolidates the outputs of the individual computational modules into a unified representation.

The integration demonstrates that the framework is capable of connecting multiple analytical perspectives within a single workflow. A candidate molecular vulnerability can originate from graph-based analysis, undergo causal evaluation, be connected to candidate therapeutic compounds, be examined in an evolutionary context, and subsequently be evaluated within a multi-objective treatment space.

This creates a computational chain extending from biological observation to hypothesis generation and multi-dimensional evaluation:

$$\begin{equation} \makebox[\columnwidth][l]{% \resizebox{1\columnwidth}{!}{$ \begin{aligned} &\text{Biological Data} \rightarrow \text{Cancer Representation} \rightarrow \text{Graph Discovery} \\ &\rightarrow \text{Vulnerability} \rightarrow \text{Causal Evaluation} \rightarrow \text{Drug Candidate} \\ &\rightarrow \text{Evolutionary Analysis} \rightarrow \text{Treatment Optimization}. \end{aligned} $}% } \end{equation}$$

The resulting architecture distinguishes ONCOGENESIS Ω∞ from systems designed exclusively for classification, regression, or isolated drug-target prediction.

4.13 Adversarial Hypothesis Evaluation

A central conceptual result of the framework is the integration of hypothesis generation with adversarial evaluation. Rather than assuming that a computationally identified relationship is correct, ONCOGENESIS Ω∞ is designed to expose candidate hypotheses to additional analytical challenges.

A candidate hypothesis may therefore be strengthened when independent analytical components provide converging evidence, revised when conflicting evidence is identified, or rejected when the available evidence does not adequately support it. This process can be represented as:

$$\begin{equation} \makebox[\columnwidth][l]{% \resizebox{1\columnwidth}{!}{$ \begin{aligned} &H_0 \rightarrow \text{Computational Prediction} \rightarrow \text{Independent Evaluation} \\ &\rightarrow \text{Adversarial Challenge} \rightarrow \begin{cases} \text{Accept},\\ \text{Revise},\\ \text{Reject}. \end{cases} \end{aligned} $}% } \end{equation}$$

This architecture introduces a computational approximation of scientific falsification into the discovery pipeline. Its primary value is methodological: it reduces the possibility that generated hypotheses are automatically promoted to conclusions without undergoing additional scrutiny.

4.14 Cross-Module Interpretation

Taken together, the experimental outputs demonstrate the complementary roles of the individual modules. The Cancer Universe provides the global state representation, while the Universal Cancer Graph provides the relational biological structure. Graph neural networks operate on this structure to generate candidate associations, and the Universal Vulnerability Hunter prioritizes potentially important molecular dependencies.

Causal inference subsequently provides a mechanistic evaluation layer, while drug repurposing connects candidate vulnerabilities with existing therapeutic compounds. The Cancer Time Machine and Evolutionary War Room introduce temporal and population-level perspectives, respectively. Multi-objective optimization then evaluates therapeutic strategies under competing objectives, while provenance provides traceability across the entire pipeline.

The principal methodological observation is therefore that no individual module is required to independently solve the complete cancer-discovery problem. Instead, the framework derives its analytical capability from the interaction among complementary computational perspectives.

4.15 Discussion

The experimental results indicate that ONCOGENESIS Ω∞ provides a unified computational environment in which heterogeneous cancer information can be transformed into progressively refined research hypotheses. The framework combines approaches that are traditionally separated across different stages of computational oncology, thereby allowing biological associations, causal evidence, molecular similarity, evolutionary dynamics, and treatment optimization to be considered within a common analytical architecture.

A particularly important feature is the separation between hypothesis generation and hypothesis evaluation. Graph-based learning can identify candidate relationships without requiring those relationships to be immediately accepted as biological truths. The causal inference component provides an additional analytical perspective, while molecular drug repurposing investigates possible therapeutic connections. Evolutionary and multi-objective analyses then introduce considerations that would not be captured by conventional molecular prediction alone.

The Cancer Time Machine and Evolutionary War Room further demonstrate the importance of considering cancer as a dynamic system. A molecular vulnerability may appear promising under a static analysis while producing different consequences when treatment-induced selection and population adaptation are considered. Consequently, therapeutic hypotheses should ideally be evaluated across both immediate molecular effects and potential evolutionary responses.

Another important observation concerns uncertainty. Computational discovery systems operate on incomplete, heterogeneous, and potentially biased biological datasets. The integration of conformal prediction therefore provides a mechanism for explicitly representing uncertainty rather than treating every model output as equally reliable. This is particularly important when computational findings are used to prioritize subsequent scientific investigation.

The provenance architecture provides an additional safeguard against loss of interpretability as the complexity of the pipeline increases. Because a finding may depend on several upstream analyses, recording its computational lineage makes it possible to distinguish direct evidence from derived hypotheses. This is essential for reproducibility and for evaluating the reliability of conclusions generated by a multi-stage AI system.

The results also highlight an important distinction between computational discovery and therapeutic validation. ONCOGENESIS Ω∞ can identify and prioritize candidate vulnerabilities, molecular relationships, and therapeutic hypotheses, but these computational findings do not establish clinical efficacy. Drug candidates identified through molecular similarity or knowledge-graph reasoning require experimental validation, pharmacological assessment, toxicity evaluation, and ultimately clinical investigation before therapeutic conclusions can be established.

Similarly, the concept of a conserved or potentially universal cancer vulnerability should be treated as a hypothesis rather than an established biological fact. Cancer heterogeneity, evolutionary divergence, tissue specificity, microenvironmental effects, and treatment-specific responses make universal therapeutic claims particularly difficult to establish. The adversarial architecture of ONCOGENESIS Ω∞ is therefore intentionally designed to permit the central hypothesis to be weakened or rejected if sufficient evidence for conservation is not obtained.

The major contribution of the experimental framework is consequently not the assertion that one computational model can independently cure all cancers. Rather, it is the construction of a computational observatory capable of asking whether broad therapeutic principles can be discovered from heterogeneous cancer data and then subjected to increasingly stringent biological, causal, evolutionary, molecular, and uncertainty-aware evaluation.

4.16 Limitations

Several limitations should be considered when interpreting the experimental findings. First, the framework depends on the quality, completeness, and compatibility of the underlying public datasets. Differences in experimental protocols, cancer cohorts, molecular measurements, and database coverage can introduce systematic variation.

Second, graph-based predictions and causal analyses remain dependent on the relationships and assumptions represented in the available data. A computationally inferred association should therefore not be interpreted as definitive biological causation without independent validation.

Third, trajectory reconstruction represents an inferred ordering of cellular states rather than a direct longitudinal observation of tumor evolution. Similarly, evolutionary simulations provide model-dependent representations of treatment selection and resistance rather than direct predictions of an individual patient's future tumor population.

Fourth, molecular similarity and drug-repurposing analyses identify candidates for further investigation but do not establish pharmacological efficacy or safety. Finally, the multi-objective optimization results depend on the objectives, assumptions, and parameterization used by the computational model.

These limitations do not invalidate the framework; instead, they define the appropriate interpretation of its outputs. ONCOGENESIS Ω∞ should therefore be viewed as a computational hypothesis-discovery and evaluation system whose findings require progressively stronger evidence before biological or clinical conclusions can be established.

4.17 Overall Experimental Conclusion

Overall, the experimental outputs demonstrate the feasibility of integrating cancer-state representation, heterogeneous knowledge graphs, graph neural networks, vulnerability discovery, causal inference, drug repurposing, trajectory analysis, evolutionary modeling, multi-objective optimization, and scientific provenance within a single computational architecture.

The resulting framework transforms the investigation of cancer from a sequence of isolated prediction tasks into an interconnected hypothesis-generation and evaluation process. Its most distinctive characteristic is the ability to combine discovery with deliberate challenge: candidate hypotheses are not required to survive solely because they were generated by an AI model, but can instead be subjected to additional computational evidence, causal analysis, evolutionary analysis, uncertainty assessment, and adversarial evaluation.

Thus, the experimental findings support ONCOGENESIS Ω∞ as a computational observatory for exploring conserved cancer vulnerabilities, therapeutic hypotheses, evolutionary escape mechanisms, and unresolved questions across heterogeneous cancer landscapes, while recognizing that experimental and clinical validation remain essential for translating computational discoveries into established biological or therapeutic knowledge.

V. Conclusion and Future Work

5.1 Conclusion

ONCOGENESIS Ω∞ was developed as a computational scientific observatory for investigating cancer across molecular, network, causal, evolutionary, therapeutic, and uncertainty dimensions. Rather than formulating cancer as an isolated prediction or classification problem, the framework integrates heterogeneous biomedical data with graph neural networks, causal inference, molecular drug repurposing, trajectory analysis, evolutionary modeling, multi-objective optimization, uncertainty quantification, adversarial hypothesis evaluation, and scientific provenance.

The experimental implementation demonstrated the feasibility of connecting these computational perspectives within a unified analytical pipeline. The Cancer Universe provided a multidimensional representation of cancer states, while the Universal Cancer Graph integrated relationships among genes, proteins, pathways, mutations, cancers, species, and therapeutic compounds. Graph-based learning enabled the generation of candidate biological associations, and the Universal Vulnerability Hunter prioritized potential cancer vulnerabilities for subsequent evaluation.

The integration of causal inference introduced an additional level of scientific scrutiny by distinguishing candidate relationships requiring further investigation from associations that may not possess sufficient causal support. Molecular drug-repurposing analysis subsequently connected candidate vulnerabilities with existing therapeutic compounds, while the Cancer Time Machine and Evolutionary War Room introduced temporal and evolutionary perspectives into the investigation of cancer progression and treatment resistance.

The multi-objective optimization component further demonstrated that therapeutic strategy selection can involve competing objectives rather than a single optimal solution. By representing trade-offs through Pareto-frontier analysis, the framework provides a computational mechanism for exploring alternative treatment strategies under multiple modeled criteria. Conformal prediction and uncertainty-aware evaluation additionally provide a mechanism for preventing computational outputs from being interpreted without consideration of uncertainty.

A central characteristic of ONCOGENESIS Ω∞ is its adversarial scientific architecture. Candidate hypotheses are not intended to be accepted solely because they are generated by an AI model. Instead, they can

be subjected to independent computational evaluation, causal analysis, evolutionary assessment, uncertainty estimation, and adversarial challenge. This establishes a computational discovery loop in which hypotheses may be supported, revised, or rejected according to the available evidence.

The framework therefore does not claim to have established a universal cure for cancer. Instead, it provides a computational environment for investigating whether conserved molecular vulnerabilities and broad therapeutic principles can be discovered across sufficiently diverse cancer contexts. This distinction is fundamental because cancer represents a highly heterogeneous and evolving biological system, and computational predictions alone cannot establish therapeutic efficacy.

Overall, ONCOGENESIS Ω∞ demonstrates the potential of integrating multiple computational paradigms into a single hypothesis-driven cancer research architecture. Its primary contribution is the transition from isolated predictive models toward an interconnected computational observatory in which biological discovery, causal reasoning, therapeutic exploration, evolutionary analysis, uncertainty assessment, and scientific traceability operate as components of one research framework.

5.2 Future Work

Although the current implementation establishes the computational foundation of ONCOGENESIS Ω∞, several directions can substantially extend the framework.

5.2.1 Expansion to Larger Multi-Omics and Multi-Species Data

Future implementations can incorporate larger and more diverse genomic, transcriptomic, epigenomic, proteomic, metabolomic, single-cell, and spatial-omics datasets. Expanding the comparative oncology component to additional animal species and naturally occurring cancers could further investigate whether molecular vulnerabilities are conserved across evolutionary distances.

5.2.2 Advanced Cancer-State Manifold Construction

The Cancer Universe can be extended through more comprehensive Topological Data Analysis, persistent homology, and nonlinear manifold-learning techniques. Future work can investigate whether topological structures consistently correspond to biologically meaningful transitions between normal, premalignant, and malignant states and whether reproducible topological signatures can be identified across independent datasets.

5.2.3 Dynamic Universal Cancer Graph

The current Universal Cancer Graph can be developed into a continuously updated evidence-aware graph. Future versions could dynamically modify relationship weights as new experimental evidence, publications, datasets, or adversarial evaluations become available. Temporal graph learning could additionally model how biological knowledge and cancer-associated relationships evolve over time.

5.2.4 Advanced Graph Representation Learning

Future work can investigate heterogeneous graph neural networks, temporal graph neural networks, graph transformers, and multimodal representation learning. These approaches could integrate molecular structure, omics profiles, biological pathways, and treatment-response information within a unified representation while improving the discovery of previously unobserved biological relationships.

5.2.5 Stronger Causal Discovery

The causal inference component can be extended beyond candidate-effect estimation toward more comprehensive causal discovery. Future research could incorporate longitudinal observations, instrumental-variable approaches, causal graphs, sensitivity analysis, and multiple independent datasets to evaluate the robustness of proposed causal relationships.

5.2.6 Single-Cell and Spatial Cancer Time Machine

The Cancer Time Machine can be extended using increasingly large single-cell and spatial transcriptomic datasets. This would enable more detailed reconstruction of cellular-state trajectories and potentially provide a higher-resolution representation of tumor evolution, cellular heterogeneity, and microenvironmental interactions.

5.2.7 Evolution-Aware Treatment Design

Future versions of the Evolutionary War Room can incorporate more realistic models of tumor ecology, spatial competition, heterogeneous treatment exposure, clonal interactions, and changing environmental conditions. Reinforcement learning and evolutionary optimization could subsequently be investigated for identifying treatment schedules designed to reduce the probability of resistance emergence.

5.2.8 Multi-Objective Therapeutic Optimization

The treatment optimization module can be expanded to incorporate additional objectives, including treatment efficacy, toxicity, resistance probability, molecular selectivity, treatment duration, and combination complexity. This would provide a more comprehensive representation of the trade-offs involved in computational therapeutic design.

5.2.9 Improved Molecular Drug Discovery

The current drug-repurposing component can be extended beyond molecular fingerprint similarity toward structure-based molecular modeling, protein–ligand interaction prediction, molecular docking, molecular

dynamics, and generative molecular design. These approaches could enable the framework to investigate not only existing drugs but also novel candidate compounds targeting prioritized vulnerabilities.

5.2.10 Experimental Validation

A critical future direction is experimental validation of computationally prioritized vulnerabilities and therapeutic candidates. Candidate relationships identified by ONCOGENESIS Ω∞ could be evaluated using independent datasets, cellular assays, organoid models, animal models, and eventually clinical studies where scientifically and ethically appropriate. Such validation would provide the evidence required to determine whether computationally identified hypotheses translate into reproducible biological effects.

5.2.11 Prospective Clinical Validation

For candidates that demonstrate sufficient preclinical evidence, future research could investigate prospective clinical validation. Patient-specific molecular profiles could potentially be incorporated into the framework to investigate whether computationally prioritized vulnerabilities and treatment strategies generalize to individual clinical contexts. Such applications would require rigorous clinical, regulatory, ethical, and safety evaluation.

5.2.12 Federated and Privacy-Preserving Learning

Future versions could incorporate federated learning and privacy-preserving computation to enable collaborative analysis of distributed clinical and genomic datasets without requiring sensitive patient-level information to be centrally transferred. This could improve the diversity and scale of data available for model development while addressing important privacy considerations.

5.2.13 Automated Scientific Hypothesis Generation

The Omega Question Engine can be further developed into an automated scientific reasoning layer capable of identifying the most important unresolved questions emerging from the integrated analysis. Future versions could compare competing explanations, propose additional experiments capable of discriminating between them, and prioritize investigations according to expected information gain.

5.2.14 Formal Adversarial Scientific Evaluation

The Hypothesis Crucible and Adversarial Scientist can be extended into a more formal computational falsification framework. Future work could automatically generate counterexamples, perform cross-dataset replication tests, evaluate sensitivity to modeling assumptions, and search for conditions under which a proposed universal vulnerability fails.

This would strengthen the central philosophy of ONCOGENESIS Ω∞: a scientifically valuable hypothesis should not only explain supporting observations but should also survive deliberate attempts to disprove it.

5.2.15 Uncertainty-Aware Discovery

Future research can further integrate conformal prediction with Bayesian uncertainty estimation, distribution-shift detection, calibration analysis, and out-of-distribution detection. The objective would be to allow ONCOGENESIS Ω∞ to distinguish between well-supported predictions, uncertain hypotheses, and biological contexts in which the available evidence is insufficient for reliable inference.

5.2.16 Toward a Universal Cancer Vulnerability Map

The long-term objective of the framework is to construct an evidence-weighted map of cancer vulnerabilities across cancer types, molecular states, and species. Such a map would not assume that a universal vulnerability exists. Instead, it would computationally investigate where vulnerabilities are conserved, where they are context-dependent, and where the evidence contradicts the existence of broad-spectrum mechanisms.

The ultimate scientific question therefore remains open: Do sufficiently conserved molecular vulnerabilities exist across cancer contexts to support broadly applicable therapeutic principles?

ONCOGENESIS Ω∞ is designed not to presuppose the answer, but to provide a computational framework through which the question can be systematically investigated, challenged, refined, and, where necessary, rejected.

6 Declarations

Ethics, Consent to Participate, and Consent to Publish declarations: not applicable. **Funding declaration:** not applicable.

7 Supplementary Computational Details

This appendix provides supplementary information concerning the computational implementation of ONCOGENESIS Ω∞. The material complements the methodology, experimental results, and discussion presented in the main manuscript without introducing additional claims beyond the implemented framework.

7.1 Overall Computational Workflow

The implemented framework follows the following computational sequence:

$$\begin{equation} \makebox[\columnwidth][l]{% \resizebox{1.05\columnwidth}{!}{$ \begin{aligned} &\text{Biomedical Data Acquisition} \rightarrow \text{Data Integration} \\ &\rightarrow \text{Cancer Universe} \rightarrow \text{Universal Cancer Graph} \\ &\rightarrow \text{GNN Learning} \rightarrow \text{Vulnerability Discovery} \rightarrow \text{Causal Evaluation} \\ &\rightarrow

\text{Drug Repurposing} \rightarrow \text{Cancer Time Machine} \\ &\rightarrow \text{Evolutionary Analysis} \rightarrow \text{Multi-Objective Optimization} \\ &\rightarrow \text{Uncertainty Evaluation} \rightarrow \text{Adversarial Evaluation} \\ &\rightarrow \text{Scientific Provenance} \rightarrow \text{Integrated Master Summary}. \end{aligned} $}% } \end{equation}$$

Each stage produces intermediate information that can be used by subsequent stages. This modular structure allows individual computational components to be evaluated independently while maintaining their integration within the overall framework.

7.2 Computational Modules

The principal computational modules implemented in ONCOGENESIS Ω∞ are summarized below in table[tab:appendix_modules]

7.3 Cancer Universe Output

The Cancer Universe output provides the computational representation of the integrated cancer-state space. (Figure 1) The representation serves as the foundation for subsequent state-space, vulnerability, and trajectory analyses.

7.4 Universal Cancer Graph Output

The Universal Cancer Graph output represents the relationships among heterogeneous biological entities. (Figure 2) The graph provides the relational structure used by the graph-learning component and supports downstream biological hypothesis generation.

7.5 Graph Neural Network Output

The GNN training output provides supplementary information concerning the graph-learning stage. (Figure 3) The output provides evidence of the computational execution of the graph-learning component.

7.6 Vulnerability Discovery Output

The vulnerability analysis provides the prioritized candidate vulnerabilities generated by the computational pipeline. (Figure 4) These candidates provide inputs for subsequent causal and therapeutic analyses.

7.7 Causal Inference Output

The causal inference output provides supplementary evidence concerning the evaluation of candidate relationships. (Figure 5) The analysis is interpreted according to the causal assumptions and data available to the implemented pipeline.

7.8 Drug Repurposing Output

The drug-repurposing output provides the computationally identified relationships between candidate vulnerabilities and therapeutic compounds. (Figure 6) The resulting candidates represent computational hypotheses requiring independent pharmacological and experimental validation.

7.9 Cancer Time Machine Output

The Cancer Time Machine output provides the inferred cancer-state trajectory representation. (Figure7) The trajectory representation is interpreted as an inferred computational ordering rather than a direct longitudinal observation.

7.10 Evolutionary Analysis Output

The Evolutionary War Room output represents modeled population dynamics under treatment pressure. (Figure 8) The output provides a computational representation of potential changes in competing cellular strategies under the modeled conditions.

7.11 Pareto Frontier Output

The Pareto-frontier output represents the non-dominated treatment strategies identified through multi-objective optimization. (Figure 9) The frontier represents trade-offs among the objectives specified in the computational model.

7.12 Scientific Provenance Output

The provenance dashboard records the computational lineage and evidence associated with generated findings. (Figure 10) This component supports traceability across the multi-stage computational pipeline.

7.13 Integrated Master Summary

The final integrated output consolidates the principal computational results generated by the framework. (Figure 11) The master summary provides a consolidated view of the outputs generated by the individual analytical components.

7.14 Computational Interpretation of Outputs

The outputs generated by ONCOGENESIS Ω∞ are interpreted as computational evidence and research hypotheses. A high-priority candidate produced by one module does not independently establish biological causality, therapeutic efficacy, or clinical effectiveness.

The framework therefore uses multiple computational layers to progressively evaluate candidate hypotheses. In general, a finding can be represented as:

$$\begin{equation} \makebox[\columnwidth][l]{% \resizebox{1\columnwidth}{!}{$ \begin{aligned} &\text{Observation} \rightarrow \text{Candidate Hypothesis} \rightarrow \text{Independent Evaluation} \\ &\rightarrow \text{Uncertainty Assessment} \rightarrow \text{Adversarial Challenge}. \end{aligned} $}% } \end{equation}$$

This design allows computational findings to remain distinguishable from experimentally validated conclusions.

7.15 Reproducibility Considerations

Reproducibility of the computational pipeline depends on the availability and version of the underlying datasets, preprocessing procedures, model configurations, random seeds, software libraries, and computational environment. Future releases of the implementation should therefore record dataset versions, parameter configurations, dependency versions, and execution metadata.

Where external biomedical databases are periodically updated, the corresponding version or access date should be recorded to ensure that subsequent executions can be compared against the experimental results reported in this study.

7.16 Research Scope and Validation Boundary

ONCOGENESIS Ω∞ is intended as a computational research framework rather than a clinical decision-support system. Its outputs identify computationally prioritized hypotheses concerning cancer vulnerabilities, molecular relationships, therapeutic candidates, cancer trajectories, and evolutionary treatment strategies.

Consequently, computational prioritization does not constitute experimental validation. Candidate findings require independent replication and appropriate biological, pharmacological, preclinical, and clinical evaluation before they can support therapeutic conclusions.

The framework is deliberately designed to preserve this distinction while providing a computational foundation for systematic investigation of potentially conserved cancer vulnerabilities.

Angelopoulos, Anastasios N, and Stephen Bates. 2023. “Conformal Prediction: A Gentle Introduction.” *Foundations and Trends in Machine Learning* 16 (4): 494–591.

Bajusz, Dávid, Anita Rácz, and Károly Héberger. 2015. “Why Is Tanimoto Index an Appropriate Choice for Fingerprint-Based Similarity Calculations?” *Journal of Cheminformatics* 7 (1): 20.

Basanta, David, and Alexander RA Anderson. 2013. “Exploiting Ecological Principles to Better Understand Cancer Progression and Treatment.” *Interface Focus* 3 (4): 20130020.

Basanta, David, Jacob G Scott, Mayer N Fishman, Gustavo Ayala, Simon W Hayward, and Alexander RA Anderson. 2012. “Investigating Prostate Cancer Tumour–Stroma Interactions: Clinical and Biological Insights from an Evolutionary Game.” *British Journal of Cancer* 106 (1): 174–81.

Behan, Fiona M, Francesco Iorio, Gabriele Picco, et al. 2019. “Prioritization of Cancer Therapeutic Targets Using CRISPR–Cas9 Screens.” *Nature* 568 (7753): 511–16.

Clough, Emily, Tanya Barrett, Stephen E Wilhite, et al. 2024. “NCBI GEO: Archive for Gene Expression and Epigenomics Data Sets: 23-Year Update.” *Nucleic Acids Research* 52 (D1): D138–44.

Consortium, UniProt. 2019. “UniProt: A Worldwide Hub of Protein Knowledge.” *Nucleic Acids Research* 47 (D1): D506–15.

Cui, Chen, Xiaoyu Ding, Dingyan Wang, et al. 2021. “Drug Repurposing Against Breast Cancer by Integrating Drug-Exposure Expression Profiles and Drug–Drug Links Based on Graph Neural Network.” *Bioinformatics* 37 (18): 2930–37.

Cunningham, Jessica J, Robert A Gatenby, and Joel S Brown. 2011. “Evolutionary Dynamics in Cancer Therapy.” *Molecular Pharmaceutics* 8 (6): 2094.

Deb, Kalyanmoy. 2001. “Nonlinear Goal Programming Using Multi-Objective Genetic Algorithms.” *Journal of the Operational Research Society* 52 (3): 291–302.

Deb, Kalyanmoy, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. 2002. “A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II.” *IEEE Transactions on Evolutionary Computation* 6 (2): 182–97.

Doshi, Siddhant, and Sundeep Prabhakar Chepuri. 2022. “A Computational Approach to Drug Repurposing Using Graph Neural Networks.” *Computers in Biology and Medicine* 150: 105992.

Edgar, Ron, Michael Domrachev, and Alex E Lash. 2002. “Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository.” *Nucleic Acids Research* 30 (1): 207–10.

Gatenby, Robert A, Ariosto S Silva, Robert J Gillies, and B Roy Frieden. 2009. “Adaptive Therapy.” *Cancer Research* 69 (11): 4894–903.

Gaulton, Anna, Louisa J Bellis, A Patricia Bento, et al. 2012. “ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery.” *Nucleic Acids Research* 40 (D1): D1100–1107.

Gaulton, Anna, Anne Hersey, Michał Nowotka, et al. 2017. “The ChEMBL Database in 2017.” *Nucleic Acids Research* 45 (D1): D945–54.

Gerstung, Moritz, Clemency Jolly, Ignaty Leshchiner, et al. 2020. “The Evolutionary History of 2,658 Cancers.” *Nature* 578 (7793): 122–28.

Gillespie, Marc, Bijay Jassal, Ralf Stephan, et al. 2022. "The Reactome Pathway Knowledgebase 2022." *Nucleic Acids Research* 50 (D1): D687–92.
Hamilton, Will, Zhitao Ying, and Jure Leskovec. 2017. "Inductive Representation Learning on Large Graphs." *Advances in Neural Information Processing Systems* 30.
Hanahan, Douglas et al. 2022. "Hallmarks of Cancer: New Dimensions." *Cancer Discov* 12 (1): 31–46.
Hanahan, Douglas, and Robert A Weinberg. 2011. "Hallmarks of Cancer: The Next Generation." *Cell* 144 (5): 646–74.
Hernan, MA, and JM Robins. 2020. "Chapman & Hall." *Boca Raton*.
Himmelstein, Daniel Scott, Antoine Lizee, Christine Hessler, et al. 2017. "Systematic Integration of Biomedical Knowledge Prioritizes Drugs for Repurposing." *Elife* 6: e26726.
Jiang, Wei, Weicai Ye, Xiaoming Tan, and Yun-Juan Bao. 2025. "Network-Based Multi-Omics Integrative Analysis Methods in Drug Discovery: A Systematic Review." *BioData Mining* 18 (1): 27.
Kipf, Thomas N, and Max Welling. 2016. "Semi-Supervised Classification with Graph Convolutional Networks." *arXiv Preprint arXiv:1609.02907*.
Knox, Craig, Mike Wilson, Christen M Klinger, et al. 2024. "DrugBank 6.0: The DrugBank Knowledgebase for 2024." *Nucleic Acids Research* 52 (D1): D1265–75.
LeBlanc, Amy K, and Christina N Mazcko. 2020. "Improving Human Cancer Therapy Through the Evaluation of Pet Dogs." *Nature Reviews Cancer* 20 (12): 727–42.
Li, Jianwei, Bing Li, Xukun Zhang, Xuxu Ma, and Ziyu Li. 2025. "MDMNI-DGD: A Novel Graph Neural Network Approach for Druggable Gene Discovery Based on the Integration of Multi-Omics Data and the Multi-View Network." *Computers in Biology and Medicine* 185: 109511.
Lundberg, Scott M, and Su-In Lee. 2017. "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems* 30.
McFarland, James M, Zandra V Ho, Guillaume Kugener, et al. 2018. "Improved Estimation of Cancer Dependencies from Large-Scale RNAi Screens Using Model-Based Normalization and Data Integration." *Nature Communications* 9 (1): 4610.
McIntyre, Stephen D, Virendra Kakade, Yoichiro Mori, and Elena G Tolkacheva. 2014. "Heart Rate Variability and Alternans Formation in the Heart: The Role of Feedback in Cardiac Dynamics." *Journal of Theoretical Biology* 350: 90–97.
Morgan, Harry L. 1965. "The Generation of a Unique Machine Description for Chemical Structures-a Technique Developed at Chemical Abstracts Service." *Journal of Chemical Documentation* 5 (2): 107–13.
Nowak, Martin A. 2006. *Evolutionary Dynamics: Exploring the Equations of Life*. Vol. 1115. Belknap Press of Harvard University Press Cambridge, MA.
Nowell, Peter C. 1976. "The Clonal Evolution of Tumor Cell Populations: Acquired Genetic Lability Permits Stepwise Selection of Variant Sublines and Underlies Tumor Progression." *Science* 194 (4260): 23–28.
"Pan-Cancer Analysis of Whole Genomes." 2020. *Nature* 578 (7793): 82–93.
Pearl, Judea et al. 2003. "Causality: Models, Reasoning, and Inference." *Econometric Theory* 19 (675-685): 46.
Perdomo-Quinteiro, Pablo, Emre Guney, and Alberto Belmonte-Hernández. 2026. "Generating Explainable Hypotheses for Drug Repurposing with Graph Neural Networks." *Scientific Reports*.
Rogers, David, and Mathew Hahn. 2010. "Extended-Connectivity Fingerprints." *Journal of Chemical Information and Modeling* 50 (5): 742–54.
Romano, Yaniv, Evan Patterson, and Emmanuel Candes. 2019. "Conformalized Quantile Regression." *Advances in Neural Information Processing Systems* 32.
Saelens, Wouter, Robrecht Cannoodt, Helena Todorov, and Yvan Saeys. 2019. "A Comparison of Single-Cell Trajectory Inference Methods." *Nature Biotechnology* 37 (5): 547–54.
Sharma, Amit, and Emre Kiciman. 2020. "Dowhy: An End-to-End Library for Causal Inference." *arXiv Preprint arXiv:2011.04216*.
Sharma, Amit, Vasilis Syrgkanis, Cheng Zhang, and Emre Kıcıman. 2021. "Dowhy: Addressing Challenges in Expressing and Validating Causal Assumptions." *arXiv Preprint arXiv:2108.13518*.
Street, Kelly, Davide Risso, Russell B Fletcher, et al. 2018. "Slingshot: Cell Lineage and Pseudotime Inference for Single-Cell Transcriptomics." *BMC Genomics* 19 (1): 477.
Szklarczyk, Damian, Annika L Gable, Katerina C Nastou, et al. 2021. "The STRING Database in 2021: Customizable Protein–Protein Networks, and Functional Characterization of User-Uploaded Gene/Measurement Sets." *Nucleic Acids Research* 49 (D1): D605–12.
Tate, John G, Sally Bamford, Harry C Jubb, et al. 2019. "COSMIC: The Catalogue of Somatic Mutations in Cancer." *Nucleic Acids Research* 47 (D1): D941–47.
Trapnell, Cole, Davide Cacchiarelli, Jonna Grimsby, et al. 2014. "Pseudo-Temporal Ordering of Individual Cells Reveals Dynamics and Regulators of Cell Fate Decisions." *Nature Biotechnology* 32 (4): 381.

Tsherniak, Aviad, Francisca Vazquez, Phil G Montgomery, et al. 2017. "Defining a Cancer Dependency Map." *Cell* 170 (3): 564–76.
Veličković, Petar, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. "Graph Attention Networks." *arXiv Preprint arXiv:1710.10903*.
Vovk, Vladimir, Alexander Gammerman, and Glenn Shafer. 2005. *Algorithmic Learning in a Random World*. Springer.
Weinstein, John N, Eric A Collisson, Gordon B Mills, et al. 2013. "The Cancer Genome Atlas Pan-Cancer Analysis Project." *Nature Genetics* 45 (10): 1113–20.
West, Jeffrey, Li You, Jingsong Zhang, et al. 2020. "Towards Multidrug Adaptive Therapy." *Cancer Research* 80 (7): 1578–89.
Xu, Tao, Jinying Zhao, and Momiao Xiong. 2023. "Graphical Learning and Causal Inference for Drug Repurposing." *medRxiv*.
Zhang, Huaiwu, Xinliang Sun, Jianxin Wang, Min Li, and Jing Tang. 2026. "A Causal Inference Framework for Identifying Essential Genes to Enhance Drug Synergy Prediction." *Bioinformatics* 42 (5): btag010.
Zhang, Junming, Shuwen Xiong, Yugui Xu, and Yongqing Zhang. 2025. "Prediction of Cancer Drug Response Based on Heterogeneous Graph Neural Networks and Multi-Omics Data." *Neural Networks*, 108001.
Zhou, Jie, Ganqu Cui, Shengding Hu, et al. 2020. "Graph Neural Networks: A Review of Methods and Applications." *AI Open* 1: 57–81.