



BioSLATE: Biomarker Selection and Synthetic Lethality Analysis for
Therapeutic Exploration in HGSOC

Student Name: Faith Ogundimu

Grant Number: SUMSCH037

Supervisor: Associate Professor Colm Ryan

Department of Medicine

Table of Contents

Table of Contents.....	2
List of Abbreviations.....	4
Scientific Abstract.....	6
Introduction.....	7
Clinical Challenges of High-Grade Serous Ovarian Cancer.....	7
Synthetic Lethality: A Precision Oncology Approach.....	7
Integrative Framework for Biomarker Discovery and Synthetic Lethal Screening.....	8
Materials and Methods.....	9
Data Acquisition.....	9
Candidate Biomarker Filtering in TCGA Tumours.....	10
CNA–Protein Integration and Statistical Modelling.....	10
Validation of CNA–Protein Biomarkers in Ovarian Cancer Cell Lines.....	10
Regression-Based Synthetic Lethality Screening in HGSOC Cell Lines.....	11
Figure 1: Workflow for Synthetic Lethality Discovery in Amplified HGSOC Genes.....	12
AI-Based Synthetic Lethality Literature Mining.....	13
Agent Design and Purpose.....	13
Input Parameters and Execution Modes.....	13
Multi-Agent Architecture.....	13
Output and Scoring Framework.....	14
Software and Tools.....	14
Figure 2: Architecture of the OncoSynth Multi-Agent Literature Mining System... 15	
Protein–Protein Interaction Validation of Synthetic Lethal Pairs (BioGRID).....	16
Network and Pathway Analysis of Potent Synthetic Lethal Targets (STRING & g:Profiler). 16	
Benchmarking Synthetic Lethal Pairs Against Public Databases.....	17
Clinical and Genomic Data Processing.....	17
Survival Modelling.....	18
Synthetic Lethal Target Prioritisation and Drug Tractability Analysis.....	18
Drug-Target Mapping and GDSC Integration.....	18
Copy Number-Drug Sensitivity Correlation Analysis.....	18
Results.....	19
Identification of CNA–Protein Associations in TCGA HGSOC.....	19
Cell Line Support for Amplified Biomarkers.....	19
Table 1: Sequential Filtering Steps for TCGA–DepMap Integration and Candidate Biomarker Selection.....	20
Identification of Potent Synthetic Lethal Interactions.....	20
Figure 3: Regression-Based Synthetic Lethality Screening in HGSOC Cell Lines..	

21	
Limited Direct Interactions but Widespread Network Overlap.....	22
STRING Support and Pathway Enrichment of SL Hits.....	22
Figure 4: Protein-Protein Interaction Network Analysis Validates Synthetic Lethal Gene Pair Relationships.....	23
Predominantly Unestablished Synthetic Lethal Interactions.....	23
Clinical Cohort Characteristics.....	24
Cox Regression Outcomes.....	24
Kaplan–Meier Analyses.....	24
Summary.....	24
Figure 5: Biomarker Survival Analysis In High-Grade Serous Ovarian Carcinoma.	25
Therapeutic Tractability Landscape of Synthetic Lethal Targets.....	26
Limited Drug Availability Constrains Pharmacological Validation.....	26
Proof-of-Concept Validation of CHD7-CDK4 Synthetic Lethality.....	26
Figure 6: Translational Validation and Tractability of Synthetic Lethal Biomarkers.....	27
Discussion.....	28
Copy Number Alterations as a Distinct Genomic Lens for Synthetic Lethality.....	28
Contextualising Putative Synthetic Lethal Pairs vs. Canonical Synthetic Lethal Vulnerabilities.....	28
Transport-Enriched Biomarkers and Oncogenic-Enriched Targets Reflect Canonical Synthetic Lethal Network Logic.....	29
Meaningful yet Modest Survival Signals.....	30
From Synthetic Lethality to Therapeutic Translation.....	31
Strengths, Limitations and Future Directions.....	31
Multi-Cancer Pipeline Applicability.....	32
Conclusion.....	32
Bibliography.....	33
Appendices.....	36
Supplementary Figure 1.....	36
Supplementary Figure 1: Distribution of GISTIC CNA Scores in TCGA HGSOC..	36
Supplementary Table 1.....	36
Supplementary Table 1: Annotation of Datasets Used in This Study.....	36
Supplementary Figure 2.....	38
Supplementary Figure 2: Cross-Validation of TCGA CNA–Protein Hits in HGSOC Cell Lines.....	38
Supplementary Figure 3.....	39
Supplementary Figure 3: Top shared protein interactors across synthetic-lethal pairs.....	39

Supplementary Table 2.....	40
Supplementary Table 2: Tractability Classification of Synthetic Lethal Target Genes.....	40

List of Abbreviations

- **ACLY** - ATP Citrate Lyase
- **ACTN4** - Actinin Alpha 4
- **AI** - Artificial Intelligence
- **API** - Application Programming Interface
- **ATR** - ATR Serine/Threonine Kinase
- **AUC** - Area Under the Curve
- **BioGRID** - Biological General Repository for Interaction Datasets
- **BRCA** - Breast Cancer gene
- **CCNE1** - Cyclin E1
- **CDK4** - Cyclin Dependent Kinase 4
- **CHD7** - Chromodomain Helicase DNA Binding Protein 7
- **CHK1** - Checkpoint Kinase 1
- **CI** - Confidence Interval
- **CLI** - Command Line Interface
- **CNA** - Copy Number Alteration
- **CRISPR** - Clustered Regularly Interspaced Short Palindromic Repeats
- **CSV** - Comma-Separated Values
- **DDR** - DNA Damage Response
- **DepMap** - Cancer Dependency Map
- **FDR** - False Discovery Rate
- **FET** - Fisher's Exact Test
- **GDSC** - Genomics of Drug Sensitivity in Cancer
- **GISTIC** - Genomic Identification of Significant Targets in Cancer
- **GO:BP** - Gene Ontology Biological Process
- **HC3** - Heteroskedasticity-Consistent (type 3)
- **HGNC** - HUGO Gene Nomenclature Committee

- **HGSOC** - High-Grade Serous Ovarian Cancer
- **HR** - Hazard Ratio
- **ITPR2** - Inositol 1,4,5-Trisphosphate Receptor Type 2
- **JSON** - JavaScript Object Notation
- **KRAS** - KRAS Proto-Oncogene
- **MAPK1** - Mitogen-Activated Protein Kinase 1
- **MECOM** - MDS1 And EVI1 Complex Locus
- **MYC** - MYC Proto-Oncogene
- **NAALADL2** - N-Acetylated Alpha-Linked Acidic Dipeptidase Like 2
- **NF1** - Neurofibromin 1
- **OLS** - Ordinary Least Squares
- **ORF** - Open Reading Frame
- **OS** - Overall Survival
- **PARP** - Poly(ADP-ribose) polymerase
- **PFS** - Progression-Free Survival
- **PIK3CA** - Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
- **PPI** - Protein-Protein Interaction
- **RAF1** - Raf-1 Proto-Oncogene
- **RB1** - RB Transcriptional Corepressor 1
- **RPPA** - Reverse-Phase Protein Array
- **SCNA** - Somatic Copy Number Alteration
- **SL** - Synthetic Lethal/Synthetic Lethality
- **STRING** - Search Tool for the Retrieval of Interacting Genes/Proteins
- **TCGA** - The Cancer Genome Atlas
- **TNBC** - Triple-Negative Breast Cancer
- **TP53** - Tumor Protein P53
- **TXNRD1** - Thioredoxin Reductase 1
- **URI1** - URI1, Prefoldin Like Chaperone

Scientific Abstract

High-grade serous ovarian cancer (HGSOC) is the most prevalent and aggressive subtype of ovarian cancer, characterised by late-stage diagnosis and poor prognosis due to frequent relapse following chemotherapy and lack of effective screening strategies. There is an urgent global need for novel, targeted, and personalised therapeutic strategies to improve patient outcomes. This project employs a synthetic lethality framework to identify biomarkers and therapeutic vulnerabilities specific to HGSOC. Synthetic lethality exploits gene pairs whose concurrent disruption selectively induces cancer cell death while sparing normal cells, thereby revealing potential precision drug targets.

We integrate multi-omics data from ovarian cancer cell lines and patient tumours to identify recurrent genetic alterations indicative of cancer-specific weaknesses. Leveraging large-scale CRISPR screening datasets, we systematically identify essential genes whose perturbation is lethal in the context of these alterations. Subsequently, candidate therapeutic targets are prioritised based on existing drug availability or feasibility of drug development.

The outcome will be a curated list of clinically actionable targets and combination therapy candidates designed to inform precision oncology approaches for HGSOC. By exploiting cancer-specific vulnerabilities, this research aims to contribute to improved treatment paradigms, enhanced quality of life and prolonged survival for patients afflicted with high-grade serous ovarian cancer.

Introduction

Clinical Challenges of High-Grade Serous Ovarian Cancer

High-grade serous ovarian cancer (HGSOC) is the most common and lethal subtype of ovarian cancer (Punzón-Jiménez *et al.*, 2022). Due to vague early symptoms, most HGSOC cases present at advanced stages (III/IV), resulting in an aggressive disease course with poor prognosis (Punzón-Jiménez *et al.*, 2022). Standard treatment consists of maximal cytoreductive surgery followed by platinum/taxane chemotherapy. While this approach often induces an initial remission, the cancer recurs in ~70–80% of patients and eventually develops resistance to chemotherapy (Pignata *et al.*, 2017). Consequently, the five-year survival rate for advanced HGSOC lingers around 30%, a statistic that has improved little in recent decades (He, Li and Zhang, 2023). Molecularly, HGSOC tumours are characterised by pervasive chromosomal instability and extensive copy number alterations, leading to heightened inter- and intra-tumoural heterogeneity (Kleinmanns and Bjørge, 2024). These clinical and molecular challenges underscore an urgent need for new therapeutic strategies that go beyond the one-size-fits-all approach of surgery and cytotoxic chemotherapy. In particular, precision oncology approaches are sought to exploit tumour-specific vulnerabilities and improve outcomes for this deadly gynaecological cancer.

Synthetic Lethality: A Precision Oncology Approach

One promising avenue for mechanistically grounded, tumour-selective therapy is the concept of synthetic lethality. Synthetic lethality describes a scenario in which the concurrent perturbation of two genes is lethal to cells, whereas disruption of either gene alone is survivable (Kaelin, 2005; Shieh, 2022). In cancer, this implies that if a tumour harbours a particular genetic alteration (mutation, deletion or amplification), inhibition of a partner gene essential only in that altered context can trigger selective cancer cell death while sparing normal cells (Kaelin, 2005; Shieh, 2022). This approach directly targets “cancer vulnerabilities” rooted in the tumour’s genotype. A hallmark example is the lethal interplay between *BRCA1/2* loss and poly(ADP-ribose) polymerase (*PARP*) inhibition, tumours with *BRCA* mutations (deficient in homologous recombination DNA repair) are sensitive to *PARP* inhibitors, which induce irreparable DNA damage in the absence of functional *BRCA*, killing the cancer while normal cells with intact *BRCA* are unharmed (Helleday, 2011; Shieh, 2022). This synthetic lethal

strategy led to the first FDA-approved targeted therapy for HGSOC – *PARP* inhibitors in patients with *BRCA* mutations or homologous recombination deficiency (Ragupathi *et al.*, 2023). This validated the broader principle that leveraging tumour-specific genetic defects can yield effective, less toxic treatments. Building on this success, multiple synthetic lethal interactions are under active investigation. For instance, inhibiting DNA damage response kinases like *ATR* and *CHK1* can preferentially sensitise cancer cells to DNA damage or *PARP* blockade and such agents are being tested in combination trials (Kleinmanns and Bjørge, 2024). Overall, the synthetic lethality paradigm offers a rational framework to discover actionable cancer vulnerabilities and thus, by pinpointing gene pairs where one is frequently altered in HGSOC and the other can be targeted, we can identify novel therapeutic opportunities grounded in tumour biology.

Integrative Framework for Biomarker Discovery and Synthetic Lethal Screening

In light of HGSOC’s genomic complexity, our research deploys an integrative multi-omics strategy to systematically uncover and validate synthetic lethal interactions with translational potential. We begin by identifying candidate biomarkers, which are genes recurrently amplified or deleted in HGSOC, through analysis of large patient tumour datasets (e.g. TCGA copy number and proteomic profiles). Such aberrations are hypothesised to represent cancer-specific vulnerabilities which may be exploited. We then overlay functional genomic data from large-scale CRISPR-Cas9 knockout screens in HGSOC cancer cell lines to find genes that become essential only in the presence of those HGSOC-specific alterations. This unbiased in-silico screening approach, similar to prior large-scale dependency analyses, enables the nomination of putative synthetic lethal gene pairs (Zhan *et al.*, 2019). To prioritise pairs with clinical relevance, we incorporate pharmacogenomic and druggability data, favouring targets for which small-molecule inhibitors or biologics exist or can be developed.

Notably, we introduce a novel multi-agent literature mining platform, “*OncoSynth*,” to rigorously validate and contextualise each candidate synthetic lethal interaction. OncoSynth autonomously gathers evidence from biomedical literature, drug databases and clinical trial registries for a given gene pair, integrating information on any reported co-lethality, known drug targets and clinical studies, and then ranks the pair’s therapeutic potential. This automated, AI-driven tool is a key innovation of our framework and aids in identifying whether shortlisted synthetic lethal

pairs are supported by existing evidence and mechanistic rationale, reinforcing their translational promise.

This project, funded by Breakthrough Cancer Research, exemplifies a bench-to-bedside approach. By uniting genomic biomarkers with functional dependency screens, we aim to create a pipeline for discovering HGSOC-specific lethal gene interactions that can be readily translated into precision therapies. In the following report, we detail this approach and highlight how harnessing synthetic lethality can expand the therapeutic arsenal against HGSOC's otherwise treatment-refractory biology, ultimately striving to improve survival and hope for patients with this aggressive cancer.

Materials and Methods

Data Acquisition

To identify candidate synthetic lethal interactions in HGSOC, we integrated clinical, genomic, proteomic and functional dependency datasets from The Cancer Genome Atlas (TCGA) and the Cancer Dependency Map (DepMap). TCGA data were downloaded from cBioPortal (study ID: [ov_tcga_pan_can_atlas_2018](#)), including gene-level somatic copy number alterations (CNAs) inferred using GISTIC2 (scores: -2 to +2) (see [Supplementary Figure 1](#)) and matched reverse-phase protein array (RPPA) profiles.

The RPPA data were validated to originate from the iTRAQ-based CPTAC proteomics dataset described by (Zhang *et al.*, 2016). Clinical data is also present in the [ov_tcga_pan_can_atlas_2018](#) folder, named [data_clinical_patient.txt](#).

To assess gene essentiality, we used CRISPR-Cas9 dependency scores from the DepMap 24Q1 Chronos release, spanning approximately 1100 cell lines. A curated panel of 23 ovarian cancer cell lines was obtained from the DepMap Context resource for High-Grade Serous Ovarian Cancer. Of these, only 21 cell lines had available copy number data and were retained for CNA-based analyses (OVCAR4 and CAOVS3 were not present in the CNA matrix). For CRISPR-based dependency analysis, only 18 of the 23 lines had available gene effect scores (SNU119, OVSAHO, OVCAR4, FUV1, and OVKATE were not present in the Chronos gene effect matrix). An annotation of all datasets used in this study can be viewed in [Supplementary Table 1](#).

Candidate Biomarker Filtering in TCGA Tumours

Somatic copy number alterations (CNAs) were obtained as discrete GISTIC2 scores ranging from -2 (deep deletion) to $+2$ (high-level amplification). Genes were retained for downstream analysis if they were amplified or deleted in at least 5% of TCGA HGSOC samples.

To enable integration with proteomic profiles, the CNA gene list was aligned with RPPA protein expression data using Entrez Gene IDs. Protein identifiers in the RPPA matrix (HGNC symbols and aliases) were mapped to Entrez IDs using the HGNC [gene_with_protein_product.txt](#) reference file from HGNC. Genes lacking a valid Entrez mapping or protein expression data were excluded.

CNA–Protein Integration and Statistical Modelling

To evaluate the functional relevance of somatic CNAs, we assessed whether gene amplifications or deletions were associated with consistent changes in protein expression across TCGA HGSOC samples. CNA scores and RPPA values were merged for each gene and sample, retaining only genes with matched data in at least six patients. For each gene, a univariate linear regression model was fitted:

$$\text{Protein} \sim \beta_0 + \beta_1 \times \text{CNA} + \varepsilon$$

where CNA was treated as an ordinal variable (-2 to $+2$). The slope coefficient β_1 estimated the direction and strength of the association and statistical significance was assessed using the corresponding p-value. Genes were classified as significant if the CNA–protein linear regression returned a p-value < 0.05 , irrespective of slope direction. This approach captures any consistent association between CNA state and protein abundance.

Validation of CNA–Protein Biomarkers in Ovarian Cancer Cell Lines

To assess whether the CNA–protein associations observed in patient tumours were recapitulated in experimental models, we analysed absolute copy number data from the DepMap 24Q4 release. Twenty-one HGSOC cell lines were retained after filtering, with gene identifiers harmonised by Entrez IDs to enable integration with TCGA-derived candidates. Thresholds for deep deletion ($\text{CN} < 1$) and high-level amplification ($\text{CN} \geq 6$) were applied. Initially, genes were considered altered if at least 5% of independent cell lines ($n \geq 2$) exhibited the event, however to increase

statistical rigor, this was increased to 3 independent cell lines. These lists were then intersected with the tumour-derived CNA–protein hits to obtain a subset of biomarkers with support in both primary tumours and cell line models.

All steps involved in candidate gene filtering from TCGA and DepMap datasets, including CNA thresholding, recurrence criteria and proteomic alignment are visually illustrated in [Figure 1](#).

Regression-Based Synthetic Lethality Screening in HGSOC Cell Lines

Synthetic lethal interactions were inferred by modelling CRISPR gene dependency scores as a function of biomarker copy number status in DepMap HGSOC cell lines. To reduce redundancy, biomarkers with identical CNA profiles across all lines were clustered. The first gene in a cluster was temporarily selected as the representative gene of that cluster.

For each biomarker–target pair, ordinary least squares (OLS) regression was fitted with gene dependency as the outcome and biomarker CNA as the predictor. Regression coefficients were estimated with heteroskedasticity-consistent (HC3) variance adjustment. Effect size was calculated as the slope β standardised by the variance of dependency scores, reflecting the magnitude of dependency change per unit CNA (i.e. Cohen’s D).

A multiple-testing correction was applied using the Benjamini–Hochberg procedure. Pairs were classified as candidate SL interactions if they met three conditions: $FDR < 0.05$, effect size < 0 , and predicted dependency at $CNA = 2 > -0.6$ (to exclude broadly essential genes). A potency filter was then applied, requiring the difference between observed mean dependency at $CNA \geq 6$ and the regression-predicted dependency at $CNA = 2$ (delta effect) to be ≤ -0.2 , ensuring that only amplified contexts that showed enhanced vulnerability were retained. The equation can be viewed below.

$$Dependency_i = \beta_0 + \beta_1 \cdot CNA_i + \varepsilon_i$$

where $Dependency_i$ is the CRISPR gene effect score in cell line i , CNA_i is the copy number value of the biomarker in the same line.

Delta is defined as:

$$\Delta = mean(Dependency)_{CNA \geq 6} - \hat{y}_{CNA=2}$$

with $\hat{y}_{CNA=2} = \beta_0 + 2\beta_1$ being the model-predicted dependency at $CNA = 2$.

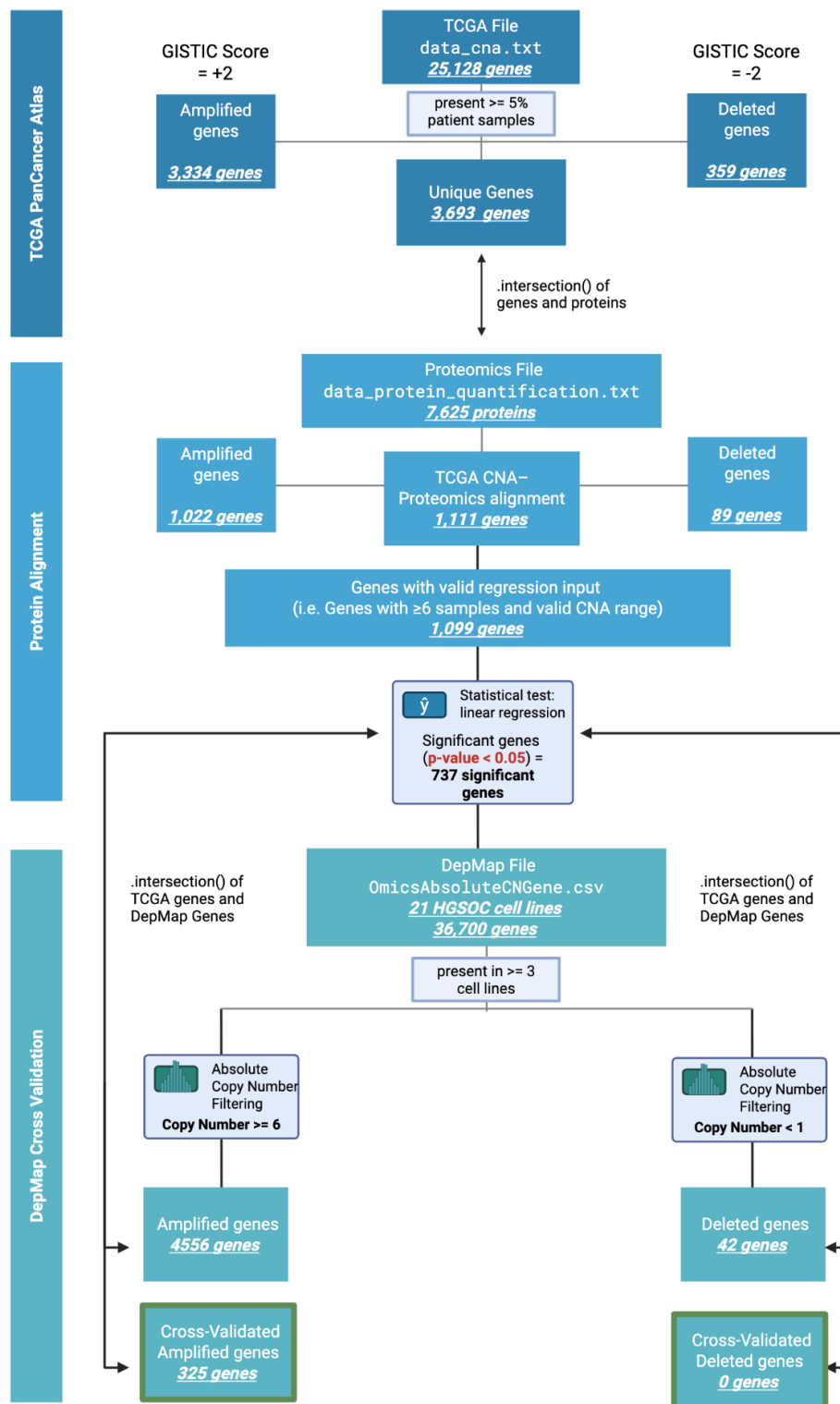


Figure 1: Workflow for Synthetic Lethality Discovery in Amplified HGSOC Genes.

A schematic overview of the analytical pipeline used to identify synthetic lethal interactions in high-grade serous ovarian cancer (HGSOC). The workflow integrates TCGA-derived copy number and proteomic data to identify amplification-linked genes, which are then validated in DepMap HGSOC cell lines. Candidate biomarkers are filtered by recurrence, protein expression association and cross-cohort concordance.

AI-Based Synthetic Lethality Literature Mining

Agent Design and Purpose

To systematically evaluate whether candidate synthetic lethal interactions identified *in silico* are supported by existing biomedical evidence, we developed *OncoSynth*, a multi-agent framework for automated literature and database mining. The design goal was to prioritise biomarker–target gene pairs with the strongest therapeutic and translational potential by integrating evidence across PubMed, Open Targets and ClinicalTrials.gov. Unlike single-source approaches, the multi-agent framework allows parallelised evidence gathering followed by structured synthesis and confidence scoring, ensuring both breadth and reproducibility in the evaluation of candidate pairs.

Input Parameters and Execution Modes

OncoSynth accepts gene pairs defined as a biomarker and its putative synthetic lethal partner. Two execution modes were implemented. In *batch* mode, gene pairs are imported from a structured CSV file and processed sequentially, generating individual reports for each pair. In *interactive* mode, biomarker–target pairs can be entered directly via the command line interface (CLI), enabling real-time exploration of individual hypotheses. Both modes route inputs through the same underlying pipeline, ensuring consistency of evidence retrieval and report generation.

Multi-Agent Architecture

The system was constructed in Python using the CrewAI framework, with each module encapsulated as a specialised agent assigned a discrete role. A PubMed search agent interrogates the biomedical literature for co-occurrence of the biomarker and target, prioritising abstracts explicitly mentioning synthetic lethality. Two other independent PubMed literature agents assess the broader oncogenic context of each gene individually, with emphasis on ovarian cancer. Drug annotation is performed by an Open Targets agent, which retrieves tractability scores, known inhibitors, mechanisms of action and clinical status. Clinical relevance is further evaluated by a ClinicalTrials.gov agent, which identifies interventional trials and extracts study phase, recruitment status and condition type.

The outputs of these retrieval agents are synthesised by a biomedical analyst agent, which integrates the evidence with attention to cancer specificity and mechanistic plausibility. A

deterministic confidence scoring tool then applies a weighted rubric, assigning up to 40 points for direct synthetic lethality evidence, 30 for druggability, 15 for clinical trial support and 15 for cancer relevance. Reports are finally authored by a technical writer agent, which converts the structured JSON evidence into a markdown report formatted for clinical and research audiences.

Output and Scoring Framework

For each gene pair, OncoSynth produces a structured report containing background information, supporting synthetic lethality evidence, drug target data, clinical trial annotations, translational implications, and references with PubMed identifiers. Reports are prefixed with a confidence score ranging from 0 to 100, which is calculated reproducibly by the deterministic scoring agent. Scores below 50 are labelled as low-confidence, while scores of 50 or higher are designated high-confidence.

Software and Tools

The agent was implemented in Python using the following packages: `crewAI`, `Bio.Entrez`, `langchain-openai`, `pydantic` and `requests`. The PubMed interface uses NCBI Entrez E-utilities, the drug module uses the Open Targets v4 GraphQL API, and clinical trials are fetched via ClinicalTrials.gov v2 REST API.

All source code, including execution scripts for batch (`batch.py`), interactive (`interactive.py`) modes, deterministic scoring functions and logging utilities, is available upon request. The full agent architecture, including input-output flow and modular agent roles, is summarised in [Figure 2](#).

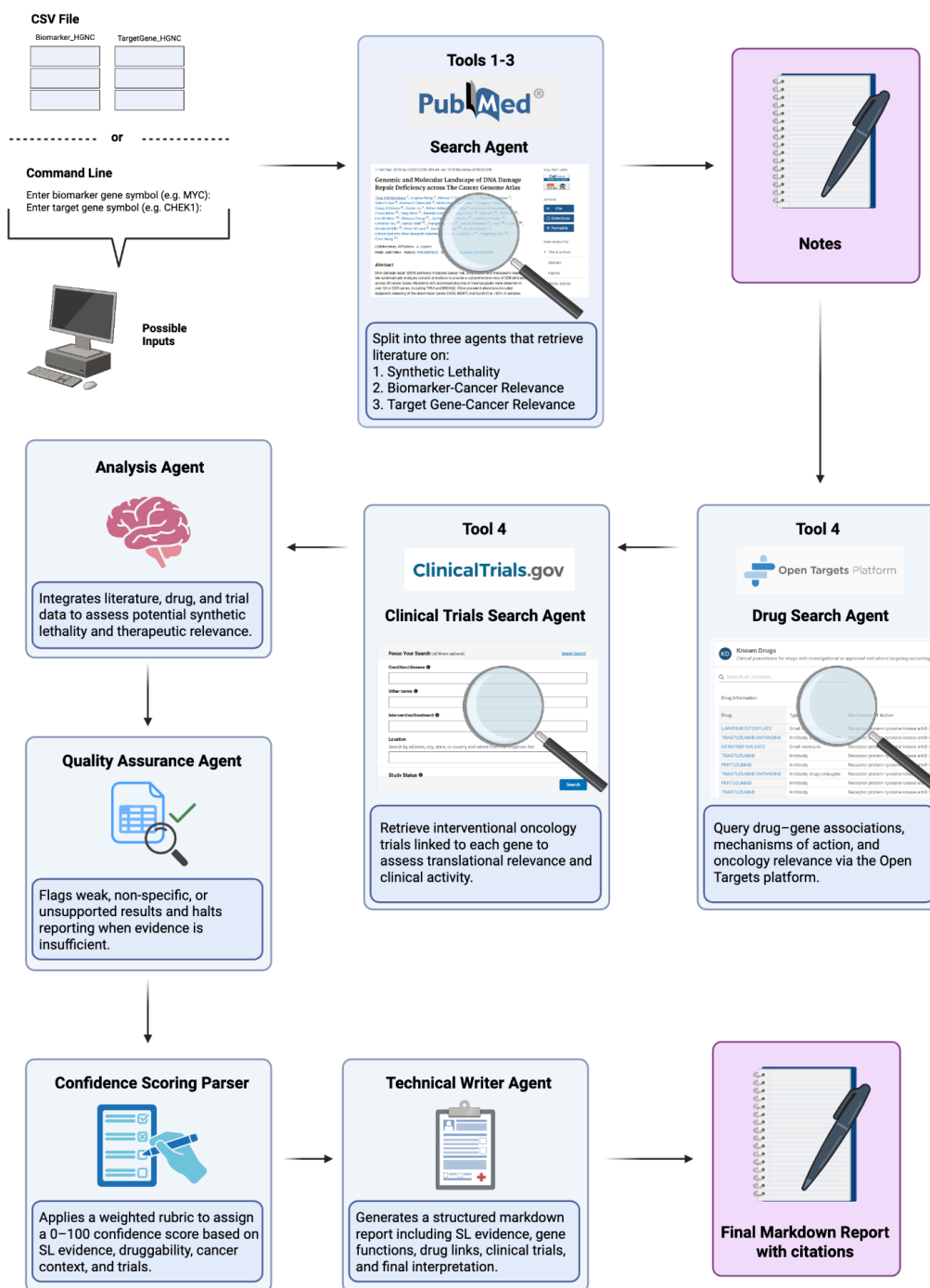


Figure 2: Architecture of the OncoSynth Multi-Agent Literature Mining System.

Schematic overview of the OncoSynth framework for synthetic lethality evidence retrieval and integration. Independent agents query PubMed for biomarker–target co-occurrence and cancer-specific literature, Open Targets for druggability and therapeutic annotations, and ClinicalTrials.gov for trial evidence. Outputs from these retrieval modules are passed to an analyst agent, which synthesises the findings, and a deterministic scoring agent, which assigns a reproducible confidence score (0–100) across four evidence domains. A technical writer agent generates the final structured markdown report, which is output in batch or interactive execution modes.

Protein–Protein Interaction Validation of Synthetic Lethal Pairs (BioGRID)

To assess whether synthetic lethal pairs identified from regression screening were biologically supported by protein–protein interaction (PPI) networks, we integrated results with BioGRID (release 4.4.241). BioGRID tab-delimited interaction data were filtered to retain only physical interactions between human proteins (taxon ID 9606) and to exclude self-interactions. The resulting dataset comprised 862,452 unique interaction pairs involving 19,995 unique human genes.

For each biomarker–target pair identified as a potent SL hit, we calculated several measures of network connectivity. First, we determined whether the biomarker and target directly interacted in BioGRID. Second, we computed the number of total interactors for each gene, the number of shared interactors and the Jaccard index of overlap. To evaluate whether observed overlap exceeded random expectation, a Fisher’s exact test (FET) was applied to a contingency table of shared versus unique interactors, yielding a log-transformed p-value as the “FET PPI overlap” statistic. Biomarker clusters containing multiple co-amplified genes were resolved by selecting a single representative gene per cluster, prioritising non-ORF symbols and genes with higher PPI connectivity. The representative biomarker, along with PPI-derived overlap metrics, was merged back into the potent SL hit set, producing a final “PPI-validated” SL dataset.

Network and Pathway Analysis of Potent Synthetic Lethal Targets (STRING & g:Profiler)

To investigate whether potent synthetic lethal (SL) pairs were supported by known functional networks in a second database, we evaluated both direct protein–protein interactions (PPIs) and biological pathway enrichment in STRING. Potent SL hits with HGNC annotations were cross-referenced with the STRING v12.0 database. STRING protein links were filtered to retain human interactions with a combined confidence score ≥ 400 , corresponding to medium confidence. For each biomarker–target pair, we determined whether a direct PPI was supported under this threshold and summarised the number of pairs validated.

To assess biological coherence of SL genes, we conducted pathway enrichment analysis using g:Profiler (v1.0). Separate enrichment runs were performed for amplified biomarkers and for SL targets, testing against Gene Ontology Biological Process (GO:BP) terms. Significant terms were defined as those meeting g:Profiler’s multiple-testing correction (adjusted $p < 0.05$). To visualise enrichment patterns, we network-based chord diagrams where pathways and genes were

collapsed into broader functional categories (e.g. cell cycle, chromatin organisation, mitochondrial translation).

Benchmarking Synthetic Lethal Pairs Against Public Databases

To assess whether the potent synthetic lethal (SL) pairs identified from regression screening in HGSOC cell lines had prior evidence, we benchmarked them against two public SL resources: SynLethDB and ISLE. SynLethDB aggregates experimentally validated, literature-curated, and computationally predicted SL interactions, providing a comprehensive knowledgebase. In contrast, ISLE is a computational pipeline that infers clinically relevant SL pairs by integrating tumour molecular profiles, patient survival data and evolutionary conservation (Lee *et al.*, 2018; Wang *et al.*, 2022).

For each biomarker–target pair, gene symbols were normalised to uppercase HGNC identifiers. Pairs were then queried in both databases in either direction (biomarker–target or target–biomarker). Binary flags were assigned for presence in SynLethDB and ISLE. A final novelty indicator was created for pairs absent from both resources. The merged benchmarking results were exported for integration into downstream translational reporting.

Clinical and Genomic Data Processing

We analysed The Cancer Genome Atlas high-grade serous ovarian carcinoma (TCGA-OV, PanCanAtlas 2018) cohort to assess the prognostic relevance of amplification events in SL biomarkers. The unique biomarker candidates from the biomarker-target pairs were mapped to TCGA GISTIC2 copy-number calls using Entrez ID–HGNC cross-referencing. Deep amplifications were defined as GISTIC score = 2, consistent with the binary threshold applied in prior analyses. Clinical annotations were retrieved from cBioPortal, including overall survival (OS) and progression-free survival (PFS) time and status. Patient IDs were harmonised across genomic and clinical datasets and only patients with both CNA and clinical data available were retained. Thereafter, only cases with age information were retained for Cox regression modelling.

Survival Modelling

Univariate and age-adjusted Cox proportional hazards models were fitted separately for OS and PFS. Models were restricted to biomarkers with ≥ 50 patients and at least five individuals in both amplified and non-amplified groups. Hazard ratios (HRs), 95% confidence intervals (CIs) and p-values were extracted for each biomarker. Benjamini–Hochberg correction was applied across all tests to control the false discovery rate (FDR).

Synthetic Lethal Target Prioritisation and Drug Tractability Analysis

We assessed the therapeutic potential of unique synthetic lethal target genes from the biomarker-target pairs using the Open Targets Platform API to systematically evaluate drug tractability across multiple therapeutic modalities. Tractability scores were computed for small molecules (8 buckets: approved drugs, clinical precedence, discovery precedence, predicted tractable), antibodies (9 buckets: clinical precedence, high/medium confidence based on protein localisation), and other clinical modalities (3 buckets). Each gene received binary scores across tractability categories, which were aggregated into bucket sums and classified into hierarchical categories: Clinical Precedence > Discovery Precedence > Predicted Tractable > Unknown.

Drug-Target Mapping and GDSC Integration

Known drugs for tractable targets were extracted from Open Targets and cross-referenced with the Genomics of Drug Sensitivity in Cancer (GDSC) pharmacogenomics database. Drug names were manually curated to resolve nomenclature inconsistencies between databases (e.g., "SORAFENIB TOSYLATE" → "SORAFENIB"). We filtered unique compounds tested in HGSOC cell lines, identifying overlapping drugs available for correlation analysis with synthetic lethal pairs.

Copy Number-Drug Sensitivity Correlation Analysis

For validation of synthetic lethal relationships, we correlated biomarker copy number amplification (DepMap absolute copy number ≥ 6) with drug sensitivity (GDSC AUC values) using Pearson correlation. Cell line identifiers (ARXSPAN IDs) were standardised across datasets and analysis was restricted to pairs with ≥ 3 overlapping cell lines. Statistical

significance was assessed using two-tailed tests, with correlation strength thresholds of $|r| > 0.7$ for strong relationships.

Results

Identification of CNA–Protein Associations in TCGA HGSOC

To prioritise candidate biomarkers, we first applied a prevalence filter requiring copy number alterations in at least 5% of tumours in the TCGA HGSOC cohort ($n = 572 \times 0.05 \approx 29$). This yielded 3,334 amplified and 359 deleted genes, corresponding to 3,693 unique candidates. Integration with RPPA profiles was performed by harmonising gene identifiers through HGNC-to-Entrez mappings, resulting in 1,111 genes with matched CNA and protein expression data across 119 overlapping patient samples.

We next evaluated the impact of CNAs on protein abundance using univariate linear regression models, treating the discrete GISTIC scores (–2 to +2) as ordinal predictors of protein levels. Of the 1,111 evaluable genes, 737 displayed significant CNA–protein associations at a nominal threshold of $p < 0.05$. These genes, in which copy number state consistently predicted protein abundance, represent the filtered set of functionally supported biomarkers that were carried forward for downstream cross-validation in ovarian cancer cell line datasets.

Cell Line Support for Amplified Biomarkers

Across the 21 HGSOC cell lines, this filtering identified 4,556 amplified genes and 42 deleted genes. Comparison with the 737 tumour-derived candidates revealed 325 overlapping amplified genes, whereas no deletions passed the prevalence threshold in both datasets (see [Supplementary Figure 2](#)). These 325 amplified genes were designated as the cross-validated biomarker set and prioritised for downstream synthetic lethality screening. The filtering pipeline and resulting genes can be viewed in [Table 1](#).

Table 1: Sequential Filtering Steps for TCGA–DepMap Integration and Candidate Biomarker Selection

Filtering Step	Number of Genes Remaining
Initial amplified genes ($\geq 5\%$ patients) - TCGA	3,334
Initial deleted genes ($\geq 5\%$ patients) - TCGA	359
Total unique candidate genes	3,693
Genes with matched proteomics data	1,111
DepMap significant genes (regression: $p < 0.05$)	737
Initial amplified genes ($\geq 5\%$ cell lines) - DepMap	4556
Initial deleted genes ($\geq 5\%$ cell lines) - DepMap	42
TCGA-validated amplified genes cross-checked in DepMap	325
TCGA-validated deleted genes cross-checked in DepMap	0

Identification of Potent Synthetic Lethal Interactions

Across all biomarker–target combinations tested ($n \approx 5.2 \times 10^5$), we identified 3,476 pairs with $\text{FDR} < 0.05$, of which 1,601 showed negative effect sizes consistent with synthetic lethal interactions. Filtering to exclude targets broadly essential in diploid contexts reduced this to 1,075 “selective” hits. Imposing the potency criterion ($\text{delta effect} \leq -0.2$) yielded a final set of **735** potent synthetic lethal interactions. These results demonstrate that regression modelling of copy number variation against CRISPR gene effect scores can robustly capture context-specific dependencies in HGSOc. The global distribution of regression results and examples of biomarker–target interactions are shown in [Figure 3](#).

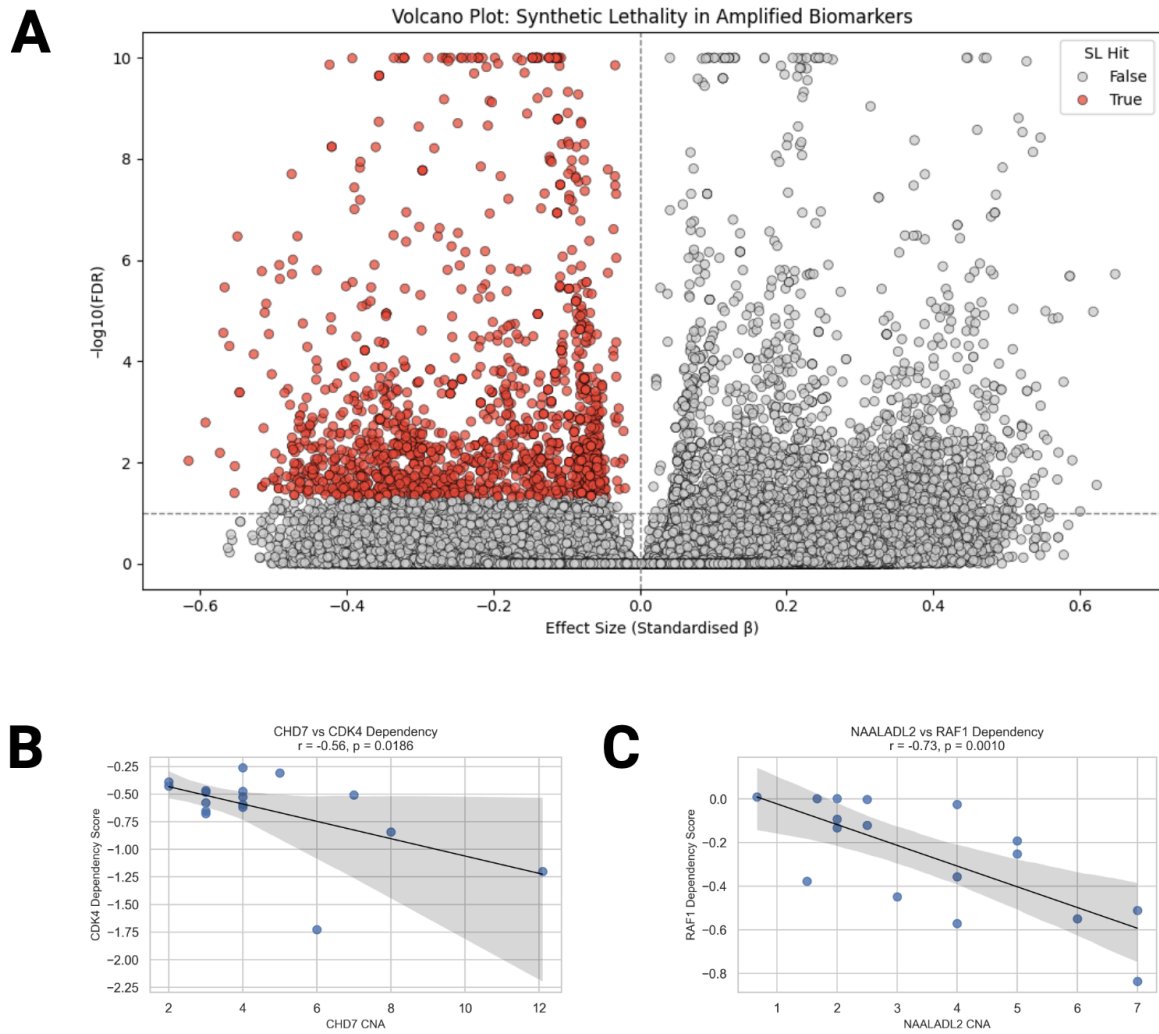


Figure 3: Regression-Based Synthetic Lethality Screening in HGSOC Cell Lines.

(A) Volcano plot showing the distribution of regression results across ~521,000 biomarker–target tests. The x-axis indicates effect size (standardised regression slope, β), and the y-axis shows $-\log_{10}(\text{FDR})$. Red points denote significant hits ($\text{FDR} < 0.05$ and effect size < 0).

(B) Example regression plot of CHD7 copy number versus CDK4 dependency score, illustrating a significant synthetic lethal interaction. Each point represents a DepMap HGSOC cell line, with regression fit shown in black.

(C) Example regression plot of NAALADL2 copy number versus RAF1 dependency score, highlighting a second representative interaction.

Limited Direct Interactions but Widespread Network Overlap

The processed BioGRID dataset encompassed over 850,000 unique protein interactions. Intersection with the potent synthetic lethal set revealed that only 24 biomarker–target pairs (3.3%) were supported by direct physical interactions. However, when expanded to representative biomarker–target pairs ($n = 735$), 620 pairs (84.4%) shared at least one common interactor, indicating extensive indirect connectivity within the protein interaction network. These shared interactors included canonical tumour suppressors and oncogenic regulators such as TP53 and MYC, respectively, which recurred across multiple synthetic lethal relationships, [Supplementary Figure 3](#).

Fisher's exact test identified 124 pairs (16.9%) with statistically significant enrichment of shared protein interactions ($p < 0.05$), suggesting that synthetic lethal dependencies are underpinned by coherent network modules rather than random associations. The median number of shared interactors amongst connected pairs was 6, with Jaccard similarity indices averaging 0.029, indicating moderate but meaningful network overlap. Representative biomarker selection within amplification clusters ensured that poorly characterised genes did not dominate the analysis whilst maintaining biological relevance.

A panel of key plots can be viewed in [Figure 4](#).

STRING Support and Pathway Enrichment of SL Hits

At the STRING confidence threshold of 400, 15 biomarker–target pairs were supported by direct physical interactions between the two genes.

Pathway enrichment analysis revealed strong functional clustering of both biomarkers and targets. Amplified biomarkers were enriched for pathways associated with molecular transport and RNA metabolism, consistent with roles in protein localisation and transcript processing. Synthetic lethal target genes were enriched in canonical cancer hallmarks, including cell cycle progression, checkpoint regulation, DNA repair and synthesis, chromatin organisation, and mitochondrial translation. Additional clusters included pathways related to protein synthesis and degradation and stress response and apoptosis.

Protein-Protein Interaction Analysis of Synthetic Lethal Gene Pairs

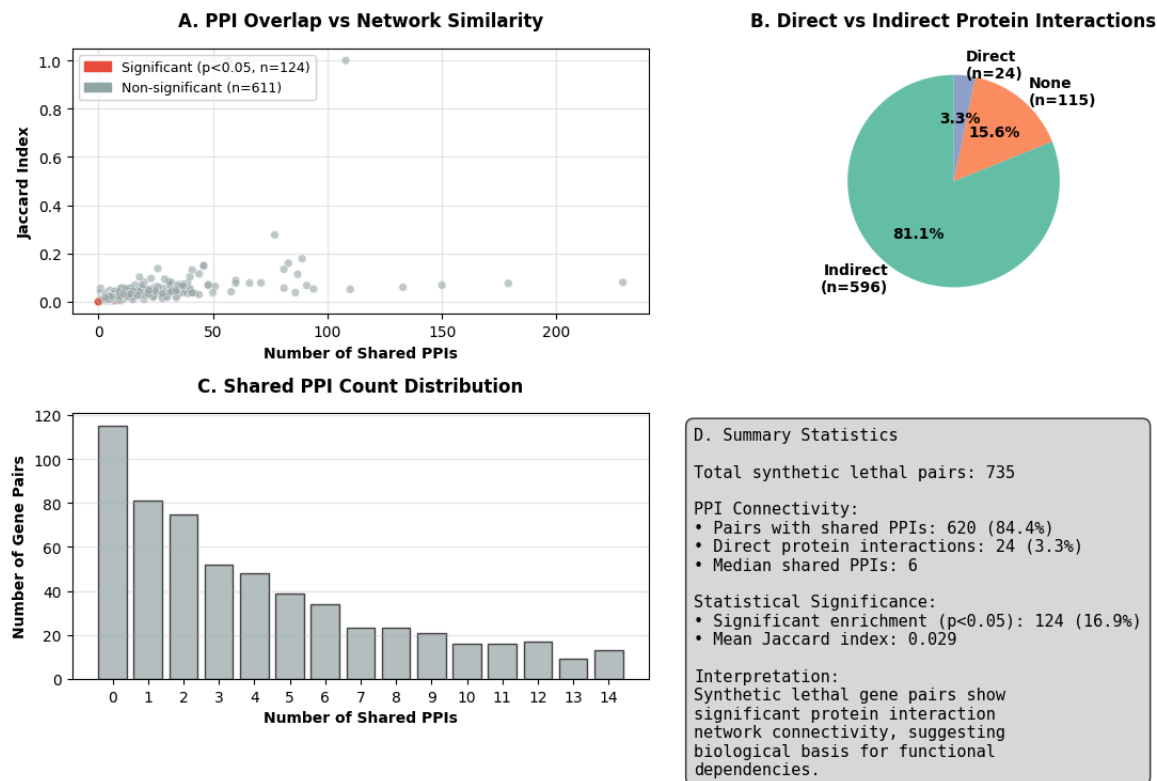


Figure 4: Protein-Protein Interaction Network Analysis Validates Synthetic Lethal Gene Pair Relationships.

(A) Scatter plot showing the relationship between number of shared protein-protein interactions (PPIs) and Jaccard similarity index for 735 representative biomarker–target pairs. Points are coloured by statistical significance (red, Fisher's exact test $p < 0.05$; grey, non-significant).

(B) Proportion of synthetic lethal pairs with direct versus indirect protein interactions.

(C) Frequency distribution of shared PPI counts, showing the number of gene pairs for each level of network connectivity (top 15 categories shown).

(D) Summary statistics for the complete dataset. Analysis demonstrates that 84.4% of synthetic lethal pairs share protein interactions, with 16.9% showing statistically significant enrichment. Network-mediated relationships (indirect interactions) predominate over direct protein–protein contacts (84.4% versus 4.2%), supporting a model wherein synthetic lethal dependencies arise through disruption of interconnected functional modules rather than direct physical associations.

Predominantly Unestablished Synthetic Lethal Interactions

Of the 735 potent SL pairs tested, 3 were found exclusively in SynLethDB, 0 were unique to ISLE, and 1 pair was reported in both databases. The remaining 731 pairs (99.5%) were absent from both resources. This benchmarking step demonstrated that the vast majority of the candidate interactions identified in our HGSOC-focused screen represent previously unreported synthetic lethal relationships.

Clinical Cohort Characteristics

Among the 558 patients with matched biomarker and clinical data, 330 OS events (59.1%) were observed, with a median OS of 33.3 months. For PFS, 407 events (71.3%) occurred, with a median of 14.7 months. Biomarker amplification frequencies had a median of 8.4% across the cohort.

Cox Regression Outcomes

Across 201 biomarkers tested, no associations with OS or PFS remained significant after FDR correction ($\text{FDR} < 0.05$). However, 9 biomarkers showed nominal significance (adjusted $p < 0.05$) in the age-adjusted OS models, including *URII* ($\text{HR} = 1.41$, $p = 0.022$), *ACTN4* ($\text{HR} = 1.53$, $p = 0.028$), and *CCNE1* ($\text{HR} = 1.33$, $p = 0.0498$). For PFS, 9 biomarkers reached nominal significance, the most notable being *ITPR2* ($\text{HR} = 1.58$, $p = 0.014$) and *URII* ($\text{HR} = 1.39$, $p = 0.021$).

Kaplan–Meier Analyses

Exploratory Kaplan–Meier survival curves were generated for three key biomarkers. *CCNE1* amplification ($n=110$) was associated with reduced OS (median 38.0 vs 48.1 months, log-rank $p = 0.0010$). *ACTN4* amplification ($n=51$) similarly correlated with worse OS (36.4 vs 47.6 months, log-rank $p = 0.0110$). *URII* amplification ($n=101$) was also strongly associated with inferior outcomes (38.0 vs 47.7 months, log-rank $p = 0.0001$).

Summary

Although no biomarkers achieved FDR-significant associations, several candidates demonstrated consistent nominal signals across OS and PFS, with hazard ratios in the range of 1.3–1.5 and amplification frequencies of 8–18%. These include *CCNE1*, a well-established oncogene in HGSOC, along with *URII* and *ACTN4*, which warrant prioritisation for further validation, see [Figure 5](#).

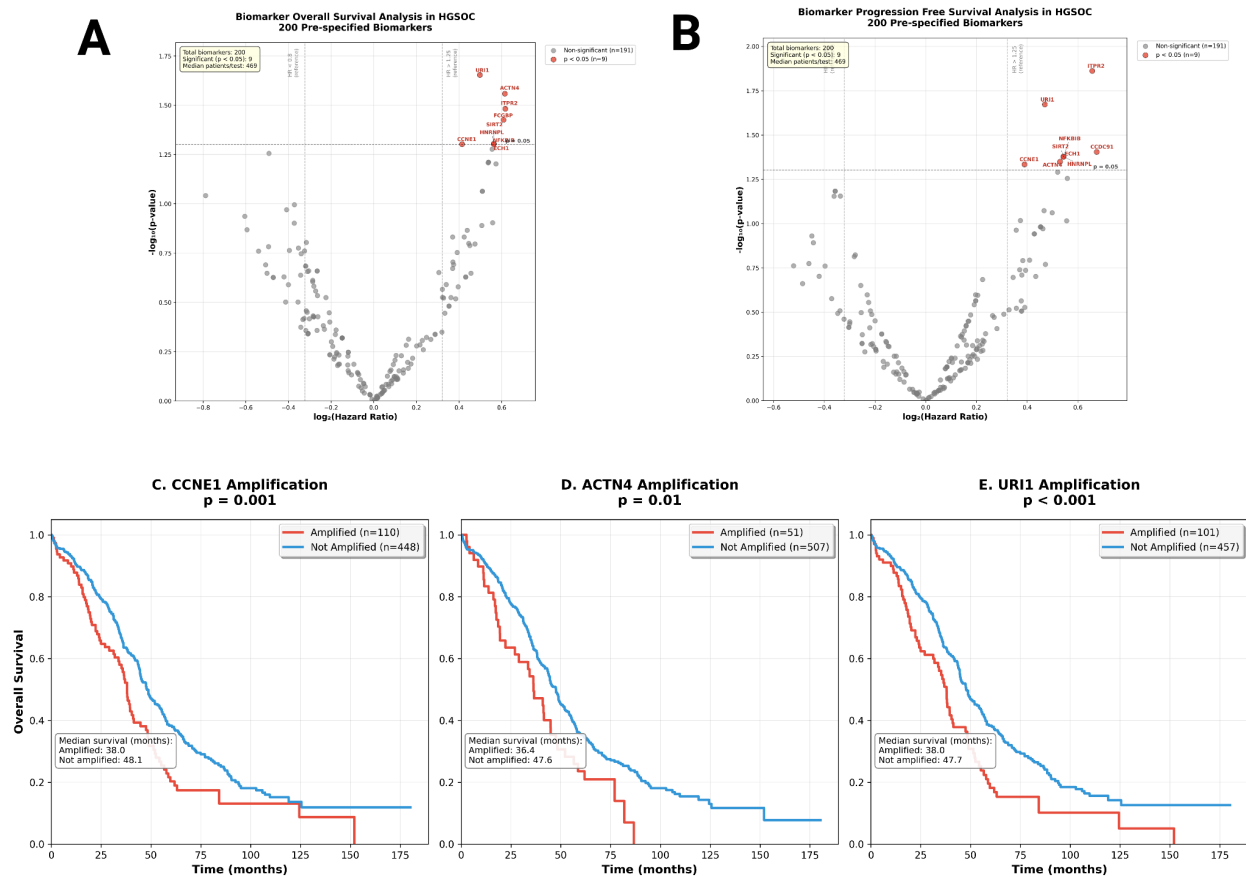


Figure 5: Biomarker Survival Analysis In High-Grade Serous Ovarian Carcinoma.

(A) Volcano plot showing progression-free survival (PFS) analysis of 200 pre-specified biomarkers. Points represent individual biomarkers plotted by $\log_2(\text{hazard ratio})$ versus $-\log_{10}(\text{p-value})$. Red points indicate biomarkers with nominal significance ($p < 0.05$), grey points represent non-significant biomarkers. Horizontal dashed line indicates $p = 0.05$ significance threshold. Vertical dashed lines represent hazard ratio reference thresholds ($\text{HR} = 0.8$ and $\text{HR} = 1.25$). Significant biomarkers are labelled with gene names. Statistics box shows total biomarkers tested, number achieving nominal significance, and median patient sample size.

(B) Volcano plot showing overall survival (OS) analysis of the same 200 biomarkers. Layout and statistical thresholds are identical to panel A. Gene labels indicate biomarkers achieving nominal significance for overall survival outcomes.

(C) Kaplan-Meier curves showing overall survival stratified by CCNE1 amplification status. Red curves represent patients with amplified biomarkers, blue curves represent patients without amplification. P-values were calculated using the log-rank test. Sample sizes for each group are indicated in the legend. Median survival times are displayed in the text box, with "NR" indicating not reached. Shaded areas represent 95% confidence intervals.

(D) Kaplan-Meier curves showing overall survival stratified by ACTN4 amplification status. Format and statistical methods identical to panel C.

(E) Kaplan-Meier curves showing overall survival stratified by UBR1 amplification status. Format and statistical methods identical to panel C.

Grid lines are included in all panels to aid in reading survival probabilities, hazard ratios, and p-values. All analyses were performed on the same HGSOc patient cohort with biomarker amplification data available.

Therapeutic Tractability Landscape of Synthetic Lethal Targets

Of 229 synthetic lethal target genes analyzed, 148 (64.6%) demonstrated tractability evidence across at least one therapeutic modality. Small molecule tractability was most prevalent, with 7 targets having approved drugs, 13 showing clinical precedence, and 89 exhibiting discovery precedence. For antibody-based therapeutics, 51 targets showed high confidence tractability and 30 showed medium confidence based on protein localisation and accessibility. Notably, tier 1 targets with the highest synthetic lethal effect sizes (*KRAS*, *CDK4*, *RAFI*, *ACLY*, *TXNRD1*) all demonstrated strong small molecule tractability with approved drugs, indicating immediate therapeutic relevance.

Limited Drug Availability Constrains Pharmacological Validation

Cross-referencing of Open Targets drug annotations with GDSC revealed substantial gaps in available pharmacological data. Of 28 unique drugs targeting synthetic lethal candidates, only 2 (7.1%) were available in GDSC: PALBOCICLIB and SORAFENIB. This limitation severely constrained our ability to perform systematic correlation analyses between biomarker amplification and drug sensitivity. Most absent compounds were recent clinical candidates (SOTORASIB, ADAGRASIB, REVUMENIB) or specialized therapeutics not yet incorporated into large-scale screening panels.

Proof-of-Concept Validation of CHD7-CDK4 Synthetic Lethality

We successfully demonstrated proof-of-concept validation for the *CHD7-CDK4* synthetic lethal relationship using PALBOCICLIB sensitivity data. In 4 high-grade serous ovarian cancer cell lines with CHD7 amplification (copy number ≥ 6), we observed a strong negative correlation ($r = -0.763$, $p = 0.237$) between CHD7 copy number and PALBOCICLIB AUC values, indicating that higher CHD7 amplification associates with increased CDK4 inhibitor sensitivity. While underpowered due to limited sample size (correlation of $|r| > 0.950$ required for significance at $N=4$), this analysis demonstrates the feasibility of orthogonal validation using independent pharmacogenomics datasets and supports the clinical translation potential of CNA-based synthetic lethal screening approaches. [Figure 6](#) contains a panel comprising the tractable genes and associated pathways.

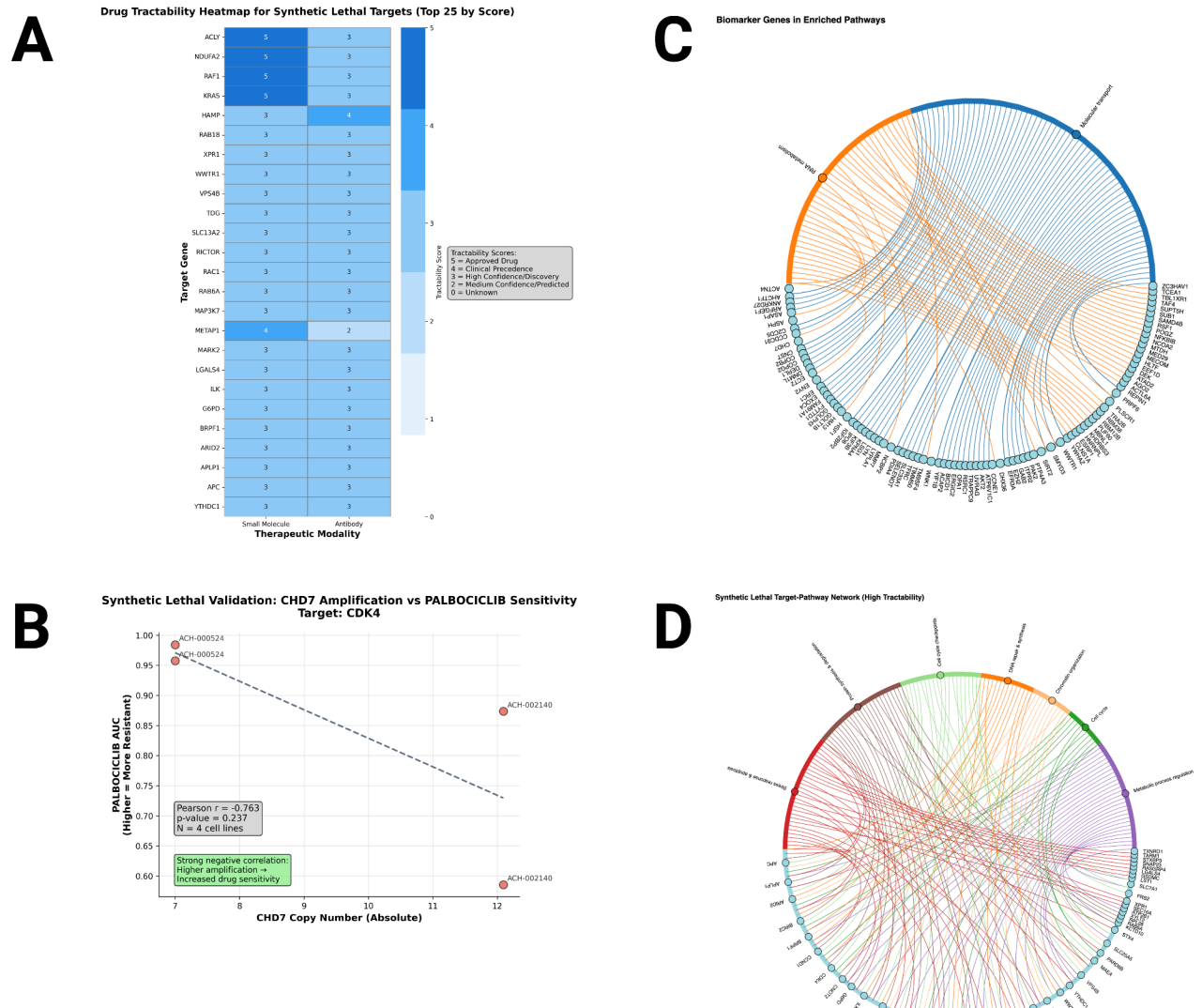


Figure 6: Translational Validation and Tractability of Synthetic Lethal Biomarkers.

(A) Ranked distribution of the top 25 tractability scores derived from Open Targets, highlighting genes with existing approved drugs or high-confidence tractability predictions.

(B) Validation plot showing the association between CHD7 amplification and sensitivity to palbociclib (target: CDK4). A strong negative correlation was observed ($r = -0.763$), although it did not reach statistical significance ($p = 0.237$) due to the limited number of available HGSOC cell lines ($n = 4$).

(C) Bokeh chord plot mapping amplified biomarkers to their enriched biological pathways, as derived from pathway collapse and g:Profiler analysis.

(D) Bokeh chord plot of tractable synthetic lethal targets and their associated pathways, focusing on genes with clinical or therapeutic evidence of tractability.

Discussion

Copy Number Alterations as a Distinct Genomic Lens for Synthetic Lethality

Our identified synthetic lethal interactions extend beyond the canonical BRCA–PARP paradigm, providing a benchmark to contextualise the novelty and potential impact of the HGSOc-specific, copy-number-driven vulnerabilities uncovered here. Because this work is copy-number focused, tumour suppressors such as TP53 and BRCA1/2, typically inactivated through mutation or promoter methylation rather than copy number alteration were not central to the analysis (Eskander and Tewari, 2014). This shift in emphasis allowed us to uncover new vulnerabilities through a distinct genomic lens.

This study focused on genes passing the stringent TCGA GISTIC2 thresholds for deep amplification (+2) or homozygous deletion (−2). Consequently, tumour suppressors more commonly observed in the shallow deletion range (−1), such as *PTEN*, were not retained in our analysis, despite their biological relevance in HGSOc. Nevertheless, this filtering captured the key recurrent alterations reported by TCGA, including amplifications of *MECOM*, *MAPK1*, *KRAS*, *CCNE1* and *PIK3CA*, and deletions of *NF1* and *RBI* (Bell *et al.*, 2011).

Contextualising Putative Synthetic Lethal Pairs vs. Canonical Synthetic Lethal Vulnerabilities

Canonical synthetic-lethal paradigms, such as *SMARCA4*-deficient cells exhibiting strong depletion of *SMARCA2* (Ehrenhöfer-Wölfer *et al.*, 2019), or *STAG2*-deficient isogenic cells undergoing uniform cell death upon *STAG1* knockdown (Mondal *et al.*, 2019), are consistently described in the literature as acute, lethal, or strongly cytotoxic. By contrast, our copy-number-driven model yields modest but statistically robust regression slopes (median $d \approx -0.26$) across 17 cell lines, alongside substantial potency deltas (median $\Delta \approx -0.36$) and strong statistical confidence (median FDR ≈ 0.005).

We recognise that the scales and readouts differ fundamentally: our study employs multi-line regression analysis of high-level amplification states (GISTIC score +2), while canonical studies use acute knockout or complete gene depletion in single or isogenic systems. Thus, comparisons should be taken as qualitative indicators of biological plausibility rather than numeric equivalences. Importantly, our amplification-driven approach captures a distinct biological scenario, the effects of oncogene overexpression rather than complete gene loss, which naturally leads to more graded dependency effects. Importantly, amplification-linked dependencies may

carry greater translational weight in HGSOC, where recurrent oncogene gains such as *CCNE1*, *MECOM*, and *KRAS* are far more prevalent than complete gene deletions (Harbers *et al.*, 2021). More broadly, meta-analyses have observed that oncogene-related synthetic lethal interactions often manifest with smaller effect sizes, highlighting that modest but consistent dependencies can still be biologically meaningful and therapeutically relevant (Lord, Quinn and Ryan, 2020). Nevertheless, the consistency of "strong" vulnerability labels in established SL models provides a qualitative benchmark for assessing the biological significance and translational potential of our copy-number-specific hits.

Transport-Enriched Biomarkers and Oncogenic-Enriched Targets Reflect Canonical Synthetic Lethal Network Logic

In our dataset, the divergence between transport-centred enrichment among biomarkers (e.g., protein/nitrogen compound transport, intracellular/Golgi vesicle transport, cellular localisation) and oncogenic pathway enrichment among targets (mitotic cell cycle, chromatin organisation/remodelling, DNA repair) reflects the underlying biology of copy-number-driven HGSOC, (see [Figure 6C & 6D](#)).

Extensive somatic copy-number alterations (SCNAs) in HGSOC create broad amplification blocks that inevitably carry “passenger” genes whose dosage shifts perturb homeostatic processes such as trafficking and localisation, yielding transport-biased biomarker signals (Beroukhi *et al.*, 2010; Zack *et al.*, 2013).

By contrast, the genes emerging as tractable targets cluster in cell-cycle control and DNA-damage response (DDR) pathways, the same axes that genomically unstable tumours become dependent upon to buffer replication stress and maintain viability (Kaelin, 2005; O’Connor, 2015; McGranahan and Swanton, 2017; Lecona and Fernandez-Capetillo, 2018).

This interpretation is further reinforced by our tractability analysis, which highlights that key cell-cycle regulators such as *CCNE1* (Discovery Precedence), *CCND1* and *CDK4* (Clinical Precedence), together with chromatin remodellers including *APC*, *ARID2* and *BRPF1* (high-confidence antibody tractability), are categorised as druggable, (see [Figure 6A](#)). This convergence of synthetic-lethal logic with tractability evidence underscores the plausibility of targeting these dependencies in HGSOC.

This pattern aligns with established synthetic-lethality logic: genome-wide dosage imbalance creates collateral vulnerabilities in essential maintenance circuits, while oncogene-addicted states

concentrate selective pressure onto a small set of survival pathways (Kaelin, 2005; Weinstein and Joe, 2006). In HGSOC specifically, TCGA showed that high-grade serous tumours are dominated by SCNAs, with recurrent amplifications (e.g., *CCNE1*, *MECOM*, *KRAS*) and frequent pathway disruptions in *RB*, *PI3K/RAS*, homologous recombination and *FOXMI* networks, exactly the domains where SL targets are expected to sit (Bell *et al.*, 2011).

Furthermore, the predominance of network-mediated over direct protein interactions (81.1% versus 3.3%, see [Figure 4B](#)) supports a model wherein synthetic lethal relationships arise through disruption of interconnected functional modules.

Taken together, transport-skewed biomarkers are best interpreted as signatures of CNA-induced cellular burden, whereas the enrichment of targets in cell-cycle/DDR/chromatin modules reflects actionable dependencies predicted by, and repeatedly validated in synthetic-lethal biology, including the mutation-centred *BRCA-PARP* archetype (Helleday, 2011; Eskander and Tewari, 2014).

Meaningful yet Modest Survival Signals

The observed hazard ratios in our HGSOC biomarker analysis, ranging from 1.31 to 1.58 for nominally significant ($p\text{-value} < 0.05$) associations, represent clinically meaningful effect sizes that warrant further investigation despite falling below the threshold typically associated with strong prognostic markers (Azuerro, 2016). These moderate hazard ratios are consistent with the polygenic nature of copy-number-driven oncogenesis in HGSOC, where individual amplified genes contribute incrementally to disease progression rather than exerting dominant effects (Beroukhi *et al.*, 2010). The convergence of several biomarkers around the 1.4-1.5 HR range suggests a coherent biological signal, particularly notable given that genes like *CCNE1*, a well-established driver in HGSOC, achieved statistical significance with an HR of 1.33, validating our analytical approach (Etemadmoghadam *et al.*, 2009). In the context of HGSOC's genomic complexity, where therapeutic options remain limited and prognosis is challenging to stratify, biomarkers with HRs exceeding 1.3 could provide clinically actionable information for patient stratification and treatment planning, especially when combined in multi-gene prognostic models (Waldron *et al.*, 2014).

From Synthetic Lethality to Therapeutic Translation

Among the tractable targets, a subset such as *CDK4* and *RAF1* aligns with existing drug portfolios and near-term translational opportunities, whereas others represent long-term exploratory candidates requiring novel therapeutic development. Translation of these findings into practice will also demand companion diagnostics capable of robustly identifying amplification-positive patients. Crucially, our tractability analysis demonstrates that 7 high-priority targets already have approved drugs (*KRAS*, *CDK4*, *RAF1*, *ACLY*, *TXNRD1*, *RPL28*, *NDUFA2*), positioning them for near-term translational testing. A further 13 genes with clinical precedence expand this into a realistic medium-term pipeline, while the remaining novel candidates (including kinases like *MAP3K7* and *MARK2*) represent the long-horizon exploratory space. This stratification frames a rational drug development roadmap: immediate translation with repurposing opportunities, medium-term investment in clinically advanced targets, and long-term innovation to capture currently intractable biology, see [Supplementary Table 2](#) for a complete gene to translational impact annotation.

Strengths, Limitations and Future Directions

This study leverages integrative in-silico pipelines and large-scale functional datasets, yet its reliance on cell line models and absence of orthogonal pharmacological validation underscores the need for experimental follow-up to substantiate clinical relevance.

While our computational framework identified 735 potent synthetic lethal interactions, the limited availability of corresponding drugs in pharmacogenomic databases (only 2 of 28 compounds in GDSC) constrained systematic validation. Future experimental work should prioritise functional CRISPR screens in HGSOC cell lines with defined amplification profiles, followed by drug sensitivity assays using available inhibitors.

An additional translational avenue is the use of zebrafish xenograft models, which provide a rapid and tractable in vivo system to test synthetic lethal dependencies under physiological conditions. With their optical transparency, high fecundity, and established oncology applications, zebrafish larvae permit real-time monitoring of tumour growth, dissemination, and angiogenesis following implantation of human ovarian cancer cells. While few of the pharmacological agents implicated in our screen have been systematically evaluated for tolerability in zebrafish, prior studies have successfully applied the model to interrogate

pathways relevant to our targets (Wei *et al.*, 2022; Madakashira *et al.*, 2024). This suggests that zebrafish assays could provide a pragmatic intermediate step, allowing rapid prioritisation of tractable synthetic lethal pairs under controlled drug exposure, generating functional evidence to de-risk subsequent mammalian validation studies.

Multi-Cancer Pipeline Applicability

While the pipeline was optimised for HGSOC, its modular design suggests it can be readily generalised to other cancers characterised by recurrent copy number alterations and synthetic lethal dependencies, such as triple-negative breast cancer, squamous lung carcinoma, and high-grade glioma, where chromosomal instability and DNA-damage response vulnerabilities are similarly pervasive (Mirchia *et al.*, 2019; Joshi *et al.*, 2021; Silvestri *et al.*, 2022).

Conclusion

This work establishes a copy-number-centred framework for uncovering synthetic lethal vulnerabilities in HGSOC, moving beyond canonical mutation-driven paradigms to highlight oncogene amplifications and collateral dependencies as therapeutically actionable features. By integrating SCNA–protein regression, CRISPR dependency profiling, tractability scoring and survival analysis, the pipeline delineates a spectrum of candidates ranging from near-term repurposing opportunities (e.g. *CDK4*, *RAF1*, *KRAS*) to longer-term exploratory targets requiring novel drug development. Although the effect sizes observed are modest, they are biologically coherent within a polygenic CNA landscape and clinically relevant when aggregated into multi-gene models. The limitations inherent to in-silico analyses underscore the importance of orthogonal validation, including functional CRISPR assays and zebrafish xenografts as intermediate models. More broadly, the modular design enables extension of this framework to other SCNA-dominated cancers such as TNBC, squamous lung carcinoma, and glioma, offering a generalisable route to prioritise synthetic lethal interactions for translational oncology.

Bibliography

Azuero, A. (2016) ‘A note on the magnitude of hazard ratios’, *Cancer*, 122(8), pp. 1298–1299. Available at: <https://doi.org/10.1002/cncr.29924>.

Bell, D. *et al.* (2011) ‘Integrated genomic analyses of ovarian carcinoma’, *Nature*, 474(7353), pp. 609–615. Available at: <https://doi.org/10.1038/nature10166>.

Beroukhi, R. *et al.* (2010) ‘The landscape of somatic copy-number alteration across human cancers’, *Nature*, 463(7283), pp. 899–905. Available at: <https://doi.org/10.1038/nature08822>.

Ehrenhöfer-Wölfer, K. *et al.* (2019) ‘SMARCA2-deficiency confers sensitivity to targeted inhibition of SMARCA4 in esophageal squamous cell carcinoma cell lines’, *Scientific Reports*, 9(1), p. 11661. Available at: <https://doi.org/10.1038/s41598-019-48152-x>.

Eskander, R.N. and Tewari, K.S. (2014) ‘PARP inhibition and synthetic lethality in ovarian cancer’, *Expert Review of Clinical Pharmacology*, 7(5), pp. 613–622. Available at: <https://doi.org/10.1586/17512433.2014.930662>.

Etemadmoghadam, D. *et al.* (2009) ‘Integrated Genome-Wide DNA Copy Number and Expression Analysis Identifies Distinct Mechanisms of Primary Chemoresistance in Ovarian Carcinomas’, *Clinical Cancer Research*, 15(4), pp. 1417–1427. Available at: <https://doi.org/10.1158/1078-0432.CCR-08-1564>.

Harbers, L. *et al.* (2021) ‘Somatic Copy Number Alterations in Human Cancers: An Analysis of Publicly Available Data From The Cancer Genome Atlas’, *Frontiers in Oncology*, 11. Available at: <https://doi.org/10.3389/fonc.2021.700568>.

He, T., Li, H. and Zhang, Z. (2023) ‘Differences of survival benefits brought by various treatments in ovarian cancer patients with different tumor stages’, *Journal of Ovarian Research*, 16(1), p. 92. Available at: <https://doi.org/10.1186/s13048-023-01173-7>.

Helleday, T. (2011) ‘The underlying mechanism for the PARP and BRCA synthetic lethality: Clearing up the misunderstandings’, *Molecular Oncology*, 5(4), pp. 387–393. Available at: <https://doi.org/10.1016/j.molonc.2011.07.001>.

Joshi, Asim *et al.* (2021) ‘Molecular characterization of lung squamous cell carcinoma tumors reveals therapeutically relevant alterations’, *Oncotarget*, 12(6), pp. 578–588. Available at: <https://doi.org/10.18632/oncotarget.27905>.

Kaelin, W.G. (2005) ‘The Concept of Synthetic Lethality in the Context of Anticancer Therapy’, *Nature Reviews Cancer*, 5(9), pp. 689–698. Available at: <https://doi.org/10.1038/nrc1691>.

Kleinmanns, K. and Bjørge, L. (2024) ‘Enhancing precision oncology in high-grade serous carcinoma: the emerging role of antibody-based therapies’, *npj Women's Health*, 2(1), p. 7. Available at: <https://doi.org/10.1038/s44294-024-00010-6>.

Lecona, E. and Fernandez-Capetillo, O. (2018) ‘Targeting ATR in cancer’, *Nature Reviews Cancer*, 18(9), pp. 586–595. Available at: <https://doi.org/10.1038/s41568-018-0034-3>.

Lee, J.S. *et al.* (2018) ‘Harnessing synthetic lethality to predict the response to cancer treatment’, *Nature Communications*, 9(1), p. 2546. Available at: <https://doi.org/10.1038/s41467-018-04647-1>.

Lord, C.J., Quinn, N. and Ryan, C.J. (2020) ‘Integrative analysis of large-scale loss-of-function screens identifies robust cancer-associated genetic interactions’, *eLife*. Edited by M.E. Murphy, 9, p. e58925. Available at: <https://doi.org/10.7554/eLife.58925>.

Madakashira, B.P. *et al.* (2024) ‘DNA hypomethylation activates Cdk4/6 and Atr to induce DNA replication and cell cycle arrest to constrain liver outgrowth in zebrafish’, *Nucleic Acids Research*, 52(6), pp. 3069–3087. Available at: <https://doi.org/10.1093/nar/gkae031>.

McGranahan, N. and Swanton, C. (2017) ‘Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future’, *Cell*, 168(4), pp. 613–628. Available at: <https://doi.org/10.1016/j.cell.2017.01.018>.

Mirchia, K. *et al.* (2019) ‘Total copy number variation as a prognostic factor in adult astrocytoma subtypes’, *Acta Neuropathologica Communications*, 7(1), p. 92. Available at: <https://doi.org/10.1186/s40478-019-0746-y>.

Mondal, G. *et al.* (2019) ‘A requirement for STAG2 in replication fork progression creates a targetable synthetic lethality in cohesin-mutant cancers’, *Nature Communications*, 10(1), p. 1686. Available at: <https://doi.org/10.1038/s41467-019-09659-z>.

O’Connor, M.J. (2015) ‘Targeting the DNA Damage Response in Cancer’, *Molecular Cell*, 60(4), pp. 547–560. Available at: <https://doi.org/10.1016/j.molcel.2015.10.040>.

Pignata, S. *et al.* (2017) ‘Treatment of recurrent ovarian cancer’, *Annals of Oncology*, 28, pp. viii51–viii56. Available at: <https://doi.org/10.1093/annonc/mdx441>.

Punzón-Jiménez, P. *et al.* (2022) ‘Molecular Management of High-Grade Serous Ovarian Carcinoma’, *International Journal of Molecular Sciences*, 23(22), p. 13777. Available at: <https://doi.org/10.3390/ijms232213777>.

Ragupathi, A. *et al.* (2023) ‘Targeting the BRCA1/2 deficient cancer with PARP inhibitors: Clinical outcomes and mechanistic insights’, *Frontiers in Cell and Developmental Biology*, 11, p. 1133472. Available at: <https://doi.org/10.3389/fcell.2023.1133472>.

Shieh, G.S. (2022) ‘Harnessing Synthetic Lethal Interactions for Personalized Medicine’, *Journal of Personalized Medicine*, 12(1), p. 98. Available at: <https://doi.org/10.3390/jpm12010098>.

Silvestri, M. *et al.* (2022) ‘Copy number alterations analysis of primary tumor tissue and circulating tumor cells from patients with early-stage triple negative breast cancer’, *Scientific Reports*, 12(1), p. 1470. Available at: <https://doi.org/10.1038/s41598-022-05502-6>.

Waldron, L. *et al.* (2014) ‘Comparative Meta-analysis of Prognostic Gene Signatures for Late-Stage Ovarian Cancer’, *JNCI: Journal of the National Cancer Institute*, 106(5), p. dju049. Available at: <https://doi.org/10.1093/jnci/dju049>.

Wang, J. *et al.* (2022) ‘SynLethDB 2.0: a web-based knowledge graph database on synthetic lethality for novel anticancer drug discovery’, *Database*, 2022, p. baac030. Available at: <https://doi.org/10.1093/database/baac030>.

Wei, G. *et al.* (2022) ‘The influence of sunitinib and sorafenib, two tyrosine kinase inhibitors, on development and thyroid system in zebrafish larvae’, *Chemosphere*, 308, p. 136354. Available at: <https://doi.org/10.1016/j.chemosphere.2022.136354>.

Weinstein, I.B. and Joe, A.K. (2006) ‘Mechanisms of disease: Oncogene addiction--a rationale for molecular targeting in cancer therapy’, *Nature Clinical Practice. Oncology*, 3(8), pp. 448–457. Available at: <https://doi.org/10.1038/ncponc0558>.

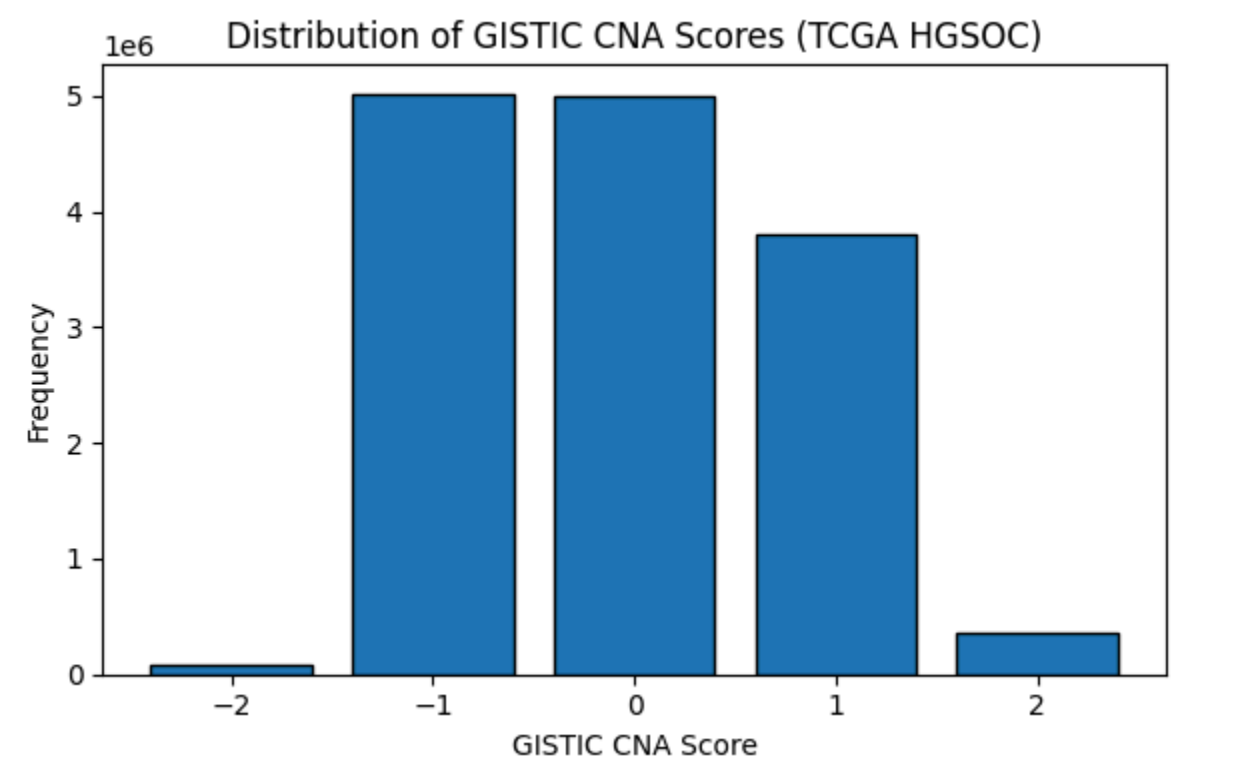
Zack, T.I. *et al.* (2013) ‘Pan-cancer patterns of somatic copy number alteration’, *Nature Genetics*, 45(10), pp. 1134–1140. Available at: <https://doi.org/10.1038/ng.2760>.

Zhan, T. *et al.* (2019) ‘CRISPR/Cas9 for cancer research and therapy’, *Seminars in Cancer Biology*, 55, pp. 106–119. Available at: <https://doi.org/10.1016/j.semcancer.2018.04.001>.

Zhang, H. *et al.* (2016) ‘Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer’, *Cell*, 166(3), pp. 755–765. Available at: <https://doi.org/10.1016/j.cell.2016.05.069>.

Appendices

Supplementary Figure 1



Supplementary Figure 1: Distribution of GISTIC CNA Scores in TCGA HGSOC.
Bar plot showing the frequency of copy number alteration (CNA) events across all genes in the TCGA high-grade serous ovarian cancer (HGSOC) cohort. CNA states were derived from GISTIC2.0 scores: -2 (deep deletion), -1 (shallow deletion), 0 (diploid), +1 (low-level gain), and +2 (high-level amplification). The majority of events fall within shallow deletion (-1), diploid (0), and low-level gain (+1) categories, with fewer deep deletions and high-level amplifications observed.

Supplementary Table 1

Supplementary Table 1: Annotation of Datasets Used in This Study

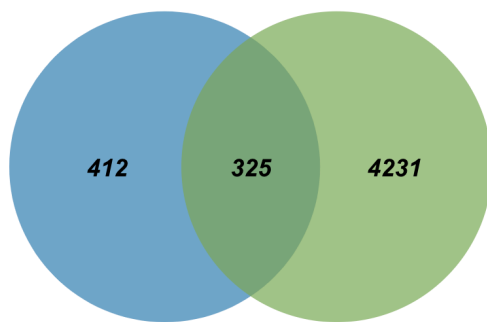
Dataset	Source	File Name	Description	Cell Lines Available
TCGA Copy Number Alterations	TCGA PanCancer Atlas (via cBioPortal)	data_cna.txt	Gene-level GISTIC2.0 scores (-2 to +2) across 562 HGSOC patient samples.	N/A

TCGA Protein Expression (RPPA)	TCGA PanCancer Atlas (via cBioPortal)	data_protein_quantification.txt	Normalised RPPA expression for selected proteins in HGSOC tumours	N/A
Cell Line Metadata	DepMap Annotations	cell_lines_in_High-Grade_Serous_Ovarian_Cancer.csv	Used to ensure only HGSOC cell lines were subsetted from DepMap files.	JHOS2, OVMIU , OVCAR5, SNU119, TYKNU, HEY, CAO4, NIHOVCAR3, COV318, HEYA8, OVSAHO, OVCAR8, FUOV1, KURAMOCHI, OAW28, 59M, COV362, JHOS4, ONCODG1, PEA1, OVKATE, OVCAR3, CAO3
PureCN Values (DepMap)	DepMap Public 24Q4	OmicsAbsoluteCNGene.csv	Absolute CNA data for genes in HGSOC tumours	JHOS2, OVMIU, OVCAR5, SNU119, TYKNU, HEY, CAO4, NIHOVCAR3, COV318, HEYA8, OVSAHO, OVCAR8, FUOV1, KURAMOCHI, OAW28, 59M,

				COV362, JHOS4, ONCODG1, PEA1, OVKATE
CRISPR Dependency Scores	DepMap (DEMETER2/A vana)	CRISPRGeneEffect.csv	Gene dependency scores from pooled CRISPR screens in cancer cell lines	JHOS2, OVMIU, OVCAR5, TYKNU, HEY, CAOV4, NIHOVCAR3, COV318, HEYA8, OVCAR8, KURAMOCHI, OAW28, 59M, COV362, JHOS4, CAOV3, ONCODG1, PEA1

Supplementary Figure 2

TCGA Significant Genes (regression: <0.05 ; $n=737$)
 \cap
 DepMap Amplifications (CN ≥ 6 ; $n=4556$)

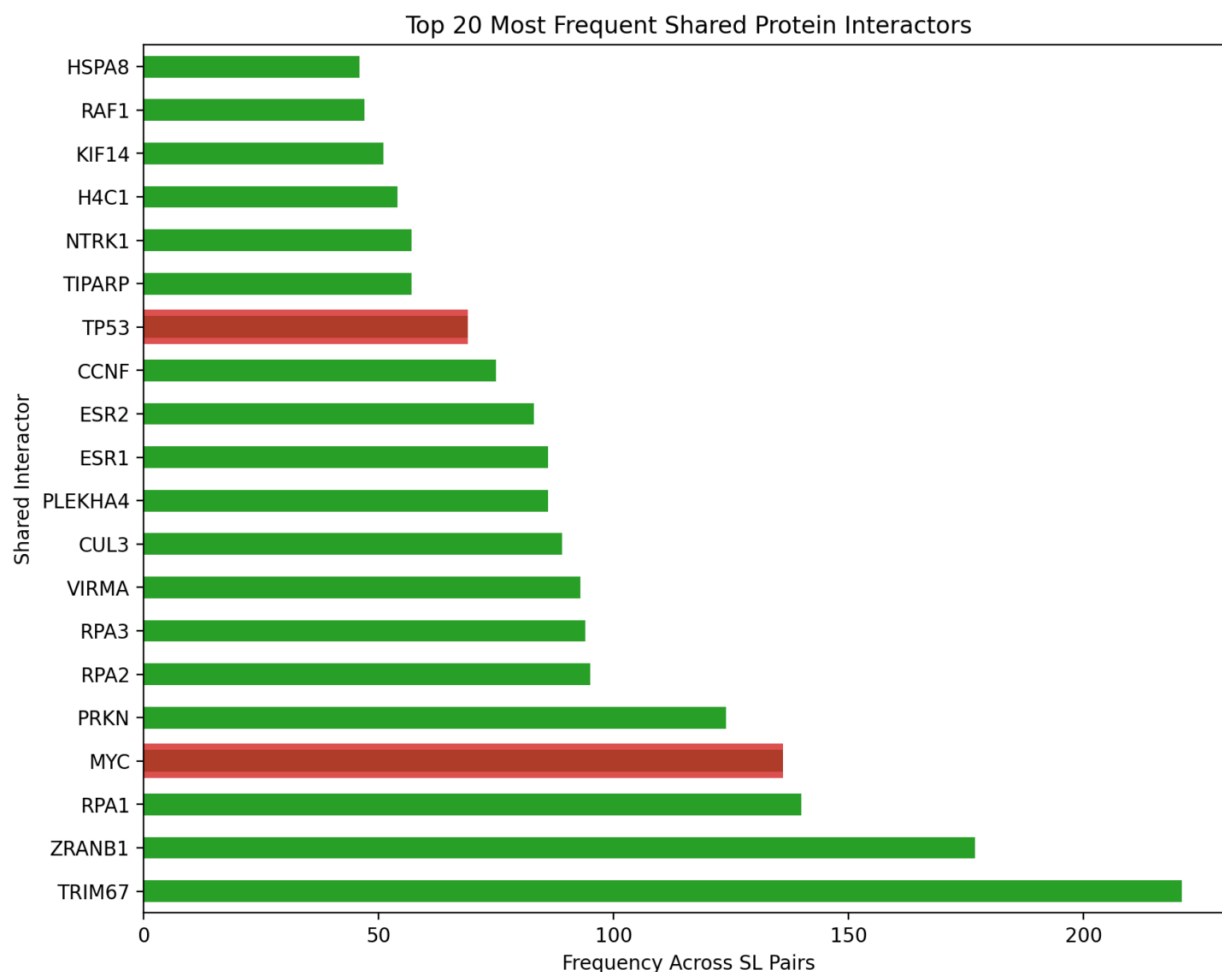


TCGA Significant Genes (regression: <0.05 ; $n=737$)
 \cap
 DepMap Deletions (CN < 1 ; $n=42$)

**Supplementary Figure 2: Cross-Validation of TCGA CNA–Protein Hits in HGSOC Cell Lines.**

Venn diagrams comparing the 737 TCGA genes with significant CNA–protein associations ($p < 0.05$) to genes altered in DepMap HGSOC cell lines. (Left) Amplifications were defined as absolute copy number ≥ 6 , observed in ≥ 3 lines ($n = 4,556$). This analysis identified 325 overlapping genes supported in both TCGA tumours and cell lines. (Right) Deletions were defined as absolute copy number < 1 , observed in ≥ 3 lines ($n = 42$). No overlap with the TCGA significant set was detected.

Supplementary Figure 3



Supplementary Figure 3: Top shared protein interactors across synthetic-lethal pairs.

Bar chart showing the 20 most frequent shared protein interactors among the representative biomarker–target pairs ($n = 735$). Bar length denotes the number of SL pairs in which each interactor is shared. TP53 and MYC, canonical tumour suppressor and oncogenic regulator are highlighted, reflecting their recurrent appearance across multiple SL relationships. Using BioGRID ($>850k$ PPIs), only 24 pairs (3.3%) had a direct physical interaction, whereas 620 pairs (84.4%) shared ≥ 1 indirect interactor, indicating extensive network-level connectivity among SL candidates.

Supplementary Table 2

Supplementary Table 2: Tractability Classification of Synthetic Lethal Target Genes

Tier 1: Clinical Translation Ready (7 targets)*Genes with approved drugs demonstrating immediate therapeutic potential*

Gene Symbol	Gene Name	Primary Function	Drug Status	Representative Drugs	SM Tractability	AB Tractability
KRAS	KRAS proto-oncogene	Oncogenic signalling	Approved	AMG 510 (Sotorasib), MRTX849 (Adagrasib)	Clinical Precedence	High Confidence
CDK4	Cyclin-dependent kinase 4	Cell cycle regulation	Approved	Palbociclib, Ribociclib, Abemaciclib	Clinical Precedence	Unknown
RAF1	Raf-1 proto-oncogene	MAPK pathway kinase	Approved	Sorafenib	Clinical Precedence	High Confidence
ACLY	ATP citrate lyase	Metabolic enzyme	Approved	Bempedoic acid	Clinical Precedence	High Confidence
TXNRD1	Thioredoxin reductase 1	Redox homeostasis	Approved	Auranofin	Clinical Precedence	Unknown
NDUFA2	NADH dehydrogenase subunit	Mitochondrial respiration	Approved	Various	Clinical Precedence	High Confidence
RPL28	Ribosomal protein L28	Protein synthesis	Approved	Homoharringtonine	Clinical Precedence	Unknown

Tier 2: Pipeline Ready (6 targets)

Genes with strong clinical precedence in advanced trials

Gene Symbol	Gene Name	Primary Function	Clinical Status	SM Tractability	AB Tractability	Strategic Value
RELA	RelA (p65 NF-κB subunit)	Transcription factor	Advanced Clinical	Clinical Precedence	Unknown	Inflammation-cancer nexus targeting
BIRC2	Baculoviral IAP repeat 2	Apoptosis inhibitor	Advanced Clinical	Clinical Precedence	Unknown	Direct apoptosis pathway modulation
CCND1	Cyclin D1	Cell cycle regulator	Advanced Clinical	Clinical Precedence	Unknown	G1/S checkpoint control
MEN1	Menin	Tumour suppressor	Phase 1 Clinical	Clinical Precedence	Unknown	Epigenetic regulation
KMT2A	Lysine methyltransferase 2A	Histone methylation	Phase 1 Clinical	Clinical Precedence	Unknown	Chromatin remodelling target
METAP1	Methionyl aminopeptidase 1	Protein processing	Advanced Clinical	Clinical Precedence	Medium-Low Confidence	Novel mechanism exploitation

Tier 3: Discovery Highlights (Multi-Modal Opportunities) (27 targets)

Genes tractable through multiple modalities with high biological significance

Tier 3A: High-Confidence Multi-Modal (20 targets)

Both small molecule and high-confidence antibody tractable

Gene Symbol	Gene Name	Primary Function	SM Tractability	AB Tractability	Biological Significance
MAP3K7	Mitogen-activated protein kinase kinase kinase 7	Kinase signalling	Discovery Precedence	High Confidence	Stress response pathway regulation
MARK2	MAP/microtubule affinity regulating kinase 2	Protein kinase	Discovery Precedence	High Confidence	Microtubule dynamics control
ILK	Integrin-linked kinase	Cell adhesion kinase	Discovery Precedence	High Confidence	Cell survival and adhesion
G6PD	Glucose-6-phosphate dehydrogenase	Metabolic enzyme	Discovery Precedence	High Confidence	Glucose metabolism regulation
RICTOR	RPTOR independent companion of MTOR	mTOR complex component	Discovery Precedence	High Confidence	Growth signaling control
RAC1	Rac family small GTPase 1	Small GTPase	Discovery Precedence	High Confidence	Cytoskeleton and motility
WWTR1	WW domain containing transcription regulator 1	Transcription cofactor	Discovery Precedence	High Confidence	Hippo pathway regulation
VPS4B	Vacuolar protein sorting 4 homolog B	Vesicle trafficking	Discovery Precedence	High Confidence	Autophagy regulation
ARID2	AT-rich interaction domain 2	Chromatin remodelling	Discovery Precedence	High Confidence	SWI/SNF complex component
BRPF1	Bromodomain and PHD finger containing 1	Epigenetic regulator	Discovery Precedence	High Confidence	Histone acetylation

XPR1	Xenotropic and polytropic retrovirus receptor 1	Phosphate exporter	Discovery Precedence	High Confidence	Phosphate homeostasis
YTHDC1	YTH domain containing 1	RNA binding protein	Discovery Precedence	High Confidence	mRNA processing
TDG	Thymine DNA glycosylase	DNA repair enzyme	Discovery Precedence	High Confidence	Base excision repair
TARM1	T cell receptor associated transmembrane adaptor 1	Membrane adaptor	Unknown	High Confidence	T cell signalling
STXBP3	Syntaxin binding protein 3	Vesicle trafficking	Unknown	High Confidence	SNARE complex regulation
STX4	Syntaxin 4	Vesicle fusion	Unknown	High Confidence	Membrane trafficking
SNAP23	Synaptosome associated protein 23	SNARE protein	Unknown	High Confidence	Vesicle exocytosis
SLC4A7	Solute carrier family 4 member 7	Ion transporter	Unknown	High Confidence	pH homeostasis
SLC7A1	Solute carrier family 7 member 1	Amino acid transporter	Unknown	High Confidence	Arginine transport
SLC44A3	Solute carrier family 44 member 3	Choline transporter	Unknown	High Confidence	Choline metabolism

Tier 3B: Promising Discovery Targets (7 targets)

High biological significance with tractability potential

Gene Symbol	Gene Name	Primary Function	SM Tractability	AB Tractability	Notes
CCNE1	Cyclin E1	Cell cycle regulator	Discovery Precedence	Unknown	S-phase progression control

APC	APC regulator of WNT signalling	Tumour suppressor	Discovery Precedence	High Confidence	Wnt pathway regulation
ALDOC	Aldolase, fructose-bisphosphate C	Glycolytic enzyme	Predicted Tractable	High Confidence	Metabolic reprogramming
KDM2A	Lysine demethylase 2A	Histone demethylase	Discovery Precedence	Unknown	Chromatin modification
KMT5A	Lysine methyltransferase 5A	Histone methyltransferase	Discovery Precedence	Unknown	H4K20 methylation
L3MBTL3	L3MBTL histone methyl-lysine binding protein 3	Chromatin reader	Discovery Precedence	Unknown	Epigenetic regulation
SMARCA4	SWI/SNF related chromatin remodelling complex subunit	Chromatin remodeller	Discovery Precedence	Unknown	SWI/SNF complex ATPase

Legend:

- **SM:** Small Molecule tractability
- **AB:** Antibody tractability
- **Clinical Precedence:** Compounds in clinical trials or approved
- **Discovery Precedence:** Known binding sites or structural information
- **Predicted Tractable:** Computational prediction of druggability
- **High/Medium/Low Confidence:** Based on subcellular localisation and accessibility scores

Data Source: Open Targets Platform tractability assessment for synthetic lethal target genes identified from HGSOC dependency screens.