BOGO: A Proteome-Wide Gene Overexpression Platform for Discovering Rational Cancer Combination Therapies

Kyeong Beom Jo[1,2,3,*], Mohammed M. Alruwaili[3,*], Da-Eun Kim[4,*], Yongjun Koh[5,*], Hyeyeon Kim[2,*], Kwontae You[6], Ji-Sun Kim[2], Saba Sane[1,2], Yanqi Guo[3], Jacob P. Wright[3], Hyobin Julianne Lim[3,7], Maricris N. Naranjo[3], Atina G. Coté[8], Frederick P. Roth[9], David E. Hill[10,11,12], Jung-Hyun Choi[13], Hunsang Lee[14], Kenneth A. Matreyek[15], Kyle K.-H. Farh[6], Jong-Eun Park[5,#], Hyunkyung Kim[,#], Andrei V. Bakin[3,#], Dae-Kyum Kim[1,2,3,16,17,18,#].

[1]Division of Surgical and Interventional Sciences, Department of Surgery, McGill University, Montreal, Quebec, Canada

[2]Cancer Research Program, The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada

[3]Department of Cancer Genetics and Genomics, Roswell Park Comprehensive Cancer Center, Buffalo, New York, USA

[4]Department of Biochemistry and Molecular Biology, College of Medicine, Korea University, Seoul, Korea

[5]Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea

[6]Illumina Artificial Intelligence Laboratory, Illumina Inc., San Diego, California, USA

[7]Tetrad Graduate Program, University of California, San Francisco, California, USA

[8]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, Canada

[9]Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, USA

[10]Center for Cancer Systems Biology (CCSB), Dana-Farber Cancer Institute, Boston, Massachusetts, USA

[11]Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, Massachusetts, USA

[12]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA

[13]Department of Biochemistry, Chungbuk National University, Cheongju, Korea

[14]Department of Life Sciences, School of Life Sciences and Biotechnology, Korea University, Seoul, Korea

[15]Department of Pathology, School of Medicine, Case Western Reverse University, Cleveland, Ohio, USA

[16]Division of Thoracic and Upper Gastrointestinal Surgery, Department of Surgery, Faculty of Medicine and Health Sciences, McGill University, Montreal, Quebec, Canada

[17]Montreal General Hospital Foundation, McGill University Health Centre, Montreal, Quebec, Canada

[18]Aune Foundation, Vancouver, British Columbia, Canada

*Co-first

#Co-corresponding

Abstract

Cancer drug resistance remains a major barrier to durable treatment success, often leading to relapse despite advances in precision oncology. While combination therapies are being increasingly investigated, such as chemotherapy with small molecule inhibitors, predicting drug response and identifying rational drug combinations based on resistance mechanisms remain major challenges. Therefore, a proteome-wide, single-gene overexpression screening platform is essential for guiding rational therapy selection. Here, we present **BOGO** (**B**xb1-landing pad human **O**RFeome-integrated system for a proteome-wide **G**ene **O**verexpression), a robust, scalable, and reproducible screening platform that enables single-copy, site-specific integration and overexpression of ~19,000 human ORFs across cancer cell models. Using BOGO, we identified drug-specific response drivers for 16 chemotherapeutic agents and integrated clinical datasets to uncover proliferation and resistance-associated genes with prognostic potential. Drug response similarity networks revealed both shared and unique mechanisms, highlighting key pathways such as autophagy, apoptosis, and Wnt signaling, and notable resistance-associated genes including BCL2, POLD2, and TRADD. In particular, we proposed a synergistic combination of the BCL2 family inhibitor ABT-263 (Navitoclax®) and the DNA analog TAS-102 (Lonsurf®), which revealed that lysosomal modulation is a key mechanism driving DNA analog resistance. This combination therapy selectively enhanced cytotoxicity in colorectal and pancreatic cancer cells *in vitro*, and demonstrated therapeutic benefit *in vivo* in both cell line-derived xenograft (CDX) and patient-derived xenograft (PDX) models. Together, these findings establish BOGO as a powerful gene overexpression perturbation platform for systematically identifying chemoresistance and chemosensitization drivers, and for discovering rational combination therapies. Its scalability and reproducibility position BOGO as a broadly applicable tool for functional genomics and therapeutic discovery beyond cancer resistance.

Introduction

Cancer therapies have evolved from traditional chemotherapy to more precise approaches such as targeted and cell therapy treatments[1]. Despite these advances, chemotherapy remains a first-line option for many cancers, often due to cost-benefit considerations and the lack of well-defined molecular targets. However, its clinical effectiveness is frequently limited by the emergence of drug resistance, which is a major cause of treatment failure and cancer-related mortality[2]. To address this challenge and enhance the impact of chemotherapy, combining it with targeted inhibitors offers a promising strategy to overcome resistance mechanisms and improve therapeutic efficacy[3]. Combining chemotherapy agents such as doxorubicin with epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor, Gefitinib, has shown potential in disrupting key signaling pathways, thereby restoring sensitivity in resistant cancer cells and improving treatment outcomes[4]. It is essential to understand the systemic molecular dynamics of chemoresistance and chemosensitization for the rational design of combinatorial strategies.

Contemporary approaches to precision oncology have primarily relied on genomic profiling, such as whole genome and whole exome sequencing, to identify actionable mutations and guide targeted therapy[5]. While such strategies have proven transformative for personalized treatment, they have shown limited predictive value for responses to conventional chemotherapeutics[6]. Genomic and transcriptomic data, including mRNA sequencing, often provide only static snapshots of tumor biology and fail to capture the functional and adaptive properties of cancer cells under drug pressure. Critical determinants of drug sensitivity, such as pathway activity, protein abundance, and cellular plasticity, are not readily revealed by those sequencing-based analyses[7]. To address these challenges, functional genomics has emerged as a powerful complementary approach to static genetic profiling[8]. By experimentally perturbing gene expression and measuring phenotypic outcomes, functional screening can uncover key regulators of drug response that remain invisible to static profiling.

The most common functional genomics method is loss-of-function screening, such as CRISPR-Cas9 knockout screenings[9], which have provided valuable insights into gene dependencies in cancer. While highly informative, knockout screening has some challenges in confirming the knockout effect of already silenced genes, which is ~50% of the whole protein-coding genes per cell line or tissue[10]. Complementing knockout screening, gain-of-function approaches such as CRISPR activation (CRISPRa)[11] and open reading frame (ORF) library screens[12], most often delivered via lentiviral systems[13], have offered an additional perspective. Gene overexpression screening is especially well suited to identify resistance mechanisms driven by gene amplification or transcriptional upregulation, thereby revealing the “escape routes” cancer cells use to evade therapeutic pressure and sustain proliferation[14]. This may provide a unique and rationalized approach for proposing targeted therapy as a complementary partner to chemotherapy. While overexpression screenings using CRISPRa[15] or lentivirus[16] are useful, they have limitations in capturing the full complexity of tumor heterogeneity. These include heterogeneous gene expression due to random genomic integration[17], limited proteome-wide coverage, defined here as coverage across protein-coding genes, caused by chromatin accessibility dynamics[18], and false-positive driver candidates identified from activation of multiple genes per cell[19]. Moreover, since 90% of FDA-approved

drugs that can be used for combinatorial therapy are known to target proteins, functional genomics on protein-coding genes[20] could provide more immediate clinical insights for drug repurposing[21]. Therefore, there is a need for an overexpression screening platform that enables a “proteome-wide”, “single gene per cell” manner from a defined genomic locus to find the main “drivers” for chemoresistance.

This study introduces BOGO, a new, robust, scalable gene overexpression screening platform that has precisely integrated the comprehensive human ORFeome (hORFeome) collection[22,23] using the Bxb1 landing pad system[24] [**B**xb1-landing pad human **O**RFeome-integrated system for a proteome-wide **G**ene **O**verexpression (**BOGO**)]. BOGO enables systematic investigation of how gene overexpression contributes to drug resistance in cancer (**Figure 1A**). The ORFeome collection provides near-complete coverage of proteome-wide genes of ~19,000 ORFs. The BOGO system ensures precise integration of a single ORF at the AAVS1 locus, eliminating the variability and insertional artifacts commonly associated with traditional gene overexpression methods and guaranteeing a single gene overexpression per cell. We applied BOGO in the HeLa cell line, which is the most widely used cancer cell model[25], and exposed them to 16 clinically relevant chemotherapeutic agents, uncovering both established and novel resistance mechanisms, particularly involving the BCL2 binding domain, autophagy, and apoptosis pathways. Thymidine-based nucleoside analog chemotherapy, TAS-102, activated lysosomal activity while the BCL2 inhibitor, ABT-263, suppressed BCL2 family overexpression and disrupted lysosomal function. Combined treatment of TAS-102 and ABT-263 resulted in increased cytotoxicity in colon and pancreatic cancer cell lines, as well as enhanced tumor suppression in colon cancer CDX and PDX models. This platform predicts resistance mechanisms, informs treatment selection, and reveals oncogenic drivers. BOGO identifies genes that drive resistance and supports the rational design of combination therapies by mapping drug-gene interactions.

Results

The BOGO system represents a significant advance in functional genomics by providing a reliable platform for gain-of-function screening at a massive scale.

BOGO facilitates stable, single-ORF-per-cell overexpression through precise integration of a Bxb1 landing pad donor vector into the AAVS1 locus via homologous recombination[26] (**Figure 1B**). Gateway cloning was used to ensure accurate bulk overexpression of protein-coding genes, with each ORF expressed under the EF1α promoter. The resulting cell pool encompassed ~19,000 protein-coding genes, derived from hORFeome 9.1[22,23] and additional clones (ORFeome Delta) provided by collaborators. HeLa cells engineered with the Bxb1 landing pad system were transfected with the hORFeome library (**Figure 1B**). The Bxb1 integrase enabled precise, single-ORF insertion at the AAVS1 site, minimizing variability and ensuring uniform gene expression. Cells were sorted based on mCherry reporter signal and expanded to form the experimental pool, representing the full hORFeome landscape. fluorescence-activated cell sorting (FACS) analysis revealed that ~20% of cells were mCherry-positive/mTagBFP2-negative (**Supplementary Figure 1A, left panel**), achieving 150× coverage with over 3 million sorted cells, indicating robust representation of the ORFeome (**Supplementary Figure 1A, right panel**). Extensive sequencing of the initial cell pool confirmed that ORF sequences were concordant with their corresponding coding DNA sequences (CDS), and approximately 90% of ORFs were successfully integrated (**Figure 1C**). To assess reproducibility, we tested five independent batches in large-scale drug screens. Results showed minimal batch effects, with high correlations between technical replicates across treatments (**Figure 1D**), confirming the stability and reliability of ORFeome representation after freeze-thaw.

Leveraging the established HeLa ORFeome cell pool, we conducted two initial functional screens to investigate: i) drug response, and ii) ORF silencing, both of which are critical aspects of cancer biology and gene expression regulation. To evaluate the impact of individual ORFs on drug response, we quantified the relative abundance of ORF-expressing cell populations across three conditions: i) baseline (untreated library cells), ii) DMSO-treated cells (vehicle control), and iii) chemotherapy-treated cells (experimental condition). Amplicons derived from genomic DNA were subjected to next-generation sequencing (NGS) to measure ORF abundance. The $\log_2$ fold change ($\log_2$FC) between drug-treated and vehicle control conditions was used to assess differential representation. ORFs with $\log_2FC > 0.5$ or < -0.5 and $p < 0.05$ were defined as drug response effectors (**Supplementary Table 1**). We classified these enriched ORFs as drug resistance drivers, while depleted ORFs were designated as drug sensitization drivers.

To verify the reproducibility and functional relevance of our perturbation system, we performed individual overexpression experiments for seven selected ORFs, comparing their effects to mCherry fluorescence control cells. These validations confirmed the phenotypic outcomes observed in the BOGO drug response screen. For example, overexpression of BCL2 resulted in increased cell abundance under Trifluridine (TFT) treatment, consistent with its classification as a drug resistance driver in our screen (**Figure 1E**). Conversely, overexpression of CDK6, identified as a drug sensitization driver, led to a marked reduction in cell population following TFT treatment (**Figure 1E**), aligning with previous findings[27] that CDK6 overexpression sensitizes cells

to DNA damage. This strong concordance between phenotypic changes in single-ORF overexpression lines and proteome-wide perturbation profiles provides direct evidence for the functional validity and reproducibility of the BOGO screening methodology.

To demonstrate the complementary value of the BOGO screening platform relative to conventional transcriptomic approaches, we performed bulk mRNA sequencing on wild-type HeLa cells treated with doxorubicin, a widely used chemotherapeutic agent with broad clinical applications[28]. Transcriptional responses were profiled at two time points: 120 hours, corresponding to the duration of our ORFeome screen, and 24 hours, to capture acute gene expression changes. We compared the sets of differentially expressed genes (DEGs) identified in these mRNA-seq datasets (**Supplementary Table 2**) with the doxorubicin response effectors uncovered through our drug response screen. This analysis revealed both overlapping and distinct gene sets (**Figure 1F**). Notably, the BOGO screen identified approximately five times more genes than bulk mRNA-seq at the 120-hour time point, with only 2% overlap. At 24 hours, the BOGO screen detected ten times more genes, yet the overlap with DEGs was even lower, at 0.5%. These differences likely stem from the inherent limitations of bulk mRNA sequencing, which captures endogenous gene expression subject to dynamic regulatory mechanisms. In contrast, the BOGO platform utilizes constitutive ORF overexpression, enabling the identification of functional drug resistance effectors that may be masked or underrepresented in natural transcriptional profiles. Thus, overexpression-based screening offers a powerful and complementary approach to conventional transcriptomics, revealing biologically and clinically relevant genes that may be overlooked by standard methods.

To validate and further characterize the hits identified in our drug response screen, we performed high-saturation retesting (HSR), a targeted approach involving smaller, selectively curated subsets of the ORFeome library analyzed at increased coverage to ensure reproducibility and robustness. Approximately 10% of the original ORF pool, comprising a mix of proliferation- and drug response-associated ORFs, was retested at >1,000-fold coverage. We compared the results of the initial drug response screen with those from the HSR for doxorubicin (**Supplementary Table 3**). First, we confirmed high representation of the selected ORFs in the HSR dataset in which 2,184 ORFs (98.5%) from the cherry-picked pool were successfully detected (**Supplementary Figure 1E, left Venn diagram**). Among the doxorubicin effector ORFs, 93 (96.9%) were recovered in the HSR, with only 3 ORFs (3.1%) not detected (**Supplementary Figure 1E, right Venn diagram**), confirming the fidelity and coverage of the retesting process. Importantly, the HSR validated the directional trends observed in the initial screen (**Figure 1G**). ORFs previously identified as drug sensitization drivers, such as HIT1H3B, exhibited negative average fold changes in the HSR, consistent with enhanced drug sensitivity. Conversely, drug resistance drivers, such as PITX2, showed positive average fold changes, confirming their resistance phenotype (**Supplementary Figure 1B**). ORFs classified as non-respondents consistently displayed near-zero fold changes, reinforcing their lack of drug response. These results demonstrate the statistical robustness and reproducibility of the BOGO platform, supporting its reliability in identifying biologically and clinically relevant drug response effectors.

In parallel, we addressed the challenge of ORF silencing, an inherent limitation of overexpression systems. To monitor silencing, an IRES-mCherry cassette was inserted downstream of each ORF, allowing

discrimination between active and silenced ORF expression. By comparing ORF abundance in mCherry-positive versus mCherry-negative cell populations, we quantified the proportion of active expression. ORFs with $\log_2$FC > -1 or < -1 and $p < 0.05$ were defined as silencing effectors. Positive $\log_2$FC values indicated non-silencing ORFs, while negative values indicated silencing ORFs. Analysis revealed that over 60% of ORFs were retained as non-silencing, while fewer than 30% exhibited measurable silencing effects (**Supplementary Figure 1C**). Furthermore, five selected genes maintained stable expression over five days of culture, with no observable silencing, corroborating their classification as non-silencing ORFs in our screen. In contrast, GPER and BRD9 exhibited silencing effects (**Supplementary Figure 1D**), consistent with their silencing profiles in the drug response screen (**Supplementary Figure 1E**). Notably, GPER, a membrane protein known to induce cellular toxicity via inclusion body formation, has been previously reported to suppress cell growth through this mechanism[29], explaining its lack of detectable representation. These results confirm the stability and reliability of gene expression within the BOGO system and demonstrate its utility in identifying functional drug response effectors and assessing gene silencing tendencies.

Functional classification of drug-response effectors reveals their distinct molecular features and predicts clinical outcomes in cancer.

Our drug response screen comprised 16 chemotherapeutic agents, encompassing a broad range of cytotoxic mechanisms. These chemotherapeutics were grouped into four distinct functional categories based on their mechanisms of action, including 6 antimetabolite activity-based drugs, 4 DNA cross-linking agents, 4 strand-break agents, and 2 microtubule-arresting drugs, serving as a reference map of key cytotoxic mechanisms[30] (**Supplementary Figure 2A**). We systematically classified 18,388 human open reading frames (ORFs) into distinct categories based on their influence on chemotherapeutic drug responses (**Figure 2A**). Due to low or undetectable representation in cell populations, 3,882 ORFs were excluded from further analysis. The remaining 14,506 ORFs were categorized according to their differential representation patterns across different chemotherapeutic treatments, resulting in five functional groups: Core, Categorical, Multidrug, Peripheral, and Non-respondent (**Figure 2A**).

To assess the landscape of drug-response-modulating ORFs (dORFs), we first examined their cumulative distribution across all tested chemotherapeutic agents. The number of unique dORFs increased progressively with the inclusion of additional drugs (**Supplementary Figure 2B**), a trend that significantly exceeded expectations from random permutation, thereby validating the sensitivity and breadth of our screening platform in capturing gene-drug interactions. A histogram of drug response effects revealed a decline in the number of dORFs as the number of drugs increased, with only a small subset of dORFs exerting broad regulatory effects across up to 13 agents (**Supplementary Figure 2C**). Based on this distribution, dORFs affecting responses to more than 7 drugs, approximately half of the maximum observed, were designated as Core ORFs. To define thresholds for broader versus singular drug responses, we modeled the distribution of drug-response counts per ORF using a mixture of gamma and normal distributions. The intersection corresponded to 2.64 drugs (**Supplementary Figure 2D**), which was used to define Multidrug ORFs as those influencing responses to between 2 and 7 drugs. Peripheral ORFs were defined as those modulating responses to a single

chemotherapeutic agent (**Supplementary Figure 2E**), reflecting the specificity of certain gene-drug interactions. Within the Categorical ORFs, defined as ORFs that selectively affect responses to drugs with shared mechanisms of action, dORFs for DNA strand break inducers were most prevalent, although other drug categories also exhibited distinct dORF profiles (**Supplementary Figure 2F**). Non-respondent ORFs are those that showed no significant impact on drug responses across all tested agents. The mutual exclusivity of these categories was confirmed via Venn diagram analysis (**Supplementary Figure 2G**).

We next investigated whether the defined ORF categories exhibit distinct intrinsic properties that may underlie their differential drug response profiles. Among these, Core ORFs demonstrated the most pronounced divergence across multiple molecular and functional features, consistent with their broad association with chemoresistant phenotypes. Specifically, Core ORFs exhibited significantly elevated z-scores for GC content, predicted protein disorder, Gene Ontology (GO) term enrichment, publication frequency (PubCount), and associations with Mendelian diseases (**Figure 2B**). Conversely, features such as tissue-specific upregulation (Tissue Count), evolutionary age, expression breadth (number of tissues in which the gene is expressed), expression abundance, and the number of associated protein complexes (as annotated in the CORUM database) were markedly reduced in Core ORFs (**Figure 2B**). Notably, Core ORFs also tended to be shorter in length compared to other categories, whereas Categorical ORFs exhibited the longest sequences. These observations suggest that ORFs with low baseline expression may be more susceptible to phenotypic modulation upon artificial overexpression, potentially accounting for their prominence in our screen. In contrast, genes with high endogenous expression may be less responsive to perturbation, thereby limiting their detectability as drug response effectors. While this bias may influence the representation of Core ORFs, our screening platform remains robust in identifying functionally relevant drivers of drug response and disease. Taken together, these findings indicate that Core ORFs represent a distinct subset of underappreciated regulatory elements that acquire functional significance upon upregulation, differing from other ORF categories in both structural and functional attributes.

To evaluate the clinical relevance of the categorized ORFs, we assessed their overlap with cancer-specific biomarkers curated in TheMarker database[31]. For each ORF category, we quantified three metrics: (1) the percentage of ORFs overlapping with known cancer biomarkers (% of ORFs), (2) the proportion of biomarkers associated with chemotherapeutic agents included in our drug response screen (% of Therapeutics), and (3) the contribution of each ORF category to the overall biomarker pool (% of Biomarkers; **Figure 2C, left**). Core ORFs exhibited significantly elevated z-scores across all three metrics, indicating strong enrichment for clinically validated biomarkers and highlighting their broad relevance to chemotherapy response. In contrast, Categorical ORFs showed lower z-scores, suggesting that their functional roles may be more mechanism-specific and less represented in current biomarker databases. These trends were clearly visualized in heatmaps of z-score distributions, which highlighted the translational potential of distinct ORF subsets (**Figure 2C, left**). Further analysis of biomarker types revealed a predominance of predictive biomarkers across all ORF categories, supporting the hypothesis that individual ORFs may influence favorable or unfavorable outcomes in response to specific chemotherapeutic agents (**Figure 2C, right**). This integrative approach, combining functional

categorization with biomarker overlap, demonstrates the utility of our screening platform in identifying ORF candidates with high translational relevance.

To further assess the utility of the BOGO platform in identifying prognostic clinical markers and predicting drug resistance, we compared the effects of individual ORFs on cell proliferation with cancer prognostic indicators derived from the Human Protein Atlas (HPA) database[20] and The Cancer Genome Atlas (TCGA)[32]. Differentially represented ORFs were identified by comparing their cell populations at baseline and after nine days of competitive growth. ORFs with significant changes ($\log_2 FC > 1$ or < -1; $p < 0.05$) were classified as cell proliferation regulators, with enriched ORFs designated as cell growth upregulators and depleted ORFs as cell growth downregulators. Using z-score rankings, we evaluated the predictive performance of our ORF list in identifying clinically relevant markers. Precision-recall analysis revealed that ranking ORFs in descending order by proliferation score (placing upregulators at the top) yielded high precision for identifying markers associated with unfavorable cancer prognosis (**Figure 2D**). Conversely, ranking ORFs in ascending order (placing downregulators at the top) enriched for markers of favorable prognosis, although precision declined more steeply, reaching a comparable minimum of approximately 40% (**Figure 2E**). To illustrate the clinical significance of these findings in cervical cancer (CESC), we selected representative ORFs from both ends of the prognostic spectrum. Overexpression of Phosphoglucomutase 1 (PGM1), a top-ranked upregulator, was strongly associated with reduced survival probability (**Figure 2F**), consistent with its role in promoting cell proliferation. In contrast, overexpression of SRA Stem-Loop Interacting RNA Binding Protein (SLIRP), a top-ranked downregulator, correlated with improved patient outcomes (**Figure 2G**), suggesting a suppressive effect on cancer proliferation. Together, these precision-recall and survival analyses underscore the clinical relevance of our screening approach and its capacity to identify promising cell proliferation regulators with prognostic value.

We hypothesized that drug response effectors would exhibit strong correlations with cell proliferation regulators, based on the premise that robust resistance mechanisms may confer enhanced proliferative capacity in cancer cells. To test this, we compared our drug response screen results with a manually curated drug resistance dataset from TCGA[33], evaluating the ability of our platform to identify known resistance drivers. Precision-recall analysis of our ranked drug response dataset revealed high precision among top-ranked genes for both cisplatin (**Figure 2H**) and 5-fluorouracil (5-FU) resistance (**Figure 2I**), indicating effective enrichment of established resistance mechanisms. For instance, overexpression of Src homology 3 domain-binding protein 2 (SH3BP2), a top cisplatin resistance candidate, was strongly associated with reduced survival in patients with pancreatic adenocarcinoma (PAAD) (**Figure 2J**). Similarly, Coiled-coil domain-containing protein 47 (CCDC47), identified as a 5-FU resistance driver, correlated with poor prognosis in hepatocellular carcinoma (LIHC) (**Figure 2K**). Both genes also demonstrated drug resistance-driving effects in our screening data (**Supplementary Figure 2H**). This integrative analysis, combining functional screening with clinical datasets, underscores the capability of the BOGO platform to uncover novel drug resistance genes with direct biological and clinical relevance in cancer.

Multi-modal analysis of ORFeome drug response profiles identifies BCL2-related pathways as a convergent mechanism of chemoresistance.

To investigate how genetic perturbations induced by human ORFeome overexpression influence cellular responses to chemotherapeutic stress, we performed dimensionality reduction analysis on the 16 ORFeome drug response profiles. Uniform Manifold Approximation and Projection (UMAP) plots revealed distinct clustering patterns corresponding to individual drug treatments (**Figure 3A**). DMSO controls from five independent batches clustered tightly along the UMAP1 axis, indicating minimal batch effects. Furthermore, biological replicates consistently clustered by chemotherapeutic agents, underscoring the robustness and reproducibility of the screening data. Notably, drugs with shared mechanisms of action generally clustered in close proximity, demonstrating the ability of dimensionality reduction to capture broad biological responses. For example, mitomycin C clustered separately from other alkylating agents, likely reflecting its unique mechanism as a DNA synthesis inhibitor. Similarly, although TFT and Fluorouracil (5-FU) are both classified as antimetabolites, their divergent clustering patterns suggest distinct resistance profiles and underlying mechanisms of action[34].

To uncover shared patterns of the 16 drug response profiles, we performed an all-by-all correlation analysis of ORF drug response profiles, resulting in over 200 million pairwise comparisons (**Figure 3B**). High-confidence interactions, defined by a Pearson correlation coefficient exceeding |0.9|, were extracted and visualized as a similarity network, revealing clusters of ORFs with comparable response behaviors across similar sets of chemotherapeutic agents[35] (**Figure 3B, left panel**). In this network, shorter distances between ORFs indicate greater similarity in drug response regulation following overexpression, suggesting shared biological functions and underlying regulatory pathways. To functionally annotate these clusters, we applied Spatial Analysis of Functional Enrichment (SAFE)[36], which identified key biological processes associated with chemotherapeutic response (**Figure 3B, right panel**). Notably, we observed enrichment of BCL2-related pathways, including BCL2 Homology Domain Binding and Autophagy & Apoptosis. These pathways are well-established mediators of both resistance and sensitivity to a number of chemotherapeutics[37,38], yet have not previously been linked to DNA analog drugs such as TAS-102 and TFT. This novel association highlights the sensitivity of our approach in detecting both known and previously unrecognized mechanisms of drug response. Together, these findings demonstrate the power of the BOGO platform to reveal functional gene networks and uncover key regulatory pathways underlying chemotherapeutic responses, reinforcing its utility for identifying actionable targets in cancer therapy.

To further investigate the mechanistic relevance of the BCL2-related pathways identified in our drug response similarity network, specifically, the clusters annotated as BH domain binding and Autophagy and Apoptosis (**Figure 3B**), we conducted an in-depth network analysis to evaluate their potential as targets for combination therapy. We performed STRING network analysis[39] on the sub-network comprising ORFs from these two clusters to elucidate protein-protein interactions among the identified genes (**Figure 3C**). The 20 ORFs from the SAFE-enriched clusters were analyzed using STRING to map experimentally supported and predicted interactions (**Supplementary Table 4**). This analysis revealed 12 key genes involved in critical signaling pathways and cellular damage responses, highlighted as diamond-shaped nodes in **Figure 3C**. The right panel

of the figure illustrates the interactions between these protein nodes and their associated GO clusters, reflecting strong functional connectivity and biological relevance. Among the enriched pathways were vesicular transport, apoptotic signaling, histone modification, and Notch receptor signaling, all of which are functionally linked to the BCL2-related genes identified in the two clusters. These findings provide a compelling mechanistic rationale for targeting BCL2-associated pathways in chemotherapeutic resistance and underscore the utility of the BOGO platform in uncovering actionable biological networks.

To validate the biological relevance of our screen, we performed gene set enrichment analysis of ORF drug-response regulators across ORF categories (**Supplementary Figure 3A-B**). Core and Multidrug ORFs were enriched for canonical cancer pathways, including Wnt, TGF-β, PI3K-AKT, and Hippo signaling, as well as apoptosis and DNA damage response, consistent with mechanisms of chemoresistance. Non-respondent ORFs showed limited enrichment in cancer-related pathways, mapping instead to fatty acid metabolism and tight junction processes, suggesting minimal contributions to drug response. Silenced ORFs were uniquely enriched for kinase-related pathways, including AMPK and MAP kinase signaling, highlighting that some resistance drivers remain undetected in the BOGO screen (**Supplementary Figure 3A**).

Drug-specific analysis revealed mechanistically coherent patterns in individual chemotherapy (**Supplementary Figure 3B**). Drug response effectors for Cisplatin was enriched for DNA repair and regulated necrosis, consistent with its mechanism. Effectors for Paclitaxel were associated with microtubule activity and cellular senescence, reflecting its role in mitotic arrest. TFT/TAS-102 effectors converged on NF-κB signaling, pyrimidine metabolism, and uniquely on KIT signaling and MYC activation, highlighting distinct mechanisms. Doxorubicin effector were enriched in apoptosis and stress-response pathways, aligning with its induction of DNA damage and cellular stress. These enrichment signatures align with correlation and STRING network. The consistent appearance of autophagy-lysosomal activity, ferroptosis, necrosis, and pyroptosis across both categorical and drug-specific analyses reinforces their central role in mediating cancer drug resistance and validates the BOGO platform's ability to capture both known and novel resistance mechanisms.

To further validate the BOGO overexpression screening platform, we compared drug response profiles with transcriptional data obtained from single-cell RNA sequencing (scRNA-seq). scRNA-seq was performed on cells treated with TFT, cisplatin, doxorubicin, and paclitaxel, each representing distinct mechanistic categories within our chemotherapy panel. Approximately 32,000 cells were captured across five treatment conditions, including a DMSO control, with cell number variation maintained within 10%. Clustering analysis revealed distinct transcriptional profiles for each treatment (**Figure 3D, left panel**). Paclitaxel and TFT treatments exhibited unique clustering patterns, while cisplatin and doxorubicin showed partial overlap, consistent with their converging mechanisms of action. Gene enrichment analysis of treatment-specific clusters revealed significant enrichment for pathways commonly associated with chemoresistance, including autophagy, apoptosis, and Wnt signaling, echoing findings from our drug response profile similarity network (**Figure 3D, right panel**).

Notably, genes such as BCL2, TP53, CDK6, and TYMS (the target of TFT) were associated with all four treatments and showed specific dysregulation within the autophagy and apoptosis cluster (**Supplementary**

Figures 3B-E), suggesting that this cluster may represent a population of chemoresistant cells. Differential gene expression analysis (**Supplementary Table 5**) further identified significant downregulation of TRADD, a gene implicated in apoptotic signaling within the same GO cluster[40,41], and upregulation of POLD2, a Wnt pathway gene linked to cell proliferation[42] (**Figure 3D**). Importantly, the regulation of these genes was consistent across both the scRNA-seq data and the drug-ORF network analysis. Additionally, z-score analysis of $\log_2$ fold changes highlighted key resistance-associated genes, including YBX3 (cisplatin), ZNF143 (doxorubicin), ZNF217 (TFT), and GATA2 (paclitaxel), alongside several less-characterized candidates with potential roles in drug resistance or sensitization[43–46] (**Supplementary Figure 3F**). This strong concordance across platforms reinforces the ability of the BOGO screen to capture relevant drug response mechanisms at single-cell resolution and positions BCL2 as a promising therapeutic target, particularly in the context of DNA analog drugs such as TFT and TAS-102.

In summary, the convergence of evidence from dimensionality reduction, drug-gene network analysis, and single-cell transcriptomics strongly supports the pivotal role of BCL2-related pathways in mediating chemotherapeutic resistance across diverse classes of agents. These findings underscore the robustness and sensitivity of our ORFeome overexpression screening platform in uncovering both established and previously unrecognized resistance mechanisms. The consistent identification of BCL2 as a central regulatory node across multiple analytical frameworks provides a compelling rationale for its further investigation as a candidate for combination therapy, particularly in the context of DNA analog drugs such as TFT and TAS-102.

BCL2 inhibition overcomes chemoresistance by disrupting lysosomal function, a mechanism revealed through functional genomic screening.

Building on our identification of BCL2 as a central regulator and the enrichment of autophagy and apoptosis pathways in chemotherapeutic resistance, we evaluated the therapeutic relevance of combining a BCL2 inhibitor with a chemotherapeutic agent. We selected TAS-102, an oral formulation comprising TFT and tipiracil hydrochloride (TPI), due to its clinical relevance and favorable pharmacokinetics, including oral bioavailability[47]. To target BCL2-related resistance mechanisms, we employed ABT-263 (Navitoclax®), a broad-spectrum BCL2 family inhibitor known to antagonize BCL2, BCL2L1, and BCL2L2[48], all of which were identified as drug resistance drivers in our network analysis (**Figure 3C**). This combination strategy was designed to disrupt BCL2-mediated survival signaling and enhance the cytotoxic efficacy of TAS-102.

To identify physiologically relevant cancer models for evaluating the TAS-102 and BCL2 inhibitor combination therapy, we investigated whether TAS-102 treatment modulates endogenous BCL2 protein expression. The well-established association between BCL2 signaling, autophagy, and chemoresistance in colorectal and pancreatic cancers[49,50] provided a strong rationale for selecting cell lines representative of these malignancies. Western blot analysis revealed a pronounced upregulation of BCL2 following TAS-102 treatment in colorectal cancer cell lines HCT 116 and RKO, and in the pancreatic cancer cell line MIA PaCa-2. A modest increase was observed in HeLa cells, while HT-29 cells showed no detectable BCL2 expression (**Supplementary Figure 4A**). These differential expression patterns informed the selection of HCT 116 as the

primary model for subsequent combination therapy experiments, due to its robust BCL2 induction and relevance to colorectal cancer.

Given the role of BCL2 in modulating apoptosis and autophagy[51], we examined protein markers associated with lysosomal activity and autophagy following single-agent and combination treatments. In HCT 116 cells, treatment with TAS-102, either alone or in combination with ABT-263, resulted in a pronounced increase in the expression of LAMP1 and TFEB, key markers of lysosomal biogenesis and function[52] (**Supplementary Figure 4B**). In contrast, autophagy markers LC3B and p62[53] exhibited relatively stable expression across all treatment conditions. Similar expression patterns were observed in HeLa cells, although the effects were more modest, particularly for TFEB following combination treatment. These results suggest that while autophagy activity remains stable, lysosomal modulation is dynamically regulated in response to TAS-102 and ABT-263, implicating lysosomal pathways as potential contributors to chemotherapeutic resistance mechanisms.

To further investigate the impact of TAS-102 and ABT-263 on lysosomal regulation, we examined TFEB localization using immunofluorescence staining. Treatment with TAS-102 led to increased nuclear localization of TFEB, consistent with activation of lysosomal biogenesis. However, combination treatment with ABT-263 significantly reduced TFEB nuclear localization (**Figure 4A; Supplementary Figure 4C**), suggesting disruption of TFEB-mediated lysosomal activation. This observation aligns with the concurrent elevation of TFEB and LAMP1 protein levels, in the absence of corresponding degradation of p62 or LC3B (**Supplementary Figure 4B**), indicating that while autophagosome maturation may proceed, lysosomal degradation is impaired. These data suggest that although total TFEB levels increase under TAS-102 treatment, nuclear translocation and functional activation of lysosomes are disrupted when ABT-263 is co-administered, pointing to lysosomal modulation as a key mechanism underlying the observed drug response.

To further investigate the functional consequences of lysosomal modulation, we assessed lysosomal proteolytic activity using DQ-BSA staining[54], which emits red fluorescence upon proteolysis. In HCT 116 cells, TAS-102 treatment significantly increased red fluorescence intensity, indicating enhanced lysosomal activity. However, concurrent treatment with ABT-263 markedly reduced fluorescence, suggesting impaired proteolysis (**Figure 4B; Supplementary Figure 4D**). Consistent with these findings, lysosomal pH measurements revealed acidification under TAS-102 treatment alone, while combination therapy led to increased pH, indicative of disrupted lysosomal acidification (**Figure 4C**). These results point to a potential resistance mechanism involving lysosomal dysfunction, wherein TAS-102 promotes lysosomal activation, but BCL2 inhibition compromises lysosomal function. To assess the functional relevance of this disruption, we co-treated cells with TAS-102 and Bafilomycin A1, a known lysosomal inhibitor. This combination significantly reduced cell proliferation compared to TAS-102 alone after 48 hours of treatment (**Figure 4D**), supporting the role of lysosomal integrity in modulating chemotherapeutic response.

To identify lysosomal modulators contributing to chemotherapeutic resistance, we developed a follow-up gene overexpression perturbation screen by integrating the BOGO platform with Lysotracker[55]-based FACS

(**Figure 4E**). HeLa cells expressing the ORFeome library from the previous drug response screen were stained with Lysotracker and sorted into populations with high and low lysosomal activity (top and bottom 5% based on green fluorescence intensity). Subsequent next-generation sequencing (NGS) enabled the identification of ORFs associated with lysosomal modulation. ORFs enriched in the top 5% were classified as lysosome activators, while those enriched in the bottom 5% were designated lysosome repressors ($\log_2$FC > 0.5 or < -0.5; $p < 0.05$). UMAP analysis confirmed consistent clustering across biological triplicates for the top 5%, bottom 5%, and presort samples, supporting the robustness and reproducibility of the BOGO screening platform (**Supplementary Figure 4E**). Notably, BCL2, previously identified as a TAS-102-induced resistance driver, also exhibited strong lysosomal activation in this screen (**Supplementary Table 6**). Pathway enrichment analysis of lysosomal modulators revealed significant involvement of TGFβ signaling (**Figure 4F**), implicating its role in autophagy and lysosomal regulation. To further refine the analysis, we intersected lysosomal modulators with TAS-102-specific resistance and sensitization-driving ORFs, identifying 109 ORFs as lysosomal activators, 78 as lysosomal repressors, and 91 overlapping with both lysosomal modulation and TAS-102 response profiles (**Supplementary Figure 4F**). These findings highlight a subset of lysosome-linked resistance drivers specific to TAS-102, offering mechanistic insight and potential therapeutic targets.

Among the 109 identified lysosomal activators, we selected GHSR and CCDC86 for further analysis based on three criteria: (i) strong lysosomal activation ($\log_2$FC > 1), (ii) novelty, and (iii) pronounced TAS-102 resistance. Kaplan-Meier survival analysis using data from the TCGA PanCancer Atlas revealed that high GHSR expression was significantly associated with poorer overall survival (**Figure 4G, left panel**), particularly in colorectal, lung, and esophagogastric cancers (**Figure 4G, right panel**). Similarly, elevated CCDC86 expression correlated with poor prognosis over a 10-year period (**Figure 4H, left panel**), across multiple cancer types, including colorectal, lung, and bile duct cancers (**Figure 4H, right panel**). These findings support the relevance of GHSR and CCDC86 as lysosome-associated resistance markers and potential therapeutic targets. Collectively, our functional experiments demonstrate that BCL2 inhibition in combination with TAS-102 alters lysosomal function and that the BOGO platform enables phenotypic functional screening to uncover mechanistic pathways and identify candidate genes with implications for therapy response and cancer prognosis.

***In Vitro* and *In Vivo* preclinical validation demonstrates the synergistic and p53-dependent anti-tumor activity of combining BCL2 inhibition with TAS-102.**

To assess the clinical applicability of TAS-102 in combination with ABT-263, we investigated their cytotoxic effects in both p53-dependent and cancer-specific contexts. STRING network analysis revealed an indirect interaction between BCL2 and TYMS, the primary target of TAS-102, mediated through p53 (**Supplementary Figure 5A**), suggesting that p53 may serve as a critical hub linking BCL2 and TYMS pathways. To test this hypothesis, we employed p53 knockout models and compared their response to combination therapy with wild-type cells (**Supplementary Figure 5B)**. Dosages of ABT-263 were selected based on IC_{50} values derived from dose-response curves in colorectal cancer cell lines (**Supplementary Figure 5C**). Cytotoxicity assays revealed a substantial reduction in cell viability in p53 wild-type HCT 116 cells following combination treatment with TAS-102 and ABT-263, compared to TAS-102 alone (**Figure 5A**). Notably, the median lethal dose (LD_{50}) of TAS-102

decreased by up to 1,000-fold when combined with 50 nM ABT-263 (**Figure 5A**). This sensitizing effect was significantly attenuated in p53-knockout HCT 116 cells (**Figure 5B**), suggesting a p53-dependent mechanism underlying the interaction between BCL2 signaling and TAS-102 responsiveness.

Additionally, we evaluated the synergistic effects of TAS-102 and ABT-263 in a pancreatic cancer cell line MIA PaCa-2 and the normal human pancreatic ductal epithelial cell line hTERT-HPNE, to assess cancer-specific cytotoxicity, which is essential for minimizing off-target effects and improving therapeutic safety profiles[56]. Similar to HCT 116, MIA PaCa-2 cells exhibited enhanced sensitivity to TAS-102 when sequentially treated with ABT-263, resulting in a lower effective dose to reach LD_{50} (**Figure 5C**). In contrast, hTERT-HPNE cells showed no additive or synergistic cytotoxic response (**Figure 5D**), supporting the cancer-specific nature of the combination therapy. These findings reinforce the therapeutic potential of BCL2 inhibition to enhance TAS-102 efficacy, particularly in p53-proficient and epithelial-like cancer models, and highlight the importance of cellular context in predicting combination treatment outcomes.

On the other hand, RKO cells, which share microsatellite instability-high (MSI-H) status with HCT116, exhibited no synergistic effect in either p53 wild-type or knockout conditions (**Supplementary Figures 5D-E**). While morphological differences between epithelial HCT116 and mesenchymal-like RKO cells may influence drug responses, somatic mutations such as KRAS or BRAF could modulate sensitivity. HCT116 and MIA PaCa2, both KRAS-mutant, are sensitive to the ABT-263 combination therapy[57,58] due to KRAS-driven upregulation of BCL-XL, the target of ABT-263, and greater reliance on lysosomal activity[59], whereas RKO cells could resist apoptosis through MCL-1 upregulation downstream of BRAF-V600E, reducing sensitivity to ABT-263[60].

To validate the in vitro findings in a more physiologically relevant context, we replicated the TAS-102 and ABT-263 combination treatment in both cell-derived xenograft (CDX) and patient-derived xenograft (PDX) models using colorectal cancer systems. In the CDX model, combination therapy consistently recapitulated the enhanced anti-tumor effects observed in vitro, resulting in significantly reduced tumor growth compared to either monotherapy (**Figure 5E**). Similarly, the PDX model demonstrated slowed tumor volume progression under combination treatment relative to TAS-102 alone, confirming that the synergistic efficacy of the therapy is reproducible *in vivo* (**Figure 5F**). In addition, both CDX and PDX models showed a significant reduction in tumor weight under combination treatment compared to TAS-102 alone. (**Figure 5G-H**). Importantly, body weight remained stable, and white and red blood cell counts showed no significant changes In both CDX and PDX models (**Supplementary Figures 5F-H and 5J-L**); however, platelet levels decreased in both models under the combination treatment (**Supplementary Figures 5I and 5M**), consistent with previously reported adverse effects of TAS-102-induced thrombocytopenia and ABT-263-associated platelet suppression[61,62]. These results confirm that the synergistic interaction between TAS-102 and BCL2 inhibition is both effective and reproducible in vivo, reinforcing the translational potential of this combination strategy to overcome TAS-102 chemoresistance and enhance chemosensitization in colorectal cancer.

Discussion

In this study, we introduce BOGO, a novel, proteome-wide gene overexpression screening platform that enables precise, single-gene-per-cell perturbations to identify translational combination therapies. By integrating the comprehensive human ORFeome via the Bxb1 landing pad system, BOGO overcomes key limitations of traditional lentiviral and CRISPRa-based approaches, such as heterogeneous expression and multi-gene activation (**Figure 1**). Applied to a panel of 16 chemotherapeutic agents, BOGO systematically identified gene subsets associated with chemoresistance and chemosensitization, uncovering clinically relevant drivers of cancer resistance and prognosis (**Figure 2**). The platform revealed both known and novel drug response mechanisms, including BCL2-mediated resistance to DNA analog drugs such as TAS-102 (**Figure 3**). Mechanistic validation demonstrated that combination therapy with TAS-102 and the BCL2 inhibitor, ABT-263, disrupts lysosomal function (**Figure 4**), selectively enhances cytotoxicity, and exhibits synergistic efficacy in colorectal and pancreatic cancer models, both in vitro and in vivo (**Figure 5**). These findings underscore BOGO's translational potential for discovering rational combination therapies and advancing functional genomics in cancer.

Despite its robust design and high-throughput capabilities, BOGO has limitations that warrant future optimization. First, undetected ORFs may still hold biological significance, potentially due to selection bias introduced by the IRES-mCherry cassette, which may disadvantage longer ORFs. Second, the platform does not account for non-coding RNAs or alternative isoforms, as the hORFeome comprises only canonical coding sequences. Third, although single-copy integration at the AAVS1 locus with the EF1α promoter ensures uniform expression, it may not fully recapitulate the copy number variation observed in clinical tumors.

Additionally, the platform's reliance on HeLa cells at one time point may limit its ability to capture lineage-specific resistance mechanisms relevant to other cancer types. The current configuration also lacks consideration of microenvironmental factors and cell-cell interactions, which are critical for drug response. Moreover, the Bxb1 landing pad mouse model and immune system interactions have not yet been evaluated in the context of combination therapy. Future integration with *in vivo* models[63], patient-derived organoids[64], and expanded screening across diverse cancer lineages[65] could enhance BOGO's translational relevance and mechanistic depth. The ability to identify patient-specific resistance mechanisms could revolutionize precision oncology and extend to other domains of personalized medicine. Recognizing that drug resistance evolves over time, time-course BOGO screens could distinguish between early survival genes and those mediating long-term adaptation, enabling stage-specific interventions and improving therapeutic durability.

BOGO also uncovered novel chemoresistance pathways beyond apoptosis, including lysosomal dysfunction and Wnt signaling. Further mechanistic studies of newly identified resistance genes, such as CCDC86, GHSR, and POLD2, are needed to elucidate their mechanism in drug response and cell survival, and to evaluate their potential as therapeutic targets. Integration of BOGO with single-cell resolution technologies (e.g., Perturb-seq[66]) could reveal heterogeneity in resistance mechanisms, enabling identification of rare but clinically significant subpopulations and refining our understanding of tumor evolution and treatment response.

Given that many BOGO-identified drug response effectors are linked to pathways targeted by existing drugs, there is significant potential for drug repurposing. Systematic screening of FDA-approved compounds[67] against BOGO hits could accelerate the development of combination therapies, reducing time and cost associated with novel drug development. This strategy could be particularly impactful in resource-limited settings or for rare diseases with limited treatment options.

Beyond cancer, BOGO offers opportunities for expansion into non-cancer disease models, such as neurodegenerative disorders, autoimmune conditions, and metabolic syndromes, which often involve gene amplification or transcriptional dysregulation[68]. These applications position BOGO as a versatile platform for identifying drug response mechanisms and uncovering novel therapeutic targets. Together, our findings demonstrate the broad applicability, mechanistic insight, and clinical relevance of the BOGO system, supporting its use across diverse biological systems and disease models.

Figure legends

Figure 1. BOGO Enables Functional Proteome-Wide Screening

(**A**) Schematic overview of the study design. (**B**) Diagram of the BOGO platform illustrating ORF integration into the AAVS1 locus via Bxb1 landing pad. (**C**) Venn diagram showing overlap between verified ORFeomes and genes included in our downstream analysis. (**D**) Correlation matrix of drug response profile depicting batch and replicate variation across independent chemotherapeutic drug screenings, demonstrating high reproducibility among technical replicates (N=2-3). (**E**) Bar graph of individual competition assays under TFT treatment (N=3). (**F**) UpSet plot showing overlap of differentially expressed genes (DEGs) between the doxorubicin drug response screen and the HeLa wild-type transcriptomic screen at 24 and 120 hrs of doxorubicin treatment (±Log2 fold change 0.5, $p < 0.05$). (**G**) Bar graph of average fold change in High-Saturation Retesting (HSR) for "Non-respondent," "Sensitizing," and "Resistant" ORF categories from the drug response screen. Data in bar graphs are shown as mean ± Standard Error of the Mean (SEM). Statistical analysis by one or two-way ANOVA with Tukey's multiple comparisons; ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

Figure 2. Biological and Clinical Relevance of BOGO Screening for Cell Proliferation and Drug Resistance

(**A**) Sankey diagram categorizing ORF numbers into drug response groups; undetected ORFs defined by median log2TMM < -1. (**B**) Heatmap of Z-scores for intrinsic ORF characteristics across categories. (**C**) Left: Heatmap showing overlap between ORFs and cancer biomarkers from TheMarker database; Right: Distribution of biomarker types across ORF categories. (**D-E**) Precision-recall curves ranking ORFs by cell proliferation effects, (**D**) comparing top upregulators with unfavorable prognosis, and (**E**) top downregulators with favorable prognosis from the HPA dataset. (**F-G**) Kaplan-Meier survival curves for most promising examples: (**F**) PGM1 unfavorable) and (**G**) SLIRP (favorable) in cervical cancer. (**H-I**) Precision-recall curves evaluating (**H**) cisplatin and (**I**) 5-FU resistance drivers from the ORFeome screen. (**J-K**) Kaplan-Meier survival curves for (**J**) SH3BP2 (pancreatic cancer) and (**K)** CCDC47 (liver cancer) that are associated with reduced survival. Survival curve comparisons were performed using the log-rank test.

Figure 3. Comprehensive Systems Analysis of Drug Response Screening Highlighting BCL2-Associated Functional Pathways

(**A**) UMAP plot of ORF perturbation profiles for 16 chemotherapeutic agents, color-coded by one of four drug mechanisms of action. (**B**) Functional enrichment network from drug-ORF correlations, showing ORF as node and Pearson correlation (cut-off: < 0.9) as edge (left), functional clusters using SAFE (right). (**C**) STRING protein-protein interaction (PPI) network of two BCL2-related SAFE clusters ("BH domain binding" and "Autophagy & Apoptosis"), highlighting genes involved in vesicular transport, apoptotic signaling, histone modification, and Notch receptor pathways. The network is based on high-confidence experimental interactions (score > 0.9), <50 interactors in the first and second shells. 12 diamond-shaped nodes highlight ORFs originating from the SAFE network clusters. (**D**) UMAP plots showing Single-cell RNA-seq treatment-specific ORF-expressing cell

population changes for four different chemotherapeutics (left) and enriched pathways in distinct cell clusters identified by the Leiden algorithm and analyzed using DAVID (right).

Figure 4. Functional Validation of BCL2 Inhibition and Autophagy-Lysosomal Modulation

(**A**) Merged fluorescent images of TFEB (green) co-stained with Hoechst (blue), showing nuclear localization changes after drug treatments. The bar chart quantifies nuclear TFEB fluorescence differences across four treatments. (**B**) Merged fluorescent images of DQ-BSA lysosomal proteolysis co-stained with Hoechst, showing lysosomal activity; bar chart quantifies fluorescence intensity differences. (**C**) Bar chart showing lysosomal pH changes after drug treatments. **(D**) Cell proliferation assay of colorectal cancer cells treated with TAS-102 and the lysosomal inhibitor Bafilomycin A1. (**E**) Schematic workflow of high-throughput Lysotracker-based ORFeome overexpression screen with FACS sorting. (**F**) Diagram of TGF-β signaling pathway highlighting lysosomal activator and repressor ORFs identified in the Lysotracker-FACs screening. (**G-H**) Kaplan-Meier survival curves depicting overall survival for patients with (**G**) GHSR and (**H**) CCDC86 expression across multiple cancer types from the TCGA PanCancer Atlas, accompanied by stacked bar graphs showing the number of cancer patients stratified by expression levels, with studies denoted by the study abbreviations. Data in bar graphs are shown as mean ± Standard Deviation (SD). Statistical analysis was performed using one or two-way ANOVA with Tukey's multiple comparisons; * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$. Survival curve comparisons were performed using the log-rank test.

Figure 5. BCL2 Inhibition Synergizes with TAS-102 to Suppress Cancer Growth *In Vitro* and *In Vivo*

(**A-C**) Synergistic growth inhibition of TAS-102 and ABT-263 in (**A**) HCT 116 p53 wild-type (WT) cells, (**B**) p53 knockout (KO) cells, and (**C**) MIA PaCa-2 pancreatic cancer cells; all pairwise comparisons reached **** $p \leq 0.0001$. (**D**) Cytotoxicity assay in hTERT-HPNE. (**E-F**) Tumor volume (mm^3) progression in (**E**) HCT 116 cell-derived xenografts (CDX) and (**F**) colorectal patient-derived xenografts (PDX), after treatment with DMSO, TAS-102, ABT-263, and TAS-102 + ABT-263 combination therapy. * $p \leq 0.05$. (**G-H**) Tumor weight in (**G**) CDX and (H) PDX models, showing significant reduction under combination treatment compared to TAS-102 alone. Data in panels A-D are presented as mean ± standard deviation (SD), whereas panels E-H show mean ± standard error of the mean (SEM). Statistical analysis for panels A-F was performed using two-way ANOVA with Tukey's multiple comparisons; $p \leq 0.05$, and panels G-H were analyzed using Welch's t test; *$p \leq 0.05$.

Supplementary Figure 1. Validation and Characterization of the BOGO System and Screen Results

(**A**) Efficiency of ORF integration in HeLa cells measured by mCherry and mTagBFP2 (BFP) fluorescence via FACS (left). Stacked bar graph showing the number of sorted and unsorted cells via mCherry fluorescence (right). (**B**) Venn diagrams illustrating detection of selected ORFs in total HSR genes (left) and doxorubicin-specific drivers (right). The bar graph shows increased log2 fold change in PITX2 and HIST1H3B. Arrows indicate an increase or decrease from the drug response screen to the HSR screen (right top). (**C**) Proportion of ORFs that underwent gene silencing as measured by the percentage of mCherry-negative over mCherry-positive. Positive control indicates mCherry-positive cells (**D**) Bar plot showing percentage of cell populations affected by ORF silencing. (**E**) Bar plot showing silencing ratio for the selected 7 ORFs in the drug response screen.

Supplementary Figure 2. Classification of Drugs used for Screening, Categorization of ORFs with respect to drug response effect, and ORF Candidates as Resistance Biomarker

(**A**) Overview of mechanisms and cellular targets of chemotherapeutic agents used in drug response screens. (**B**) Cumulative number of unique drug response effector ORFs with respect to the number of drugs. (**C**) Histogram showing the number of chemotherapeutic drugs responded to effector ORFs. (**D**) Mixture distribution analysis to define a cutoff for multi-drug responders. A dotted line was drawn at 2.64. (**E**) Number of unique ORFs for an individual chemotherapeutic. (**F**) Unique Categorical ORF linked to Chemotherapeutics Categories. (**G**) Venn diagram illustrating the distinctness of gene subsets. Numbers indicate the number of ORFs in five categories. 0 indicates there is no overlap between subsets. (**H**) Drug response effector expression of SH3BP2 and CCDC47 in four chemotherapeutics from the drug response screen.

Supplementary Figure 3. Drug Response Screening with Enrichment Analysis and Single-Cell Validation

(A-B) Gene Set Enrichment Analysis (GSEA) bubble plots of **(A)** core, categorical (antimetabolite, DNA-cross linking agent, DNA strand break agent and microtubule inhibitor), multidrug, non-respondent, and silencing ORFs; **(B)** individual chemotherapies from peripheral ORFs. **(C-F)** UMAP projections showing single-cell ORF expression of key resistance markers: **(C)** *TYMS*, **(D)** *BCL2*, **(E)** *TP53*, and **(F)** *CDK6*. **(G)** Heatmap of normalized z-scores displaying ORF expression across one DMSO and four chemotherapeutic conditions, highlighting the top five resistance-associated ORFs per treatment cluster.

Supplementary Figure 4. Molecular and Cellular Validation of BCL2 Inhibition and Autophagy-Lysosomal Dynamics

(**A**) Western blot analysis of BCL2 expression in cancer cell lines post-TAS-102 treatment (0.6 μM for RKO, 0.8 μM for HeLa, 2 μM for MIA PaCa-2, HT-29, HCT 116). (**B**) Western blot analysis of lysosomal markers (LAMP1, TFEB) and autophagy markers (LC3B, p62) in HeLa and HCT 116 cells under treatment. (**C**) Separate fluorescent images of TFEB (green) immunofluorescence and Hoechst nuclear staining across treatments (2 μM TAS-102, or 0.5 μM ABT-263, 2 μM TAS-102 + 0.5 μM ABT-263). (**D**) Separate fluorescent images of DQ-BSA (red) lysosomal proteolysis and Hoechst nuclear staining across treatment (2 μM TAS-102, or 0.5 μM ABT-263, 2 μM TAS-102 + 0.5 μM ABT-263). (**E**) UMAP plot showing clustering of pre-sort and Lysotracker-sorted; top and bottom 5% green fluorescence (GF) from HeLa ORFeome cell. Each dot indicates biological replicates (N=3). (**F**) Upset plot showing overlap between TAS-102 drug response effector ORF and lysosomal modulator ORFs from Lysotracker staining.

Supplementary Figure 5. Synergistic Efficacy and Safety Profile of TAS-102 and ABT-263 Combination

(**A**) STRING PPI network showing an indirect p53-mediated link between BCL2 and TYMS, medium confidence (<0.5; protein interaction score from STRING), <5 interactors in first and second shells. (**B**) Dose-response curves for ABT-263 in colorectal cancer lines HCT 116, RKO (p53 WT and KO), and HT-29 using varying concentrations 0, 0.01, 0.05, 0.1, 0.5, 1, 4, 10, 25, and 50 μM (**C-D**). Synergistic cytotoxicity of TAS-102 and ABT-263 in (**C**) RKO p53 WT (**D**) and p53 KO cells. (**E-F**) Mouse body weight data from (**E**) CDX and (**F**) PDX

models, showing stable weight across treatments. (**G-K**) Complete blood count (CBC) parameters, including Red blood cell, white blood cell, Hemoglobin, Monocyte, and Granulocyte, with no significant differences across treatments. (**L**) Platelet counts showed a significant reduction of * $p < 0.05$ in combination therapy compared to ABT-263 alone. Data in panels C-D are presented as mean ± standard deviation (SD), whereas panels B, E-L show mean ± standard error of the mean (SEM). Statistical analysis was performed using one-way or two-way ANOVA with Tukey's multiple comparisons.

Supplementary Table

Supplementary Table 1:

Dataset of the 16 chemotherapeutic drug response screens in HeLa ORFeome cells.

Supplementary Table 2:

mRNA transcriptomics datasets on doxorubicin treatment on HeLa wild-type cells for 24 hrs and 120 hrs.

Supplementary Table 3:

Dataset of selected ~2K rearrayed genes overexpression for doxorubicin-treated screening in HeLa ORFeome cells.

Supplementary Table 4:

Neighbourhood score derived from SAFE: Gene included in BH binding domain and Autophagy & Apoptosis.

Supplementary Table 5:

Dataset of scRNA-seq, clustered using the Leiden algorithm with 0.5 resolution.

Supplementary Table 6:

Dataset of TAS-102 drug response screen with Lysotracker staining for FACS, the top and bottom 5% of green fluorescence.

Acknowledgment

This research was supported by a startup grant from Aune Foundation via Montreal General Hospital Foundation, and the Research Institute of the McGill University Health Centre (D.-K.K.), NIH R21CA259719 and DoD BC220542 (A.V.B), National Research Foundation (NRF) funded by the Korean government (NRF-2021M3A9I4024447, J.P.; 2022R1A2C2010940, H. K) and Chungbuk National University Glocal30 project 2025 (J.-H.C.).

Methods

Cell culturing

T-REx™-HeLa Cell Line (HeLa cells) was maintained as previously described[69], using Dulbecco's Modified Eagle's Medium (DMEM, Corning, 10-017-CV). Before the screening, they were tested for mycoplasma contamination (ATCC, 30-1012K), and cells were confirmed using short tandem repeat (STR) profiling at Genome Diagnostics Centre of The Hospital for Sick Children (SickKids)[70]. HeLa Flp-In T-REx cells engineered with the Bxb1-landing pad at the AAVS1 locus with EF1α promoter (HeLa EF1α cells) were generously provided by the Roth Lab at the University of Toronto, currently moved to the University of Pittsburgh[71].

The human cell lines, colorectal cancer HT29 (RRID: CVCL_0320), pancreatic cancer MIA PaCa-2 (RRID: CVCL_0428), and pancreatic ductal cell hTERT HPNE (RRID: CVCL_C466) were cultured as recommended by ATCC. Human colorectal cancer HCT 116 and HCT 116 p53(-/-; KO) cell lines were shared by Dr. Bert Vogelstein at Johns Hopkins University[72]. Human colorectal cancer RKO and RKO p53(-/-; KO) cell lines were shared by Dr. Alessandro Carugo at MD Anderson Cancer Center[73]. Those cell lines were maintained in DMEM and authenticated again using STR profiling by ATCC or the Roswell Park Genomic Shared Resource within the last three years. The cells were routinely screened for mycoplasma; all studies used mycoplasma-free cells. Human cell cultures were maintained in media supplemented with 10% heat-inactivated fetal bovine serum (FBS; Corning, 35-015-CV) and 1% penicillin/streptomycin (P/S, Corning, 30-002-CI) at 37°C with 5-10% CO_2 in a humidified incubator.

En masse LR cloning of the Bxb1-landing pad ORFeome library

Entry clones in Human Open Reading Frame Collection (ORFeome) v9.1[22,74] with ~1000 additional clones donated by the Center for Cancer Systems Biology (CCSB) at Dana-Farber Cancer Institute (Delta space) were verified by next-generation sequencing to confirm the authenticity of their coding DNA sequences. Entry clones were stamped from 384-well plates onto corresponding LB agar plates containing either kanamycin (25 µg/mL; Thermo Fisher, AC611290050) or spectinomycin (50 µg/mL; Thermo Fisher, S05845G). After incubation at 37°C for 48 hours (hrs), the LB agar plates were scraped and collected for pooled plasmid preparation using the Qiagen Plasmid Mini Prep kit (Qiagen, 27104) - one miniprep per one 384-well plate. The amount of DNA proportional to the number of clones was pooled and mixed into a single tube to create a complete set of entry clones, where each ORF was represented in equivalent ratios. Based on these pools of Entry clones, we performed *En masse* LR cloning as previously described[22,23,75]. 150 ng of the entry pool was incubated with 5x LR Clonase (Invitrogen, 11791020) and 150 ng of destination vector (pDEST-HC-REC, Roth Lab) for 5 days, starting with a 5 µL reaction and supplementing 1 µL of 5x LR Clonase and 150 ng of destination vector each day. We used 1 µL of the LR reaction mix to perform electroporations 8 times (total 8 µL used) with 10-beta electrocompetent *E. coli* (New England Biolabs, C3020K) to generate 20 million Colony Forming Units, which achieves approximately 1000x library representation. In detail, the electroporated bacteria were plated onto LB agar containing carbenicillin (Fisher, BP26485, 100 µg/ml) plates and incubated at 37°C for 18 hrs. Colonies

were collected by gentle scraping, and the final destination library was isolated using a PureLink ™ Fast Low-Endotoxin Maxi Plasmid Purification Kit (Invitrogen, A35892).

Large-scale overexpression perturbation by constructing a HeLa cell pool with the ORFeome library

HeLa EF1α cells were passaged 3 times after thawing before transfection, and 1.3 million cells were seeded in each of twelve 10 cm dishes. We transfected 10 µg of Bxb1-recombinase expression vector (pCAG-NLS-HA-coBxb, Roth Lab) per dish using Lipofectamine 3000 (Invitrogen, L3000015) according to the manufacturer's instructions. The media was refreshed after 12 hrs to remove the transfection mix. After 36 hrs, we transfected 10 µg of the ORFeome destination library prepared in the previous section. The media was refreshed 12 hrs later and incubated for an additional 24 hrs to allow the mCherry fluorescence expression to appear. The cells that have been transfected with the library were sorted for mCherry using Fluorescence-activated cell sorting (FACS) by Sony MA9000 at Flow & Immune Analysis Shared Resource (FIASR) at Roswell Park Comprehensive Cancer Center (RPCCC), and grown in 15 cm dishes for 3 more days until the residual Blue Fluorescent Protein (mTagBFP2) signal diminished. The cells integrated with the library were selected using FACS through two times, at pre-sort on the second day and post-sort on the fourth day, over 5 days post-transfection, and a 150x representation was achieved (Pre-sort: ~15 million, Post-sort: ~11 million cells). To minimize batch effect resulting from variability within cell pools, we expanded the pool of cells with ORFeome integrations to one hundred 15 cm dishes. The cell pool was then trypsinized, and 3 million ORFeome-integrated HeLa cells (HeLa ORFeome cells) were cryopreserved per vial, with 200 vials for subsequent drug selection screens.

Drug dose-response assays

Drug dose-response assays were carried out using the modified protocol from the previous study[12]. HeLa cells were thawed and grown in 10 cm dishes under standard culturing conditions before an assay. We seeded 3000 cells into 96-well plates at 100 µL volume and incubated them for 24 hrs to allow for attachment. Each drug was diluted to a 50 or 100 µM concentration and was serially diluted at a 1:3 ratio using culture medium to cover a range of eight concentrations. 100 µL of the diluted drug was added in triplicate and incubated for 72 hrs. The media were then removed, and viability was measured using AlamarBlue (Invitrogen, A50101) reagent according to the manufacturer's instructions. OD_{595nm} (Absorbance at 595 nm) reading was measured after 2 hr of reagent incubation, and the absolute Inhibitory Concentration at 50% (IC_{50}) was calculated using GraphPad Prism 9 software (Version 9.1.0).

Chemotherapeutic Screening

We selected 16 chemically synthesized cytotoxic chemotherapeutic agents that primarily target the cell cycle or exert non-specific effects through DNA damage. These drugs were chosen to represent a broad range of mechanisms of action commonly employed in adult and pediatric cancer treatments, as previously described[76]. To streamline the analysis, we categorized the chemotherapeutics based on their primary mechanisms of action. In some cases, drugs were grouped under broader mechanistic categories. For instance, doxorubicin, a topoisomerase II inhibitor, and bleomycin, an antitumor antibiotic, were classified under agents that induce DNA

strand breaks, reflecting their shared outcome despite differing molecular mechanisms. Based on these mechanistic classifications, the 16 chemotherapeutics were grouped into the following four categories:

- *Antimetabolites*: Fluorouracil (5-FU, Sigma, F5130), Trifluridine (TFT, Med Chem Express, HY-A0061), 5-ethynyl-2′-deoxyuridine (EdU, Toronto Research Chem - Sigma, E9386), Floxuridine (FdU, Adooq Bioscience, A10392-500), 6-mercaptopurine (Sigma, 852678), TAS-102 (Adooq Bioscience, 733030-01-8)
- *DNA Cross-Linking Agents*: Mitomycin C (Sigma, M5353), Cisplatin (Cayman, 13119), Carboplatin (British Pharmacopeia Reference - Sigma, BP711), Carmustine (European Pharmacopoeia Reference - Sigma, Y0002128)
- *DNA Strand Break Agents*: Doxorubicin (Tocris, 2252), Bleomycin (Sigma, B1141000), Etoposide (Sigma, PHR2717), Irinotecan (Sigma, 341205)
- *Microtubule Inhibitors:* Paclitaxel (Sigma, T7402) and Vinblastine (Sigma, V0300000)

Chemotherapeutic drug	**IC_{50}**
Cisplatin	0.333 uM
Carboplatin	3.37 uM
Carmustine	7 uM
Mitomycin C	14 nM
TFT	0.75 uM
5-FU	0.1925 uM
EdU	0.875 uM
FdU	25 nM
6-mercaptopurine	2 uM
TAS-102	0.8 uM
Doxorubicin	5 nM
Bleomycin	3.3 uM
Irinotecan	0.8 uM
Etoposide	0.625 uM
Paclitaxel	2 nM
Vinblastine	0.3 nM

Using IC_{50} values obtained through the drug dose-response assay, we determined an appropriate dose to treat the cells in a 15 cm dish format. We seeded at least 3 million cells and treated them with drugs at their IC_{50} dose for 72 hrs. DMSO was selected as the vehicle control. Screens were conducted in five independent batches, each including baseline (often referred to as T0 in CRISPR screening)[77] and its own set of vehicle controls to account for batch-specific variation. To ensure full library representation, we thawed 10 cryopreserved ORFeome-overexpressing pool vials and expanded them for 3 days, recovering a minimum of 3 million viable cells in total to maintain 150x coverage. The cells were then seeded at 3 million per dish and treated with the

respective drugs in triplicate, including corresponding vehicle controls in each batch. After 72 hrs of treatment, cells were trypsinized, and 3 million cells were passaged into fresh dishes with continued drug treatment. After 48 additional hrs, the drug-containing medium was removed, and cells were allowed to recover in regular culture medium for 3 more days or until confluency. Recovered cells were trypsinized and pelleted (~10 million cells per pellet) for subsequent genomic DNA extraction.

Polymerase Chain Reaction (PCR) amplification of ORFs from genomic DNA

According to the manufacturer's instructions, genomic DNA was extracted from the cell pellets using the Blood and Cell Culture DNA Midi Kit (Qiagen, 13343). We performed two sets of PCRs (nested PCR) using the Q5 High-Fidelity 2X Master Mix (New England Biolabs, M0544S) and primers (Integrated DNA Technologies) that bind to i) landing pad locus and ii) Gateway-cloning sites, to amplify the full-length Open Reading Frame (ORF) from the landing pad with improved selectivity and reduced genomic DNA contaminations. A total of 18 µg of DNA was used in the first PCR in twelve 50 µL PCR reactions, which were pooled to represent the full 150x coverage. The second PCR was performed using 2.5 µL of the previous PCR reaction as the template. Two 50 µL PCR2 reactions were performed and pooled before preparing the sequencing library. For NGS library preparation using the TruSeq platform, custom primers were used that included a 5′ NNNNN sequence to secure the complexity of the first 5 bp during sequencing[78].

Primers	Type	Sequences
BxbAmp-F	Landing pad locus	5'-CACCCACACAAAGGAAAAGGG-3'
BxbAmp-R	Landing pad locus	5'-CACATTGCCAAAAGACGGCA -3
Bxb-ngs-F	Gateway-cloning site	5'-NNNNNCTCACAAGTTTGTACAAAAAAG-3'
Bxb-ngs-R	Gateway-cloning site	5'-NNNNNTGACCACTTTGTACAAGAAAG-3'

Sequencing library preparation

The quality of the amplicons was confirmed on Tapestation (Agilent, G2991BA) at the Genomics Shared Resource (GSR) of Roswell Park Comprehensive Cancer Center. With samples that passed the quality control, primer dimers were first removed using an Ampure XP bead clean-up kit (Beckman-Coulter, A63882). Then, amplicons were sheared using a Covaris E220 Focused-ultrasonicator (Covaris, COV-500239) to sizes averaging 250-300 bp. These sheared amplicons were ligated for TruSeq adaptors and were indexed using Kapa hyperprep for sample barcodes (Illumina, 20015963; Roche 0796237100). The resulting sequencing library was finally removed from primer dimers using an Ampure kit. Finally, DNA was quantified using qPCR and loaded into NovaSeq 6000 (Illumina, 20012850), and a minimum of 3 million paired-end reads were generated for each sample.

ORFeome Bioinformatic Analysis

Bioinformatic analysis was processed in a Python Jupyter Notebook script (Python 3.11.0), which is shared as supplementary material. FASTQ files were trimmed using Cutadapt[79] to remove the remaining TruSeq adapter sequence according to Illumina, further trimmed with Trimmomatic[80] with a sequence Phred score (LEADING:20, TRAILING:20, SLIDINGWINDOW:4:15, MINLEN:25). For the ORFeome sequence, all sequences were checked

and assigned appropriate gene symbols with the NCBI gene database[81] (downloaded on August 11th, 2023). The analyzed files were aligned to the reference of human ORFeome sequences v9.1 and delta space using the Bowtie2 tool[82,83] (v2.5.2). Post to the alignment, post-processed FASTQ files, and aligned SAM files were evaluated using FASTQC (v0.12.1) and multiQC (v1.27.1)[84,85]. SAM files were restored in BAM file format with a series of post-processing steps using Samtools[86], and the raw read counts were generated using Pysam (0.23.0), including concordant and multiple reads with a MAPQ threshold of 30[86]. The parameters of processes were chosen based on the Cut&Run flow used in the peer ORFeome screening publication and best suited for short read alignment[87,88]. The raw readouts were normalized in the Trimmed Mean of M-value (TMM) using the EdgeR (v4.0.16)[89,90]. The normalized read counts were selected based on the calculation of sequences with a median log2TMM of higher than -1 to remove noise read counts. The full dataset was further normalized using the quantile-normalization method and batch-corrected to create an equal distribution of the dataset to analyze the difference between samples using EdgeR (v4.0.16) and Limma (v3.58.1) [91]. For sample-to-sample comparison, a TMM dataset was processed using dimensional reduction analysis of Uniform Manifold Approximation and Projection (UMAP) (UMAP-learn[92]; v0.5.7). The efficiency of ORFeome silencing was assessed by quantifying fluorescence levels in cells expressing mCherry and mTagBFP2. Silencing was evaluated using the Silencing Ratio, defined as $log_2(E/S)$ where E represents the number of expressing cells ($mCherry^+$ $mTagBFP2^-$) and S denotes the number of silenced cells ($mCherry^-$ $mTagBFP2^-$). This ratio reflects the difference in relative fluorescence units between cells actively expressing the ORF ($mCherry^+$ $mTagBFP2^-$) and those in which the ORF is likely silenced ($mCherry^-$ $mTagBFP2^-$). ORFs with a Silencing Ratio less than -1 (log2 fold change < -1) were classified as successfully silenced.

For more accurate statistical testing, any samples' underlying ORF population of less than half of the maximum expressed ORF population in total samples was excluded before TMM normalization [93]. Post-TMM normalization, the dataset was winsorized using DescTools[94] (0.99.57) with a 0.001% quantile range to remove extreme data points. Then, upper-quantile normalization was performed to normalize the dataset further using EdgeR[90] (v4.0.16). A batch correction was processed using a design matrix, including control experimental sample information, multiple batches, and altered next-generation sequencing runs to accurately adjust the batch effect using Limma[91]. Further, log2 fold change of TMM counts between experimental drug condition and the corresponding batch's DMSO in the collection and statistical testing was processed using permutation testing (n=10,000). Adjusted p-values were computed using two-tailed tests in a combination of t-values, log2 mean difference, and Z-transformed rank-sum statistics, scaling with Liptak Stouffer's z method[95]. Data was visualized using Matplotlib[96] and Seaborn[97].

Automated entry clone selections for High Saturation Retesting (HSR)

Approximately 10% of the ORFeome pool underwent retesting at higher coverage (>1,000 fold) to assess differentially represented ORFs. The selection of these ORFs relied on T-score statistical analysis of log2-fold changes for individual drugs, grouped into eight baskets. These selections were based on core resistance or sensitizing ORFs across all drugs, individual drugs, and controls. The selection criteria for HSR ORFeome focused on genes that demonstrated high T-scores, indicative of strong or weak proliferation phenotypes. In total,

2,217 ORFs were selected and cherry-picked from the human ORFeome using a customized robotic platform (S&P Robotics Inc., BM3-BC). These selected ORFs were then pooled and prepared within the HeLa EF1α cells following the same *en masse* LR cloning[75]. Doxorubicin screening was carried out with matching dosages of 5 nM as in the original screening, employing identical PCR ORF amplification methods and previously mentioned ORFeome bioinformatics analysis in this method. The selection of "Resistant", "Sensitizing", and "Non-respondent" was based on the same log2 fold change parameter as drug response screening. The graph was generated using GraphPad Prism 9, and statistical analysis was performed using one-way ANOVA followed by Tukey's multiple comparisons test.

Single ORF overexpression assay

To generate single cell lines for verification, specific entry clones corresponding to ORFs such as TYMS, RABL3, BCL2, MUTYH, CDK6, GPER, and BRD9 were grown overnight and isolated using the NucleoSpin Plasmid kit (Macherey-Nagel, NC9405546) from the ORFeome 9.1 collection[22]. 150 ng of entry vector, 150 ng of destination vector (pDEST-HC-REC shared by Roth Lab), and 1 µL of LR Clonase were incubated for 1 hour and transformed into 10-beta chemically competent cells (New England Biolabs, C3019) using heat shock [98]. Transformants were grown overnight in LB containing carbenicillin (Fisher, 10177012) and used for plasmid extractions. Sequences were confirmed through Sanger reactions. One day before transfection, HeLa EF1α cells were seeded into 6-well plates at 350K cells/well. The correctly cloned plasmids were co-transfected with the Bxb1 recombinase plasmid at 1 µg each into HeLa EF1α cells using Lipofectamine 3000.. After 5 days, cells were sorted for mCherry-positive populations to establish stable single-ORF-overexpressing cell lines. For the cellular competition assays, the protocol was adapted from the previous study[99]. 30,000 mCherry-positive cells (experimental) were co-cultured with 60,000 BFP-positive (control) cells in each well of a 6-well plate. After the 24 hrs incubation, three wells were treated with 0.75 µM TFT. After a 5-day treatment period, relative cell populations were quantified by reading 20,000 events per well using FACS. Results from the competition assay were compared to fold change data from the initial TFT screening. For silencing validation, phenotypic changes were assessed by quantifying their total population percentage after 5 days without any drug treatment using FACS. Following the recovery, specific populations were sorted via FACS for subsequent library sequencing. The alignment between phenotypic changes and transcriptomic response was evaluated by comparing the population percentages with Silencing Ratios using a previously mentioned ORF bioinformatics pipeline. The cell population was normalized by the mCherry control. The graph was generated using GraphPad Prism 9, and statistical analysis was performed using one-way ANOVA followed by Tukey's multiple comparisons test.

mRNA transcriptomics analysis

The treatment duration and the protocol are adapted from a previous study on finding drug combinations[100]. HeLa Cells were seeded at 3 million in 15 cm dishes and treated with an IC_{50} of 5nM Doxorubicin for 24 hrs and 120 hrs. After the treatment, cells were pelleted, and each cell pellet consisted of a minimum of 1 million cells per control and experimental set. Each cell pellets were preserved in TRIzol (Qiagen, #79306) and stored at -20℃ before being sent to Genomics Shared Resources (GSR) at Roswell Park Comprehensive Cancer Center.

Purified RNA was prepared using the miRNeasy micro kit (Qiagen, 217084), and quality control of RNA integrity number (RIN) was measured using Agilent Bioanalyzer 2100 (Agilent, G2939BA). Sequencing libraries were prepared with the KAPA mRNA HyperPrep Kit (Roche, KK8581). The final RNA-seq libraries were sequenced on Illumina NovaSeq 6000. Nextflow version 24.10.4[101] with nf-core/rna-seq v3.18.0[102] was utilized using Docker (v27.5.1). The sequencing files were aligned to the GRCh38.p14 human genome[103]. The workflow was configured to perform adapter trimming with Trim Galore[104], followed by alignment using STAR[105] (v2.7.10a) and quantification using the RSEM[106] (v1.3.1). Ribosomal RNA reads were removed using SORTMERNA[107] (v4.3.7) before quantification. Downstream analysis used count-based quantification, matching the configuration of the ORFeome bioinformatic pipeline. Paired t-tests were performed using SciPy's stats.ttest_rel function[108] (v1.15.2), and empirical p-values were calculated via permutation testing (n = 10,000) to assess statistical significance.

ORF Gene categorization

According to HeLa ORFeome cell sequencing, reads were aligned to 18,388 ORFeome libraries. Of these, 14,506 ORFs were considered well-represented, defined by a minimum median $\log_2$ TMM expression > -1[109]. ORFs below this cutoff were regarded as noise and excluded from downstream analysis. Within the screened 14,506 ORFeomes, those ORFs were categorized into five categorical subsets. Among the categorized ORFs, those with a log2 fold change threshold of ±0.5 and a p-value less than 0.05 were identified as differentially represented ORFs. Within this set, enriched ORFs (positive log2 fold change) are interpreted as putative resistance drivers, conferring a survival advantage under treatment, while depleted ORFs (negative log2 fold change) suggest sensitization drivers, indicating increased drug sensitivity. First, 197 core genes were selected based on any ORFs with the filter of more than 7 drugs (more than half of the total chemotherapeutics - 1). Second, 1,072 categorical genes were assigned based on the recalculation of the median(log2{drug categories}) - median(log{rest of categories}) of previously mentioned statistical testing and filtered to differentially represented ORFs. Third, 3,832 peripheral genes were identified as ORFs significantly altered (absolute log2 fold change ≥ 0.5 and p-value ≤ 0.05) in response to specific chemotherapeutic drugs, but not in others. For each drug, genes meeting these criteria were selected and then filtered to exclude any genes that also met the same significance criteria in other drugs, thereby ensuring drug-specificity. Fourth, 6,556 multidrug genes were chosen based on a filter in a sum of chemotherapies between $2 \leq x < 7$. Two chemotherapies were selected to a minimum number, based on the mixture distribution of gamma, and the normal distribution shows a cutoff value of 2.64 in filtered ORFs in the sum of chemotherapies. Lastly, 2,849 non-respondent genes were picked with ORFs, with the filter for no enriched or depleted ORFs. To avoid any overlapping of ORFs between categories, the selection was prioritized sequentially from "Core" to "Non-respondent" genes. Gene categorization overlaps for comparison with ORFeome screening results were determined using the Venn library (v0.1.3) in Python (3.11).

Biomarker Analysis

To assess the clinical relevance of gene subsets identified in our screening, we performed a multi-tiered biomarker overlap analysis using TheMarker database[31], a curated resource of disease, biomarker, and drug associations. Gene matching was based on NCBI Gene identifiers, and only verified, non-silent ORFs from the

screening dataset were included in the analysis. Verified, non-silent ORFs were intersected with cancer-specific biomarker entries in the database, and for each gene subset, the proportion of genes overlapping with cancer-associated biomarkers, classed as "ICD-02: Benign/in-situ/malignant neoplasm," was calculated to evaluate enrichment for clinically validated cancer-related genes. We then examined the association between biomarker-annotated ORFs and chemotherapeutic drugs; among the 16 drugs profiled in the screen, 7 were represented in the database, and we quantified how many biomarker-annotated ORFs within each subset were linked to one or more of these drugs, connecting screening hits to existing pharmacogenomic knowledge. Finally, we assessed the representation of each gene subset within the total biomarker pool by calculating the proportion of ORFs present in the database relative to the total unique biomarkers catalogued. All overlap proportions were normalized using z-scores to enable comparisons across subsets and highlight differential enrichment patterns. The z-scores were visualized as heatmaps, and the numbers of four biomarker types within each ORF category were calculated as log10-transformed counts and displayed in stacked bar graphs. The stacked bars represent log-transformed counts for each category separately and do not correspond to the sum of total biomarker numbers.

Gene properties analysis

Length of each ORF (ORF Length) and its GC percentage (GC Content) were both derived from human ORFeome v9.1 sequences and delta space references after alignment as described above. We adopted some of the gene properties from our previous study[22,23]. Genotype-Tissue Expression project (GTEx)[110] v6 transcriptome data (downloaded on March 15th, 2016) was processed with R v3.5.1 to extract normalized log read count. In this dataset, 16 brain subregions were collapsed into three brain tissues: basal ganglia, cerebellum, and others. The median expression of each gene across all tissue samples was calculated, and a cutoff of >5 was applied. Cell line samples were excluded, and the analysis was restricted to the set of protein-coding genes. We considered these values as a reference for gene expression level (Abundance), which were calculated as mentioned in our previous study [22]. The same data was used to calculate the number of tissues each gene was expressed in (Broadness). Genes' age information (Age) was adopted from the Protein Historian database [111], with PPODv4_Jaccard as the family database and Dollo parsimony model as the reconstruction algorithm. The D2P2 database of disordered protein predictions[112] was used to calculate the fraction of disordered regions relative to full protein length (Disorder). We extracted all human D2P2 IDs mapped to Ensembl IDs from "genomes.protein.gz". Using this list and the "consensus_ranges.tsv.gz" file, containing the start and end positions of each predicted disordered region, the total disordered length for each protein was computed. The length of each protein was calculated using the "genomes.fa.bz2" file, which stored the full protein sequences. The fraction was calculated by merging the data by gene IDs. The DAVID functional annotation tool[113] (v2023q4) was used to retrieve the information for other properties. The list of Entrez IDs was submitted, and the Functional Annotation Table for the following databases was downloaded on January 16th, 2024: OMIM_DISEASE, Gene Ontology (GO) - GOTERM_BP_DIRECT (Biological Process), GOTERM_CC_DIRECT (Cellular Components), GOTERM_MF_DIRECT (Molecular Function), PUBMED_ID, and UP_TISSUE. Terms associated with each dataset were counted. The number of publications (PubCount) was determined by the number of unique PubMed

IDs per gene. Similarly, we used Gene Ontology to obtain the number of associated terms for each gene (GO Terms), UP_TISSUE[114] to assess the number of tissues in which each gene's expression level was upregulated (Tissue Count), and OMIM to find the number of diseases per gene (OMIM Terms). As a proxy for genes' importance and using the CORUM database version 4.1 (downloaded on January 23rd, 2024), we calculated the number of protein complexes associated with each gene (CORUM Complexes).

Prognostic Evaluation of Proliferation-Associated ORFs Using z-score Ranking, Precision–Recall Metrics, and Survival Analysis

To assess the clinical relevance of cancer proliferation-related ORF populations, we employed a combined z-score ranking approach. For each experimental batch, changes in ORF abundance were calculated as the log2 difference between mean expression under DMSO treatment and baseline expression, following a strategy adapted from CRISPR screening methodologies[115]. These differences, representing perturbation responses, were computed for each batch independently. The resulting difference matrix was then z-score normalized batch-wise, transforming each batch's distribution. To aggregate signals across multiple experimental batches, an inverse z-score transformation method from a previous study[116] was applied per ORF. Specifically, each z-score was converted back to an approximated original scale using the Inverse z-score Transformation. The combined z-score for each ORF was then calculated by averaging these inverse-transformed values across all batches. ORFs were ranked in descending order based on their combined z-scores. The ORF list was integrated with prognostic annotations from the Human Protein Atlas (HPA)[117] (v24.0) through a two-step merge using Entrez gene IDs and gene symbols. Genes annotated with a favorable and unfavorable prognosis (defined as being linked to poor outcome in at least one cancer type) were designated as "Hits". Based on the first hit, the first rank was computed. To assess performance, precision was calculated as the proportion of Hits among the top-ranked genes, and recall as the proportion of total Hits identified up to a given rank. In addition, the maximum precision at or beyond each rank was computed using a cumulative maximum function applied in reverse order. Both the maximum and recall curves were visualized to evaluate the model's ability to prioritize the proliferation-related ORF associated with poor prognosis. A parallel analytical pipeline was applied to assess the association between ranked ORFs and drug resistance, using manually curated resistance from The Cancer Genome Atlas (TCGA), following criteria from prior work[118]. For survival analysis, we used TCGA-derived RNA-seq data from HPA. Patients were stratified into high and low expression groups for each gene using the best expression cut-off, provided by HPA, defined as the Fragments Per Kilobase of transcript per Million mapped reads (FPKM) threshold that yields the lowest log-rank P-value between the two groups.

Spatial Analysis of Functional Enrichment (SAFE) network analysis

Using the log2 fold change of samples, overall perturbation profile similarity was calculated using the Pearson Correlation Coefficient (PCC) using Scipy[108] (v1.15.2). Paired ORFs with a PCC of 0.9 and a p-value <0.01 were selected as significant ORF perturbation similarity. Visualization of a selected paired ORF network was marked using the Cytoscape tool (v3.4.0)[119]. To visualize this novel network, we used Spatial Analysis of Functional Enrichment (SAFE; v1.5) to determine, visualize, and predict significant functional modules in the network[36]. The

network layouts were generated using the edge-weighted spring-embedded layout. GO terms for each ORF were extracted from FuncAssociate (v3 - GO updated in February 2018)[120]. The SAFE analysis will be run using MATLAB with the default setting file options: neighborhoodRadius = 1, groupEnrichmentMinSize = 5, plotNetwork = 0 (off), and THRESHOLD_ENRICHMENT = 0.3.

Protein-Protein interaction network analysis

We selected 21 ORF genes from two key clusters identified by SAFE ("BH domain binding" and "Autophagy & Apoptosis") based on SAFE's first neighborhood enrichment scores and domain annotation scores. These genes were analyzed using the STRING database[59] (accessed March 10th, 2025) to construct a protein–protein interaction network focused on their predominant GO domains. In STRING, only experimentally supported interactions with a high-confidence score of ≥0.9 were included. To capture relevant protein interactions beyond the selected ORFs, the maximum number of interactors was limited to 50 in both the first and second interaction shells. The resulting network was clustered using k-means clustering[121] with the minimum number of clusters, with cluster names derived from their GO term descriptions. The network layout was arranged in a grid format and further visualized in Cytoscape[119].

Gene Set Enrichment Analysis

Enrichment analysis was performed on ORF categorizations and drug response effector ORFs for individual chemotherapies using GSEApy (v1.1.9) via the Enrichr API[122]. Enrichment was tested against multiple curated gene sets, including KEGG_2021_Human, Reactome_Pathways_2024, WikiPathways_2024_Human, GO_Biological_Process_2025, and GO_Molecular_Function_2025. Only the individual chemotherapy gene was further analyzed, including BioPlanet_2019. To improve robustness, enriched terms with fewer than three associated genes or p-values < 0.05 were filtered out. For ORF categorizations, GO terms were manually curated to retain those relevant to cancer-associated mechanisms. For individual chemotherapies, GO terms shared by multiple drugs within the same category (present in more than two drugs) were identified by grouping terms by drug and category. These results were then manually reviewed to retain mechanistically meaningful and cancer-associated terms, incorporating both commonly shared GO terms and individual drug-specific terms. Curated GO terms were visualized using Matplotlib, plotting odds ratio and –log10(p-value) for each term.

Preparation for single-cell RNA-seq

Cryopreserved HeLa cell stocks after drug selection screen were thawed in a 37 °C water bath for 1 min and transferred to DMEM (Gibco, 11965092) containing 10% FBS (Gibco, 16000044), followed by centrifugation at 300 ×g for 5 min at room temperature. Cells were further washed twice with cold DPBS (Gibco, 14190144) containing 0.04% BSA (Miltenyi Biotec, 130-091-376) and filtered through 40 µm cell strainers (Flowmi, BAH136800040). Cell suspensions were then loaded, targeting recovery of 10,000 cells per channel to construct single-cell RNA sequencing libraries using Chromium Next GEM Single Cell 5' v2 kit (10x Genomics, PN-1000263) according to the manufacturer's instructions. Generated libraries were sequenced on an Illumina NovaSeq 6000 at a paired-end, 150 bp reads configuration, aiming for 400 million reads per library.

Single-cell RNA-seq bioinformatic Analysis

Reads were aligned to the GRCh38 human genome (2020-A, 10x Genomics) using Cell Ranger (v6.0.2, 10x Genomics) with an include-introns parameter. Filtered count matrices were further processed with Scanpy (v1.9.1)[123]. Low-quality cells with < 2,000 total counts, < 500 number of genes, > 10% mitochondrial gene counts, and > 0.4 Scrublet (v0.2.3)[124] doublet scores were filtered out. Raw counts were normalized per cell to 10,000, and ln(x+1) was transformed and further used as gene expression values. The top 2,000 highly variable genes were selected with scanpy.pp.highly_variable_genes function with seurat_v3 method[125] for each dataset, and principal component analysis (PCA) was performed with the selected highly variable genes. The neighborhood graph was built from the top 50 principal components and visualized using UMAP[126]. Clustering was performed with the Leiden method[127] with 0.5 resolution, and the Wilcoxon rank-sum test[128] calculate marker genes of each cluster. Cell cycle scores were calculated with previously defined cell cycle genes[129] using scanpy.tll.score_genes_cell_cycle function.

Drug-response assay for Lysotracker screening

Cells were treated with TAS-102 at concentrations ranging from 0.049 to 108 μM in a three-fold dilution series for 24, 48, or 72 hrs to establish the optimal conditions for induced lysosomal activation. DMSO vehicle control and single ORF overexpressing cell lines of CDK6 and BCL2 were included as drug sensitivity and resistance positive controls, respectively. Following TAS-102 treatment, lysosomal activity was assessed by staining cells with 50 nM LysoTracker Green DND-26 (Thermo Fisher Scientific Inc., L7526) for 20 minutes at 37°C. After two washes with DPBS, cells were analyzed using a CytoFlex flow cytometer (Beckman Coulter, C09766) to measure green fluorescence as an indicator of lysosomal activity.

Lysotracker screening

The FACS method followed the protocol from the previous study[130]. HeLa ORFeome cells were thawed and pooled to maximize ORF library coverage, then allowed to recover for three days before the experiment. For the screening, 3 million cells were seeded per 15 cm dish for the DMSO control, and 20 million cells were seeded for TAS-102 treatment to ensure sufficient representation of potential resistant populations. Drug treatment was initiated while cells were still in suspension, before full adherence. Following 72 hrs of TAS-102 treatment, lysosomal activity was assessed by staining cells with 50 nM LysoTracker Green for 20 minutes at 37°C. After two DPBS washes, cells were sorted on a Sony MA900 flow cytometry cell sorter, gating for mCherry positivity and then for enhanced lysosome activity (High green fluorescence, upper 5%) and diminished lysosome activity (Low green fluorescence, lower 5%) populations. Sorted cells were collected, seeded in 6-well plates, and expanded for one week before being cultured to full confluence in 15 cm dishes. Genomic DNA extraction, PCR amplification, and sequencing were performed as previously described for chemotherapeutic screening. Sequencing data from the sorted populations were analyzed using the ORFeome bioinformatics pipeline to prioritize candidate genes and compare the findings with those from the TAS-102 chemoresistance screen. Survival analysis and patient stratification based on the selected genes were performed using data from the TCGA PanCancer Atlas[32] via cBioPortal[131] (accessed on July 7th, 2025), and visualized using Matplotlib.

Western Blot Analysis for Lysosomal Pathway Protein Detection

HeLa cells were treated with DMSO, 1 μM TAS-102, or 1 μM ABT-263 (MedChemExpress, HY-10087), 1 μM TAS-102 + 1 μM ABT-263 in combination, and HCT 116 were treated with DMSO, 1 μM TAS-102, or 0.5 μM ABT-263, 1 μM TAS-102 + 1 μM ABT-263 in combination. After 48 hrs of incubation, cells were lysed in sample buffer (60 mM Tris-HCl (pH 6.8), 2% SDS, 10% glycerol, 5% β-mercaptoethanol, and 0.01% bromophenol blue) or EBC lysis buffer (50 mM Tris-HCl (pH 8.0), 200 mM NaCl and 0.5% NP-40), supplemented with a protease inhibitor cocktail (Sigma-Aldrich, 11836153001) for 30 min on ice. Total cell lysates were sonicated three times for 10 sec each and centrifuged at 15,000 ×g at 4˚C for 15 min. The supernatants were collected, and protein concentration was measured using the BCA protein assay (Thermo Fisher Scientific, Inc., 23227). Equal amounts of protein were separated by 8-15% SDS–PAGE and transferred to a PVDF membrane (Bio-Rad Laboratories, Inc., 1620177). The membranes were blocked in 1X Tris-buffered saline with 0.1% Tween 20 and 5% skim milk (CellNest, CNS109) at room temperature for 30 min and incubated with primary antibodies at 4˚C overnight or room temperature for 1 hr and secondary antibodies at room temperature for 1 hr. The signal was visualized with ImageQuant™ 800 (Cytiva™, 29399481) using an enhanced chemiluminescence detection system (GenDEPOT, LLC, W3652-020). The following antibodies were used:

Antibodies (Ab)	Ab type	Vendor	Catalog	1st Ab Dilution	2nd Ab Dilution
BCL2 (D17C4)	1st	Cell Signaling	3498; RRID:AB_190307; Lot 6	1:500	1:2500
TFEB	1st	Cell Signaling	4240; RRID AB_11220225; Lot 2	1:1000	1:2500
SQSTM1/p62	1st	Cell Signaling	5114; RRID_10624872; Lot 6	1:1000	1:2500
LAMP-1 (H4A3)	1st	Santa Cruz	Sc-20011; RRID: AB_626853; Lot C0620	1:500	1:3000
LC3B	1st	Novous Biologicals	NB100-2220; RRID: AB_10003146	1:2000	1:5000
β-Actin	1st	Sigma-Aldrich	A1978; RRID: AB_476692; Lot 0000124305	1:10000	1:10000
Peroxidase AffiniPure Goat Anti-Rabbit IgG (H+L)	2nd	Jackson ImmunoResearch	111-035-003; RRID: AB_2313567		
Peroxidase AffiniPure Goat Anti-Mouse IgG, Fcγ fragment specific	2nd	Jackson ImmunoResearch	111-035-071; RRID: AB_2338506		

Immunofluorescence assay

HCT 116 cells were seeded at a density of 1 x 10^5 cells per well in 12-well plates. On the following day, cells were treated with DMSO, 2 μM TAS-102, or 0.5 μM ABT-263, 2 μM TAS-102 + 0.5 μM ABT-263 in combination. After 24 hrs of incubation, the cells were fixed with 2% paraformaldehyde at room temperature for 10 min and permeabilized with 0.5% Triton X-100 for 5 min. Next, the cells were treated with 2% bovine serum albumin (A0100-010; GenDEPOT) for 1 h and then incubated with primary antibodies in 2% BSA at 4°C overnight. The next day, the cells were treated with Alexa Fluor 488 goat IgG antibody (A-11034; Thermo Fisher Scientific Inc.) in 2% BSA at room temperature for 1 h. Nuclei were stained with 1 μg/mL Hoechst 33342 (B2261; Sigma) for 5 min at room temperature. Finally, the coverglass slips were then mounted onto glass slides using an anti-fade fluorescence mounting medium (ab104135; Abcam). Confocal imaging (LSM800; CLSM) was subsequently performed. The primary antibodies used were against TFEB (1:500; 4240; Cell Signaling). Nucleus fluorescence was measured using ImageJ software (v1.53K; National Institutes of Health). Statistical analysis was performed using Student's T-test, followed by Bonferroni's multiple comparison testing.

DQ-BSA assay

To assess lysosomal proteolytic activity, the DQ-BSA assay was performed as previously described[132]. HCT 116 cells were seeded at 1 x 10^5 cells per well on coverglass slips in a 12-well plate. The following day, cells were treated with 2 μM TAS-102 or 0.5 μM ABT-263 or a combination of both. Following 24 hr of drug treatment, cells were incubated with 10 μg/mL DQ™ Red-BSA (D12051; Thermo Fisher Scientific Inc.) at 37°C for 24 hr in the dark. After incubation, cells were washed with DPBS and fixed in 2% formaldehyde for 10 min at room temperature. Nuclei were stained with Hoechst 33342 (1 μg/mL). The coverglass slips were then mounted onto glass slides using an anti-fade fluorescence mounting medium (ab104135; Abcam). Confocal imaging (LSM800; CLSM) was subsequently performed. Statistical analysis was performed using Student's T-test, followed by Bonferroni's multiple comparison testing.

Measurement of lysosomal pH

To assess lysosomal acidification, the lysosomal pH was measured as previously described[133]. HCT 116 cells were seeded at 1 x 10^4 cells per well into a 96-well plate. The following day, cells were incubated in complete media containing 2 μM TAS-102 or 0.5 μM ABT-263 or 2 μM TAS-102 + 0.5 μM ABT-263 in combination. Following 24 hr of drug treatment, cells were incubated with 1 μM LysoSensor™ Yellow/Blue DND-160 (L7545; Thermo Fisher Scientific Inc.) for 5 min in the dark. After incubation, cells were washed with DPBS, and then the media was replaced with HBSS supplemented with 10% FBS for 10 min for test wells. To generate a standard curve, cells were incubated with a buffer (25 mM HEPES, 115 mM KCl, 1.2 mM $MgCl_2$, 10 mM glucose, 10 μM nigericin) calibrated to pH 4.5, 5.0, 5.5, 6.0, 6.5, or 7.0 as described previously. Fluorescence was measured using a PerkinElmer EnSpire multimode plate reader with excitation/emission filters set to 329/440 nm and 380/540 nm. Statistical analysis was performed using Student's T-test, followed by Bonferroni's multiple comparison testing.

Cell proliferation assay

To assess the cell proliferation difference in drug treatment, the proliferation was measured using a method adopted from the previous protocol[134]. HCT 116 cells were seeded in a 12-well plate at a density of 5×10^4 cells per well. The following day, cells were treated with 2 μM TAS-102 or 100 nM Bafilomycin A1 or 2 μM TAS-102 + 100 nM Bafilomycin A1 in combination. Cells were harvested and counted at 0, 24, and 48 hr after drug treatment. Each counting was performed in six independent replicates for each time point, and the mean values were used to generate the proliferation graphs. Statistical analysis was performed using Student's T-test, followed by Bonferroni's multiple comparison testing.

Cytotoxicity assay

To examine drug cytotoxicity in human cells, we followed the protocol from the previous publication[135]. Cells were plated at a density of 5,000 cells/well in a 96-well plate and treated with the appropriate drugs at 0, 0.01, 0.05, 0.1, 0.5, 1, 4, 10, 25, and 50 μM concentrations for 24 hrs. The media was replenished with media containing one of 0, 25, 50, 100, 200, 500, 1000, 2000, 5000 nM concentration of ABT-263, followed by a 96-hour incubation. Cells were stained with 0.5% Methylene Blue for 30 minutes, rinsed with water, dried, solubilized in 5% SDS in DPBS, and read at 650 nm. The average of the control normalized the raw absorbance. The graph was generated using GraphPad Prism 9, and statistical analysis was performed using two-way ANOVA followed by Tukey's multiple comparisons test.

Cell Line-Derived Xenografts (CDX) Model

Female SCID/CB17 mice, aged 6 to 8 weeks, were sourced from a colony bred and maintained at Roswell Park Comprehensive Cancer Center (RPCCC). Mice were inoculated subcutaneously into the left flank with exponentially growing HCT 116 tumor cells (1×10^6 per mouse) suspended in a 1:1 mixture of DPBS and Matrigel. Mice were monitored daily, while body weight and tumor size were measured twice a week. Tumor dimensions were measured with digital calipers and calculated using the formula: (length × width^2)/2. When the tumor volume reached 50 to 100mm^3, mice were randomly divided into four treatment groups (n = 10-11 for each group): (1) Vehicle Control, (2) 50 mg/kg ABT-263, (3) 50 mg/kg TAS-102 (MedChemExpress, Cat. No. HY-16478), and (4) Combination (50 mg/kg TAS-102 + 50 mg/kg ABT-263). TAS-102 was dissolved in 10% 2-hydroxypropyl-β-cyclodextrin (HPCD) in DPBS, while ABT-263 was prepared in 10% DMSO, 40% PEG300, 5% Tween-80, and 45% saline. For the Combination group, TAS-102 was given 4 hr before ABT-263. All drugs were freshly prepared and given via oral gavage on a 5-day-on, 2-days-off schedule.

Patient-Derived Xenograft (PDX) Model

Donor and experimental colorectal patient-derived xenografts (CRC PDX-03-26-4p) were grown in male SCID/CB17 mice from RPCCC. Tumors were grown in donor mice to a volume of around 1500 mm^3, harvested rapidly after euthanasia, cut into 2×2×2 mm^3 pieces under sterile conditions, and surgically implanted into the left flank. Surgical staples were removed about ten days post-implantation. Tumor growth and mouse weights were measured twice a week. Once the average tumor volume reached 50-100 mm^3, mice were randomized into four groups like CDX model (1) Vehicle Control, (2) 50 mg/kg ABT-263, (3) 50 mg/kg TAS-102, and (4) Combination (50 mg/kg TAS-102 + 50 mg/kg ABT-263). The drugs were prepared as in the CDX experiment

and administered via the same route and dosing schedule.

Mouse Model Monitoring

Mice were housed in micro-isolator units and were provided with food and water ad libitum. Daily monitoring of animal health, body weight, and tumor growth, as well as endpoint procedures, following the protocol and guidelines approved by the Institutional Animal Care and Use Committee (IACUC). The facility is certified by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC) International and complies with regulations and standards of the US Department of Agriculture and the US Department of Health and Human Services.

At the endpoint, when the tumor diameter reached 20 mm, mice were euthanized and subjected to necropsy and organ collection. Tumor growth assessment was stopped once a mouse reached the endpoint (typically in a vehicle-treated cohort) while the survival data were still collected in the study. Intracardiac blood was collected and stored in EDTA-coated tubes at 4 °C for a complete blood count test.

Complete Blood Counts

Blood was collected by cardiac puncture into a tube containing EDTA solution at the endpoint of the study[135]. Analysis was performed with 20 µL blood samples using the Heska HemaTrue Analyzer and HeskaView Integrated Software version 2.5.2.

Statistics for *In Vivo* data

Tumor Growth Inhibition (TGI) for each mouse was calculated as the relative difference between the average tumor volume of the Vehicle Control group and the tumor volume of the individual mouse with respect to the Vehicle Control group average. Statistically significant differences were determined using one-way ANOVA for TGI and CBC and a two-way ANOVA test for tumor growth curves. Multiple testing was performed using Tukey's multiple comparisons in GraphPad Prism 9.

References

1. Letai, A. & de The, H. Conventional chemotherapy: millions of cures, unresolved therapeutic index. *Nat. Rev. Cancer* **25**, 209–218 (2025).
2. Anand, U. *et al.* Cancer chemotherapy and beyond: Current status, drug candidates, associated risks and progress in targeted therapeutics. *Genes Dis.* **10**, 1367–1401 (2023).
3. Flaherty, K. T. Chemotherapy and targeted therapy combinations in advanced melanoma. *Clin. Cancer Res.* **12**, 2366s–2370s (2006).
4. Lopez, J. P. *et al.* Gefitinib inhibition of drug resistance to doxorubicin by inactivating ABCG2 in thyroid cancer cell lines. *Arch. Otolaryngol. Head. Neck Surg.* **133**, 1022–1027 (2007).
5. Kosorok, M. R. & Laber, E. B. Precision medicine. *Annu. Rev. Stat. Appl.* **6**, 263–286 (2019).
6. Mucaki, E. J., Zhao, J. Z. L., Lizotte, D. J. & Rogan, P. K. Predicting responses to platin chemotherapy agents with biochemically-inspired machine learning. *Signal Transduct. Target. Ther.* **4**, 1 (2019).
7. Ali, M., Khan, S. A., Wennerberg, K. & Aittokallio, T. Global proteomics profiling improves drug sensitivity prediction: results from a multi-omics, pan-cancer modeling approach. *Bioinformatics* **34**, 1353–1362 (2018).
8. Menghi, F. & Liu, E. T. Functional genomics of complex cancer genomes. *Nat. Commun.* **13**, 5908 (2022).
9. Zhou, Z. *et al.* Genome-scale CRISPR-Cas9 knockout screening in nasopharyngeal carcinoma for radiosensitive and radioresistant genes. *Transl. Oncol.* **30**, 101625 (2023).
10. Ezkurdia, I. *et al.* Most highly expressed protein-coding genes have a single dominant isoform. *J. Proteome Res.* **14**, 1880–1887 (2015).
11. Zhao, W.-S. *et al.* Genome-scale CRISPR activation screening identifies a role of ELAVL2-CDKN1A axis in paclitaxel resistance in esophageal squamous cell carcinoma. *Am. J. Cancer Res.* **9**, 1183–1200 (2019).
12. Arnoldo, A. *et al.* A genome scale overexpression screen to reveal drug activity in human cells. *Genome Med.* **6**, 32 (2014).
13. Dull, T. *et al.* A third-generation lentivirus vector with a conditional packaging system. *J. Virol.* **72**, 8463–8471 (1998).
14. Alyateem, G. *et al.* Use of CRISPR-based screens to identify mechanisms of chemotherapy resistance.

Cancer Gene Ther. **30**, 1043–1050 (2023).

15. Bester, A. C. *et al.* An Integrated Genome-wide CRISPRa Approach to Functionalize lncRNAs in Drug Resistance. *Cell* **173**, 649–664.e20 (2018).
16. Neupane, M. *et al.* MECP2 Is a Frequently Amplified Oncogene with a Novel Epigenetic Mechanism That Mimics the Role of Activated RAS in Malignancy. *Cancer Discov.* **6**, 45–58 (2016).
17. Bulcha, J. T., Wang, Y., Ma, H., Tai, P. W. L. & Gao, G. Viral vector platforms within the gene therapy landscape. *Signal Transduct. Target. Ther.* **6**, 53 (2021).
18. Liu, J. *et al.* Increasing genome editing efficiency of Cas9 nucleases by the simultaneous use of transcriptional activators and histone acetyltransferase activator. *CRISPR J.* **5**, 854–867 (2022).
19. Charrier, S. *et al.* Quantification of lentiviral vector copy numbers in individual hematopoietic colony-forming cells shows vector dose-dependent effects on the frequency and level of transduction. *Gene Ther.* **18**, 479–487 (2011).
20. Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010).
21. Saranraj, K. & Kiran, P. U. Drug repurposing: Clinical practices and regulatory pathways. *Perspect. Clin. Res.* **16**, 61–68 (2025).
22. Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature* **580**, 402–408 (2020).
23. Kim, D.-K. *et al.* A proteome-scale map of the SARS-CoV-2-human contactome. *Nat. Biotechnol.* **41**, 140–149 (2023).
24. Matreyek, K. A., Stephany, J. J., Chiasson, M. A., Hasle, N. & Fowler, D. M. An improved platform for functional assessment of large protein libraries in mammalian cells. *Nucleic Acids Res.* **48**, e1 (2020).
25. Landry, J. J. M. *et al.* The genomic and transcriptomic landscape of a HeLa cell line. *G3 (Bethesda)* **3**, 1213–1224 (2013).
26. Matreyek, K. A., Stephany, J. J. & Fowler, D. M. A platform for functional assessment of large variant libraries in mammalian cells. *Nucleic Acids Res.* **45**, e102 (2017).
27. Álvarez-Fernández, M. & Malumbres, M. Mechanisms of sensitivity and resistance to CDK4/6 inhibition. *Cancer Cell* **37**, 514–529 (2020).
28. Cortes, E. P. Doxorubicin in Disseminated. *JAMA* **221**, 1132 (1972).

29. Andréll, J. & Tate, C. G. Overexpression of membrane proteins in mammalian cells for structural studies. *Mol. Membr. Biol.* **30**, 52–63 (2013).

30. Larionova, I. *et al.* Interaction of tumor-associated macrophages and cancer chemotherapy. *Oncoimmunology* **8**, 1596004 (2019).

31. Zhang, Y. *et al.* TheMarker: a comprehensive database of therapeutic biomarkers. *Nucleic Acids Res.* **52**, D1450–D1464 (2024).

32. Cancer Genome Atlas Research Network *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).

33. Moiso, E. Manual curation of TCGA treatment data and identification of potential markers of therapy response. *Oncology* (2021).

34. Gmeiner, W. H. & Okechukwu, C. C. Review of 5-FU resistance mechanisms in colorectal cancer: clinical significance of attenuated on-target effects. *Canc. Drug Resist.* **6**, 257–272 (2023).

35. Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420–aaf1420 (2016).

36. Baryshnikova, A. Spatial Analysis of Functional Enrichment (SAFE) in Large Biological Networks. *Methods Mol. Biol.* **1819**, 249–268 (2018).

37. Vogler, M. *et al.* The BCL2 family: from apoptosis mechanisms to new advances in targeted therapy. *Signal Transduct. Target. Ther.* **10**, 91 (2025).

38. Qian, S. *et al.* The role of BCL-2 family proteins in regulating apoptosis and cancer therapy. *Front. Oncol.* **12**, 985363 (2022).

39. Szklarczyk, D. *et al.* The STRING database in 2025: protein networks with directionality of regulation. *Nucleic Acids Res.* **53**, D730–D737 (2025).

40. Wu, P.-H. *et al.* Lysosomal trafficking mediated by Arl8b and BORC promotes invasion of cancer cells that survive radiation. *Commun. Biol.* **3**, 620 (2020).

41. Chen, Y. *et al.* Roles of the adaptor protein tumor necrosis factor receptor type 1-associated death domain protein (TRADD) in human diseases. *Biomed. Pharmacother.* **153**, 113467 (2022).

42. Jian, W. & Zhang, L. POLE2 silencing inhibits the progression of colorectal carcinoma cells via wnt signaling axis. *Cancer Biol. Ther.* **25**, 2392339 (2024).

43. Sun, Y. *et al.* Identification and verification of YBX3 and its regulatory gene HEIH as an oncogenic system: A multidimensional analysis in colon cancer. *Front. Immunol.* **13**, 957865 (2022).

44. Paek, A. R., Kim, S. H., Kim, S. S., Kim, K. T. & You, H. J. IGF-1 induces expression of zinc-finger protein 143 in colon cancer cells through phosphatidylinositide 3-kinase and reactive oxygen species. *Exp. Mol. Med.* **42**, 696–702 (2010).

45. Littlepage, L. E. *et al.* The transcription factor ZNF217 is a prognostic biomarker and therapeutic target during breast cancer progression. *Cancer Discov.* **2**, 638–651 (2012).

46. Alikarami, F. *et al.* GATA2 links stemness to chemotherapy resistance in acute myeloid leukemia. *Blood* **145**, 2179–2195 (2025).

47. Uboha, N. & Hochster, H. S. TAS-102: a novel antimetabolite for the 21st century. *Future Oncol.* **12**, 153–163 (2016).

48. Tse, C. *et al.* ABT-263: a potent and orally bioavailable Bcl-2 family inhibitor. *Cancer Res.* **68**, 3421–3428 (2008).

49. Sun, Y. *et al.* Hypoxia-induced autophagy reduces radiosensitivity by the HIF-1α/miR-210/Bcl-2 pathway in colon cancer cells. *Int. J. Oncol.* **46**, 750–756 (2015).

50. Yu, L. & Liu, S. Autophagy contributes to modulating the cytotoxicities of Bcl-2 homology domain-3 mimetics. *Semin. Cancer Biol.* **23**, 553–560 (2013).

51. Levine, B., Sinha, S. & Kroemer, G. Bcl-2 family members: dual regulators of apoptosis and autophagy. *Autophagy* **4**, 600–606 (2008).

52. Cheng, L. *et al.* mTOR-dependent TFEB activation and TFEB overexpression enhance autophagy-lysosome pathway and ameliorate Alzheimer's disease-like pathology in diabetic encephalopathy. *Cell Commun. Signal.* **21**, 91 (2023).

53. Schläfli, A. M., Berezowska, S., Adams, O., Langer, R. & Tschan, M. P. Reliable LC3 and p62 autophagy marker detection in formalin fixed paraffin embedded human tissue by immunohistochemistry. *Eur. J. Histochem.* **59**, 2481 (2015).

54. Frost, L. S., Dhingra, A., Reyes-Reveles, J. & Boesze-Battaglia, K. The use of DQ-BSA to monitor the turnover of autophagy-associated cargo. *Methods Enzymol.* **587**, 43–54 (2017).

55. Jeong, S.-J. *et al.* Trehalose causes low-grade lysosomal stress to activate TFEB and the autophagy-

lysosome biogenesis response. *Autophagy* **17**, 3740–3752 (2021).

56. Zhang, J. *et al.* Enabling tumor-specific drug delivery by targeting the Warburg effect of cancer. *Cell Rep. Med.* **6**, 101920 (2025).
57. Corcoran, R. B. *et al.* Synthetic lethal interaction of combined BCL-XL and MEK inhibition promotes tumor regressions in KRAS mutant cancer models. *Cancer Cell* **23**, 121–128 (2013).
58. Zhou, Y., Perez, R. E., Duan, L. & Maki, C. G. DZNep represses Bcl-2 expression and modulates apoptosis sensitivity in response to Nutlin-3a. *Cancer Biol. Ther.* **19**, 465–474 (2018).
59. Tolani, B. *et al.* Ras-mutant cancers are sensitive to small molecule inhibition of V-type ATPases in mice. *Nat. Biotechnol.* **40**, 1834–1844 (2022).
60. Kawakami, H. *et al.* Mutant BRAF upregulates MCL-1 to confer apoptosis resistance that is reversed by MCL-1 antagonism and cobimetinib in colorectal cancer. *Mol. Cancer Ther.* **15**, 3015–3027 (2016).
61. Cicero, G. *et al.* TAS-102 in metastatic colorectal cancer (mCRC): efficacy, tolerability, and quality of life in heavily pretreated elderly patients: a real-life study. *Drugs Context* **9**, (2020).
62. Mohamad Anuar, N. N., Nor Hisam, N. S., Liew, S. L. & Ugusman, A. Clinical review: Navitoclax as a pro-apoptotic and anti-fibrotic agent. *Front. Pharmacol.* **11**, 564108 (2020).
63. Low, B. E., Hosur, V., Lesbirel, S. & Wiles, M. V. Efficient targeted transgenesis of large donor DNA into multiple mouse genetic backgrounds using bacteriophage Bxb1 integrase. *Sci. Rep.* **12**, 5424 (2022).
64. Yang, H., Sun, L., Liu, M. & Mao, Y. Patient-derived organoids: a promising model for personalized cancer treatment. *Gastroenterol. Rep. (Oxf.)* **6**, 243–245 (2018).
65. Quintanal-Villalonga, Á. *et al.* Lineage plasticity in cancer: a shared pathway of therapeutic resistance. *Nat. Rev. Clin. Oncol.* **17**, 360–371 (2020).
66. Dixit, A. *et al.* Perturb-seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866.e17 (2016).
67. Madrid, P. B. *et al.* A systematic screen of FDA-approved drugs for inhibitors of biological threat agents. *PLoS One* **8**, e60579 (2013).
68. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251 (2013).
69. Youn, J.-Y. *et al.* High-density proximity mapping reveals the subcellular organization of mRNA-associated

granules and bodies. *Mol. Cell* **69**, 517–532.e11 (2018).

70. Waye, J. S. Forensic identity testing using highly polymorphic DNA markers: current status and emerging technologies. *Transfus. Med. Rev.* **7**, 193–205 (1993).
71. Chiasson, M. A. *et al.* Multiplexed measurement of variant abundance and activity reveals VKOR topology, active site and human variant impact. *Elife* **9**, (2020).
72. Sur, S. *et al.* A panel of isogenic human cancer cells suggests a therapeutic approach for cancers with inactivated p53. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3964–3969 (2009).
73. Inoue, A. *et al.* Sequential administration of XPO1 and ATR inhibitors enhances therapeutic response in TP53-mutated colorectal cancer. *Gastroenterology* **161**, 196–210 (2021).
74. Wiemann, S. *et al.* The ORFeome Collaboration: a genome-scale human ORF-clone resource. *Nat. Methods* **13**, 191–192 (2016).
75. Yachie, N. *et al.* Pooled-matrix protein interaction screens using Barcode Fusion Genetics. *Mol. Syst. Biol.* **12**, 863 (2016).
76. Mills, C. C., Kolb, E. A. & Sampson, V. B. Development of chemotherapy with cell-cycle inhibitors for adult and pediatric cancer therapy. *Cancer Res.* **78**, 320–325 (2018).
77. Dong, C. *et al.* A genome-wide CRISPR-Cas9 knockout screen identifies essential and growth-restricting genes in human trophoblast stem cells. *Nat. Commun.* **13**, 2548 (2022).
78. Ortega, P., Sanchez, A., Seldin, M. & Buisson, R. Oligo-seq protocol for mapping DNA motifs targeted by base editors. *STAR Protoc.* **6**, 103758 (2025).
79. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
80. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
81. Home - Gene - NCBI. https://www.ncbi.nlm.nih.gov/gene/.
82. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
83. Human ORFeome 2016 Release. http://horfdb.dfci.harvard.edu/.
84. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data.

http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

85. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
86. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
87. Zhu, Q., Liu, N., Orkin, S. H. & Yuan, G.-C. CUT&RUNTools: a flexible pipeline for CUT&RUN processing and footprint analysis. *Genome Biol.* **20**, 192 (2019).
88. Alerasool, N., Leng, H., Lin, Z.-Y., Gingras, A.-C. & Taipale, M. Identification and functional characterization of transcriptional activators in human cells. *Mol. Cell* **82**, 677–695.e7 (2022).
89. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
90. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
91. Smyth, G. K. limma: Linear Models for Microarray Data. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (eds. Gentleman, R., Carey, V. J., Huber, W., Irizarry, R. A. & Dudoit, S.) 397–420 (Springer New York, New York, NY, 2005).
92. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).
93. Abbas-Aghababazadeh, F., Li, Q. & Fridley, B. L. Comparison of normalization approaches for gene expression studies completed with high-throughput sequencing. *PLoS One* **13**, e0206312 (2018).
94. Signorell, A. DescTools: Tools for descriptive statistics. R package version 0.99. 20. *Cranio* (2017).
95. Hwang, D. *et al.* A data integration methodology for systems biology. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 17296–17301 (2005).
96. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
97. Waskom, M. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
98. Pope, B. & Kent, H. M. High efficiency 5 min transformation of Escherichia coli. *Nucleic Acids Res.* **24**, 536–537 (1996).
99. Girish, V. & Sheltzer, J. M. A CRISPR competition assay to identify cancer genetic dependencies. *Bio Protoc.* **10**, e3682 (2020).

100. Diaz, J. E. *et al.* The transcriptomic response of cells to a drug combination is more than the sum of the responses to the monotherapies. *Elife* **9**, (2020).

101. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).

102. Patel, H. *et al. Nf-Core/rnaseq: Nf-Core/rnaseq v3.18.0 - Lithium Lynx*. (Zenodo, 2024). doi:10.5281/ZENODO.14537300.

103. Schneider, V. A. *et al.* Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.* **27**, 849–864 (2017).

104. https://cir.nii.ac.jp/crid/1370294643762929691.

105. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

106. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

107. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).

108. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

109. Tarazona, S. *et al.* Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res.* **43**, e140 (2015).

110. Melé, M. *et al.* Human genomics. The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015).

111. Capra, J. A., Williams, A. G. & Pollard, K. S. ProteinHistorian: tools for the comparative analysis of eukaryote protein origin. *PLoS Comput. Biol.* **8**, e1002567 (2012).

112. Oates, M. E. *et al.* D^2P^2: database of disordered protein predictions. *Nucleic Acids Res.* **41**, D508–16 (2013).

113. Sherman, B. T. *et al.* DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221 (2022).

114. Zambelli, F. *et al.* RNentropy: an entropy-based tool for the detection of significant variation of gene expression across multiple RNA-Seq experiments. *Nucleic Acids Res.* **46**, e46 (2018).

115. Shifrut, E. *et al.* Genome-wide CRISPR screens in primary human T cells reveal key regulators of immune function. *Cell* **175**, 1958–1971.e15 (2018).

116. Salkovic, E., Sadeghi, M. A., Baggag, A., Salem, A. G. R. & Bensmail, H. OutSingle: a novel method of detecting and injecting outliers in RNA-Seq count data using the optimal hard threshold for singular values. *Bioinformatics* **39**, btad142 (2023).

117. Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science* **357**, (2017).

118. Moiso, E. Manual curation of TCGA treatment data and identification of potential markers of therapy response. *bioRxiv* (2021) doi:10.1101/2021.04.30.21251941.

119. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

120. Berriz, G. F., King, O. D., Bryant, B., Sander, C. & Roth, F. P. Characterizing gene sets with FuncAssociate. *Bioinformatics* **19**, 2502–2504 (2003).

121. Novikov, A. PyClustering: Data Mining Library. *J. Open Source Softw.* **4**, 1230 (2019).

122. Xie, Z. *et al.* Gene set knowledge discovery with Enrichr. *Curr. Protoc.* **1**, e90 (2021).

123. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 1–5 (2018).

124. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Systems* **8**, 281–291.e9 (2019).

125. Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).

126. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (2018).

127. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 1–12 (2019).

128. Sun, Y. & Qiu, P. Hierarchical marker genes selection in scRNA-seq analysis. *PLoS Comput. Biol.* **20**, e1012643 (2024).

129. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).

130. Kim, M. *et al.* Exogenous RNA surveillance by proton-sensing TRIM25. *Science* **388**, eads4539 (2025).

131. de Bruijn, I. *et al.* Analysis and visualization of longitudinal genomic and clinical data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. *Cancer Res.* **83**, 3861–3867 (2023).

132. Marwaha, R. & Sharma, M. DQ-Red BSA trafficking assay in cultured cells to assess cargo delivery to lysosomes. *Bio Protoc.* **7**, (2017).

133. Ma, L., Ouyang, Q., Werthmann, G. C., Thompson, H. M. & Morrow, E. M. Live-cell microscopy and fluorescence-based measurement of luminal pH in intracellular organelles. *Front. Cell Dev. Biol.* **5**, 71 (2017).

134. Morten, B. C., Scott, R. J. & Avery-Kiejda, K. A. Comparison of three different methods for determining cell proliferation in breast cancer cell lines. *J. Vis. Exp.* (2016) doi:10.3791/54350.

135. Alruwaili, M. M. *et al.* A synergistic two-drug therapy specifically targets a DNA repair dysregulation that occurs in p53-deficient colorectal and pancreatic cancers. *Cell Rep. Med.* **5**, 101434 (2024).

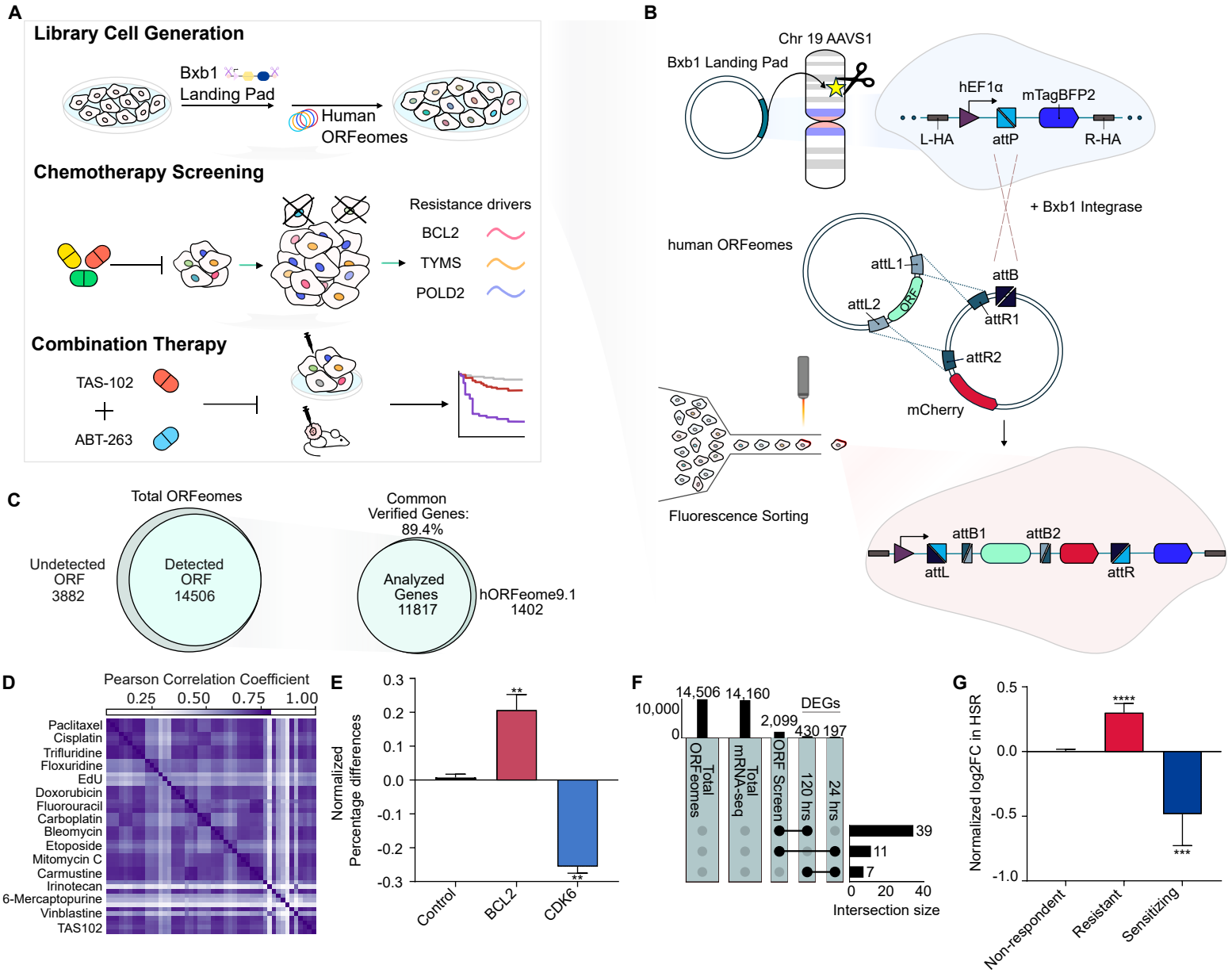


Figure 1: BOGO Enables Functional Proteome-Wide Screening

(A) Schematic overview of the study design. (B) Diagram of the BOGO platform illustrating ORF integration into the AAVS1 locus via Bxb1 landing pad. (C) Venn diagram showing overlap between verified ORFeomes and genes included in our downstream analysis. (D) Correlation matrix of drug response profile depicting batch and replicate variation across independent chemotherapeutic drug screenings, demonstrating high reproducibility among technical replicates (N=2-3). (E) Bar graph of individual competition assays under TFT treatment (N=3). (F) UpSet plot showing overlap of differentially expressed genes (DEGs) between the doxorubicin drug response screen and the HeLa wild-type transcriptomic screen at 24 and 120 hrs of doxorubicin treatment (±Log2 fold change 0.5, $p < 0.05$). (G) Bar graph of average fold change in High-Saturation Retesting (HSR) for "Non-respondent," "Sensitizing," and "Resistant" ORF categories from the drug response screening. Data in bar graphs are shown as mean ± Standard Error of the Mean (SEM). Statistical analysis by one or two-way ANOVA with Tukey's multiple comparisons; ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

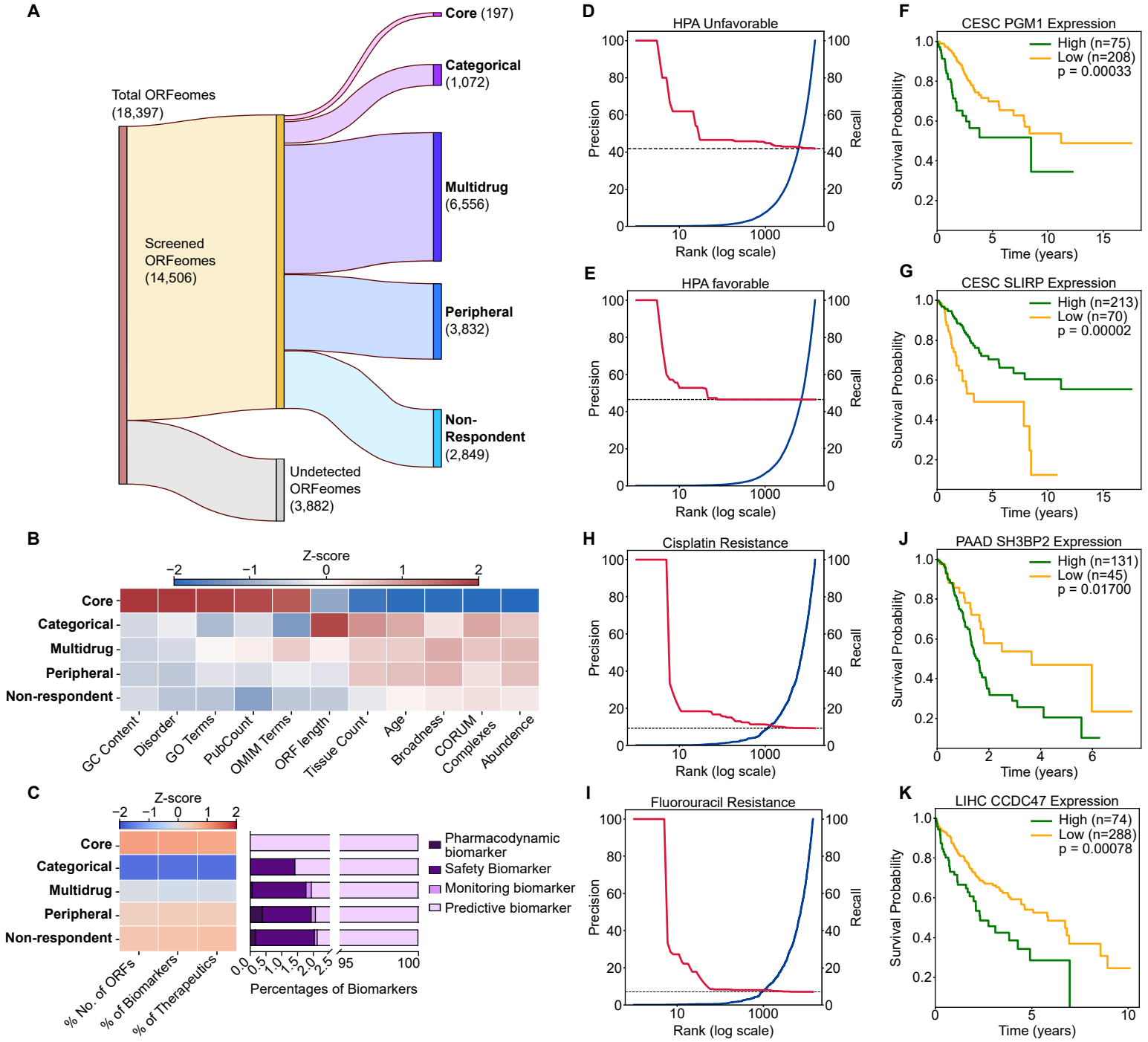


Figure 2. Biological and Clinical Relevance of BOGO Screening for Cell Proliferation and Drug Resistance
(A) Sankey diagram categorizing ORF numbers into drug response groups; undetected ORFs defined by median log2TMM < -1. (B) Heatmap of Z-scores for intrinsic ORF characteristics across categories. (C) Left: Heatmap showing overlap between ORFs and cancer biomarkers from TheMarker database; Right: Distribution of biomarker types across ORF categories. (D–E) Precision–recall curves ranking ORFs by cell proliferation effects, (D) comparing top upregulators with unfavorable prognosis, and (E) top downregulators with favorable prognosis from the HPA dataset. (F–G) Kaplan–Meier survival curves for most promising examples: (F) PGM1 unfavorable) and (G) SLIRP (favorable) in cervical cancer. (H–I) Precision–recall curves evaluating (H) cisplatin and (I) Fluorouracil resistance drivers from the ORFeome screen. (J–K) Kaplan–Meier survival curves for (J) SH3BP2 (pancreatic cancer) and (K) CCDC47 (liver cancer) that are associated with reduced survival. Survival curve comparisons were performed using the log-rank test.

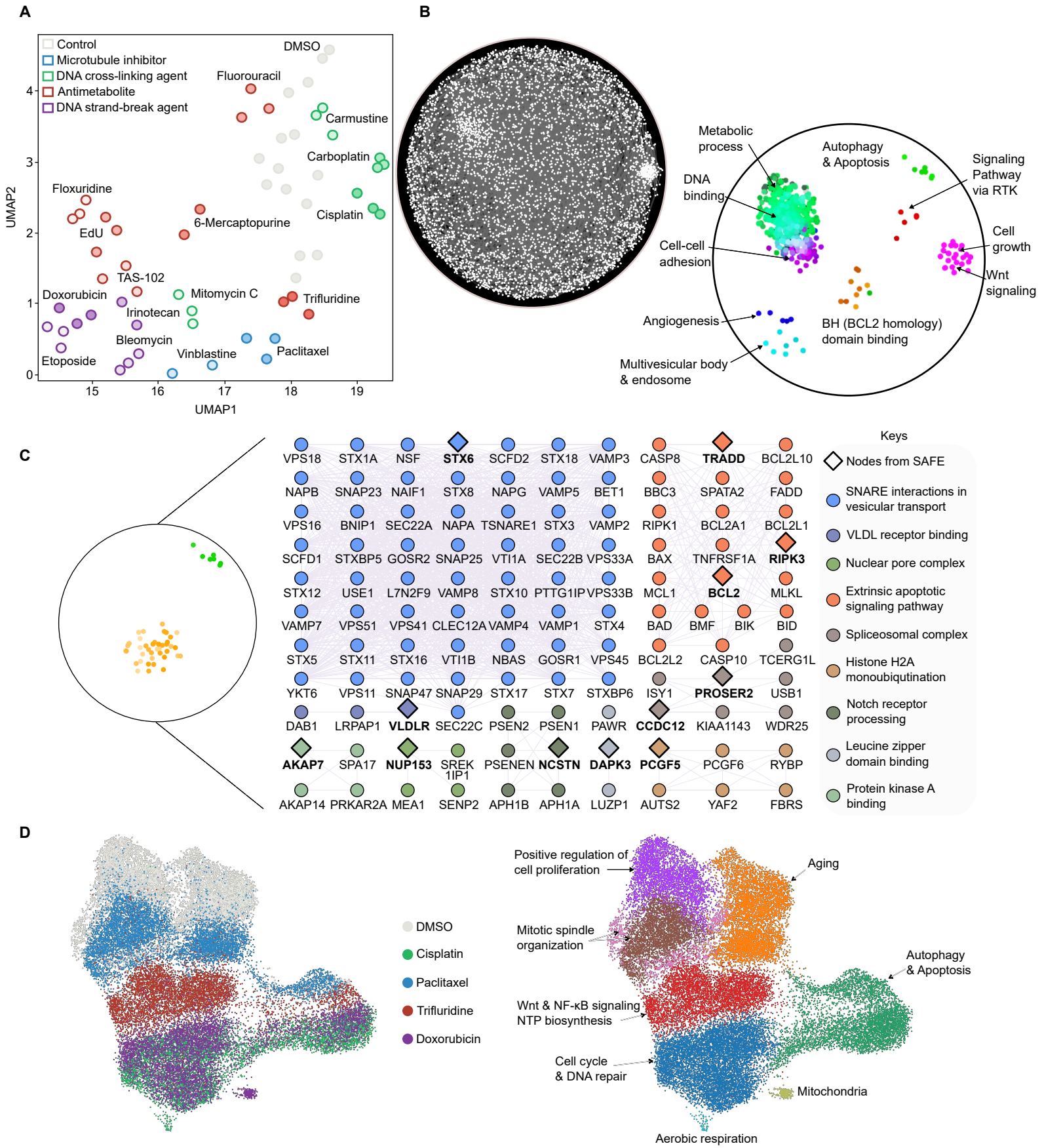


Figure 3. Comprehensive Systems Analysis of Drug Response Screening Highlighting BCL2-Associated Functional Pathways

(A) UMAP plot of ORF perturbation profiles for 16 chemotherapeutic agents, color-coded by one of four drug mechanisms of action. (B) Functional enrichment network from drug–ORF correlations, showing ORF as node and Pearson correlation (cut-off: < 0.9) as edge (left), functional clusters using SAFE (right). (C) STRING protein–protein interaction (PPI) network of two BCL2-related SAFE clusters ("BH domain binding" and "Autophagy & Apoptosis"), highlighting genes involved in vesicular transport, apoptotic signaling, histone modification, and Notch receptor pathways. The network is based on high-confidence experimental interactions (score > 0.9), <50 interactors in the first and second shells. 12 diamond shaped nodes highlight ORFs originating from the SAFE network clusters. (D) UMAP plots showing Single-cell RNA-seq treatment-specific ORF-expressing cell population changes for four different chemotherapeutics (left) and enriched pathways in distinct cell clusters identified by the Leiden algorithm and analyzed using DAVID (right).

.

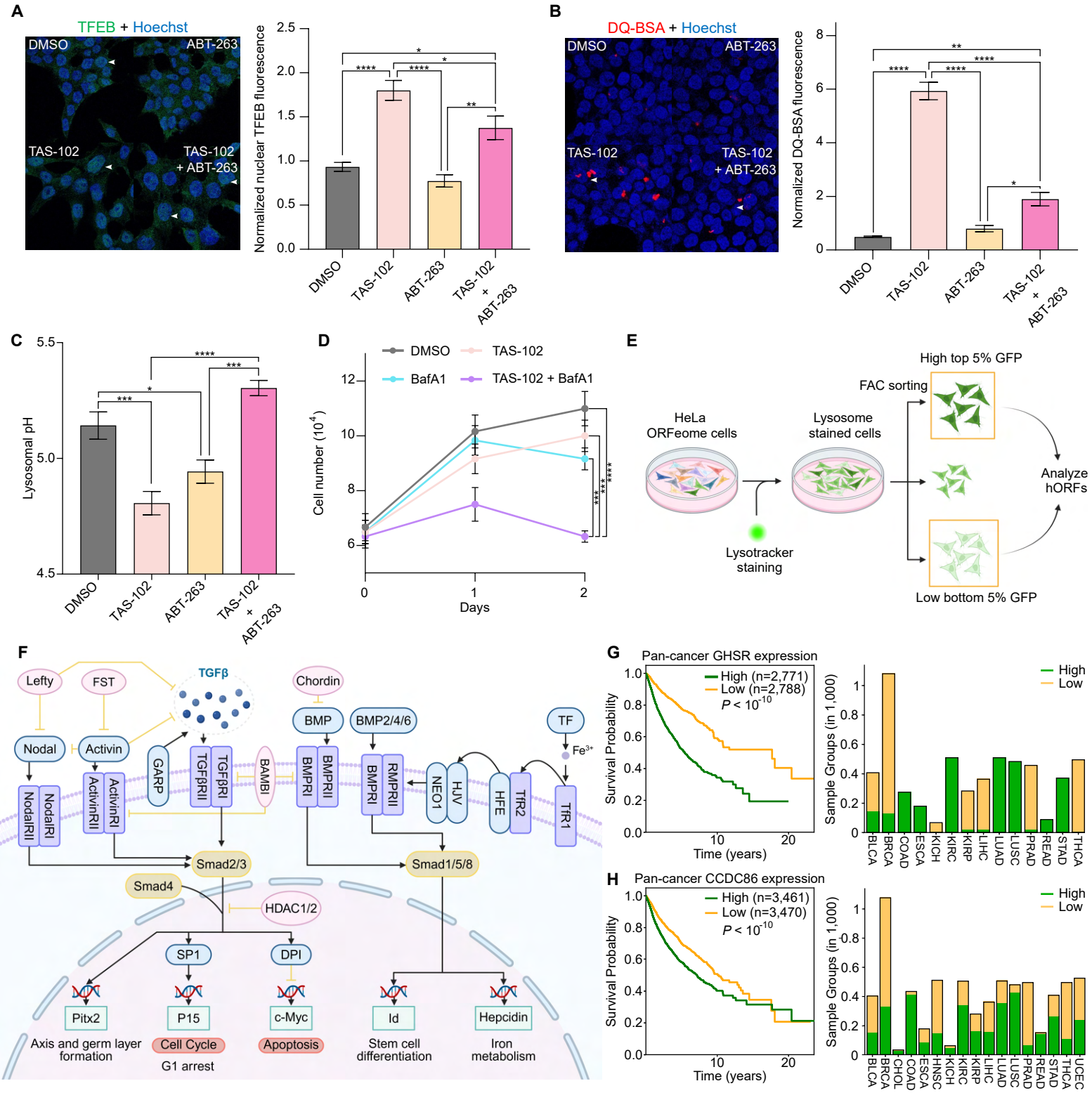


Figure 4. Functional Validation of BCL2 Inhibition and Autophagy-Lysosomal Modulation

(A) Merged fluorescent images of TFEB (green) co-stained with Hoechst (blue), showing nuclear localization changes after drug treatments. The bar chart quantifies nuclear TFEB fluorescence differences across four treatments. (B) Merged fluorescent images of DQ-BSA lysosomal proteolysis co-stained with Hoechst, showing lysosomal activity; bar chart quantifies fluorescence intensity differences. (C) Bar chart showing lysosomal pH changes after drug treatments. (D) Cell proliferation assay of colorectal cancer cells treated with TAS-102 and the lysosomal inhibitor Bafilomycin A1. (E) Schematic workflow of high-throughput Lysotracker-based ORFeome overexpression screen with FACS sorting. (F) Diagram of TGF-β signaling pathway highlighting lysosomal activator and repressor ORFs identified in the Lysotracker-FACs screening. (G) Kaplan–Meier survival curves depicting overall survival for patients with GHSR expression across multiple cancer types in the TCGA PanCancer Atlas (left), accompanied by a stacked bar graph showing the number of cancer patients stratification by expression levels (right). (H) Kaplan–Meier survival curves for CCDC86 expression across multiple cancer types in TCGA PanCancer Atlas (left), accompanied by a stacked bar graph showing the number of cancer patients stratification by expression levels (right). Data in bar graphs are shown as mean ± Standard Deviation (SD). Statistical significance was assessed using a two-tailed Student's t-test with Bonferroni multiple comparison correction ($p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$). Survival curve comparisons were performed using the log-rank test.

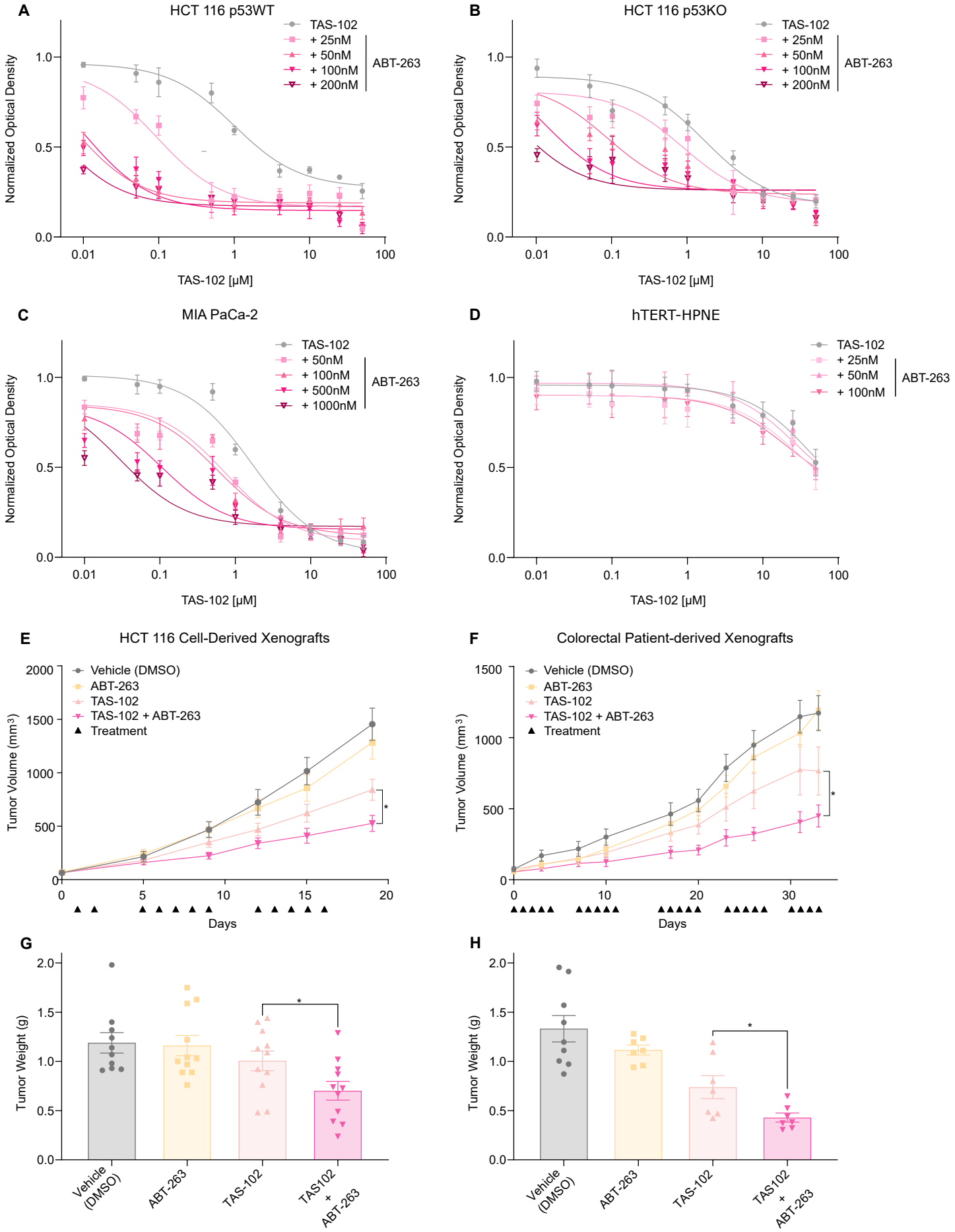
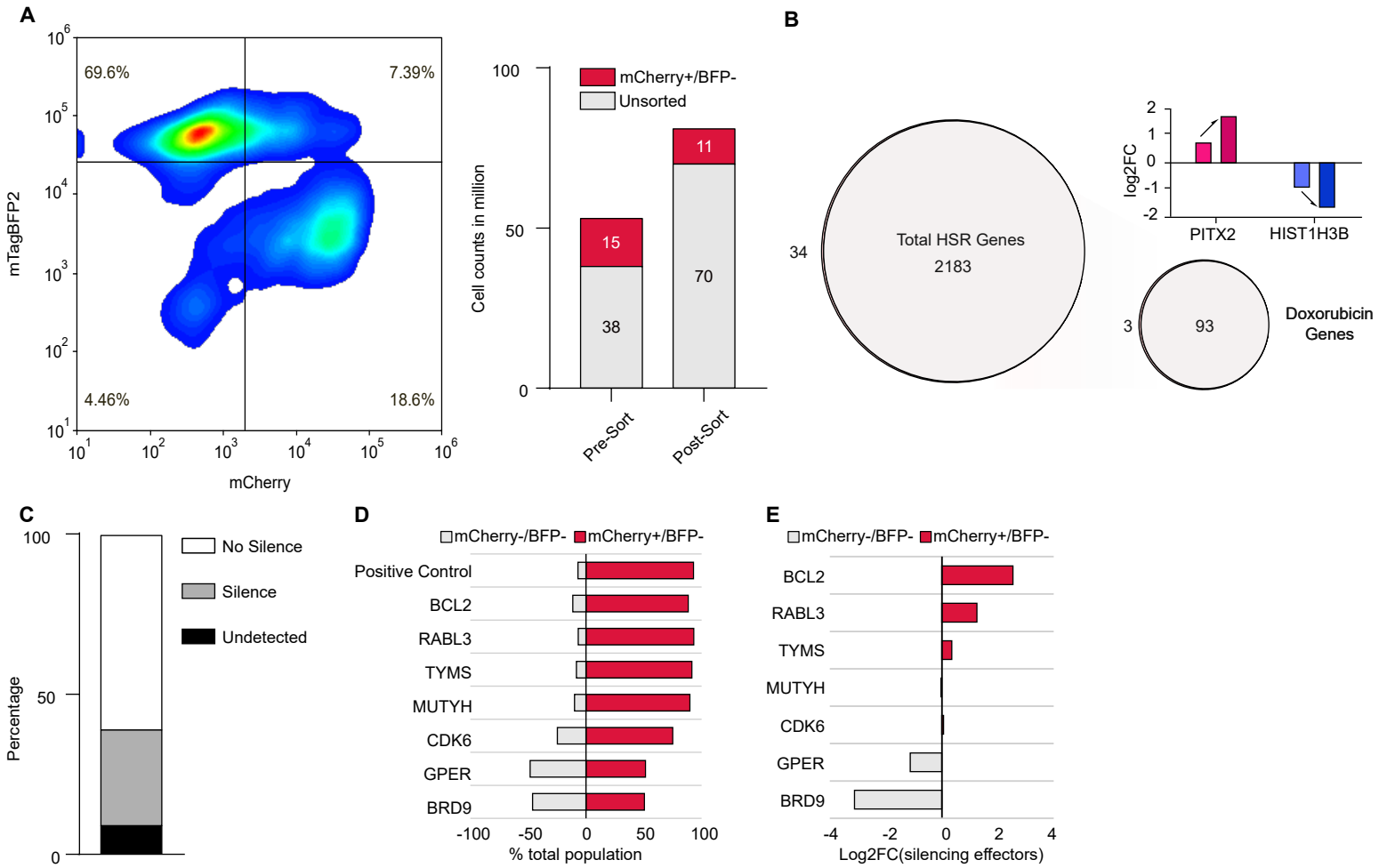
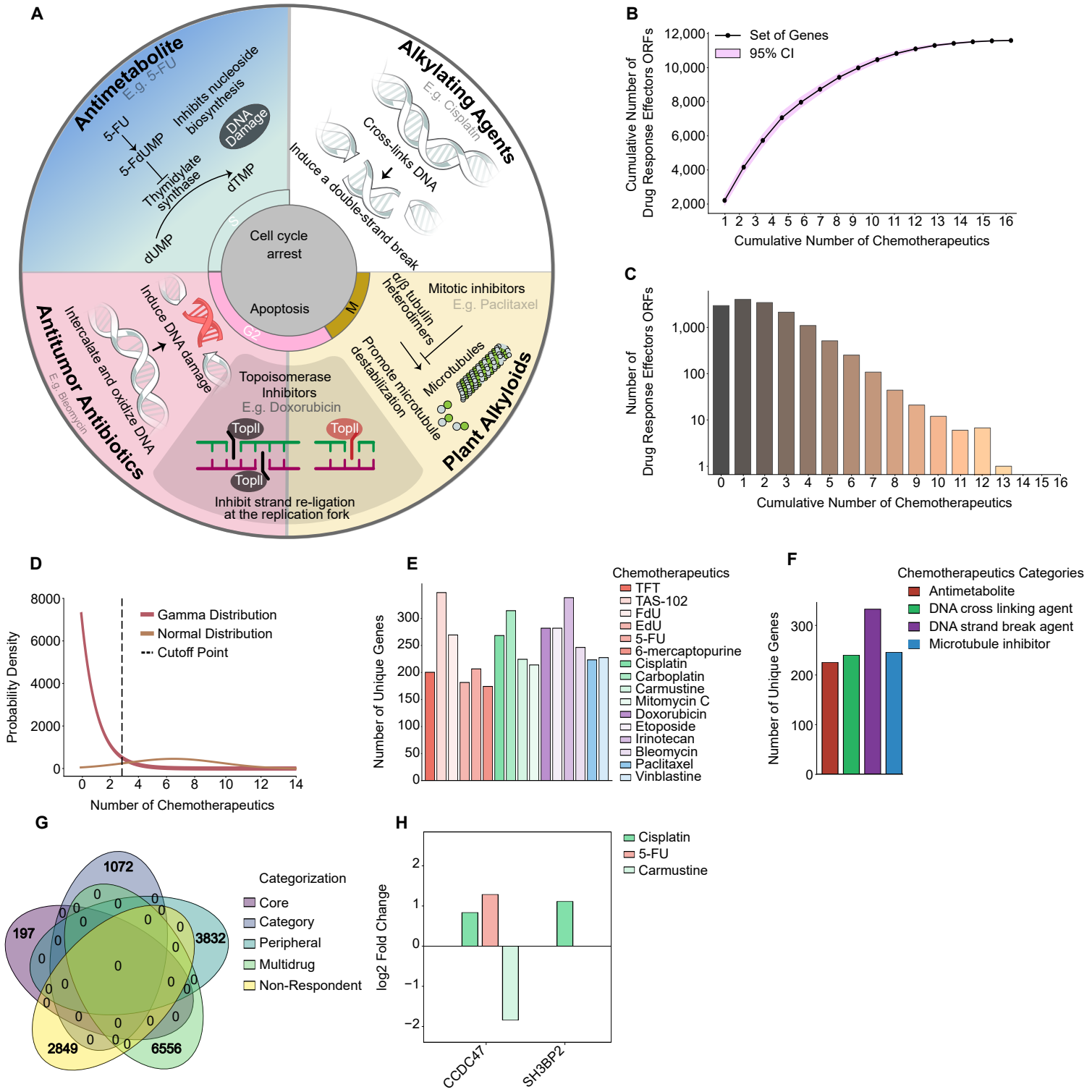


Figure 5. BCL2 Inhibition Synergizes with TAS-102 to Suppress Cancer Growth *In Vitro* and *In Vivo*
(A–C) Synergistic growth inhibition of TAS-102 and ABT-263 in (A) HCT 116 p53 wild-type (WT) cells, (B) p53 knockout (KO) cells, and (C) MIA PaCa-2 pancreatic cancer cells; all pairwise comparisons reached **** $p \leq 0.0001$. (D) Cytotoxicity assay in hTERT-HPNE. (E–F) Tumor volume (mm^3) progression in (E) HCT 116 cell-derived xenografts (CDX) and (F) colorectal patient-derived xenografts (PDX), after treatment with DMSO, TAS-102, ABT-263, and TAS-102 + ABT-263 combination therapy. * $p \leq 0.05$. (G–H) Tumor weight in (G) CDX and (H) PDX models, showing significant reduction under combination treatment compared to TAS-102 alone. Data in panels A–D are presented as mean ± standard deviation (SD), whereas panels E–H show mean ± standard error of the mean (SEM). Statistical analysis for panels A–F was performed using two-way ANOVA with Tukey's multiple comparisons; $p \leq 0.05$, and panels G–H were analyzed using Welch's t test; *$p \leq 0.05$.

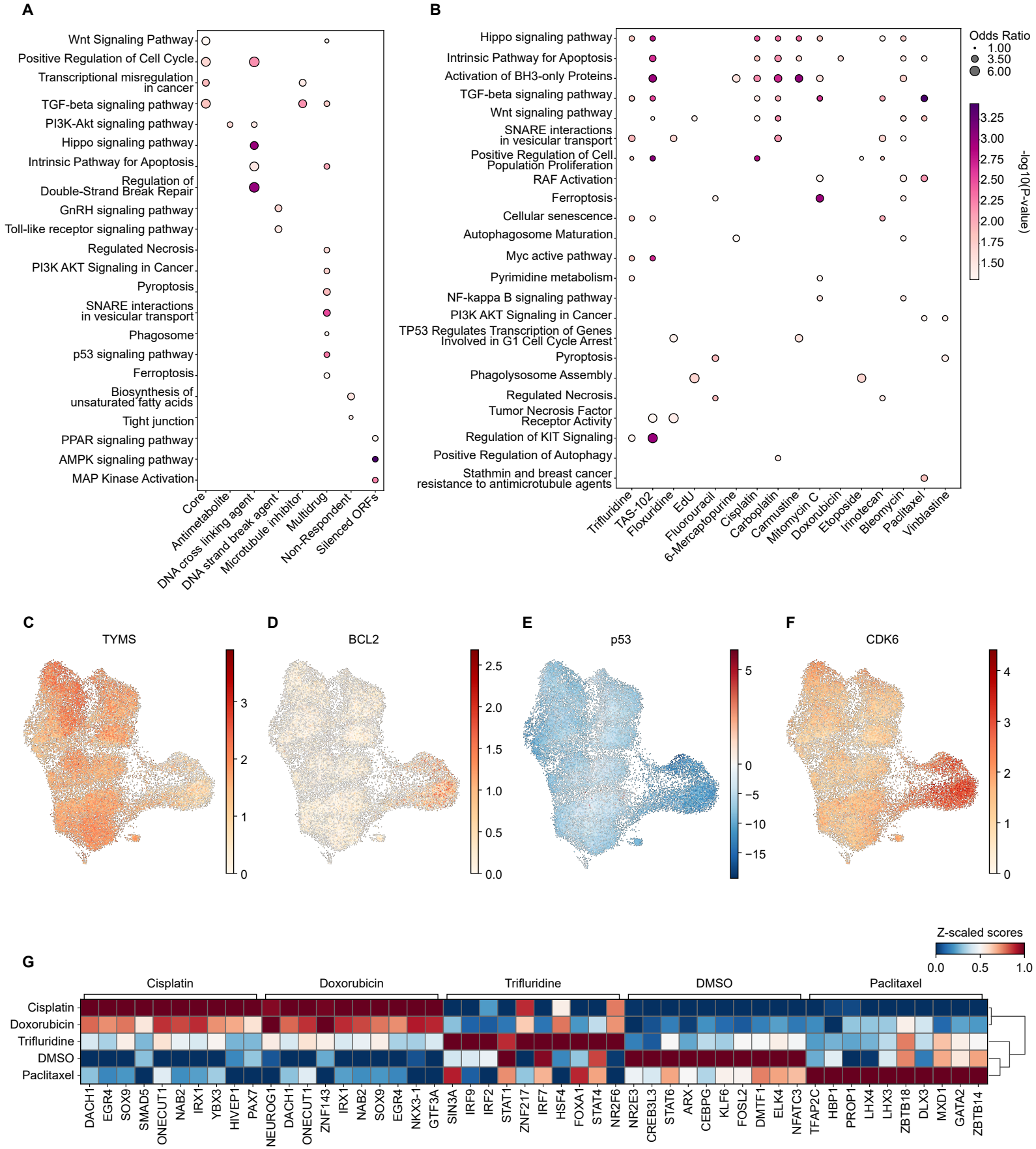


Supplementary Figure 1. Validation and Characterization of the BOGO System and Screen Results
(A) Efficiency of ORF integration in HeLa cells measured by mCherry and mTagBFP2 (BFP) fluorescence via FACS (left). Stacked bar graph showing number of sorted and unsorted cells via mCherry fluorescence (right) (B) Venn diagrams illustrating detection of selected ORFs in total HSR genes (left) and doxorubicin-specific drivers (right). The bar graph shows increased log2 fold change in PITX2 and HIST1H3B. Arrows indicate an increase or decrease from drug response screening to HSR screening (right top). (C) Proportion of ORFs that underwent gene silencing as measured by the percentage of mCherry-negative over mCherry-positive. Positive control indicates mCherry positive cells (D) Bar plot showing percentage of cell populations affected by ORF silencing. (E) Bar plot showing silencing ratio for the selected 7 ORFs in drug response screening.

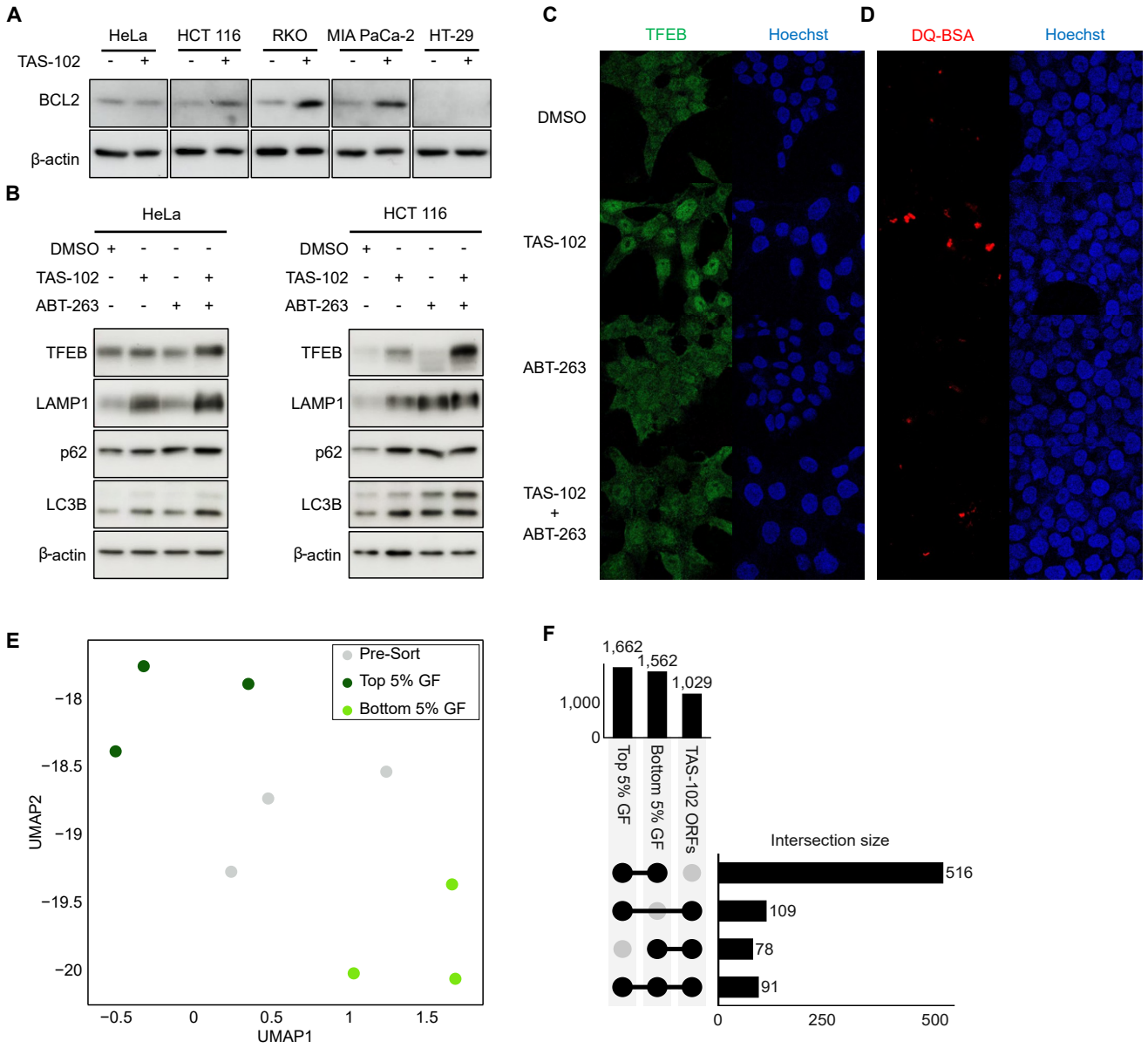


Supplementary Figure 2. Classification of Drugs used for Screening, Categorization of ORFs with respect to drug response effect, and ORF Candidates as Resistance Biomarker

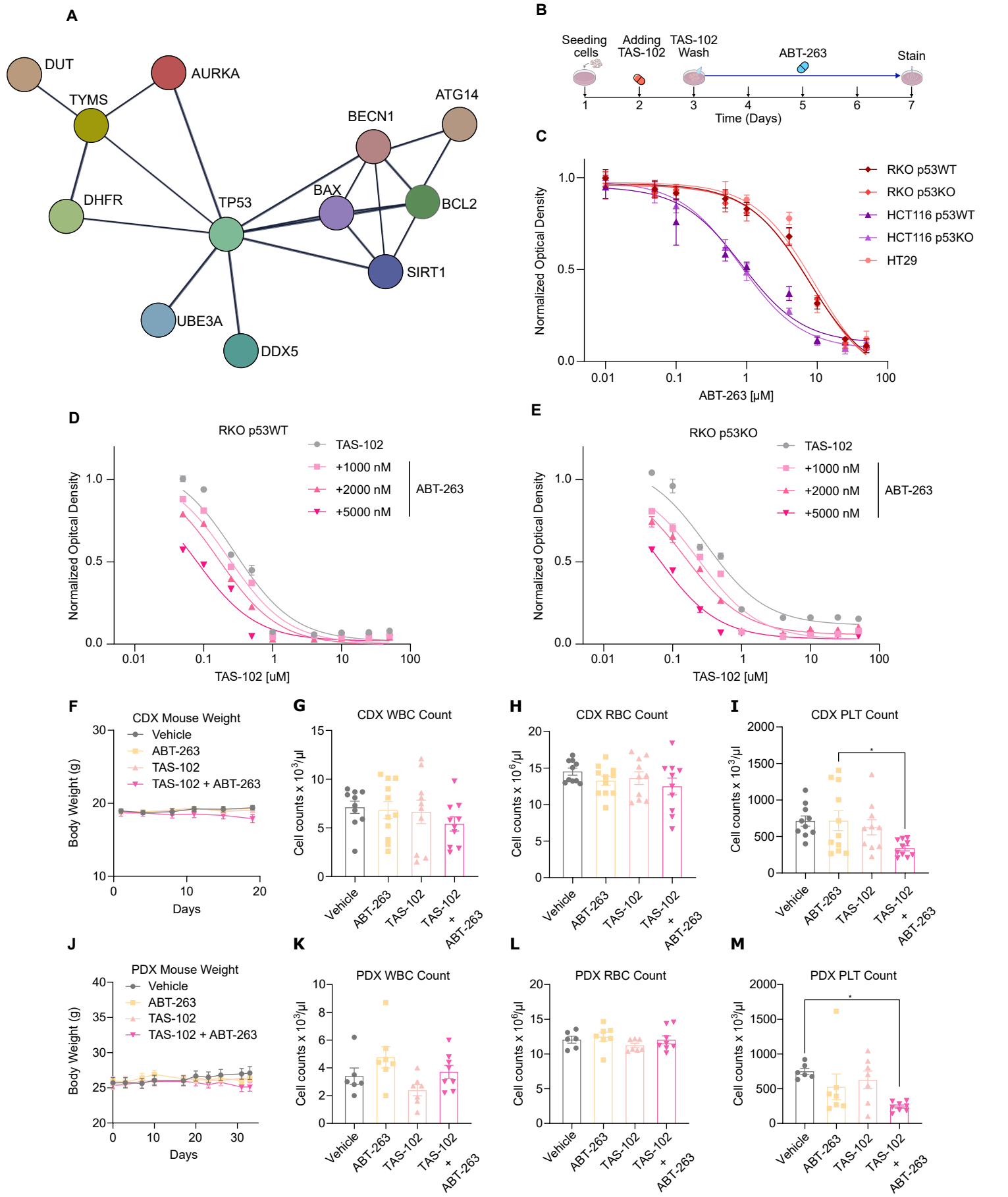
(A) Overview of mechanisms and cellular targets of chemotherapeutic agents used in drug response screening. (B) Cumulative number of unique drug response effector ORFs with respect to the number of drugs. (C) Histogram showing the number of chemotherapeutic drugs responded to effector ORFs. (D) Mixture distribution analysis to define a cutoff for multi-drug responders. A dotted line was drawn at 2.64. (E) Number of unique ORFs for an individual chemotherapeutic. (F) Unique Categorical ORF linked to Chemotherapeutics Categories. (G) Venn diagram illustrating distinctness of gene subsets. Numbers indicate the number of ORFs in five categories. 0 indicates there is no overlap between subsets. (H) Drug response effector expression of SH3BP2 and CCDC47 in four chemotherapeutics from drug response screening.



Supplementary Figure 3. Drug Response Screening with Enrichment Analysis and Single-Cell Validation
(A–B) Gene Set Enrichment Analysis (GSEA) bubble plots of (A) core, categorical (antimetabolite, DNA-cross linking agent, DNA strand break agent and microtubule inhibitor), multidrug, non-respondent, and silencing ORFs; (B) individual chemotherapies from peripheral ORFs. (C–F) UMAP projections showing single-cell ORF expression of key resistance markers: (C) TYMS, (D) BCL2, (E) TP53, and (F) CDK6. (G) Heatmap of normalized z-scores displaying ORF expression across one DMSO and four chemotherapeutic conditions, highlighting the top five resistance-associated ORFs per treatment cluster.



Supplementary Figure 4. Molecular and Cellular Validation of BCL2 Inhibition and Autophagy-Lysosomal Dynamics
(A) Western blot analysis of BCL2 expression in cancer cell lines post-TAS-102 treatment (0.6 μM for RKO, 0.8 μM for HeLa, 2 μM for MIA PaCa-2, HT-29, HCT 116). (B) Western blot analysis of lysosomal markers (LAMP1, TFEB) and autophagy markers (LC3B, p62) in HeLa and HCT 116 cells under treatment. (C) Separate fluorescent images of TFEB (green) immunofluorescence and Hoechst nuclear staining across treatments (2 μM TAS-102, or 0.5 μM ABT-263, 2 μM TAS-102 + 0.5 μM ABT-263). (D) Separate fluorescent images of DQ-BSA (red) lysosomal proteolysis and Hoechst nuclear staining across treatment (2 μM TAS-102, or 0.5 μM ABT-263, 2 μM TAS-102 + 0.5 μM ABT-263). (E) UMAP plot showing clustering of pre-sort and Lysotracker-sorted; top and bottom 5% green fluorescence (GF) from HeLa ORFeome cell. Each dot indicates biological replicates (N=3). (F) Upset plot showing overlap between TAS-102 drug response effector ORF and lysosomal modulator ORFs from Lysotracker staining.



Supplementary Figure 5. Synergistic Efficacy and Safety Profile of TAS-102 and ABT-263 Combination
(A) STRING PPI network showing indirect p53-mediated link between BCL2 and TYMS, medium confidence (<0.5; protein interaction score from STRING), <5 interactors in first and second shells. (B) Dose-response curves for ABT-263 in colorectal cancer lines HCT 116, RKO (p53 WT and KO), and HT-29 using varying concentrations 0, 0.01, 0.05, 0.1, 0.5, 1, 4, 10, 25, and 50 μM (C–D) Synergistic cytotoxicity of TAS-102 and ABT-263 in (C) RKO p53 WT (D) and p53 KO cells. (E–F) Mouse body weight data from (E) CDX and (F) PDX models, showing stable weight across treatments. (G–K) Complete blood count (CBC) parameters, including Red blood cell, white blood cell, Hemoglobin, Monocyte, and Granulocyte, with no significant differences across treatments. (L) Platelet counts showing a significant reduction of * $p < 0.05$ in combination therapy compared to ABT-263 alone. Data in panels C-D are presented as mean ± standard deviation (SD), whereas panels B, E–L show mean ± standard error of the mean (SEM). Statistical analysis was performed using one-way or two-way ANOVA with Tukey's multiple comparisons.