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

Kyeong Beom Jo[1,2,3,*], Mohammed M. Alruwaili[3,4,*], Da-Eun Kim[5,*], Yongjun Koh[6,*], Hyeyeon Kim[2,*], Kwontae You[7], Ji-sun Kim[2], Saba Sane[1,2], Yanqi Guo[3], Jacob P. Wright[3], Maricris N. Naranjo[3], Atina G. Cote[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[7], Jong-Eun Park[6,#], HyunKyung Kim[5,#], Andrei V. Bakin[3,#], Dae-Kyum Kim[1,2,3,16,17,18,#].

[1]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 Medical Laboratory Technology, College of Applied Medical Sciences, Northern Border University, Arar City, Saudi Arabia

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

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

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

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

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

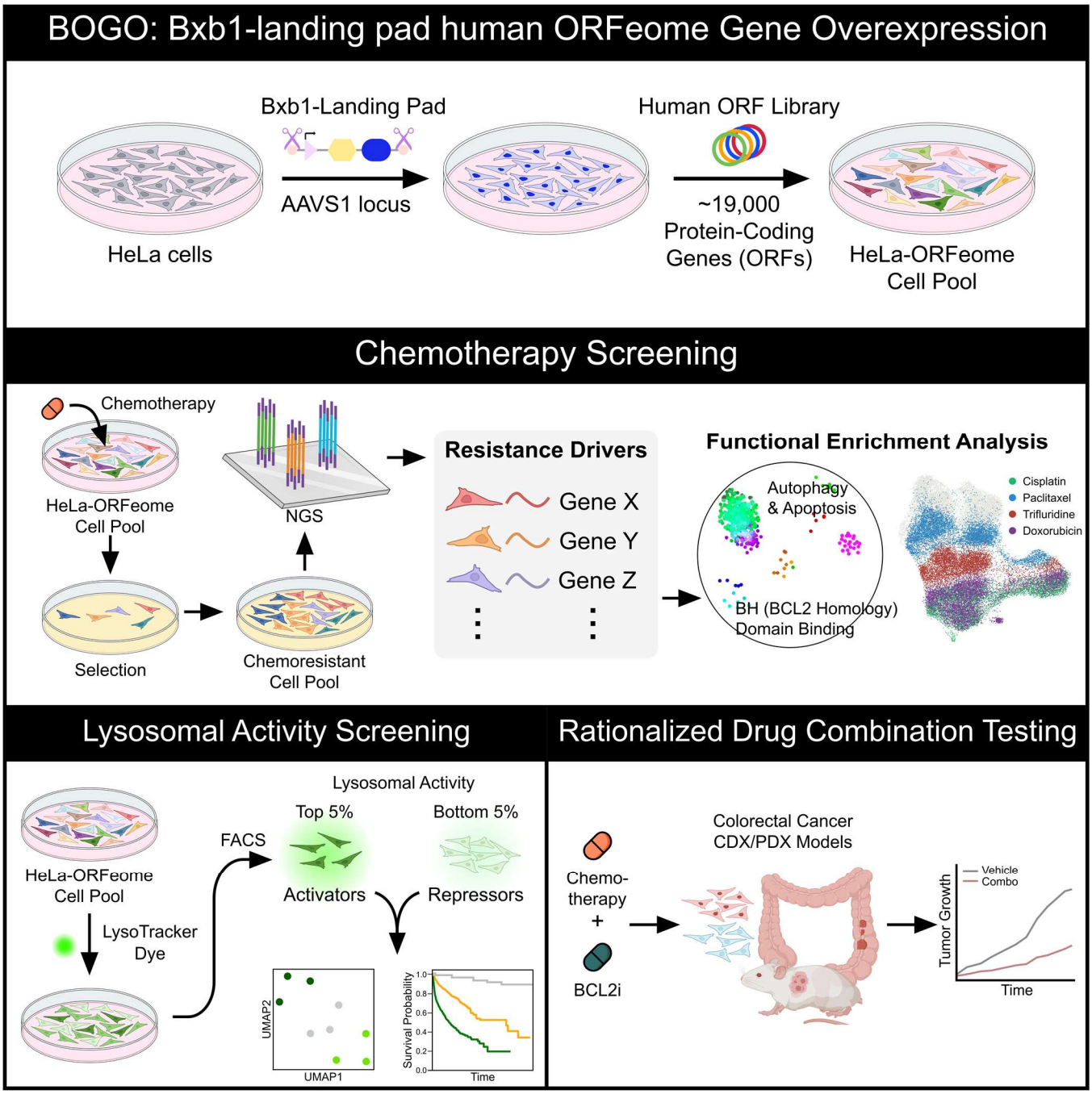
[14]Department of Life Sciences, School of Life Sciences and Biotechnology, Korea University, Seoul, Korea

[15]Department of Pathology, School of Medicine, Case Western Reserve 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

37 **Graphical Abstract**



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 open 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 scrFlupeenings[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 frames (ORFs) 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

93 complementary partner to chemotherapy. While overexpression screenings using CRISPRa[15] or
94 lentivirus[16] are useful, they have limitations in capturing the full complexity of tumor heterogeneity. These
95 include heterogeneous gene expression due to random genomic integration[17], limited proteome-wide
96 coverage, defined here as coverage across protein-coding genes, caused by chromatin accessibility
97 dynamics[18], and false-positive driver candidates identified from activation of multiple genes per cell[19].
98 Moreover, since 90% of FDA-approved drugs that can be used for combinatorial therapy are known to
99 target proteins, functional genomics on protein-coding genes[20] could provide more immediate clinical
100 insights for drug repurposing[21]. Therefore, there is a need for an overexpression screening platform that
101 enables a "proteome-wide", "single gene per cell" manner from a defined genomic locus to find the main
102 "drivers" for chemoresistance.

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

121 # Results

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

124 BOGO facilitates stable, single-ORF-per-cell overexpression through precise integration of a Bxb1
125 landing pad donor vector into the AAVS1 locus via homologous recombination[26] (**Figure 1B**). Gateway
126 cloning was used to ensure accurate bulk overexpression of protein-coding genes, with each ORF
127 expressed under the EF1α promoter. The resulting cell pool encompassed ~19,000 protein-coding genes,
128 derived from hORFeome 9.1[22,23] and additional clones (ORFeome Delta) provided by collaborators. HeLa
129 cells engineered with the Bxb1 landing pad system were transfected with the hORFeome library (**Figure**
130 **1B**). The Bxb1 integrase enabled precise, single-ORF insertion at the AAVS1 site, minimizing variability
131 and ensuring uniform gene expression. Cells were sorted based on mCherry reporter signal and
132 expanded to form the experimental pool, representing the full hORFeome landscape. fluorescence-
133 activated cell sorting (FACS) analysis revealed that ~20% of cells were mCherry-positive/mTagBFP2-
134 negative (**Supplementary Figure 1A, left panel**), achieving 150× coverage with over 3 million sorted
135 cells, indicating robust representation of the ORFeome (**Supplementary Figure 1A, right panel**).
136 Extensive sequencing of the initial cell pool confirmed that ORF sequences were concordant with their
137 corresponding coding DNA sequences (CDS), and approximately 90% of ORFs were successfully
138 integrated (**Figure 1C**). To assess reproducibility, we tested five independent batches in large-scale drug
139 screens. Results showed minimal batch effects, with high correlations between technical replicates
140 across treatments (**Figure 1D**), confirming the stability and reliability of ORFeome representation after
141 freeze-thaw.

142 Leveraging the established HeLa ORFeome cell pool, we conducted two initial functional screens
143 to investigate: i) drug response, and ii) ORF silencing, both of which are critical aspects of cancer biology
144 and gene expression regulation. To evaluate the impact of individual ORFs on drug response, we
145 quantified the relative abundance of ORF-expressing cell populations across three conditions: i) baseline
146 (untreated library cells), ii) DMSO-treated cells (vehicle control), and iii) chemotherapy-treated cells
147 (experimental condition). Amplicons derived from genomic DNA were subjected to next-generation
148 sequencing (NGS) to measure ORF abundance. The $\log_2$ fold change ($\log_2$FC) between drug-treated
149 and vehicle control conditions was used to assess differential representation. ORFs with $\log_2$FC > 0.5 or
150 < -0.5 and $p < 0.05$ were defined as drug response effectors (**Supplementary Table 1**). We classified
151 these enriched ORFs as drug resistance drivers, while depleted ORFs were designated as drug
152 sensitization drivers.

153 To verify the reproducibility and functional relevance of our perturbation system, we performed
154 individual overexpression experiments for seven selected ORFs, comparing their effects to mCherry
155 fluorescence control cells. These validations confirmed the phenotypic outcomes observed in the BOGO
156 drug response screen. For example, overexpression of BCL2 resulted in increased cell abundance under
157 Trifluridine (TFT) treatment, consistent with its classification as a drug resistance driver in our screen
158 (**Figure 1E**). Conversely, overexpression of CDK6, identified as a drug sensitization driver, led to a
159 marked reduction in cell population following TFT treatment (**Figure 1E**), aligning with previous findings[27]
160 that CDK6 overexpression sensitizes cells to DNA damage. This strong concordance between
161 phenotypic changes in single-ORF overexpression lines and proteome-wide perturbation profiles
162 provides direct evidence for the functional validity and reproducibility of the BOGO screening
163 methodology.

164 To demonstrate the complementary value of the BOGO screening platform relative to
165 conventional transcriptomic approaches, we performed bulk mRNA sequencing on wild-type HeLa cells
166 treated with doxorubicin, a widely used chemotherapeutic agent with broad clinical applications[28].
167 Transcriptional responses were profiled at two time points: 120 hours, corresponding to the duration of
168 our ORFeome screen, and 24 hours, to capture acute gene expression changes. We compared the sets
169 of differentially expressed genes (DEGs) identified in these mRNA-seq datasets (**Supplementary Table**
170 **2**) with the doxorubicin response effectors uncovered through our drug response screen. This analysis
171 revealed both overlapping and distinct gene sets (**Figure 1F**). Notably, the BOGO screen identified
172 approximately five times more genes than bulk mRNA-seq at the 120-hour time point, with only 2%
173 overlap. At 24 hours, the BOGO screen detected ten times more genes, yet the overlap with DEGs was
174 even lower, at 0.5%. These differences likely stem from the inherent limitations of bulk mRNA sequencing,
175 which captures endogenous gene expression subject to dynamic regulatory mechanisms. In contrast,
176 the BOGO platform utilizes constitutive ORF overexpression, enabling the identification of functional drug
177 resistance effectors that may be masked or underrepresented in natural transcriptional profiles. Thus,
178 overexpression-based screening offers a powerful and complementary approach to conventional
179 transcriptomics, revealing biologically and clinically relevant genes that may be overlooked by standard
180 methods.

181 To validate and further characterize the hits identified in our drug response screen, we performed
182 high-saturation retesting (HSR), a targeted approach involving smaller, selectively curated subsets of the
183 ORFeome library analyzed at increased coverage to ensure reproducibility and robustness.
184 Approximately 10% of the original ORF pool, comprising a mix of proliferation- and drug response-
185 associated ORFs, was retested at >1,000-fold coverage. We compared the results of the initial drug

186 response screen with those from the HSR for doxorubicin (**Supplementary Table 3**). First, we confirmed
187 high representation of the selected ORFs in the HSR dataset in which 2,184 ORFs (98.5%) from the
188 cherry-picked pool were successfully detected (**Supplementary Figure 1E, left Venn diagram**). Among
189 the doxorubicin effector ORFs, 93 (96.9%) were recovered in the HSR, with only 3 ORFs (3.1%) not
190 detected (**Supplementary Figure 1E, right Venn diagram**), confirming the fidelity and coverage of the
191 retesting process. Importantly, the HSR validated the directional trends observed in the initial screen
192 (**Figure 1G**). ORFs previously identified as drug sensitization drivers, such as HIT1H3B, exhibited
193 negative average fold changes in the HSR, consistent with enhanced drug sensitivity. Conversely, drug
194 resistance drivers, such as PITX2, showed positive average fold changes, confirming their resistance
195 phenotype (**Supplementary Figure 1B**). ORFs classified as non-respondents consistently displayed
196 near-zero fold changes, reinforcing their lack of drug response. These results demonstrate the statistical
197 robustness and reproducibility of the BOGO platform, supporting its reliability in identifying biologically
198 and clinically relevant drug response effectors.

199 In parallel, we addressed the challenge of ORF silencing, an inherent limitation of overexpression
200 systems. To monitor silencing, an IRES-mCherry cassette was inserted downstream of each ORF,
201 allowing discrimination between active and silenced ORF expression. By comparing ORF abundance in
202 mCherry-positive versus mCherry-negative cell populations, we quantified the proportion of active
203 expression. ORFs with $\log_2 FC > 1$ or < -1 and $p < 0.05$ were defined as silencing effectors. Positive
204 $\log_2 FC$ values indicated non-silencing ORFs, while negative values indicated silencing ORFs. Analysis
205 revealed that over 60% of ORFs were retained as non-silencing, while fewer than 30% exhibited
206 measurable silencing effects (**Supplementary Figure 1C**). Furthermore, five selected genes maintained
207 stable expression over five days of culture, with no observable silencing, corroborating their classification
208 as non-silencing ORFs in our screen. In contrast, GPER and BRD9 exhibited silencing effects
209 (**Supplementary Figure 1D**), consistent with their silencing profiles in the drug response screen
210 (**Supplementary Figure 1E**). Notably, GPER, a membrane protein known to induce cellular toxicity via
211 inclusion body formation, has been previously reported to suppress cell growth through this mechanism[29],
212 explaining its lack of detectable representation. These results confirm the stability and reliability of gene
213 expression within the BOGO system and demonstrate its utility in identifying functional drug response
214 effectors and assessing gene silencing tendencies.

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

217 Our drug response screen comprises 16 chemotherapeutic agents, encompassing a broad range of
218 cytotoxic mechanisms. These chemotherapeutics were grouped into four distinct functional categories

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

227 To assess the landscape of drug-response-modulating ORFs (dORFs), we first examined their
228 cumulative distribution across all tested chemotherapeutic agents. The number of unique dORFs
229 increased progressively with the inclusion of additional drugs (**Supplementary Figure 2B**), a trend that
230 significantly exceeded expectations from random permutation, thereby validating the sensitivity and
231 breadth of our screening platform in capturing gene-drug interactions. A histogram of drug response
232 effects revealed a decline in the number of dORFs as the number of drugs increased, with only a small
233 subset of dORFs exerting broad regulatory effects across up to 13 agents (**Supplementary Figure 2C**).
234 Based on this distribution, dORFs affecting responses to more than 7 drugs, approximately half of the
235 maximum observed, were designated as Core ORFs. To define thresholds for broader versus singular
236 drug responses, we modeled the distribution of drug-response counts per ORF using a mixture of gamma
237 and normal distributions. The intersection corresponded to 2.64 drugs (**Supplementary Figure 2D**),
238 which was used to define Multidrug ORFs as those influencing responses to between 2 and 7 drugs.
239 Peripheral ORFs were defined as those modulating responses to a single chemotherapeutic agent
240 (**Supplementary Figure 2E**), reflecting the specificity of certain gene-drug interactions. Within the
241 Categorical ORFs, defined as ORFs that selectively affect responses to drugs with shared mechanisms
242 of action, dORFs for DNA strand break inducers were most prevalent, although other drug categories
243 also exhibited distinct dORF profiles (**Supplementary Figure 2F**). Non-respondent ORFs are those that
244 showed no significant impact on drug responses across all tested agents. The mutual exclusivity of these
245 categories was confirmed via Venn diagram analysis (**Supplementary Figure 2G**).

246 We next investigated whether the defined ORF categories exhibit distinct intrinsic properties that
247 may underline their differential drug response profiles. Among these, Core ORFs demonstrated the most
248 pronounced divergence across multiple molecular and functional features, consistent with their broad
249 association with chemoresistant phenotypes. Specifically, Core ORFs exhibited significantly elevated z-
250 scores for GC content, predicted protein disorder, Gene Ontology (GO) term enrichment, publication
251 frequency (PubCount), and associations with Mendelian diseases (**Figure 2B**). Conversely, features

252 such as tissue-specific upregulation (Tissue Count), evolutionary age, expression breadth (number of
253 tissues in which the gene is expressed), expression abundance, and the number of associated protein
254 complexes (as annotated in the CORUM database) were markedly reduced in Core ORFs (**Figure 2B**).
255 Notably, Core ORFs also tended to be shorter in length compared to other categories, whereas
256 Categorical ORFs exhibited the longest sequences. These observations suggest that ORFs with low
257 baseline expression may be more susceptible to phenotypic modulation upon artificial overexpression,
258 potentially accounting for their prominence in our screen. In contrast, genes with high endogenous
259 expression may be less responsive to perturbation, thereby limiting their detectability as drug response
260 effectors. While this bias may influence the representation of Core ORFs, our screening platform remains
261 robust in identifying functionally relevant drivers of drug response and disease. Taken together, these
262 findings indicate that Core ORFs represent a distinct subset of underappreciated regulatory elements
263 that acquire functional significance upon upregulation, differing from other ORF categories in both
264 structural and functional attributes.

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

281 To further assess the utility of the BOGO platform in identifying prognostic clinical markers and
282 predicting drug resistance, we compared the effects of individual ORFs on cell proliferation with cancer
283 prognostic indicators derived from the Human Protein Atlas (HPA) database[20] and The Cancer Genome
284 Atlas (TCGA)[32]. Differentially represented ORFs were identified by comparing their cell populations at

285 baseline and after nine days of competitive growth. ORFs with significant changes ($\log_2$FC > 1 or < -1; p
286 < 0.05) were classified as cell proliferation regulators, with enriched ORFs designated as cell growth
287 upregulators and depleted ORFs as cell growth downregulators. Using z-score rankings, we evaluated
288 the predictive performance of our ORF list in identifying clinically relevant markers. Precision-recall
289 analysis revealed that ranking ORFs in descending order by proliferation score (placing upregulators at
290 the top) yielded high precision for identifying markers associated with unfavorable cancer prognosis
291 (**Figure 2D**). Conversely, ranking ORFs in ascending order (placing downregulators at the top) enriched
292 for markers of favorable prognosis, although precision declined more steeply, reaching a comparable
293 minimum of approximately 40% (**Figure 2E**). To illustrate the clinical significance of these findings in
294 cervical cancer (CESC), we selected representative ORFs from both ends of the prognostic spectrum.
295 Overexpression of Phosphoglucomutase 1 (PGM1), a top-ranked upregulator, was strongly associated
296 with reduced survival probability (**Figure 2F**), consistent with its role in promoting cell proliferation. In
297 contrast, overexpression of SRA Stem-Loop Interacting RNA Binding Protein (SLIRP), a top-ranked
298 downregulator, correlated with improved patient outcomes (**Figure 2G**), suggesting a suppressive effect
299 on cancer proliferation. Together, these precision-recall and survival analyses underscore the clinical
300 relevance of our screening approach and its capacity to identify promising cell proliferation regulators
301 with prognostic value.

302 We hypothesized that drug response effectors would exhibit strong correlations with cell
303 proliferation regulators, based on the premise that robust resistance mechanisms may confer enhanced
304 proliferative capacity in cancer cells. To test this, we compared our drug response screen results with a
305 manually curated drug resistance dataset from TCGA[33], evaluating the ability of our platform to identify
306 known resistance drivers. Precision-recall analysis of our ranked drug response dataset revealed high
307 precision among top-ranked genes for both cisplatin (**Figure 2H**) and 5-fluorouracil (5-FU) resistance
308 (**Figure 2I**), indicating effective enrichment of established resistance mechanisms. For instance,
309 overexpression of Src homology 3 domain-binding protein 2 (SH3BP2), a top cisplatin resistance
310 candidate, was strongly associated with reduced survival in patients with pancreatic adenocarcinoma
311 (PAAD) (**Figure 2J**). Similarly, Coiled-coil domain-containing protein 47 (CCDC47), identified as a 5-FU
312 resistance driver, correlated with poor prognosis in hepatocellular carcinoma (LIHC) (**Figure 2K**). Both
313 genes also demonstrated drug resistance-driving effects in our screening data (**Supplementary Figure**
314 **2H**). This integrative analysis, combining functional screening with clinical datasets, underscores the
315 capability of the BOGO platform to uncover novel drug resistance genes with direct biological and clinical
316 relevance in cancer.

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

319 To investigate how genetic perturbations induced by human ORFeome overexpression influence cellular
320 responses to chemotherapeutic stress, we performed dimensionality reduction analysis on the 16
321 ORFeome drug response profiles. Uniform Manifold Approximation and Projection (UMAP) plots revealed
322 distinct clustering patterns corresponding to individual drug treatments (**Figure 3A**). DMSO controls from
323 five independent batches clustered tightly along the UMAP1 axis, indicating minimal batch effects.
324 Furthermore, biological replicates are consistently clustered by chemotherapeutic agents, underscoring
325 the robustness and reproducibility of the screening data. Notably, drugs with shared mechanisms of
326 action generally clustered in close proximity, demonstrating the ability of dimensionality reduction to
327 capture broad biological responses. For example, mitomycin C clustered separately from other alkylating
328 agents, likely reflecting its unique mechanism as a DNA synthesis inhibitor. Similarly, although TFT and
329 5-FUlysot are both classified as antimetabolites, their divergent clustering patterns suggest distinct
330 resistance profiles and underlying mechanisms of action[34].

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

348 To further investigate the mechanistic relevance of the BCL2-related pathways identified in our
349 drug response similarity network, specifically, the clusters annotated as BH domain binding and

350 Autophagy and Apoptosis (**Figure 3B**), we conducted an in-depth network analysis to evaluate their
351 potential as targets for combination therapy. We performed STRING network analysis[39] on the sub-
352 network comprising ORFs from these two clusters to elucidate protein-protein interactions among the
353 identified genes (**Figure 3C**). The 20 ORFs from the SAFE-enriched clusters were analyzed using
354 STRING to map experimentally supported and predicted interactions (**Supplementary Table 4**). This
355 analysis revealed 12 key genes involved in critical signaling pathways and cellular damage responses,
356 highlighted as diamond-shaped nodes in **Figure 3C**. The right panel of the figure illustrates the
357 interactions between these protein nodes and their associated GO clusters, reflecting strong functional
358 connectivity and biological relevance. Among the enriched pathways were vesicular transport, apoptotic
359 signaling, histone modification, and Notch receptor signaling, all of which are functionally linked to the
360 BCL2-related genes identified in the two clusters. These findings provide a compelling mechanistic
361 rationale for targeting BCL2-associated pathways in chemotherapeutic resistance and underscore the
362 utility of the BOGO platform in uncovering actionable biological networks.

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

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

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

394 Notably, genes such as BCL2, TP53, CDK6, and TYMS (the target of TFT) were associated with
395 all four treatments and showed specific dysregulation within the autophagy and apoptosis cluster
396 (**Supplementary Figures 3B-E**), suggesting that this cluster may represent a population of
397 chemoresistant cells. Differential gene expression analysis (**Supplementary Table 5**) further identified
398 significant downregulation of TRADD, a gene implicated in apoptotic signaling within the same GO
399 cluster[40,41], and upregulation of POLD2, a Wnt pathway gene linked to cell proliferation[42] (**Figure 3D**).
400 Importantly, the regulation of these genes was consistent across both the scRNA-seq data and the drug-
401 ORF network analysis. Additionally, z-score analysis of $\log_2$ fold changes highlighted key resistance-
402 associated genes, including YBX3 (cisplatin), ZNF143 (doxorubicin), ZNF217 (TFT), and GATA2
403 (paclitaxel), alongside several less-characterized candidates with potential roles in drug resistance or
404 sensitization[43–46] (**Supplementary Figure 3F**). This strong concordance across platforms reinforces the
405 ability of the BOGO screen to capture relevant drug response mechanisms at single-cell resolution and
406 positions BCL2 as a promising therapeutic target, particularly in the context of DNA analog drugs such
407 as TFT and TAS-102.

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

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

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

427 To identify physiologically relevant cancer models for evaluating the TAS-102 and BCL2 inhibitor
428 combination therapy, we investigated whether TAS-102 treatment modulates endogenous BCL2 protein
429 expression. The well-established association between BCL2 signaling, autophagy, and chemoresistance
430 in colorectal and pancreatic cancers[49,50] provided a strong rationale for selecting cell lines representative
431 of these malignancies. Western blot analysis revealed a pronounced upregulation of BCL2 following TAS-
432 102 treatment in colorectal cancer cell lines HCT 116 and RKO, and in the pancreatic cancer cell line
433 MIA PaCa-2. A modest increase was observed in HeLa cells, while HT-29 cells showed no detectable
434 BCL2 expression (**Supplementary Figure 4A**). These differential expression patterns informed the
435 selection of HCT 116 as the primary model for subsequent combination therapy experiments, due to its
436 robust BCL2 induction and relevance to colorectal cancer.

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

447 To further investigate the impact of TAS-102 and ABT-263 on lysosomal regulation, we examined
448 TFEB localization using immunofluorescence staining. Treatment with TAS-102 led to increased nuclear

449 localization of TFEB, consistent with activation of lysosomal biogenesis. However, combination treatment
450 with ABT-263 significantly reduced TFEB nuclear localization (**Figure 4A; Supplementary Figure 4C**),
451 suggesting disruption of TFEB-mediated lysosomal activation. This observation aligns with the concurrent
452 elevation of TFEB and LAMP1 protein levels, in the absence of corresponding degradation of p62 or
453 LC3B (**Supplementary Figure 4B**), indicating that while autophagosome maturation may proceed,
454 lysosomal degradation is impaired. These data suggest that although total TFEB levels increase under
455 TAS-102 treatment, nuclear translocation and functional activation of lysosomes are disrupted when
456 ABT-263 is co-administered, pointing to lysosomal modulation as a key mechanism underlying the
457 observed drug response.

458 To further investigate the functional consequences of lysosomal modulation, we assessed
459 lysosomal proteolytic activity using DQ-BSA staining[54], which emits red fluorescence upon proteolysis.
460 In HCT 116 cells, TAS-102 treatment significantly increased red fluorescence intensity, indicating
461 enhanced lysosomal activity. However, concurrent treatment with ABT-263 markedly reduced
462 fluorescence, suggesting impaired proteolysis (**Figure 4B; Supplementary Figure 4D**). Consistent with
463 these findings, lysosomal pH measurements revealed acidification under TAS-102 treatment alone, while
464 combination therapy led to increased pH, indicative of disrupted lysosomal acidification (**Figure 4C**).
465 These results point to a potential resistance mechanism involving lysosomal dysfunction, wherein TAS-
466 102 promotes lysosomal activation, but BCL2 inhibition compromises lysosomal function. To assess the
467 functional relevance of this disruption, we co-treated cells with TAS-102 and Bafilomycin A1, a known
468 lysosomal inhibitor. This combination significantly reduced cell proliferation compared to TAS-102 alone
469 after 48 hours of treatment (**Figure 4D**), supporting the role of lysosomal integrity in modulating
470 chemotherapeutic response.

471 To identify lysosomal modulators contributing to chemotherapeuCtic resistance, we developed a
472 follow-up gene overexpression perturbation screen by integrating the BOGO platform with LysoTracker[55]-
473 based FACS (**Figure 4E**). HeLa cells expressing the ORFeome library from the previous drug response
474 screen were stained with LysoTracker and sorted into populations with high and low lysosomal activity
475 (top and bottom 5% based on green fluorescence intensity). Subsequent next-generation sequencing
476 (NGS) enabled the identification of ORFs associated with lysosomal modulation. ORFs enriched in the
477 top 5% were classified as lysosome activators, while those enriched in the bottom 5% were designated
478 lysosome repressors ($\log_2 FC > 0.5$ or < -0.5; $p < 0.05$). UMAP analysis confirmed consistent clustering
479 across biological triplicates for the top 5%, bottom 5%, and presort samples, supporting the robustness
480 and reproducibility of the BOGO screening platform (**Supplementary Figure 4E**). Notably, BCL2,
481 previously identified as a TAS-102-induced resistance driver, also exhibited strong lysosomal activation

482 in this screen (**Supplementary Table 6**). Pathway enrichment analysis of lysosomal modulators revealed
483 significant involvement of TGFβ signaling (**Figure 4F**), implicating its role in autophagy and lysosomal
484 regulation. To further refine the analysis, we intersected lysosomal modulators with TAS-102-specific
485 resistance and sensitization-driving ORFs, identifying 109 ORFs as lysosomal activators, 78 as
486 lysosomal repressors, and 91 overlapping with both lysosomal modulation and TAS-102 response
487 profiles (**Supplementary Figure 4F**). These findings highlight a subset of lysosome-linked resistance
488 drivers specific to TAS-102, offering mechanistic insight and potential therapeutic targets.

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

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

503 To assess the clinical applicability of TAS-102 in combination with ABT-263, we investigated their
504 cytotoxic effects in both p53-dependent and cancer-specific contexts. STRING network analysis revealed
505 an indirect interaction between BCL2 and TYMS, the primary target of TAS-102, mediated through p53
506 (**Supplementary Figure 5A**), suggesting that p53 may serve as a critical hub linking BCL2 and TYMS
507 pathways. To test this hypothesis, we employed p53 knockout models and compared their response to
508 combination therapy with wild-type cells (**Supplementary Figure 5B)**. Dosages of ABT-263 were
509 selected based on IC_{50} values derived from dose-response curves in colorectal cancer cell lines
510 (**Supplementary Figure 5C**). Cytotoxicity assays revealed a substantial reduction in cell viability in p53
511 wild-type HCT 116 cells following combination treatment with TAS-102 and ABT-263, compared to TAS-
512 102 alone (**Figure 5A**). Notably, the median lethal dose (LD_{50}) of TAS-102 decreased by up to 1,000-
513 fold when combined with 50 nM ABT-263 (**Figure 5A**). This sensitizing effect was significantly attenuated

514 in p53-knockout HCT 116 cells (**Figure 5B**), suggesting a p53-dependent mechanism underlying the
515 interaction between BCL2 signaling and TAS-102 responsiveness.

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

526 On the other hand, RKO cells, which share microsatellite instability-high (MSI-H) status with HCT
527 116, exhibited no synergistic effect in either p53 wild-type or knockout conditions (**Supplementary**
528 **Figures 5D-E**). While morphological differences between epithelial HCT 116 and mesenchymal-like RKO
529 cells may influence drug responses, somatic mutations such as KRAS or BRAF could modulate sensitivity.
530 HCT 116 and MIA PaCa-2, both KRAS-mutant, are sensitive to the ABT-263 combination therapy[57,58]
531 due to KRAS-driven upregulation of BCL-XL, the target of ABT-263, and greater reliance on lysosomal
532 activity[59], whereas RKO cells could resist apoptosis through MCL-1 upregulation downstream of BRAF-
533 V600E, reducing sensitivity to ABT-263[60].

534 To validate the in vitro findings in a more physiologically relevant context, we replicated the TAS-
535 102 and ABT-263 combination treatment in both cell-derived xenograft (CDX) and patient-derived
536 xenograft (PDX) models using colorectal cancer systems. In the CDX model, combination therapy
537 consistently recapitulated the enhanced anti-tumor effects observed in vitro, resulting in significantly
538 reduced tumor growth compared to either monotherapy (**Figure 5E**). Similarly, the PDX model
539 demonstrated slowed tumor volume progression under combination treatment relative to TAS-102 alone,
540 confirming that the synergistic efficacy of the therapy is reproducible *in vivo* (**Figure 5F**). In addition, both
541 CDX and PDX models showed a significant reduction in tumor weight under combination treatment
542 compared to TAS-102 alone. (**Figure 5G-H**). Importantly, body weight remained stable, and white and
543 red blood cell counts showed no significant changes In both CDX and PDX models (**Supplementary**
544 **Figures 5F-H and 5J-L**); however, platelet levels decreased in both models under the combination
545 treatment (**Supplementary Figures 5I and 5M**), consistent with previously reported adverse effects of
546 TAS-102-induced thrombocytopenia and ABT-263-associated platelet suppression[61,62]. These results

547 confirm that the synergistic interaction between TAS-102 and BCL2 inhibition is both effective and
548 reproducible in vivo, reinforcing the translational potential of this combination strategy to overcome TAS-
549 102 chemoresistance and enhance chemosensitization in colorectal cancer.

550 ## Discussion

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

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

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

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

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

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

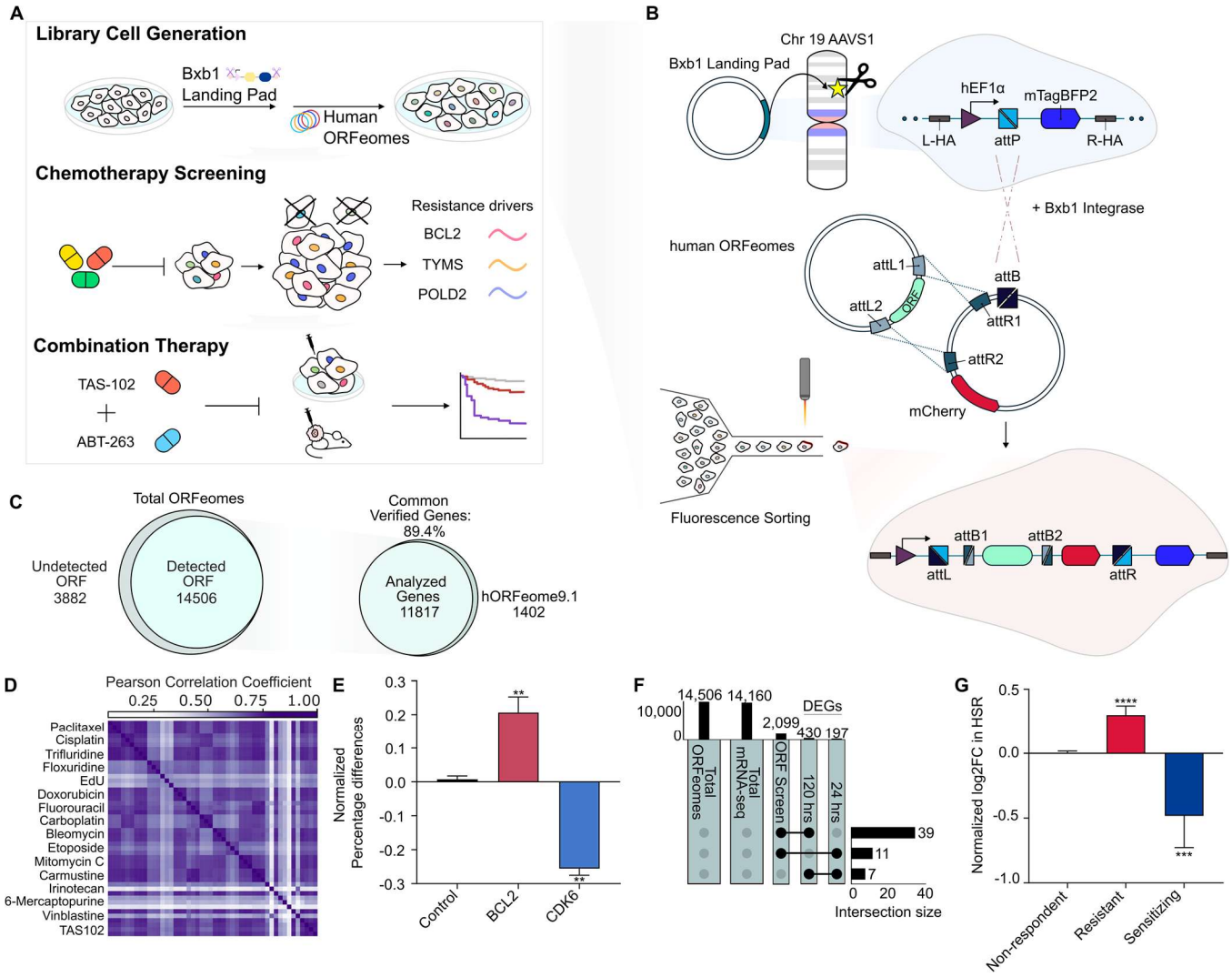


Figure 1. BOGO Enables Functional Proteome-Wide Screening

(**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 overlaps 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 (Log_2 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

612 ± Standard Error of the Mean (SEM). Statistical analysis by one or two-way ANOVA with Tukey's multiple
613 comparisons; ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

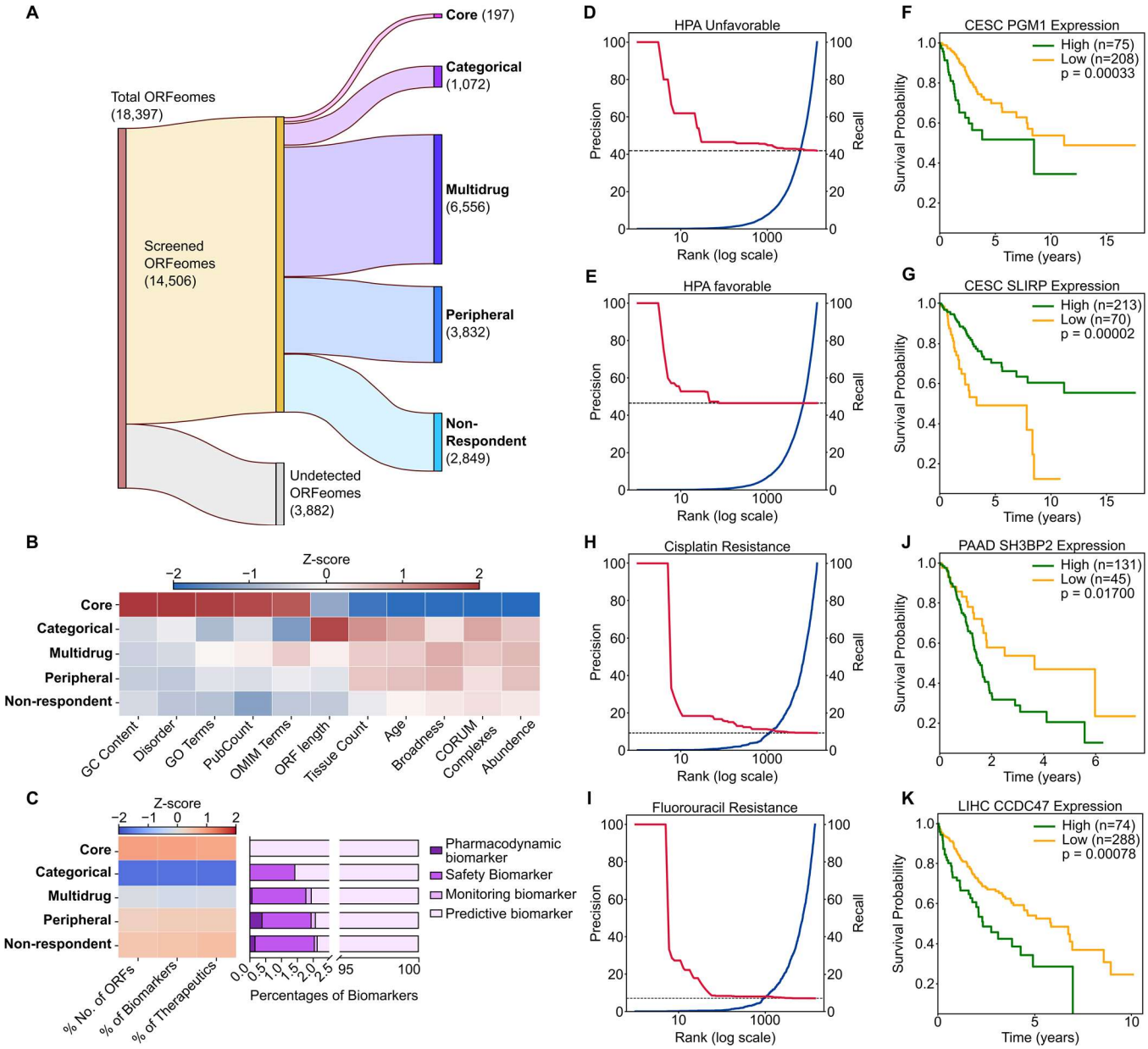


Figure 2. Biological and Clinical Relevance of BOGO Screening for Cell Proliferation and Drug Resistance

(**A**) Sankey diagram categorizing ORF numbers into drug response groups; undetected ORFs defined by median $\log_2$TMM < -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**)

623 Precision-recall curves evaluating (**H**) cisplatin and (**I**) 5-FU resistance drivers from the ORFeome screen.
624 (**J-K**) Kaplan-Meier survival curves for (**J**) SH3BP2 (pancreatic cancer) and (**K)** CCDC47 (liver cancer)
625 that are associated with reduced survival. Survival curve comparisons were performed using the log-rank
626 test.

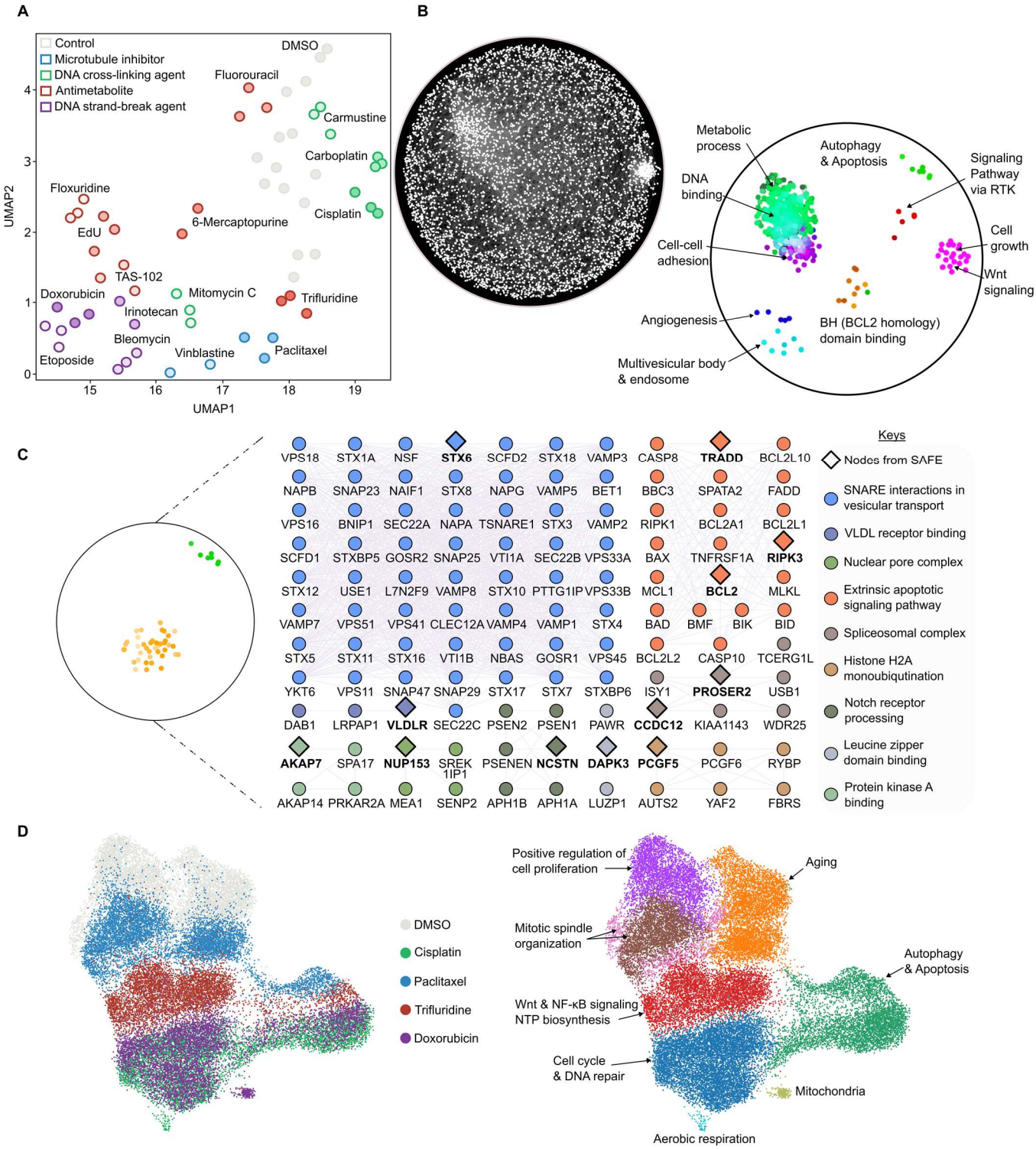


Figure 3. Comprehensive Systems Analysis of Drug Response Screening Highlighting BCL2-Associated Functional Pathways

(**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

631 ORF as node and Pearson correlation (cut-off: < 0.9) as edge (left), functional clusters using SAFE (right).
632 (**C**) STRING protein-protein interaction (PPI) network of two BCL2-related SAFE clusters ("BH domain
633 binding" and "Autophagy & Apoptosis"), highlighting genes involved in vesicular transport, apoptotic
634 signaling, histone modification, and Notch receptor pathways. The network is based on high-confidence
635 experimental interactions (score > 0.9), <50 interactors in the first and second shells. 12 diamond-shaped
636 nodes highlight ORFs originating from the SAFE network clusters. (**D**) UMAP plots showing Single-cell
637 RNA-seq treatment-specific ORF-expressing cell population changes for four different
638 chemotherapeutics (left) and enriched pathways in distinct cell clusters identified by the Leiden algorithm
639 and analyzed using DAVID (right).

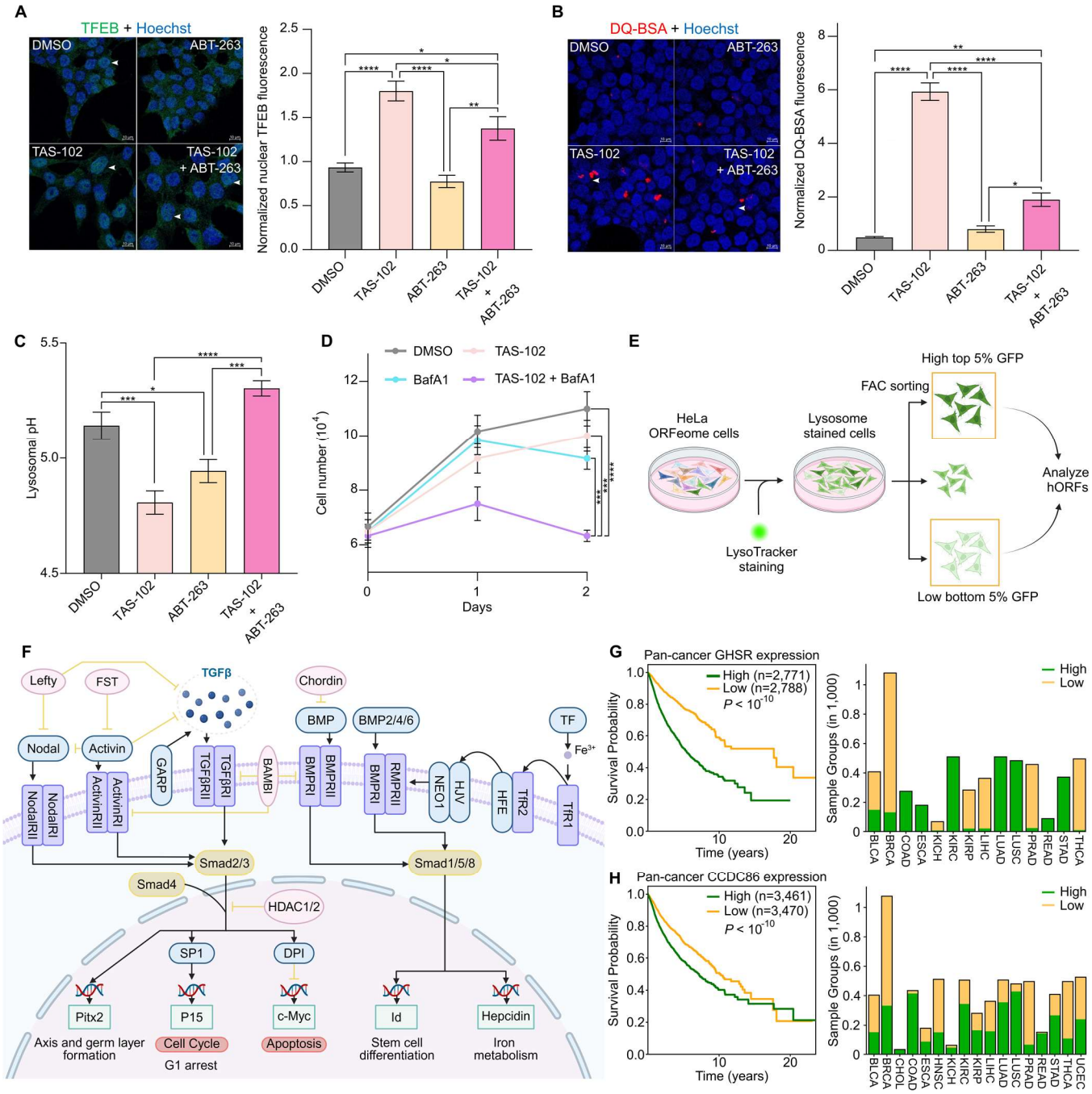
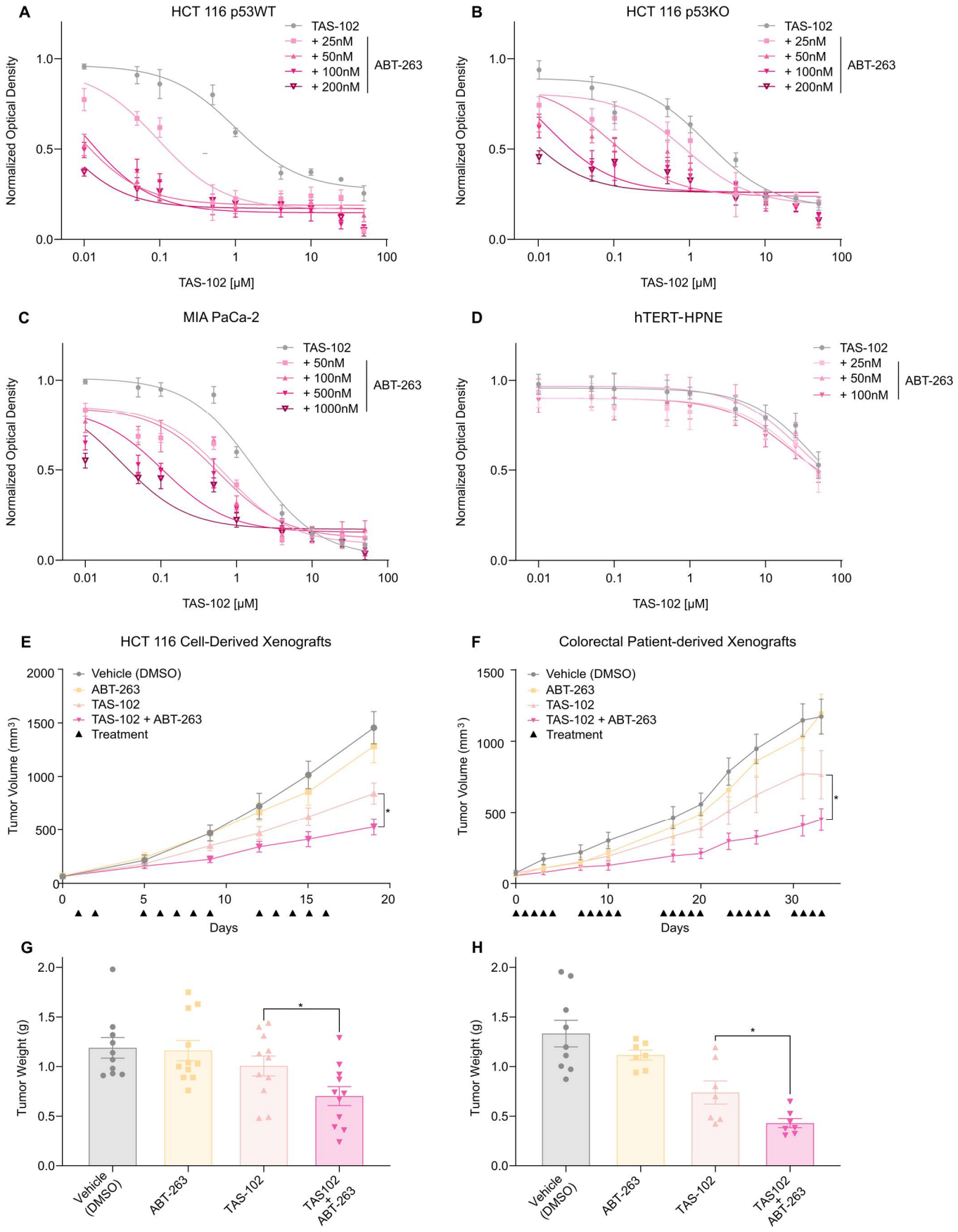


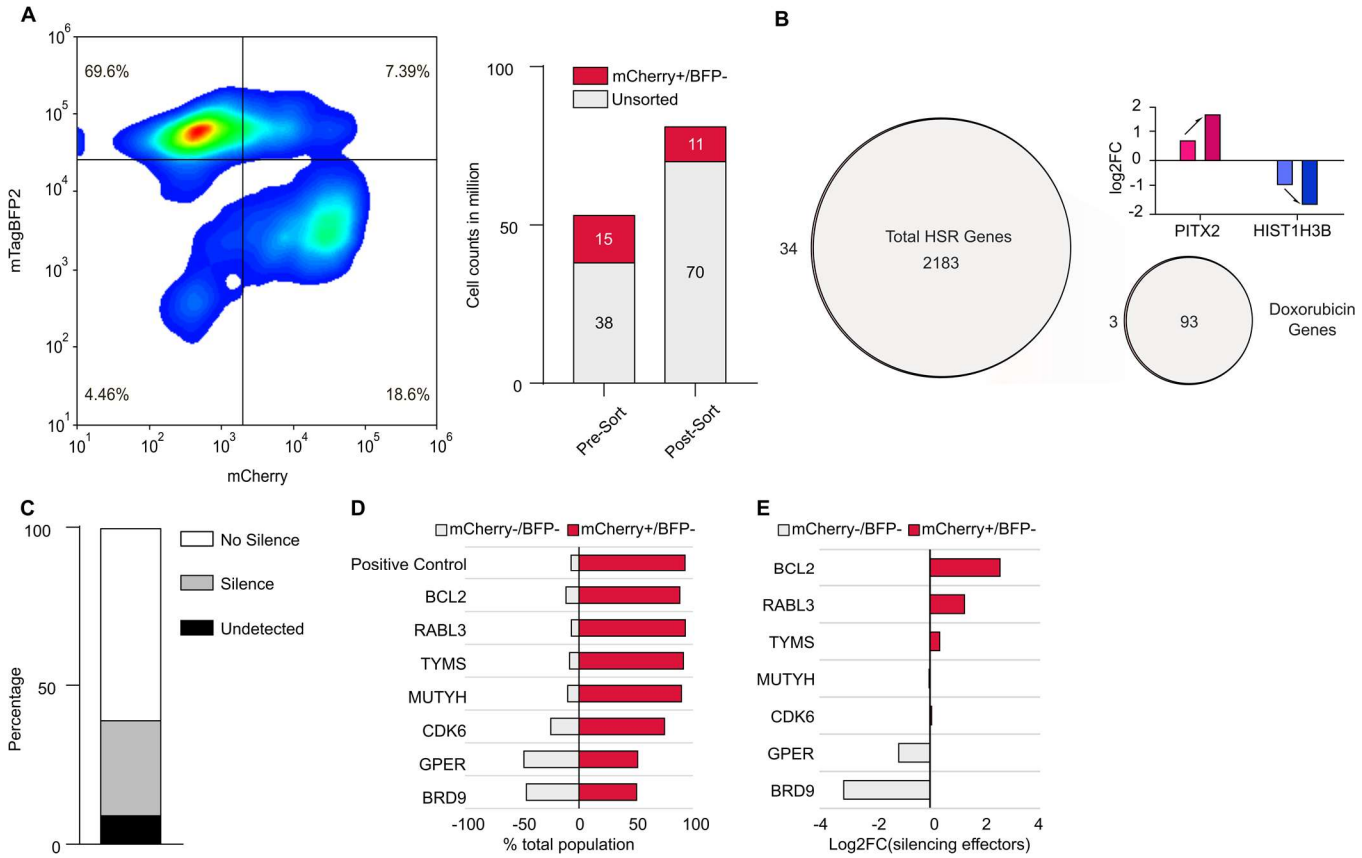
Figure 4. Functional Validation of BCL2 Inhibition and Autophagy-Lysosomal Modulation

(**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 (red) 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

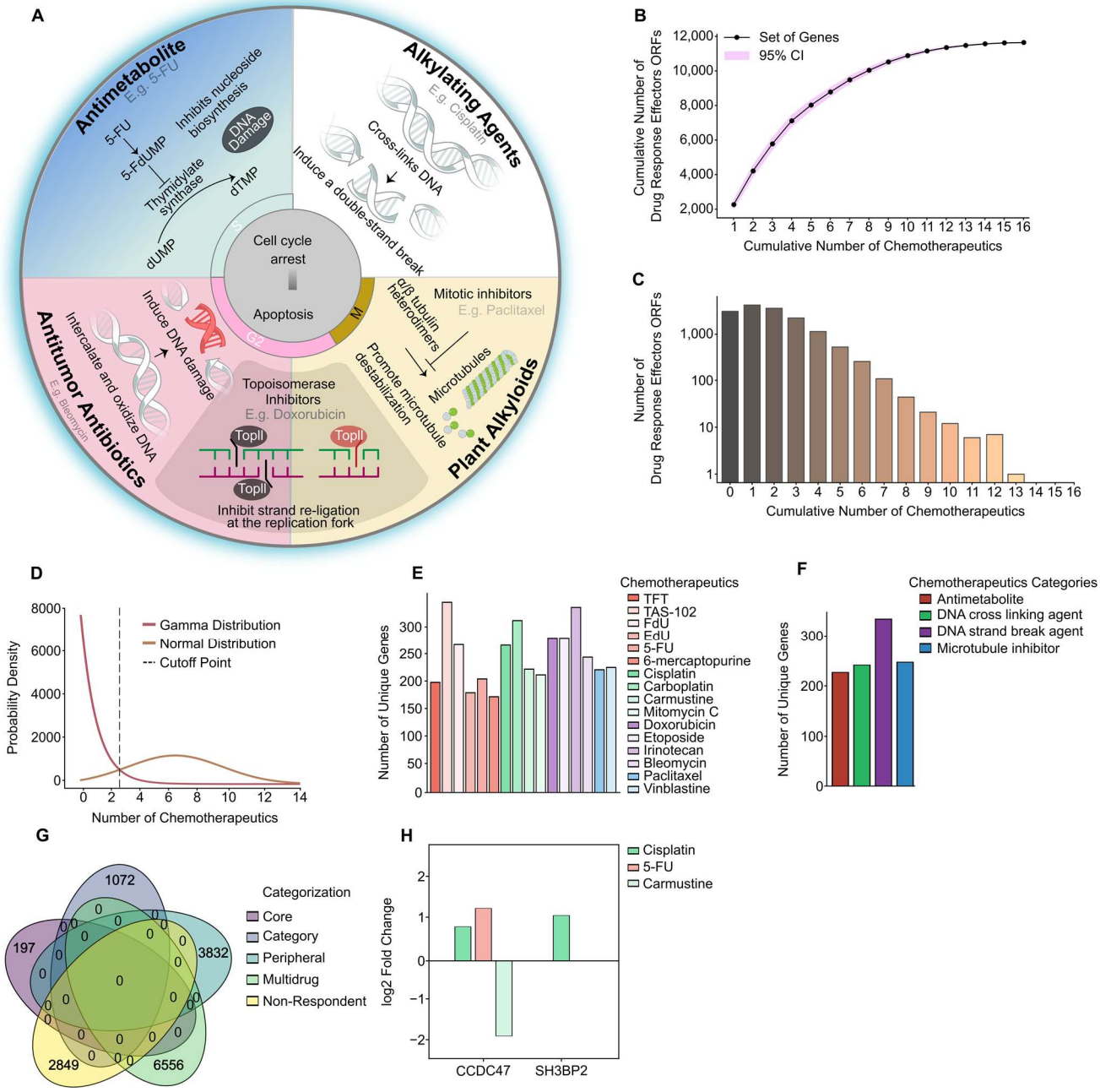
A1. (**E**) Schematic workflow of high-throughput LysoTracker-based ORFeome overexpression screen with FACS sorting. (**F**) Diagram of TGF-β signaling pathway highlighting lysosomal activator and repressor ORFs identified in the LysoTracker-FACs screening. (**G-H**) Kaplan-Meier survival curves depicting 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.


A
HCT 116 p53WT
TAS-102
+ 25nM
+ 50nM
+ 100nM
+ 200nM
ABT-263
Normalized Optical Density
TAS-102 [μM]
B
HCT 116 p53KO
C
MIA PaCa-2
+ 500nM
+ 1000nM
D
hTERT-HPNE
E
HCT 116 Cell-Derived Xenografts
Vehicle (DMSO)
ABT-263
TAS-102
TAS-102 + ABT-263
Treatment
Tumor Volume (mm³)
Days
F
Colorectal Patient-derived Xenografts
G
Tumor Weight (g)
Vehicle (DMSO)
TAS102 + ABT-263
H

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



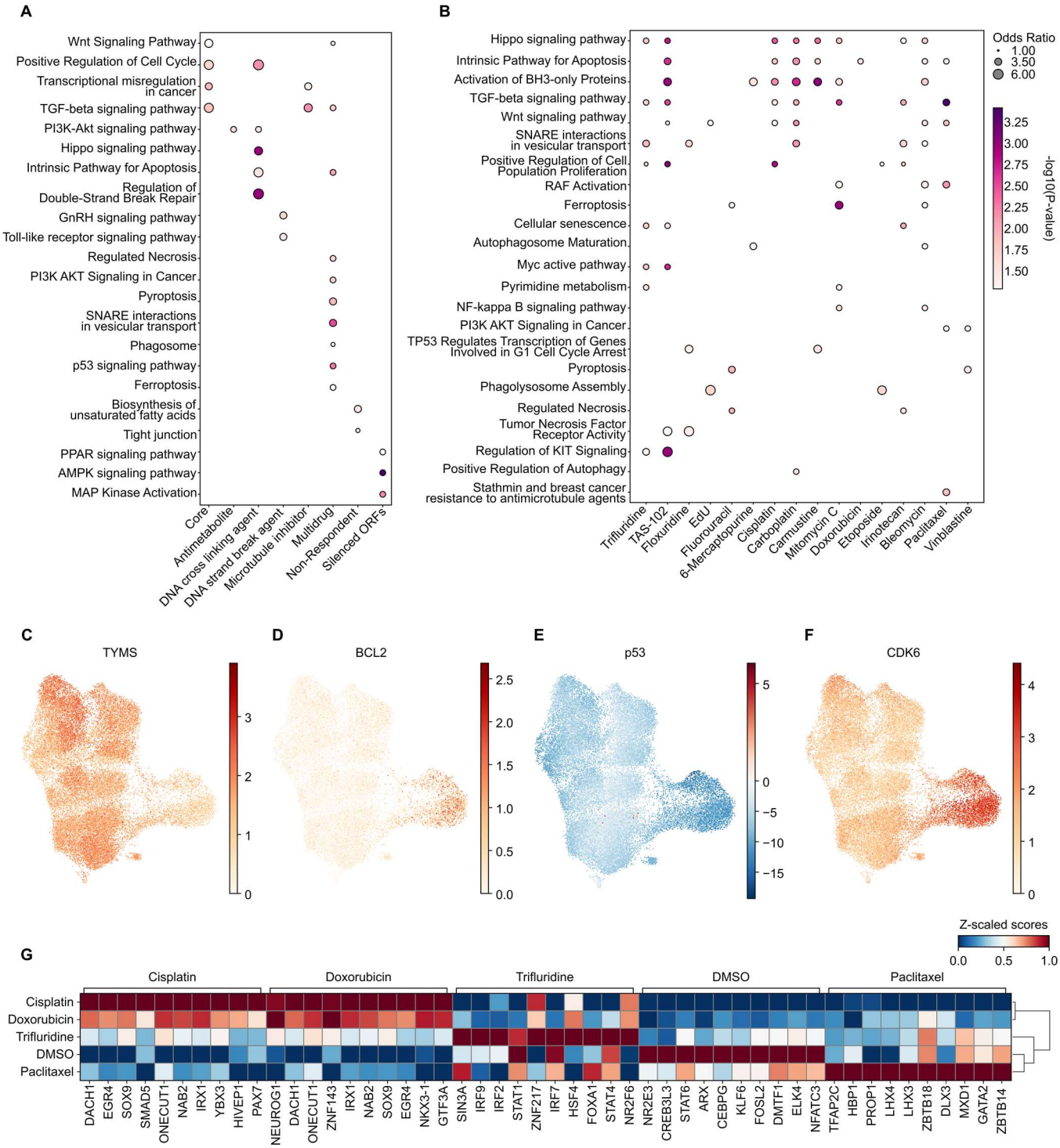
Supplementary Figure 1. Validation and Characterization of the BOGO System and Screen Results (**A**) Efficiency of ORF integration in HeLa cells measured by mCherry and mTagBFP2 (BFP) fluorescence via FACS (left). Stacked bar graph showing the number of sorted and unsorted cells via mCherry fluorescence (right). (**B**) Venn diagrams illustrate the 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 responding 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

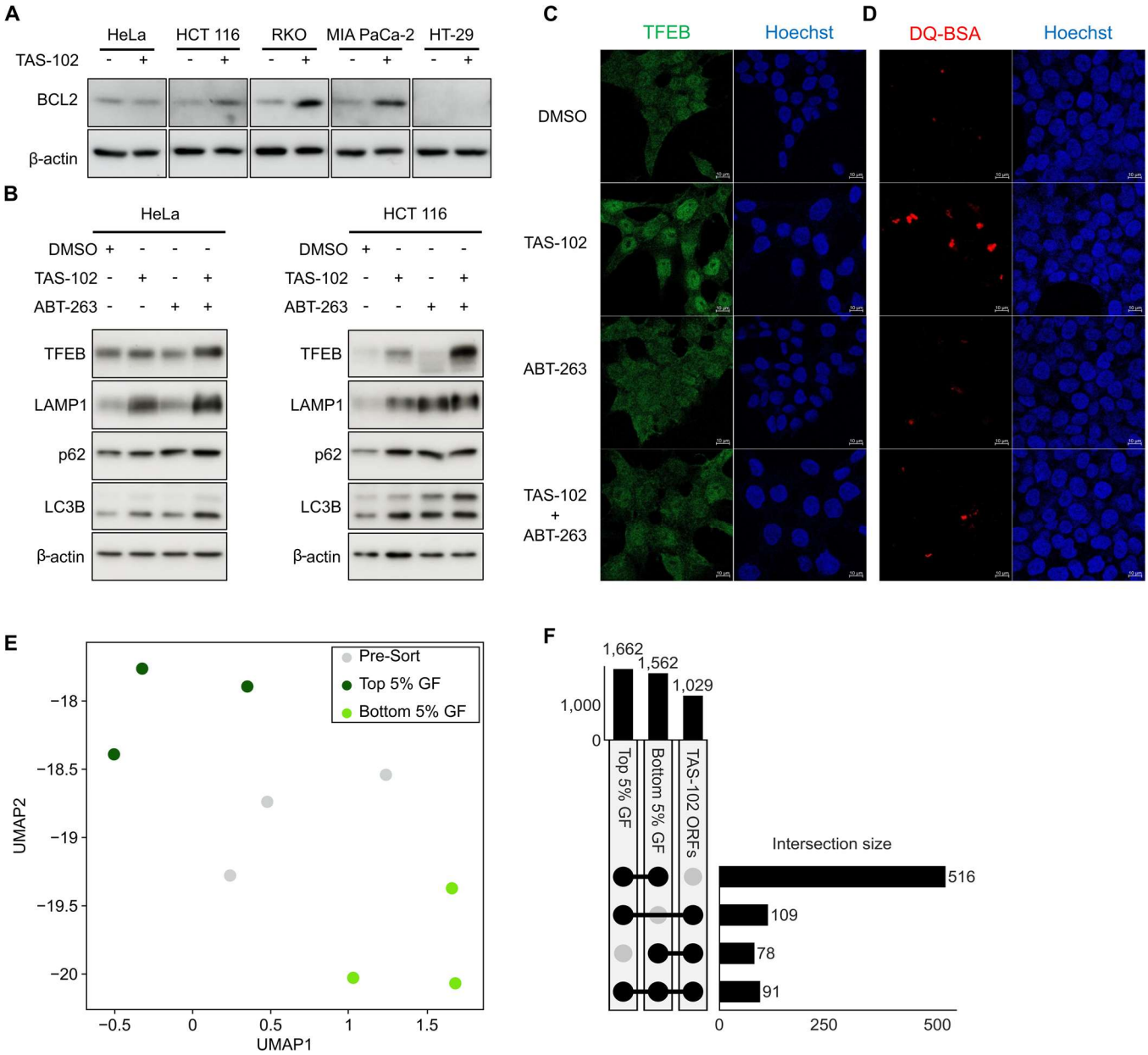
685 Chemotherapeutics Categories. (**G**) A Venn diagram illustrates the distinctness of gene subsets.
686 Numbers indicate the number of ORFs in five categories. 0 indicates there is no overlap between subsets.
687 (**H**) Drug response effector expression of SH3BP2 and CCDC47 in four chemotherapeutics from the drug
688 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

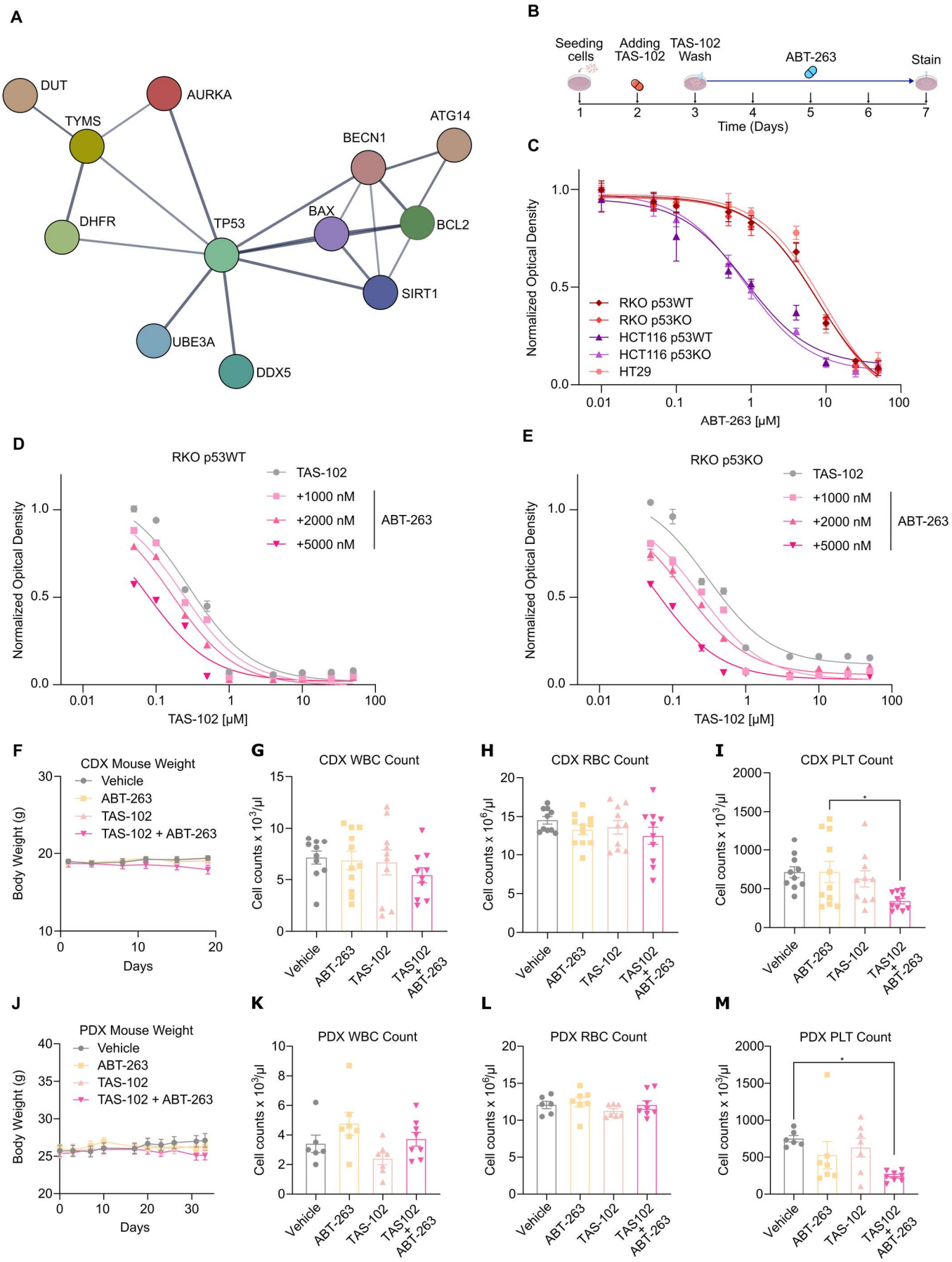
694 single-cell ORF expression of key resistance markers: **(C)** *TYMS*, **(D)** *BCL2*, **(E)** *TP53*, and **(F)** *CDK6*.
695 **(G)** Heatmap of normalized z-scores displaying ORF expression across one DMSO and four
696 chemotherapeutic conditions, highlighting the top five resistance-associated 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

706 clustering of pre-sort and LysoTracker-sorted; top and bottom 5% green fluorescence (GF) from HeLa
707 ORFeome cell. Each dot indicates biological replicates (N=3). (**F**) Upset plot showing overlap between
708 TAS-102 drug response effector ORF and lysosomal modulator ORFs from LysoTracker staining.


A
DUT
AURKA
TYMS
ATG14
BECN1
DHFR
TP53
BAX
BCL2
SIRT1
UBE3A
DDX5
B
Seeding cells
Adding TAS-102
TAS-102 Wash
ABT-263
Stain
Time (Days)
C
Normalized Optical Density
RKO p53WT
RKO p53KO
HCT116 p53WT
HCT116 p53KO
HT29
ABT-263 [μM]
D
RKO p53WT
TAS-102
+1000 nM
+2000 nM
+5000 nM
ABT-263
Normalized Opitcal Density
TAS-102 [μM]
E
RKO p53KO
TAS-102
+1000 nM
+2000 nM
+5000 nM
ABT-263
Normalized Optical Density
TAS-102 [μM]
F
CDX Mouse Weight
Vehicle
ABT-263
TAS-102
TAS-102 + ABT-263
Body Weight (g)
Days
G
CDX WBC Count
Cell counts x 10^3/μl
Vehicle
ABT-263
TAS-102
TAS102 + ABT-263
H
CDX RBC Count
Cell counts x 10^6/μl
I
CDX PLT Count
Cell counts x 10^3/μl
J
PDX Mouse Weight
Vehicle
ABT-263
TAS-102
TAS-102 + ABT-263
Body Weight (g)
Days
K
PDX WBC Count
Cell counts x 10^3/μl
L
PDX RBC Count
Cell counts x 10^6/μl
M
PDX PLT Count
Cell counts x 10^3/μl

709 **Supplementary Figure 5. Synergistic Efficacy and Safety Profile of TAS-102 and ABT-263**
710 **Combination**
711 (**A**) STRING PPI network showing an indirect p53-mediated link between BCL2 and TYMS, medium
712 confidence (<0.5; protein interaction score from STRING), <5 interactors in first and second shells. (**B**)
713 Dose-response curves for ABT-263 in colorectal cancer lines HCT 116, RKO (p53 WT and KO), and HT-
714 29 using varying concentrations 0, 0.01, 0.05, 0.1, 0.5, 1, 4, 10, 25, and 50 μM (**C-D**). Synergistic
715 cytotoxicity of TAS-102 and ABT-263 in (**C**) RKO p53 WT (**D**) and p53 KO cells. (**E-F**) Mouse body weight
716 data from (**E**) CDX and (**F**) PDX models, showing stable weight across treatments. (**G-K**) Complete blood
717 count (CBC) parameters, including red blood cell, white blood cell, hemoglobin, monocyte, and
718 granulocyte, with no significant differences across treatments. (**L**) Platelet counts showed a significant
719 reduction of * $p < 0.05$ in combination therapy compared to ABT-263 alone. Data in C-D panels are
720 presented as mean ± standard deviation (SD), whereas B and E-L panels show mean ± standard error
721 of the mean (SEM). Statistical analysis was performed using one-way or two-way ANOVA with Tukey's
722 multiple comparisons.

Code availability

The code and ORFeome reference file used to perform the analyses reported in this study can be found on GitHub (https://github.com/Harry77j/BOGO_analysis).

Data availability

The sequencing data generated in this study have been deposited in the NCBI database under BioProject SRA accession number: PRJNA1353565. The following secure link allows review of the record while it remains in private status:
https://dataview.ncbi.nlm.nih.gov/object/PRJNA1353565?reviewer=dc3jinggo7a0jli08lpi00boks
The public release data for all sequencing datasets is set for Oct 27th, 2029, or upon publication, whichever occurs first.

Acknowledgment

We thank all current and former members of our laboratories for their valuable discussions and continuous support throughout this project. We are grateful to the Genomics Shared Resources team for their expertise and assistance with sequencing, and to the Animal Facility at Roswell Park Comprehensive Cancer Center (Buffalo, United States) for supporting the in vivo studies.
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.-E. P.**; 2022R1A2C2010940, **H.K.K**) and Chungbuk National University Glocal30 project 2025 (**J.-H.C.**).

Conflict of Interest

M.M.A., **A.V.B.**, and **D.-K.K.** are inventors on a patent (WO2024233874) relating to this study.

Author Contribution

K.B.J. and **D.-K.K.** conceptualized and designed the study. **K.B.J.** performed the in vitro experiments, processed the data, analyzed the results, drafted the original manuscript, managed the project, and prepared the figures. **M.M.A.**, **Y.G.**, and **M.N.N.** performed in vitro and in vivo experiments. **D.-E.K.** conducted in vitro experiments and molecular assays. **Y.K.** processed data and analyzed results. **H.K.** and **J.-S.K.** carried out data curation, figure editing, and critical manuscript review. **S.S.** and **J.P.W.** assisted with data processing and manuscript editing. **K.Y.** provided critical resources and contributed ideas for experimental optimization. **A.G.C.**, **F.P.R.**, **K.A.M.**, **D.E.H.**, **J.-H.C.**, **K.K.-H.F.**, and **H.L.**

753 contributed to conceptualization, provided resources, experimental advice, manuscript editing, and
754 technical support. **J.-E.P.**, **H.K.K.**, **A.V.B.**, and **D.-K.K.** supervised the project, provided conceptual
755 guidance, and oversaw project administration.

756 # Supplementary Table

757 **Supplementary Table 1:**

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

759 **Supplementary Table 2:**

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

761 **Supplementary Table 3:**

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

764 **Supplementary Table 4:**

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

767 **Supplementary Table 5:**

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

769 **Supplementary Table 6:**

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

772 # Methods

773 **Cell culturing**

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

781 The human cell lines, colorectal cancer HT29 (RRID: CVCL_0320), pancreatic cancer MIA PaCa-
782 2 (RRID: CVCL_0428), and pancreatic ductal cell hTERT HPNE (RRID: CVCL_C466) were cultured as
783 recommended by ATCC. Human colorectal cancer HCT 116 and HCT 116 p53(-/-; KO) cell lines were
784 shared by Dr. Bert Vogelstein at Johns Hopkins University[72]. Human colorectal cancer RKO and RKO
785 p53(-/-; KO) cell lines were shared by Dr. Alessandro Carugo at MD Anderson Cancer Center[73]. Those
786 cell lines were maintained in DMEM and authenticated again using STR profiling by ATCC or the Roswell
787 Park Genomic Shared Resource within the last three years. The cells were routinely screened for
788 mycoplasma; all studies used mycoplasma-free cells. Human cell cultures were maintained in media
789 supplemented with 10% heat-inactivated fetal bovine serum (FBS; Corning, 35-015-CV) and 1%
790 penicillin/streptomycin (P/S, Corning, 30-002-CI) at 37°C with 5-10% CO_2 in a humidified incubator.

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

792 Entry clones in Human Open Reading Frame Collection (ORFeome) v9.1[22,74] with ~1000 additional
793 clones donated by the Center for Cancer Systems Biology (CCSB) at Dana-Farber Cancer Institute (Delta
794 space) were verified by next-generation sequencing to confirm the authenticity of their coding DNA
795 sequences. Entry clones were stamped from 384-well plates onto corresponding LB agar plates
796 containing either kanamycin (25 µg/mL; Thermo Fisher, AC611290050) or spectinomycin (50 µg/mL;
797 Thermo Fisher, S05845G). After incubation at 37°C for 48 hours (hrs), the LB agar plates were scraped
798 and collected for pooled plasmid preparation using the Qiagen Plasmid Mini Prep kit (Qiagen, 27104) -
799 one miniprep per one 384-well plate. The amount of DNA proportional to the number of clones was pooled
800 and mixed into a single tube to create a complete set of entry clones, where each ORF was represented
801 in equivalent ratios. Based on these pools of Entry clones, we performed *En masse* LR cloning as
802 previously described[22,23,75]. 150 ng of the entry pool was incubated with 5x LR Clonase (Invitrogen,
803 11791020) and 150 ng of the destination vector (pDEST-HC-REC, Roth Lab) for 5 days, starting with a

804 5 μL reaction and supplementing with 1 μL of 5x LR Clonase and 150 ng of destination vector each day.
805 We used 1 μL of the LR reaction mix to perform electroporations 8 times (total 8 μL used) with 10-beta
806 electrocompetent *E. coli* (New England Biolabs, C3020K) to generate 20 million Colony Forming Units,
807 which achieves approximately 1000x library representation. In detail, the electroporated bacteria were
808 plated onto LB agar containing carbenicillin (Fisher, BP26485, 100 μg/ml) plates and incubated at 37°C
809 for 18 hrs. Colonies were collected by gentle scraping, and the final destination library was isolated using
810 a PureLink™ Fast Low-Endotoxin Maxi Plasmid Purification Kit (Invitrogen, A35892).

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

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

829 **Drug dose-response assays**

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

837 (Absorbance at 595 nm) reading was measured after 2 hrs of reagent incubation, and the absolute
838 Inhibitory Concentration at 50% (IC_{50}) was calculated using GraphPad Prism 9 software (Version 9.1.0).

839 **Chemotherapeutic Screening**

840 We selected 16 chemically synthesized cytotoxic chemotherapeutic agents that primarily target the cell
841 cycle or exert non-specific effects through DNA damage. These drugs were chosen to represent a broad
842 range of mechanisms of action commonly employed in adult and pediatric cancer treatments, as
843 previously described[76]. To streamline the analysis, we categorized chemotherapeutics based on their
844 primary mechanisms of action. In some cases, drugs were grouped under broader mechanistic
845 categories. For instance, doxorubicin, a topoisomerase II inhibitor, and bleomycin, an antitumor antibiotic,
846 were classified under agents that induce DNA strand breaks, reflecting their shared outcome despite
847 differing molecular mechanisms. Based on these mechanistic classifications, the 16 chemotherapeutics
848 were grouped into the following four categories:

849 ● *Antimetabolites*: Fluorouracil (5-FU, Sigma, F5130), Trifluridine (TFT, Med Chem Express, HY-
850 A0061), 5-ethynyl-2′-deoxyuridine (EdU, Toronto Research Chem - Sigma, E9386), Floxuridine
851 (FdU, Adooq Bioscience, A10392-500), 6-mercaptopurine (Sigma, 852678), TAS-102 (Adooq
852 Bioscience, 733030-01-8)

853 ● *DNA Cross-Linking Agents*: Mitomycin C (Sigma, M5353), Cisplatin (Cayman, 13119), Carboplatin
854 (British Pharmacopeia Reference - Sigma, BP711), Carmustine (European Pharmacopoeia
855 Reference - Sigma, Y0002128)

856 ● *DNA Strand Break Agents*: Doxorubicin (Tocris, 2252), Bleomycin (Sigma, B1141000), Etoposide
857 (Sigma, PHR2717), Irinotecan (Sigma, 341205)

858 ● *Microtubule Inhibitors:* Paclitaxel (Sigma, T7402) and Vinblastine (Sigma, V0300000)

Chemotherapeutic drug	IC_{50}
Cisplatin	0.333 μM
Carboplatin	3.37 μM
Carmustine	7 μM
Mitomycin C	14 nM
TFT	0.75 μM
5-FU	0.1925 μM
EdU	0.875 μM
FdU	25 nM
6-mercaptopurine	2 μM

TAS-102	0.8 μM
Doxorubicin	5 nM
Bleomycin	3.3 μM
Irinotecan	0.8 μM
Etoposide	0.625 μM
Paclitaxel	2 nM
Vinblastine	0.3 nM

859 Using IC_{50} values obtained through the drug dose-response assay, we determined an appropriate dose
860 to treat the cells in a 15 cm dish format. We seeded at least 3 million cells and treated them with drugs
861 at their IC_{50} dose for 72 hrs. DMSO was selected as the vehicle control. Screens were conducted in five
862 independent batches, each including baseline (often referred to as T0 in CRISPR screening)[77] and its
863 own set of vehicle controls to account for batch-specific variation. To ensure full library representation,
864 we thawed 10 cryopreserved ORFeome-overexpressing pool vials and expanded them for 3 days,
865 recovering a minimum of 3 million viable cells in total to maintain 150x coverage. The cells were then
866 seeded at 3 million per dish and treated with the respective drugs in triplicate, including corresponding
867 vehicle controls in each batch. After 72 hrs of treatment, cells were trypsinized, and 3 million cells were
868 passaged into fresh dishes with continued drug treatment. After 48 additional hrs, the drug-containing
869 medium was removed, and cells were allowed to recover in regular culture medium for 3 more days or
870 until confluency. Recovered cells were trypsinized and pelleted (~10 million cells per pellet) for
871 subsequent genomic DNA extraction.

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

873 According to the manufacturer's instructions, genomic DNA was extracted from the cell pellets using the
874 Blood and Cell Culture DNA Midi Kit (Qiagen, 13343). We performed two sets of PCRs (nested PCR)
875 using the Q5 High-Fidelity 2X Master Mix (New England Biolabs, M0544S) and primers (Integrated DNA
876 Technologies) that bind to i) landing pad locus and ii) the Gateway-cloning sites, to amplify the full-length
877 Open Reading Frame (ORF) from the landing pad with improved selectivity and reduced genomic DNA
878 contaminations. A total of 18 μg of DNA was used in the first PCR in twelve 50 μL PCR reactions, which
879 were pooled to represent the full 150x coverage. The second PCR was performed using 2.5 μL of the
880 previous PCR reaction as the template. Two 50 μL PCR2 reactions were performed and pooled before
881 preparing the sequencing library. For NGS library preparation using the TruSeq platform, custom primers
882 were used that included a 5′ NNNNN sequence to secure the complexity of the first 5 bp during
883 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

910 normalized using the quantile-normalization method and batch-corrected to create an equal distribution
911 of the dataset to analyze the difference between samples using EdgeR (v4.0.16) and Limma (v3.58.1) [91].
912 For sample-to-sample comparison, a TMM dataset was processed using dimensional reduction analysis
913 of Uniform Manifold Approximation and Projection (UMAP) (UMAP-learn[92]; v0.5.7). The efficiency of
914 ORFeome silencing was assessed by quantifying fluorescence levels in cells expressing mCherry and
915 mTagBFP2. Silencing was evaluated using the Silencing Ratio, defined as $log_2(E/S)$ where E
916 represents the number of expressing cells ($mCherry^+$ $mTagBFP2^-$) and S denotes the number of silenced
917 cells ($mCherry^-$ $mTagBFP2^-$). This ratio reflects the difference in relative fluorescence units between
918 cells actively expressing the ORF ($mCherry^+$ $mTagBFP2^-$) and those in which the ORF is likely silenced
919 ($mCherry^-$ $mTagBFP2^-$). ORFs with a Silencing Ratio less than -1 (log2 fold change < -1) were classified
920 as successfully silenced.

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

932 **Automated entry clone selections for High Saturation Retesting (HSR)**

933 Approximately 10% of the ORFeome pool underwent retesting at higher coverage (>1,000 fold) to assess
934 differentially represented ORFs. The selection of these ORFs relied on T-score statistical analysis of
935 log2-fold changes for individual drugs, grouped into eight baskets. These selections were based on core
936 resistance or sensitizing ORFs across all drugs, individual drugs, and controls. The selection criteria for
937 HSR ORFeome focused on genes that demonstrated high T-scores, indicative of strong or weak
938 proliferation phenotypes. In total, 2,217 ORFs were selected and cherry-picked from the human
939 ORFeome using a customized robotic platform (S&P Robotics Inc., BM3-BC). These selected ORFs were
940 then pooled and prepared within the HeLa EF1α cells following the same *en masse* LR cloning[75].
941 Doxorubicin screening was carried out with matching dosages of 5 nM as in the original screening,
942 employing identical PCR ORF amplification methods and previously mentioned ORFeome bioinformatics

943 analysis in this method. The selection of “Resistant”, “Sensitizing”, and “Non-respondent” was based on
944 the same log2 fold change parameter as drug response screening. The graph was generated using
945 GraphPad Prism 9, and statistical analysis was performed using one-way ANOVA followed by Tukey’s
946 multiple comparisons test.

947 **Single ORF overexpression assay**

948 To generate single cell lines for verification, specific entry clones corresponding to ORFs such as TYMS,
949 RABL3, BCL2, MUTYH, CDK6, GPER, and BRD9 were grown overnight and isolated using the
950 NucleoSpin Plasmid kit (Macherey-Nagel, NC9405546) from the ORFeome 9.1 collection[22]. 150 ng of
951 entry vector, 150 ng of destination vector (pDEST-HC-REC shared by Roth Lab), and 1 µL of LR Clonase
952 were incubated for 1 hour and transformed into 10-beta chemically competent cells (New England Biolabs,
953 C3019) using heat shock [98]. Transformants were grown overnight in LB containing carbenicillin (Fisher,
954 10177012) and used for plasmid extractions. Sequences were confirmed through Sanger reactions. One
955 day before transfection, HeLa EF1α cells were seeded into 6-well plates at 350K cells/well. The correctly
956 cloned plasmids were co-transfected with the Bxb1 recombinase plasmid at 1 µg each into HeLa EF1α
957 cells using Lipofectamine 3000. After 5 days, cells were sorted for mCherry-positive populations to
958 establish stable single-ORF-overexpressing cell lines. For the cellular competition assays, the protocol
959 was adapted from the previous study[99]. 30,000 mCherry-positive cells (experimental) were co-cultured
960 with 60,000 BFP-positive (control) cells in each well of a 6-well plate. After the 24 hrs incubation, three
961 wells were treated with 0.75 µM TFT. After a 5-day treatment period, relative cell populations were
962 quantified by reading 20,000 events per well using FACS. Results from the competition assay were
963 compared to fold change data from the initial TFT screening. For silencing validation, phenotypic changes
964 were assessed by quantifying their total population percentage after 5 days without any drug treatment
965 using FACS. Following the recovery, specific populations were sorted via FACS for subsequent library
966 sequencing. The alignment between phenotypic changes and transcriptomic response was evaluated by
967 comparing the population percentages with Silencing Ratios using a previously mentioned ORF
968 bioinformatics pipeline. The cell population was normalized by the mCherry control. The graph was
969 generated using GraphPad Prism 9, and statistical analysis was performed using one-way ANOVA
970 followed by Tukey’s multiple comparisons test.

971 **mRNA transcriptomics analysis**

972 The treatment duration and the protocol are adapted from a previous study on finding drug
973 combinations[100]. HeLa Cells were seeded at 3 million in 15 cm dishes and treated with an IC_{50} of 5nM
974 Doxorubicin for 24 hrs and 120 hrs. After the treatment, cells were pelleted, and each cell pellet consisted
975 of a minimum of 1 million cells per control and experimental set. Each cell pellets were preserved in

976 TRIzol (Qiagen, #79306) and stored at -20℃ before being sent to Genomics Shared Resources (GSR)
977 at Roswell Park Comprehensive Cancer Center. Purified RNA was prepared using the miRNeasy micro
978 kit (Qiagen, 217084), and quality control of RNA integrity number (RIN) was measured using Agilent
979 Bioanalyzer 2100 (Agilent, G2939BA). Sequencing libraries were prepared with the KAPA mRNA
980 HyperPrep Kit (Roche, KK8581). The final RNA-seq libraries were sequenced on Illumina NovaSeq 6000.
981 Nextflow version 24.10.4[101] with nf-core/rna-seq v3.18.0[102] was utilized using Docker (v27.5.1). The
982 sequencing files were aligned to the GRCh38.p14 human genome[103]. The workflow was configured to
983 perform adapter trimming with Trim Galore[104], followed by alignment using STAR[105] (v2.7.10a) and
984 quantification using the RSEM[106] (v1.3.1). Ribosomal RNA reads were removed using SORTMERNA[107]
985 (v4.3.7) before quantification. Downstream analysis used count-based quantification, matching the
986 configuration of the ORFeome bioinformatic pipeline. Paired t-tests were performed using SciPy's
987 stats.ttest_rel function[108] (v1.15.2), and empirical p-values were calculated via permutation testing (n =
988 10,000) to assess statistical significance.

989 **ORF Gene categorization**

990 According to HeLa ORFeome cell sequencing, reads were aligned to 18,388 ORFeome libraries. Of these,
991 14,506 ORFs were considered well-represented, defined by a minimum median $\log_2$ TMM expression >
992 -1[109]. ORFs below this cutoff were regarded as noise and excluded from downstream analysis. Within
993 the screened 14,506 ORFeomes, those ORFs were categorized into five categorical subsets. Among the
994 categorized ORFs, those with a log2 fold change threshold of ±0.5 and a p-value less than 0.05 were
995 identified as differentially represented ORFs. Within this set, enriched ORFs (positive log2 fold change)
996 are interpreted as putative resistance drivers, conferring a survival advantage under treatment, while
997 depleted ORFs (negative log2 fold change) suggest sensitization drivers, indicating increased drug
998 sensitivity. First, 197 core genes were selected based on any ORFs with the filter of more than 7 drugs
999 (more than half of the total chemotherapeutics - 1). Second, 1,072 categorical genes were assigned
1000 based on the recalculation of the median(log2{drug categories}) - median(log{rest of categories}) of
1001 previously mentioned statistical testing and filtered to differentially represented ORFs. Third, 3,832
1002 peripheral genes were identified as ORFs significantly altered (absolute log2 fold change ≥ 0.5 and p-
1003 value ≤ 0.05) in response to specific chemotherapeutic drugs, but not in others. For each drug, genes
1004 meeting these criteria were selected and then filtered to exclude any genes that also met the same
1005 significance criteria in other drugs, thereby ensuring drug-specificity. Fourth, 6,556 multidrug genes were
1006 chosen based on a filter in a sum of chemotherapies between $2 \leq x < 7$. Two chemotherapies were
1007 selected to a minimum number, based on the mixture distribution of gamma, and the normal distribution
1008 shows a cutoff value of 2.64 in filtered ORFs in the sum of chemotherapies. Lastly, 2,849 non-respondent

1009 genes were picked with ORFs, with the filter for no enriched or depleted ORFs. To avoid any overlapping
1010 of ORFs between categories, the selection was prioritized sequentially from “Core” to “Non-respondent”
1011 genes. Gene categorization overlaps in comparison with ORFeome screening results were determined
1012 using the Venn library (v0.1.3) in Python (3.11).

1013 **Biomarker Analysis**

1014 To assess the clinical relevance of gene subsets identified in our screening, we performed a multi-tiered
1015 biomarker overlap analysis using TheMarker database[31], a curated resource of disease, biomarker, and
1016 drug associations. Gene matching was based on NCBI Gene identifiers, and only verified, non-silent
1017 ORFs from the screening dataset were included in the analysis. Verified, non-silent ORFs were
1018 intersected with cancer-specific biomarker entries in the database, and for each gene subset, the
1019 proportion of genes overlapping with cancer-associated biomarkers, classed as “ICD-02: Benign/in-
1020 situ/malignant neoplasm,” was calculated to evaluate enrichment for clinically validated cancer-related
1021 genes. We then examined the association between biomarker-annotated ORFs and chemotherapeutic
1022 drugs; among the 16 drugs profiled in the screen, 7 were represented in the database, and we quantified
1023 how many biomarker-annotated ORFs within each subset were linked to one or more of these drugs,
1024 connecting screening hits to existing pharmacogenomic knowledge. Finally, we assessed the
1025 representation of each gene subset within the total biomarker pool by calculating the proportion of ORFs
1026 present in the database relative to the total unique biomarkers catalogued. All overlap proportions were
1027 normalized using z-scores to enable comparisons across subsets and highlight differential enrichment
1028 patterns. The z-scores were visualized as heatmaps, and the numbers of four biomarker types within
1029 each ORF category were calculated as log10-transformed counts and displayed in stacked bar graphs.
1030 The stacked bars represent log-transformed counts for each category separately and do not correspond
1031 to the sum of total biomarker numbers.

1032 **Gene properties analysis**

1033 Length of each ORF (ORF Length) and its GC percentage (GC Content) were both derived from human
1034 ORFeome v9.1 sequences and delta space references after alignment as described above. We adopted
1035 some of the gene properties from our previous study[22,23]. Genotype-Tissue Expression project (GTEx)[110]
1036 v6 transcriptome data (downloaded on March 15th, 2016) was processed with R v3.5.1 to extract
1037 normalized log read count. In this dataset, 16 brain subregions were collapsed into three brain tissues:
1038 basal ganglia, cerebellum, and others. The median expression of each gene across all tissue samples
1039 was calculated, and a cutoff of >5 was applied. Cell line samples were excluded, and the analysis was
1040 restricted to the set of protein-coding genes. We considered these values as a reference for gene
1041 expression level (Abundance), which were calculated as mentioned in our previous study [22]. The same

1042 data was used to calculate the number of tissues each gene was expressed in (Broadness). Genes' age
1043 information (Age) was adopted from the Protein Historian database[111], with PPODv4_Jaccard as the
1044 family database and Dollo parsimony model as the reconstruction algorithm. The D2P2 database of
1045 disordered protein predictions[112] was used to calculate the fraction of disordered regions relative to full
1046 protein length (Disorder). We extracted all human D2P2 IDs mapped to Ensembl IDs from
1047 "genomes.protein.gz". Using this list and the "consensus_ranges.tsv.gz" file, containing the start and end
1048 positions of each predicted disordered region, the total disordered length for each protein was computed.
1049 The length of each protein was calculated using the "genomes.fa.bz2" file, which stored the full protein
1050 sequences. The fraction was calculated by merging the data by gene IDs. The DAVID functional
1051 annotation tool[113] (v2023q4) was used to retrieve the information for other properties. The list of Entrez
1052 IDs was submitted, and the Functional Annotation Table for the following databases was downloaded on
1053 January 16th, 2024: OMIM_DISEASE, GO - GOTERM_BP_DIRECT (Biological Process),
1054 GOTERM_CC_DIRECT (Cellular Components), GOTERM_MF_DIRECT (Molecular Function),
1055 PUBMED_ID, and UP_TISSUE. Terms associated with each dataset were counted. The number of
1056 publications (PubCount) was determined by the number of unique PubMed IDs per gene. Similarly, we
1057 used GO to obtain the number of associated terms for each gene (GO Terms), UP_TISSUE[114] to assess
1058 the number of tissues in which each gene's expression level was upregulated (Tissue Count), and OMIM
1059 to find the number of diseases per gene (OMIM Terms). As a proxy for genes' importance and using the
1060 CORUM database version 4.1 (downloaded on January 23rd, 2024), we calculated the number of protein
1061 complexes associated with each gene (CORUM Complexes).

1062 **Prognostic Evaluation of Proliferation-Associated ORFs Using z-score Ranking, Precision–Recall**
1063 **Metrics, and Survival Analysis**

1064 To assess the clinical relevance of cancer proliferation-related ORF populations, we employed a
1065 combined z-score ranking approach. For each experimental batch, changes in ORF abundance were
1066 calculated as the log2 difference between mean expression under DMSO treatment and baseline
1067 expression, following a strategy adapted from CRISPR screening methodologies[115]. These differences,
1068 representing perturbation responses, were computed for each batch independently. The resulting
1069 difference matrix was then z-score normalized batch-wise, transforming each batch's distribution. To
1070 aggregate signals across multiple experimental batches, an inverse z-score transformation method from
1071 a previous study[116] was applied per ORF. Specifically, each z-score was converted back to an
1072 approximated original scale using the Inverse z-score Transformation. The combined z-score for each
1073 ORF was then calculated by averaging these inverse-transformed values across all batches. ORFs were
1074 ranked in descending order based on their combined z-scores. The ORF list was integrated with

1075 prognostic annotations from the HPA[117] (v24.0) through a two-step merge using Entrez gene IDs and
1076 gene symbols. Genes annotated with a favorable and unfavorable prognosis (defined as being linked to
1077 poor outcome in at least one cancer type) were designated as "Hits". Based on the first hit, the first rank
1078 was computed. To assess performance, precision was calculated as the proportion of Hits among the
1079 top-ranked genes, and recall as the proportion of total Hits identified up to a given rank. In addition, the
1080 maximum precision at or beyond each rank was computed using a cumulative maximum function applied
1081 in reverse order. Both the maximum and recall curves were visualized to evaluate the model's ability to
1082 prioritize the proliferation-related ORF associated with poor prognosis. A parallel analytical pipeline was
1083 applied to assess the association between ranked ORFs and drug resistance, using manually curated
1084 resistance from The Cancer Genome Atlas (TCGA), following criteria from prior work[33]. For survival
1085 analysis, we used TCGA-derived RNA-seq data from HPA. Patients were stratified into high and low
1086 expression groups for each gene using the best expression cut-off, provided by HPA, defined as the
1087 Fragments Per Kilobase of transcript per Million mapped reads (FPKM) threshold that yields the lowest
1088 log-rank P-value between the two groups.

1089 **Spatial Analysis of Functional Enrichment (SAFE) network analysis**

1090 Using the log2 fold change of samples, overall perturbation profile similarity was calculated using the
1091 Pearson Correlation Coefficient (PCC) using Scipy[108] (v1.15.2). Paired ORFs with a PCC of 0.9 and a p-
1092 value <0.01 were selected as significant ORF perturbation similarity. Visualization of a selected paired
1093 ORF network was marked using the Cytoscape tool (v3.4.0)[118]. To visualize this novel network, we used
1094 Spatial Analysis of Functional Enrichment (SAFE; v1.5) to determine, visualize, and predict significant
1095 functional modules in the network[36]. The network layouts were generated using the edge-weighted
1096 spring-embedded layout. GO terms for each ORF were extracted from FuncAssociate (v3 - GO updated
1097 in February 2018)[119]. The SAFE analysis will be run using MATLAB with the default setting file options:
1098 neighborhoodRadius = 1, groupEnrichmentMinSize = 5, plotNetwork = 0 (off), and
1099 THRESHOLD_ENRICHMENT = 0.3.

1100 **Protein-Protein interaction network analysis**

1101 We selected 21 ORF genes from two key clusters identified by SAFE ("BH domain binding" and
1102 "Autophagy & Apoptosis") based on SAFE's first neighborhood enrichment scores and domain annotation
1103 scores. These genes were analyzed using the STRING database[59] (accessed March 10th, 2025) to
1104 construct a protein–protein interaction network focused on their predominant GO domains. In STRING,
1105 only experimentally supported interactions with a high-confidence score of ≥0.9 were included. To
1106 capture relevant protein interactions beyond the selected ORFs, the maximum number of interactors was
1107 limited to 50 in both the first and second interaction shells. The resulting network was clustered using k-

means clustering[120] 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[118].

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[121]. 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)[122]. Low-quality cells with < 2,000 total counts, < 500 number of genes, > 10% mitochondrial gene counts, and > 0.4 Scrublet (v0.2.3)[123] 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

1140 function with seurat_v3 method[124] for each dataset, and principal component analysis (PCA) was
1141 performed with the selected highly variable genes. The neighborhood graph was built from the top 50
1142 principal components and visualized using UMAP[125]. Clustering was performed with the Leiden method[126]
1143 with 0.5 resolution, and the Wilcoxon rank-sum test[127] calculate marker genes of each cluster. Cell cycle
1144 scores were calculated with previously defined cell cycle genes[128] using
1145 scanpy.tll.score_genes_cell_cycle function.

1146 **Drug-response assay for LysoTracker screening**

1147 Cells were treated with TAS-102 at concentrations ranging from 0.049 to 108 μM in a three-fold dilution
1148 series for 24, 48, or 72 hrs to establish the optimal conditions for induced lysosomal activation. DMSO
1149 vehicle control and single ORF overexpressing cell lines of CDK6 and BCL2 were included as drug
1150 sensitivity and resistance positive controls, respectively. Following TAS-102 treatment, lysosomal activity
1151 was assessed by staining cells with 50 nM LysoTracker Green DND-26 (Thermo Fisher Scientific Inc.,
1152 L7526) for 20 minutes at 37°C. After two washes with DPBS, cells were analyzed using a CytoFlex flow
1153 cytometer (Beckman Coulter, C09766) to measure green fluorescence as an indicator of lysosomal
1154 activity.

1155 **LysoTracker screening**

1156 The FACS method followed the protocol from the previous study[129]. HeLa ORFeome cells were thawed
1157 and pooled to maximize ORF library coverage, then allowed to recover for three days before the
1158 experiment. For the screening, 3 million cells were seeded per 15 cm dish for the DMSO control, and 20
1159 million cells were seeded for TAS-102 treatment to ensure sufficient representation of potential resistant
1160 populations. Drug treatment was initiated while cells were still in suspension, before full adherence.
1161 Following 72 hrs of TAS-102 treatment, lysosomal activity was assessed by staining cells with 50 nM
1162 LysoTracker Green for 20 minutes at 37°C. After two DPBS washes, cells were sorted on a Sony MA900
1163 flow cytometry cell sorter, gating for mCherry positivity and then for enhanced lysosome activity (High
1164 green fluorescence, upper 5%) and diminished lysosome activity (Low green fluorescence, lower 5%)
1165 populations. Sorted cells were collected, seeded in 6-well plates, and expanded for one week before
1166 being cultured to full confluence in 15 cm dishes. Genomic DNA extraction, PCR amplification, and
1167 sequencing were performed as previously described for chemotherapeutic screening. Sequencing data
1168 from the sorted populations were analyzed using the ORFeome bioinformatics pipeline to prioritize
1169 candidate genes and compare the findings with those from the TAS-102 chemoresistance screen.
1170 Survival analysis and patient stratification based on the selected genes were performed using data from
1171 the TCGA PanCancer Atlas[32] via cBioPortal[130] (accessed on July 7th, 2025), and visualized using
1172 Matplotlib.

1173 **Western Blot Analysis for Lysosomal Pathway Protein Detection**

1174 HeLa cells were treated with DMSO, 1 μM TAS-102, or 1 μM ABT-263 (MedChemExpress, HY-10087),
1175 1 μM TAS-102 + 1 μM ABT-263 in combination, and HCT 116 were treated with DMSO, 1 μM TAS-102,
1176 or 0.5 μM ABT-263, 1 μM TAS-102 + 1 μM ABT-263 in combination. After 48 hrs of incubation, cells were
1177 lysed in sample buffer (60 mM Tris-HCl (pH 6.8), 2% SDS, 10% glycerol, 5% β-mercaptoethanol, and
1178 0.01% bromophenol blue) or EBC lysis buffer (50 mM Tris-HCl (pH 8.0), 200 mM NaCl and 0.5% NP-40),
1179 supplemented with a protease inhibitor cocktail (Sigma-Aldrich, 11836153001) for 30 min on ice. Total
1180 cell lysates were sonicated three times for 10 sec each and centrifuged at 15,000 ×g at 4˚C for 15 min.
1181 The supernatants were collected, and protein concentration was measured using the BCA protein assay
1182 (Thermo Fisher Scientific, Inc., 23227). Equal amounts of protein were separated by 8-15% SDS–PAGE
1183 and transferred to a PVDF membrane (Bio-Rad Laboratories, Inc., 1620177). The membranes were
1184 blocked in 1X Tris-buffered saline with 0.1% Tween 20 and 5% skim milk (CellNest, CNS109) at room
1185 temperature for 30 min and incubated with primary antibodies at 4˚C overnight or room temperature for
1186 1 hr and secondary antibodies at room temperature for 1 hr. The signal was visualized with ImageQuant™
1187 800 (Cytiva™, 29399481) using an enhanced chemiluminescence detection system (GenDEPOT, LLC,
1188 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

1189 **Immunofluorescence assay**

1190 HCT 116 cells were seeded at a density of 1 x 10^5 cells per well in 12-well plates. On the following day,
1191 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
1192 combination. After 24 hrs of incubation, the cells were fixed with 2% paraformaldehyde at room
1193 temperature for 10 min and permeabilized with 0.5% Triton X-100 for 5 min. Next, the cells were treated
1194 with 2% bovine serum albumin (A0100-010; GenDEPOT) for 1 h and then incubated with primary
1195 antibodies in 2% BSA at 4°C overnight. The next day, the cells were treated with Alexa Fluor 488 goat
1196 IgG antibody (A-11034; Thermo Fisher Scientific Inc.) in 2% BSA at room temperature for 1 h. Nuclei
1197 were stained with 1 μg/mL Hoechst 33342 (B2261; Sigma) for 5 min at room temperature. Finally, the
1198 coverglass slips were then mounted onto glass slides using an anti-fade fluorescence mounting medium
1199 (ab104135; Abcam). Confocal imaging (LSM800; CLSM) was subsequently performed. The primary
1200 antibodies used were against TFEB (1:500; 4240; Cell Signaling). Nucleus fluorescence was measured
1201 using ImageJ software (v1.53K; National Institutes of Health). Statistical analysis was performed using
1202 Student's T-test, followed by Bonferroni's multiple comparison testing.

1203 **DQ-BSA assay**

1204 To assess lysosomal proteolytic activity, the DQ-BSA assay was performed as previously described[131].
1205 HCT 116 cells were seeded at 1 x 10^5 cells per well on coverglass slips in a 12-well plate. The following
1206 day, cells were treated with 2 μM TAS-102 or 0.5 μM ABT-263 or a combination of both. Following 24
1207 hrs of drug treatment, cells were incubated with 10 μg/mL DQ™ Red-BSA (D12051; Thermo Fisher
1208 Scientific Inc.) at 37°C for 24 hrs in the dark. After incubation, cells were washed with DPBS and fixed in
1209 2% formaldehyde for 10 min at room temperature. Nuclei were stained with Hoechst 33342 (1 μg/mL).
1210 The coverglass slips were then mounted onto glass slides using an anti-fade fluorescence mounting
1211 medium (ab104135; Abcam). Confocal imaging (LSM800; CLSM) was subsequently performed.
1212 Statistical analysis was performed using Student's T-test, followed by Bonferroni's multiple comparison
1213 testing.

1214 **Measurement of lysosomal pH**

1215 To assess lysosomal acidification, the lysosomal pH was measured as previously described[132]. HCT 116
1216 cells were seeded at 1 x 10^4 cells per well into a 96-well plate. The following day, cells were incubated in
1217 complete media containing 2 μM TAS-102 or 0.5 μM ABT-263 or 2 μM TAS-102 + 0.5 μM ABT-263 in

1218 combination. Following 24 hrs of drug treatment, cells were incubated with 1 μM LysoSensor™
1219 Yellow/Blue DND-160 (L7545; Thermo Fisher Scientific Inc.) for 5 min in the dark. After incubation, cells
1220 were washed with DPBS, and then the media was replaced with HBSS supplemented with 10% FBS for
1221 10 min for test wells. To generate a standard curve, cells were incubated with a buffer (25 mM HEPES,
1222 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
1223 7.0 as described previously. Fluorescence was measured using a PerkinElmer EnSpire multimode plate
1224 reader with excitation/emission filters set to 329/440 nm and 380/540 nm. Statistical analysis was
1225 performed using Student's T-test, followed by Bonferroni's multiple comparison testing.

1226 **Cell proliferation assay**

1227 To assess the cell proliferation difference in drug treatment, the proliferation was measured using a
1228 method adopted from the previous protocol[133]. HCT 116 cells were seeded in a 12-well plate at a density
1229 of 5×10^4 cells per well. The following day, cells were treated with 2 μM TAS-102 or 100 nM Bafilomycin
1230 A1 or 2 μM TAS-102 + 100 nM Bafilomycin A1 in combination. Cells were harvested and counted at 0,
1231 24, and 48 hrs after drug treatment. Each counting was performed in six independent replicates for each
1232 time point, and the mean values were used to generate the proliferation graphs. Statistical analysis was
1233 performed using Student's T-test, followed by Bonferroni's multiple comparison testing.

1234 **Cytotoxicity assay**

1235 To examine drug cytotoxicity in human cells, we followed the protocol from the previous publication[134].
1236 Cells were plated at a density of 5,000 cells/well in a 96-well plate and treated with the appropriate drugs
1237 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
1238 with media containing one of 0, 25, 50, 100, 200, 500, 1000, 2000, 5000 nM concentration of ABT-263,
1239 followed by a 96-hour incubation. Cells were stained with 0.5% Methylene Blue for 30 minutes, rinsed
1240 with water, dried, solubilized in 5% SDS in DPBS, and read at 650 nm. The average of the control
1241 normalized the raw absorbance. The graph was generated using GraphPad Prism 9, and statistical
1242 analysis was performed using two-way ANOVA followed by Tukey's multiple comparisons test.

1243 **Cell Line-Derived Xenografts (CDX) Model**

1244 Female SCID/CB17 mice, aged 6 to 8 weeks, were sourced from a colony bred and maintained at
1245 RPCCC. Mice were inoculated subcutaneously into the left flank with exponentially growing HCT 116
1246 tumor cells (1×10^6 per mouse) suspended in a 1:1 mixture of DPBS and Matrigel. Mice were monitored
1247 daily, while body weight and tumor size were measured twice a week. Tumor dimensions were measured
1248 with digital calipers and calculated using the formula: (length × width^2)/2. When the tumor volume
1249 reached 50 to 100mm^3, mice were randomly divided into four treatment groups (n = 10-11 for each group):

1250 (1) Vehicle Control, (2) 50 mg/kg ABT-263, (3) 50 mg/kg TAS-102 (MedChemExpress, Cat. No. HY-
1251 16478), and (4) Combination (50 mg/kg TAS-102 + 50 mg/kg ABT-263). TAS-102 was dissolved in 10%
1252 2-hydroxypropyl-β-cyclodextrin (HPCD) in DPBS, while ABT-263 was prepared in 10% DMSO, 40%
1253 PEG300, 5% Tween-80, and 45% saline. For the Combination group, TAS-102 was given 4 hrs before
1254 ABT-263. All drugs were freshly prepared and given via oral gavage on a 5-day-on, 2-days-off schedule.

1255 **Patient-Derived Xenograft (PDX) Model**

1256 Donor and experimental colorectal patient-derived xenografts (CRC PDX-03-26-4p) were grown in male
1257 SCID/CB17 mice from RPCCC. Tumors were grown in donor mice to a volume of around 1500 mm^3,
1258 harvested rapidly after euthanasia, cut into 2×2×2 mm^3 pieces under sterile conditions, and surgically
1259 implanted into the left flank. Surgical staples were removed about ten days post-implantation. Tumor
1260 growth and mouse weights were measured twice a week. Once the average tumor volume reached 50-
1261 100 mm^3, mice were randomized into four groups like CDX model (1) Vehicle Control, (2) 50 mg/kg ABT-
1262 263, (3) 50 mg/kg TAS-102, and (4) Combination (50 mg/kg TAS-102 + 50 mg/kg ABT-263). The drugs
1263 were prepared as in the CDX experiment and administered via the same route and dosing schedule.

1264 **Mouse Model Monitoring**

1265 Mice were housed in micro-isolator units and were provided with food and water ad libitum. Daily
1266 monitoring of animal health, body weight, and tumor growth, as well as endpoint procedures, following
1267 the protocol and guidelines approved by the Institutional Animal Care and Use Committee (IACUC). The
1268 facility is certified by the Association for Assessment and Accreditation of Laboratory Animal Care
1269 (AAALAC) International and complies with regulations and standards of the US Department of Agriculture
1270 and the US Department of Health and Human Services.

1271 At the endpoint, when the tumor diameter reached 20 mm, mice were euthanized and subjected to
1272 necropsy and organ collection. Tumor growth assessment was stopped once a mouse reached the
1273 endpoint (typically in a vehicle-treated cohort) while the survival data were still collected in the study.
1274 Intracardiac blood was collected and stored in EDTA-coated tubes at 4 °C for a complete blood count
1275 test.

1276 **Complete Blood Counts**

1277 Blood was collected by cardiac puncture into a tube containing EDTA solution at the endpoint of the
1278 study[134]. Analysis was performed with 20 μL blood samples using the Heska HemaTrue Analyzer and
1279 HeskaView Integrated Software version 2.5.2.

1280 **Statistics for *In Vivo* data**

1281 Tumor Growth Inhibition (TGI) for each mouse was calculated as the relative difference between the
1282 average tumor volume of the Vehicle Control group and the tumor volume of the individual mouse with
1283 respect to the Vehicle Control group average. Statistically significant differences were determined using
1284 one-way ANOVA for TGI and CBC and a two-way ANOVA test for tumor growth curves. Multiple testing
1285 was performed using Tukey's multiple comparisons in GraphPad Prism 9.

1286 ## References

1287 1. Letai, A. & de The, H. Conventional chemotherapy: millions of cures, unresolved therapeutic index.
1288 *Nat. Rev. Cancer* **25**, 209–218 (2025).
1289 2. Anand, U. *et al.* Cancer chemotherapy and beyond: Current status, drug candidates, associated
1290 risks and progress in targeted therapeutics. *Genes Dis.* **10**, 1367–1401 (2023).
1291 3. Flaherty, K. T. Chemotherapy and targeted therapy combinations in advanced melanoma. *Clin.*
1292 *Cancer Res.* **12**, 2366s–2370s (2006).
1293 4. Lopez, J. P. *et al.* Gefitinib inhibition of drug resistance to doxorubicin by inactivating ABCG2 in
1294 thyroid cancer cell lines. *Arch. Otolaryngol. Head. Neck Surg.* **133**, 1022–1027 (2007).
1295 5. Kosorok, M. R. & Laber, E. B. Precision medicine. *Annu. Rev. Stat. Appl.* **6**, 263–286 (2019).
1296 6. Mucaki, E. J., Zhao, J. Z. L., Lizotte, D. J. & Rogan, P. K. Predicting responses to platin
1297 chemotherapy agents with biochemically-inspired machine learning. *Signal Transduct. Target. Ther.*
1298 **4**, 1 (2019).
1299 7. Ali, M., Khan, S. A., Wennerberg, K. & Aittokallio, T. Global proteomics profiling improves drug
1300 sensitivity prediction: results from a multi-omics, pan-cancer modeling approach. *Bioinformatics* **34**,
1301 1353–1362 (2018).
1302 8. Menghi, F. & Liu, E. T. Functional genomics of complex cancer genomes. *Nat. Commun.* **13**, 5908
1303 (2022).
1304 9. Zhou, Z. *et al.* Genome-scale CRISPR-Cas9 knockout screening in nasopharyngeal carcinoma for
1305 radiosensitive and radioresistant genes. *Transl. Oncol.* **30**, 101625 (2023).
1306 10. Ezkurdia, I. *et al.* Most highly expressed protein-coding genes have a single dominant isoform. *J.*
1307 *Proteome Res.* **14**, 1880–1887 (2015).
1308 11. Zhao, W.-S. *et al.* Genome-scale CRISPR activation screening identifies a role of ELAVL2-CDKN1A
1309 axis in paclitaxel resistance in esophageal squamous cell carcinoma. *Am. J. Cancer Res.* **9**, 1183–
1310 1200 (2019).

1311 12. Arnoldo, A. *et al.* A genome scale overexpression screen to reveal drug activity in human cells.
1312 *Genome Med.* **6**, 32 (2014).
1313 13. Dull, T. *et al.* A third-generation lentivirus vector with a conditional packaging system. *J. Virol.* **72**,
1314 8463–8471 (1998).
1315 14. Alyateem, G. *et al.* Use of CRISPR-based screens to identify mechanisms of chemotherapy
1316 resistance. *Cancer Gene Ther.* **30**, 1043–1050 (2023).
1317 15. Bester, A. C. *et al.* An Integrated Genome-wide CRISPRa Approach to Functionalize lncRNAs in
1318 Drug Resistance. *Cell* **173**, 649–664.e20 (2018).
1319 16. Neupane, M. *et al.* MECP2 Is a Frequently Amplified Oncogene with a Novel Epigenetic Mechanism
1320 That Mimics the Role of Activated RAS in Malignancy. *Cancer Discov.* **6**, 45–58 (2016).
1321 17. Bulcha, J. T., Wang, Y., Ma, H., Tai, P. W. L. & Gao, G. Viral vector platforms within the gene therapy
1322 landscape. *Signal Transduct. Target. Ther.* **6**, 53 (2021).
1323 18. Liu, J. *et al.* Increasing genome editing efficiency of Cas9 nucleases by the simultaneous use of
1324 transcriptional activators and histone acetyltransferase activator. *CRISPR J.* **5**, 854–867 (2022).
1325 19. Charrier, S. *et al.* Quantification of lentiviral vector copy numbers in individual hematopoietic colony-
1326 forming cells shows vector dose-dependent effects on the frequency and level of transduction. *Gene*
1327 *Ther.* **18**, 479–487 (2011).
1328 20. Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248–1250
1329 (2010).
1330 21. Saranraj, K. & Kiran, P. U. Drug repurposing: Clinical practices and regulatory pathways. *Perspect.*
1331 *Clin. Res.* **16**, 61–68 (2025).
1332 22. Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature* **580**, 402–408
1333 (2020).
1334 23. Kim, D.-K. *et al.* A proteome-scale map of the SARS-CoV-2-human contactome. *Nat. Biotechnol.*
1335 **41**, 140–149 (2023).
1336 24. Matreyek, K. A., Stephany, J. J., Chiasson, M. A., Hasle, N. & Fowler, D. M. An improved platform

1337 for functional assessment of large protein libraries in mammalian cells. *Nucleic Acids Res.* **48**, e1
1338 (2020).
1339 25. Landry, J. J. M. *et al.* The genomic and transcriptomic landscape of a HeLa cell line. *G3 (Bethesda)*
1340 **3**, 1213–1224 (2013).
1341 26. Matreyek, K. A., Stephany, J. J. & Fowler, D. M. A platform for functional assessment of large variant
1342 libraries in mammalian cells. *Nucleic Acids Res.* **45**, e102 (2017).
1343 27. Álvarez-Fernández, M. & Malumbres, M. Mechanisms of sensitivity and resistance to CDK4/6
1344 inhibition. *Cancer Cell* **37**, 514–529 (2020).
1345 28. Cortes, E. P. Doxorubicin in Disseminated. *JAMA* **221**, 1132 (1972).
1346 29. Andréll, J. & Tate, C. G. Overexpression of membrane proteins in mammalian cells for structural
1347 studies. *Mol. Membr. Biol.* **30**, 52–63 (2013).
1348 30. Larionova, I. *et al.* Interaction of tumor-associated macrophages and cancer chemotherapy.
1349 *Oncoimmunology* **8**, 1596004 (2019).
1350 31. Zhang, Y. *et al.* TheMarker: a comprehensive database of therapeutic biomarkers. *Nucleic Acids*
1351 *Res.* **52**, D1450–D1464 (2024).
1352 32. Cancer Genome Atlas Research Network *et al.* The Cancer Genome Atlas Pan-Cancer analysis
1353 project. *Nat. Genet.* **45**, 1113–1120 (2013).
1354 33. Moiso, E. Manual curation of TCGA treatment data and identification of potential markers of therapy
1355 response. medRxiv (2021).
1356 34. Gmeiner, W. H. & Okechukwu, C. C. Review of 5-FU resistance mechanisms in colorectal cancer:
1357 clinical significance of attenuated on-target effects. *Canc. Drug Resist.* **6**, 257–272 (2023).
1358 35. Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular function.
1359 *Science* **353**, aaf1420–aaf1420 (2016).
1360 36. Baryshnikova, A. Spatial Analysis of Functional Enrichment (SAFE) in Large Biological Networks.
1361 *Methods Mol. Biol.* **1819**, 249–268 (2018).
1362 37. Vogler, M. *et al.* The BCL2 family: from apoptosis mechanisms to new advances in targeted therapy.

1363 *Signal Transduct. Target. Ther.* **10**, 91 (2025).
1364 38. Qian, S. *et al.* The role of BCL-2 family proteins in regulating apoptosis and cancer therapy. *Front.*
1365 *Oncol.* **12**, 985363 (2022).
1366 39. Szklarczyk, D. *et al.* The STRING database in 2025: protein networks with directionality of regulation.
1367 *Nucleic Acids Res.* **53**, D730–D737 (2025).
1368 40. Wu, P.-H. *et al.* Lysosomal trafficking mediated by Arl8b and BORC promotes invasion of cancer
1369 cells that survive radiation. *Commun. Biol.* **3**, 620 (2020).
1370 41. Chen, Y. *et al.* Roles of the adaptor protein tumor necrosis factor receptor type 1-associated death
1371 domain protein (TRADD) in human diseases. *Biomed. Pharmacother.* **153**, 113467 (2022).
1372 42. Jian, W. & Zhang, L. POLE2 silencing inhibits the progression of colorectal carcinoma cells via wnt
1373 signaling axis. *Cancer Biol. Ther.* **25**, 2392339 (2024).
1374 43. Sun, Y. *et al.* Identification and verification of YBX3 and its regulatory gene HEIH as an oncogenic
1375 system: A multidimensional analysis in colon cancer. *Front. Immunol.* **13**, 957865 (2022).
1376 44. Paek, A. R., Kim, S. H., Kim, S. S., Kim, K. T. & You, H. J. IGF-1 induces expression of zinc-finger
1377 protein 143 in colon cancer cells through phosphatidylinositide 3-kinase and reactive oxygen
1378 species. *Exp. Mol. Med.* **42**, 696–702 (2010).
1379 45. Littlepage, L. E. *et al.* The transcription factor ZNF217 is a prognostic biomarker and therapeutic
1380 target during breast cancer progression. *Cancer Discov.* **2**, 638–651 (2012).
1381 46. Alikarami, F. *et al.* GATA2 links stemness to chemotherapy resistance in acute myeloid leukemia.
1382 *Blood* **145**, 2179–2195 (2025).
1383 47. Uboha, N. & Hochster, H. S. TAS-102: a novel antimetabolite for the 21st century. *Future Oncol.* **12**,
1384 153–163 (2016).
1385 48. Tse, C. *et al.* ABT-263: a potent and orally bioavailable Bcl-2 family inhibitor. *Cancer Res.* **68**, 3421–
1386 3428 (2008).
1387 49. Sun, Y. *et al.* Hypoxia-induced autophagy reduces radiosensitivity by the HIF-1α/miR-210/Bcl-2
1388 pathway in colon cancer cells. *Int. J. Oncol.* **46**, 750–756 (2015).

1389 50. Yu, L. & Liu, S. Autophagy contributes to modulating the cytotoxicities of Bcl-2 homology domain-3
1390 mimetics. *Semin. Cancer Biol.* **23**, 553–560 (2013).
1391 51. Levine, B., Sinha, S. & Kroemer, G. Bcl-2 family members: dual regulators of apoptosis and
1392 autophagy. *Autophagy* **4**, 600–606 (2008).
1393 52. Cheng, L. *et al.* mTOR-dependent TFEB activation and TFEB overexpression enhance autophagy-
1394 lysosome pathway and ameliorate Alzheimer's disease-like pathology in diabetic encephalopathy.
1395 *Cell Commun. Signal.* **21**, 91 (2023).
1396 53. Schläfli, A. M., Berezowska, S., Adams, O., Langer, R. & Tschan, M. P. Reliable LC3 and p62
1397 autophagy marker detection in formalin fixed paraffin embedded human tissue by
1398 immunohistochemistry. *Eur. J. Histochem.* **59**, 2481 (2015).
1399 54. Frost, L. S., Dhingra, A., Reyes-Reveles, J. & Boesze-Battaglia, K. The use of DQ-BSA to monitor
1400 the turnover of autophagy-associated cargo. *Methods Enzymol.* **587**, 43–54 (2017).
1401 55. Jeong, S.-J. *et al.* Trehalose causes low-grade lysosomal stress to activate TFEB and the
1402 autophagy-lysosome biogenesis response. *Autophagy* **17**, 3740–3752 (2021).
1403 56. Zhang, J. *et al.* Enabling tumor-specific drug delivery by targeting the Warburg effect of cancer. *Cell*
1404 *Rep. Med.* **6**, 101920 (2025).
1405 57. Corcoran, R. B. *et al.* Synthetic lethal interaction of combined BCL-XL and MEK inhibition promotes
1406 tumor regressions in KRAS mutant cancer models. *Cancer Cell* **23**, 121–128 (2013).
1407 58. Zhou, Y., Perez, R. E., Duan, L. & Maki, C. G. DZNep represses Bcl-2 expression and modulates
1408 apoptosis sensitivity in response to Nutlin-3a. *Cancer Biol. Ther.* **19**, 465–474 (2018).
1409 59. Tolani, B. *et al.* Ras-mutant cancers are sensitive to small molecule inhibition of V-type ATPases in
1410 mice. *Nat. Biotechnol.* **40**, 1834–1844 (2022).
1411 60. Kawakami, H. *et al.* Mutant BRAF upregulates MCL-1 to confer apoptosis resistance that is reversed
1412 by MCL-1 antagonism and cobimetinib in colorectal cancer. *Mol. Cancer Ther.* **15**, 3015–3027
1413 (2016).
1414 61. Cicero, G. *et al.* TAS-102 in metastatic colorectal cancer (mCRC): efficacy, tolerability, and quality

1415 of life in heavily pretreated elderly patients: a real-life study. *Drugs Context* **9**, (2020).

1416 62. Mohamad Anuar, N. N., Nor Hisam, N. S., Liew, S. L. & Ugusman, A. Clinical review: Navitoclax as
1417 a pro-apoptotic and anti-fibrotic agent. *Front. Pharmacol.* **11**, 564108 (2020).

1418 63. Low, B. E., Hosur, V., Lesbirel, S. & Wiles, M. V. Efficient targeted transgenesis of large donor DNA
1419 into multiple mouse genetic backgrounds using bacteriophage Bxb1 integrase. *Sci. Rep.* **12**, 5424
1420 (2022).

1421 64. Yang, H., Sun, L., Liu, M. & Mao, Y. Patient-derived organoids: a promising model for personalized
1422 cancer treatment. *Gastroenterol. Rep. (Oxf.)* **6**, 243–245 (2018).

1423 65. Quintanal-Villalonga, Á. *et al.* Lineage plasticity in cancer: a shared pathway of therapeutic
1424 resistance. *Nat. Rev. Clin. Oncol.* **17**, 360–371 (2020).

1425 66. Dixit, A. *et al.* Perturb-seq: Dissecting molecular circuits with scalable single-cell RNA profiling of
1426 pooled genetic screens. *Cell* **167**, 1853–1866.e17 (2016).

1427 67. Madrid, P. B. *et al.* A systematic screen of FDA-approved drugs for inhibitors of biological threat
1428 agents. *PLoS One* **8**, e60579 (2013).

1429 68. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–
1430 1251 (2013).

1431 69. Youn, J.-Y. *et al.* High-density proximity mapping reveals the subcellular organization of mRNA-
1432 associated granules and bodies. *Mol. Cell* **69**, 517–532.e11 (2018).

1433 70. Waye, J. S. Forensic identity testing using highly polymorphic DNA markers: current status and
1434 emerging technologies. *Transfus. Med. Rev.* **7**, 193–205 (1993).

1435 71. Chiasson, M. A. *et al.* Multiplexed measurement of variant abundance and activity reveals VKOR
1436 topology, active site and human variant impact. *Elife* **9**, (2020).

1437 72. Sur, S. *et al.* A panel of isogenic human cancer cells suggests a therapeutic approach for cancers
1438 with inactivated p53. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3964–3969 (2009).

1439 73. Inoue, A. *et al.* Sequential administration of XPO1 and ATR inhibitors enhances therapeutic
1440 response in TP53-mutated colorectal cancer. *Gastroenterology* **161**, 196–210 (2021).

1441 74. Wiemann, S. *et al.* The ORFeome Collaboration: a genome-scale human ORF-clone resource. *Nat.*
1442 *Methods* **13**, 191–192 (2016).
1443 75. Yachie, N. *et al.* Pooled-matrix protein interaction screens using Barcode Fusion Genetics. *Mol. Syst.*
1444 *Biol.* **12**, 863 (2016).
1445 76. Mills, C. C., Kolb, E. A. & Sampson, V. B. Development of chemotherapy with cell-cycle inhibitors
1446 for adult and pediatric cancer therapy. *Cancer Res.* **78**, 320–325 (2018).
1447 77. Dong, C. *et al.* A genome-wide CRISPR-Cas9 knockout screen identifies essential and growth-
1448 restricting genes in human trophoblast stem cells. *Nat. Commun.* **13**, 2548 (2022).
1449 78. Ortega, P., Sanchez, A., Seldin, M. & Buisson, R. Oligo-seq protocol for mapping DNA motifs
1450 targeted by base editors. *STAR Protoc.* **6**, 103758 (2025).
1451 79. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
1452 *EMBnet.journal* **17**, 10–12 (2011).
1453 80. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data.
1454 *Bioinformatics* **30**, 2114–2120 (2014).
1455 81. Gene [Internet]. Bethesda (MD): National Library of Medicine (US), *National Center for*
1456 *Biotechnology Information*; (2004) [cited 2023 Aug 23]. Available from:
1457 https://www.ncbi.nlm.nih.gov/gene/
1458 82. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–
1459 359 (2012).
1460 83. MGC Project Team et al. The completion of the Mammalian Gene Collection (MGC). *Genome Res*.
1461 **19**, 2324–2333 (2009).
1462 84. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Available
1463 online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
1464 85. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple
1465 tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
1466 86. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079

1467 (2009).
1468 87. Zhu, Q., Liu, N., Orkin, S. H. & Yuan, G.-C. CUT&RUNTools: a flexible pipeline for CUT&RUN
1469 processing and footprint analysis. *Genome Biol.* **20**, 192 (2019).
1470 88. Alerasool, N., Leng, H., Lin, Z.-Y., Gingras, A.-C. & Taipale, M. Identification and functional
1471 characterization of transcriptional activators in human cells. *Mol. Cell* **82**, 677–695.e7 (2022).
1472 89. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of
1473 RNA-seq data. *Genome Biol.* **11**, R25 (2010).
1474 90. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential
1475 expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
1476 91. Smyth, G. K. limma: Linear Models for Microarray Data. in *Bioinformatics and Computational Biology*
1477 *Solutions Using R and Bioconductor* (eds. Gentleman, R., Carey, V. J., Huber, W., Irizarry, R. A. &
1478 Dudoit, S.) 397–420 (Springer New York, New York, NY, 2005).
1479 92. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and
1480 Projection. *J. Open Source Softw.* **3**, 861 (2018).
1481 93. Abbas-Aghababazadeh, F., Li, Q. & Fridley, B. L. Comparison of normalization approaches for gene
1482 expression studies completed with high-throughput sequencing. *PLoS One* **13**, e0206312 (2018).
1483 94. Signorell, A. DescTools: Tools for descriptive statistics. R package version 0.99. 20. *Cranio* (2017).
1484 95. Hwang, D. *et al.* A data integration methodology for systems biology. *Proc. Natl. Acad. Sci. U. S. A.*
1485 **102**, 17296–17301 (2005).
1486 96. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
1487 97. Waskom, M. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
1488 98. Pope, B. & Kent, H. M. High efficiency 5 min transformation of Escherichia coli. *Nucleic Acids Res.*
1489 **24**, 536–537 (1996).
1490 99. Girish, V. & Sheltzer, J. M. A CRISPR competition assay to identify cancer genetic dependencies.
1491 *Bio Protoc.* **10**, e3682 (2020).
1492 100. Diaz, J. E. *et al.* The transcriptomic response of cells to a drug combination is more than the sum of

1493 the responses to the monotherapies. *Elife* **9**, (2020).
1494 101. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat.*
1495 *Biotechnol.* **38**, 276–278 (2020).
1496 102. Patel, H. *et al. Nf-Core/rnaseq: Nf-Core/rnaseq v3.18.0 - Lithium Lynx*. (Zenodo, 2024).
1497 doi:10.5281/ZENODO.14537300.
1498 103. Schneider, V. A. *et al.* Evaluation of GRCh38 and de novo haploid genome assemblies
1499 demonstrates the enduring quality of the reference assembly. *Genome Res.* **27**, 849–864 (2017).
1500 104. Krueger, F., James, F., Ewels, P., Afyounian, E. & Schuster-Boeckler, B. *FelixKrueger/TrimGalore:*
1501 *v0.6.7* (Zenodo, 2021). doi:10.5281/ZENODO.5127899.
1502 105. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
1503 106. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without
1504 a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
1505 107. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in
1506 metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
1507 108. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat.*
1508 *Methods* **17**, 261–272 (2020).
1509 109. Tarazona, S. *et al.* Data quality aware analysis of differential expression in RNA-seq with NOISeq
1510 R/Bioc package. *Nucleic Acids Res.* **43**, e140 (2015).
1511 110. Melé, M. *et al.* Human genomics. The human transcriptome across tissues and individuals. *Science*
1512 **348**, 660–665 (2015).
1513 111. Capra, J. A., Williams, A. G. & Pollard, K. S. ProteinHistorian: tools for the comparative analysis of
1514 eukaryote protein origin. *PLoS Comput. Biol.* **8**, e1002567 (2012).
1515 112. Oates, M. E. *et al.* D^2P^2: database of disordered protein predictions. *Nucleic Acids Res.* **41**, D508–
1516 16 (2013).
1517 113. Sherman, B. T. *et al.* DAVID: a web server for functional enrichment analysis and functional
1518 annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221 (2022).

1519 114. Zambelli, F. *et al.* RNentropy: an entropy-based tool for the detection of significant variation of gene
1520 expression across multiple RNA-Seq experiments. *Nucleic Acids Res.* **46**, e46 (2018).
1521 115. Shifrut, E. *et al.* Genome-wide CRISPR screens in primary human T cells reveal key regulators of
1522 immune function. *Cell* **175**, 1958–1971.e15 (2018).
1523 116. Salkovic, E., Sadeghi, M. A., Baggag, A., Salem, A. G. R. & Bensmail, H. OutSingle: a novel method
1524 of detecting and injecting outliers in RNA-Seq count data using the optimal hard threshold for singular
1525 values. *Bioinformatics* **39**, btad142 (2023).
1526 117. Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science* **357**, (2017).
1527 118. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular
1528 interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
1529 119. Berriz, G. F., King, O. D., Bryant, B., Sander, C. & Roth, F. P. Characterizing gene sets with
1530 FuncAssociate. *Bioinformatics* **19**, 2502–2504 (2003).
1531 120. Novikov, A. PyClustering: Data Mining Library. *J. Open Source Softw.* **4**, 1230 (2019).
1532 121. Xie, Z. *et al.* Gene set knowledge discovery with Enrichr. *Curr. Protoc.* **1**, e90 (2021).
1533 122. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data
1534 analysis. *Genome Biol.* **19**, 1–5 (2018).
1535 123. Wolock, L. S., Lopez, R. & Klein, M. A. Scrublet: Computational Identification of Cell Doublets in
1536 Single-Cell Transcriptomic Data. *Cell Systems* **8**, 281–291.e9 (2019).
1537 124. Stuart, T. et al. Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
1538 125. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for
1539 Dimension Reduction. (2018).
1540 126. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected
1541 communities. *Sci. Rep.* **9**, 1–12 (2019).
1542 127. Sun, Y. & Qiu, P. Hierarchical marker genes selection in scRNA-seq analysis. *PLoS Comput. Biol.*
1543 **20**, e1012643 (2024).
1544 128. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-

1545 seq. *Science* **352**, 189–196 (2016).
1546 129. Kim, M. *et al.* Exogenous RNA surveillance by proton-sensing TRIM25. *Science* **388**, eads4539
1547 (2025).
1548 130. de Bruijn, I. *et al.* Analysis and visualization of longitudinal genomic and clinical data from the
1549 AACR Project GENIE Biopharma Collaborative in cBioPortal. *Cancer Res.* **83**, 3861–3867 (2023).
1550 131. Marwaha, R. & Sharma, M. DQ-Red BSA trafficking assay in cultured cells to assess cargo
1551 delivery to lysosomes. *Bio Protoc.* **7**, (2017).
1552 132. Ma, L., Ouyang, Q., Werthmann, G. C., Thompson, H. M. & Morrow, E. M. Live-cell microscopy
1553 and fluorescence-based measurement of luminal pH in intracellular organelles. *Front. Cell Dev.*
1554 *Biol.* **5**, 71 (2017).
1555 133. Morten, B. C., Scott, R. J. & Avery-Kiejda, K. A. Comparison of three different methods for
1556 determining cell proliferation in breast cancer cell lines. *J. Vis. Exp.* (2016) doi:10.3791/54350.
1557 134. Alruwaili, M. M. *et al.* A synergistic two-drug therapy specifically targets a DNA repair dysregulation
1558 that occurs in p53-deficient colorectal and pancreatic cancers. *Cell Rep. Med.* **5**, 101434 (2024).