Single-Cell Pharmacogenomic Landscapes of Epigenetic Therapy in Gastric Cancer

Chang Xu[1*], Haoran Ma[1*], Taotao Sheng[2*], Ming Hui Lee[1], Siti Aishah Binte Abdul Ghani[1], Laura Perlaza-Jimenez[3,4], Kie Kyon Huang[1], Shen Kiat Lim[1], Supriya Srivastava[5], Xuewen Ong[1], Su Ting Tay[1], Shamaine Wei Ting Ho[2], Angie Lay Keng Tan[1], Feng Zhu[5], Vincenzo Nasca[6], Giovanni Randon[6], Hassan Ashktorab[7], Alfred Sze-Lok Cheng[8], Anand D. Jeyasekharan[5,9,10,11], Shang Li[1,12], Ming Teh[13], Raghav Sundar[1,14,15], David R. Powell[4], Joseph Rosenbluh[3,16], Filippo Pietrantonio[17], Wei Peng Yong[9,10], Jimmy Bok-Yan So[11,14,15,18,19], Patrick Tan[1,2,9,12,15,20,21,#]

Affiliations

[1]Cancer and Stem Cell Biology Program, Duke-NUS Medical School, Singapore
[2]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore
[3]Cancer Research Program and Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, VIC, Australia
[4]Bioinformatics Platform, Monash University, Clayton, VIC, Australia
[5]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[6]Medical Oncology Department, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
[7]Department of Medicine, Howard University, Washington, DC, USA
[8]School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong, China
[9]Cancer Science Institute of Singapore, National University of Singapore, Singapore
[10]Department of Haematology-Oncology, National University Health System, Singapore, Singapore
[11]NUS Centre for Cancer Research, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[12]Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[13]Department of Pathology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[14]Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[15]Singapore Gastric Cancer Consortium, Singapore
[16]Functional Genomics Platform, Monash University, Clayton, VIC, 3800, Australia
[17]Gastrointestinal Oncology Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
[18]Department of Surgery, University Surgical Cluster, National University Health System, Singapore
[19]Division of Surgical Oncology, National University Cancer Institute, Singapore
[20]SingHealth/Duke-NUS Institute of Precision Medicine, National Heart Centre Singapore, Singapore
[21]Cellular and Molecular Research, National Cancer Centre, Singapore

*Equal Contribution

#Corresponding author
Prof Patrick Tan
Email: gmstanp@duke-nus.edu.sg
Cancer and Stem Cell Biology Program, Duke-NUS Medical School, Singapore

Running title: Single-Cell Pharmacogenomics of Epigenetic Therapy in GC

Grant support and Acknowledgements
This research was supported the National Medical Research Council grant MOH-000967. This work was also supported by the National Research Foundation, Singapore, and Singapore Ministry of Health's National Medical Research Council under its Open Fund-Large Collaborative Grant ("OF-LCG") (MOH-OFLCG18May-0003). This research was also supported by the Duke-NUS Core funding. We thank Duke-NUS Genome Biology Facility (DGBF) for supporting reagent procurement and preparation of sequencing libraries.

Declaration of interests
PT has stock and other ownership interests in Tempus AI, previous research funding from Kyowa Hakko Kirin and Thermo Fisher Scientific, and patents/other intellectual property through the Agency for Science and Technology Research, Singapore (all outside the submitted work). Filippo Pietrantonio reported receiving Research funding (to Institution) from Lilly, BMS, Incyte, AstraZeneca, Amgen, Agenus, Rottapharm. Personal honoraria as an invited speaker from BeiGene, Daiichi-Sankyo, Seagen, Astellas, Ipsen, AstraZeneca, Servier, Bayer, Takeda, Johnson & Johnson, BMS, MSD, Amgen, Merck-Serono, Pierre-Fabre. Advisory/Consultancy from BMS, MSD, Amgen, Pierre-Fabre, Johnson & Johnson, Servier, Bayer, Takeda, Astellas, GSK, Daiichi-Sankyo, Pfizer, BeiGene, Jazz Pharmaceuticals, Incyte, Rottapharm, Merck-Serono, Italfarmaco, Gilead, AstraZeneca, Agenus. RS reports advisory board roles for Bristol Myers Squibb, Merck, Bayer, Novartis, GSK, Astellas, Pierre-Fabre, Tavotek, Eisai, Taiho, DKSH, and MSD; honoraria for talks from Bristol Myers Squibb, Taiho, DKSH, MSD, Eli Lilly, Roche, AstraZeneca, and Ipsen; travel support from Eisai, Taiho, DKSH, Roche, AstraZeneca, Ipsen, and CytoMed Therapeutics; research funding from MSD, Natera, and Paxman Coolers; and patents from Paxman Coolers and Auristone, outside the submitted work. RS has patents pending which are licensed to Paxman and in the processing of being licensed to Auristone. All remaining authors have declared no conflicts of interest.

Abstract

Gene expression signatures ("molecular phenotypes") are extensively utilized in cancer research. Using direct capture Perturb-seq (dcPerturb-seq), we surveyed how molecular phenotypes in gastric cancer (GC) are shaped by cell-intrinsic genetic alterations interacting with cell-extrinsic therapeutic exposures. Interrogating >200 GC-related genes against 4 distinct epigenetic drug classes across multiple gastric lines, we captured 18.9 million pharmacogenomic expression interactions in 669,065 cells representing baseline and post-therapeutic molecular phenotypes. This single-cell pharmacogenomic compendium confirmed previously known gene-driven molecular phenotypes (*TP53*, *PCNA*), elucidated poorly characterized genes (*ZBTB41*), and uncovered novel gene dosage-molecular phenotype relationships. Molecular phenotypes in post-therapeutic surviving cells revealed distinct gene perturbation-induced pathways causing convergent drug resistance (EMT plasticity, cell cycle alterations, metabolic reprogramming), with *MED12*, *EP300* and *KLF4* modulating histone deacetylase inhibitor (HDACi) resistance. Mapping of *in vitro* molecular phenotypes to primary human GCs imparted prognostic information and insights into spatial heterogeneity and targeted therapy response. Our results illustrate how dcPerturb-seq can systematically and comprehensively map diverse cancer-associated molecular phenotypes across multiple gene-gene and gene-drug interactions, yielding translational insights.

Statement of Significance

By integrating single-cell perturbation screens with epigenetic drug profiling in gastric cancer, we systematically resolve how cell-intrinsic genomic pathways interact with extrinsic drug-induced responses at unprecedented resolution. Analyzing 18.9 million pharmacogenomic expression interactions across >200 cancer-related genes in 669,065 cells, our approach revealed previously unknown functional annotations, dosage-dependent effects, and convergent patterns of therapeutic resistance with direct clinical relevance. Our results provide a high-resolution roadmap contributing to a comprehensive encyclopedia of all cancer-related molecular phenotypes, accelerating translational cancer research.

Introduction

Cancer cells are dynamic cellular ecosystems exhibiting multiple functional phenotypes which are essential for malignancy, such as sustained proliferation, resistance to cell death, and evasion of host immunity [1]. Understanding how cancer phenotypes emerge from normal cells, through the complex interaction of genomic and epigenomic aberrations acquired across decades of life, represents a cornerstone of molecular cancer research. Clinically, terminal cancer phenotypes such as therapy resistance and metastasis are vital areas of investigation, representing key drivers of cancer mortality [2]. To capture the diverse functional states of cells in a structured format, transcriptome profiling has emerged as a powerful platform where high-dimensionality gene expression signatures can serve as "molecular surrogates" of functional phenotypes [3, 4]. Gene expression profiles (hereafter referred to as *molecular phenotypes*) have been used to identify clinically-relevant subtypes of cancer, predict patient prognosis, define therapeutic pathways, and dissect disease mechanisms [5]. For example, transcriptome-based consensus molecular subtypes (CMS) can classify colorectal cancers into four distinct subtypes [6]. Another important feature of molecular phenotypes is that they are often broadly "transportable" across different tumor types and between pre-clinical models and *in vivo* settings [7]. There is thus broad interest in capturing the complete repertoire of molecular phenotypes across the cancer life cycle and harnessing these for translational and clinical benefit.

The expression, composition, and coordinated modulation of molecular phenotypes is governed by many factors, both cell-intrinsic and extrinsic. Examples of cell-intrinsic factors driving specific molecular phenotypes include genomic alterations in classical oncogenes and tumor suppressor genes (eg *EGFR*, *TP53,* and *VHL* [8, 9]) and more recently chromatin modifier genes that are often mutated [10] which can elicit widespread epigenetic dysregulation [11]. Reflecting their biological coherence, molecular phenotypes arising from alterations in separate components of the same cellular process often exhibit high similarity [12], providing a basis for functional discovery. Examples of cell-extrinsic factors linked to molecular phenotypes in cancer include nutrient availability [13], hypoxia [14], interactions with the tumor microenvironment [15], niche-specific cues from metastatic sites [16], and therapeutic pressure from anti-cancer therapies [17-19]. Notably, relationships between molecular phenotypes and cell-intrinsic/cell-extrinsic factors are complex and highly non-linear, being governed by gene-gene (GxG) and gene-drug (GxD) interactions and further

influenced by cell lineage and micro-environmental context. For example, different molecular phenotypes have been shown to arise upon *KRAS* activation depending on the presence or absence of *TP53* function [20].

Despite the importance of studying these interactions, from a cost and human capital perspective, the sheer number of combinational possibilities associated with probing multiple GxG, GxD, and eventual GxGxD scenarios prohibits the use of standard single-sample bulk transcriptome profiling approaches. Recently, single-cell technologies for generating molecular phenotypes linked to gene perturbations have emerged [21]. Methodologies such as Perturb-seq and CROP-seq, which combine CRISPR genome editing with single-cell transcriptomics, have the potential for elucidating genetic perturbation-driven molecular phenotypes at scale [21-23]. These single-cell approaches have been used to study differentiation trajectory impairments in TGF-β signalling [24] and to elucidate transcriptional regulation by mSWI/SNF complex subunits [25]. To date however, few studies have applied this approach to study molecular phenotypes associated with therapy resistance [26]. Here, we applied direct-capture Perturb-seq (dcPerturb-seq) to systematically survey the landscape of therapy-associated molecular phenotypes in gastric cancer (GC), a leading cause of global cancer morbidity and mortality [27]. In this proof-of-concept study, we selected therapies targeting four different epigenetic mechanisms, due to the prominent role of epigenomics in regulating gene expression and cellular plasticity [28] and the increasing role of epigenetic drugs in GC clinical trials [29]. By mapping molecular phenotypes across 669,065 cells against 7 gastric cell lines, 226 genes, and 4 epigenetic drugs, we surveyed 4,972 gene/drug/cell line and 18.9 million expression interactions. Analysis of this single-cell pharmacogenomic compendium revealed dynamic interactions between genetic alterations, epigenetic therapies, and cellular responses, causing diverse molecular phenotypes underpinning gene-dosage effects, drug resistance, and context-specificities. We elucidated functions for poorly characterized genes, identified distinct gene-based resistance pathways to the same therapy, and transported findings from the *in vitro* setting to primary patient samples, informing clinical prognosis and response prediction. Our work establishes a framework for a comprehensive atlas of drug-perturbation molecular phenotypes, yielding insights into how molecular phenotypes are modulated by context-specific drug action and cancer-associated genomic heterogeneity.

Results

Direct Capture CRISPR Single-Cell Perturbation Sequencing (dcPerturb-seq) Captures Functional Gene Annotations from Molecular Phenotypes

To systematically map transcriptional phenotypes (ie 'molecular phenotypes') associated with gene perturbations in gastric cancer (GC), we compiled an initial list of 2,290 genes comprising known GC oncogenes, tumor suppressor genes (TSGs), cancer-related transcription factors, epigenetic regulators, and genes associated with GC progression (**Supplementary Fig. S1A**). For this proof-of-concept study, we filtered the list to 216 genes exhibiting either differential expression between primary GCs and matched normal tissues, relevance to GC supported by multiple studies, and detectible expression in gastric cell lines and single-cell RNA-seq epithelial profiles. We designed a Perturb-seq library containing four sgRNAs per target gene, also including sgRNAs targeting essential *RPL* genes as positive controls and non-targeting sgRNAs as negative controls. The total number of genes ultimately selected for screening was 226 (**Supplementary Table S1**).

We employed direct capture Perturb-seq (dcPerturb-seq) as a screening platform. Compared to standard Perturb-seq. dcPerturb-seq enables direct capture of sgRNAs, reduced lentiviral template switching, and is compatible with lineage tracing and molecular recording [30]. Using feature-barcoded lentiviral constructs, we introduced the dcPerturb-seq library into HFE145 and GES1 gastric epithelial cells for one week to evaluate short-term effects (HFE145) and four weeks to assess longer term molecular phenotypes (GES1, **Fig. 1A**). Focusing on high-quality cells with single sgRNA annotations (**Methods**), we generated transcriptomic profiles of 28,721 cells for GES1 and 43,199 cells for HFE145 across 998 sgRNAs. Transcriptomic profiles of cells with sgRNAs targeting the same gene were aggregated and normalized (**Methods**). Supporting the efficacy of dcPerturb-seq, we observed decreased expression of perturbed target genes in cells containing the target sgRNA (GES1, p value = 9.0×10^{-18}; HFE145, p value = 1.9×10^{-28}) (**Supplementary Fig. S1B** and **S1C**), likely reflecting nonsense-mediated decay (NMD) arising from target gene sgRNA disruption [31]. After four weeks, we observed significant reductions in cell populations containing sgRNAs targeting *RPL* essential genes, consistent with these sgRNAs causing cellular lethality. Similar but weaker trends were seen after 1 week of infection (**Supplementary Fig. S1D**). For the remainder of this study, we focused

on a 4-week infection period, as our primary interest was to relate molecular phenotypes to functionally measurable cell features (eg cell viability).

To assess molecular phenotypes linked to specific gene perturbations, we quantified global shifts in mean expression profiles (shift distance score) between cells with target-specific sgRNAs against non-targeting control cells using Hotelling's T^2 test [32]. To identify specific differentially expressed genes (DEGs), we applied a Anderson-Darling (AD) test [12], known for its sensitivity to detect transcriptional changes in subsets of cells. We observed positive correlations between the magnitude of global transcriptional shifts and the numbers of DEGs (average 127 cells/gene for GES1; 190 cells/gene for HFE145) (**Fig. 1B; Supplementary Fig. S1E**). We validated the robustness of this approach by repeating the analyses after down-sampling 80% and 60% of cells, observing consistent results (**Supplementary Fig. S1F**). To validate individual perturbation-associated molecular phenotypes, we evaluated candidate cancer-related genes. In sg*TP53* targeting cells, we observed upregulation of DEG signatures related to epithelial mesenchymal transition (EMT) pathways, while cell cycle related signatures were downregulated in cells with sg*PCNA* and sg*TRRAP* perturbations consistent with previous reports [33-35] (**Fig. 1C; Supplementary Fig. S2A**). We also observed similarities in molecular phenotypes for perturbed targets within the same protein complex. For example, the cell cycle kinase complex members *PCNA* and *CCND1* exhibited similar transcriptional signatures, as did *BRCA1* and *BRCA2* within the BRCC complex (**Supplementary Fig. S2B** and **S2C**).

Unsupervised clustering of the 226 perturbation-associated molecular phenotypes revealed several clusters (**Fig. 1D**). Pathway enrichment analysis identified a set of gene perturbations associated with up-regulation of inflammatory related processes - this cluster included *PTEN*, the deletion of which has been previously reported to increase NF-κB activity [36]. In another cluster, EMT, metastasis, and Notch Signaling pathways were activated in cells with perturbed EMT-related genes (eg *RHOA*, *ARID2* and *TP53*). To visualize these perturbation associations beyond pair-wise gene-gene relationships, we then performed dimensional reduction of the perturbation-associated molecular phenotypes using a machine-learning-based Minimum-Distortion Embedding (MDE) solver [37], and projected the embeddings into a 2D UMAP (Uniform Manifold Approximation and Projection) followed by unsupervised clustering (see **Methods**). We identified 21 discrete clusters and annotated each cluster using Gene Ontology (GO) pathways and

CORUM, a database of experimentally validated protein-protein interactions [38] (**Fig. 1E**). The 21 clusters showed a clear distribution of discrete biological processes including histone modification, double-strand break repair, apical junction complex and cyclin-dependent protein kinases. Genes associated with similar protein complexes clustered in close proximity such as *ARID2*/*ARID1B* in the PBAF complex and *CCND1*/*PCNA* in the cell cycle kinase complex. To ask if leveraging these perturbation clusters might facilitate the annotation of less characterized genes, we considered *ZBTB41* (Zinc Finger And BTB Domain Containing 41), a poorly annotated gene in the literature. In UMAP space, *ZBTB41* clustered with cell cycle related genes/complexes, suggesting a role in cell cycle regulation. Pathway analysis of *ZBTB41* perturbation-associated molecular phenotypes confirmed significant enrichment of cell cycle related processes including G2M checkpoint regulation, mitotic cell cycle and nuclear division processes in both GES1 and HFE145 cells (**Fig. 1F; Supplementary Fig. S2D**). To validate these findings, we independently perturbed *ZBTB41* in GES1, HFE145 and YCC1 GC cells, and again observed up-regulation of cell cycle related targets (eg *CDCA3*, *DUSP1* and *KIFC1*) (**Fig. 1G; Supplementary Fig. S2E**). Cell cycle analysis confirmed that *ZBTB41* perturbation resulted in a shift of cell cycle distributions, marked by a decreased proportion of G1-phase cells and an increased proportion of both G2/M and S-phase cells (**Supplementary Fig. S2F**). In summary, these data demonstrate the efficacy of the dcPerturb-seq platform in simultaneously capturing molecular phenotypes of multiple target genes after perturbation, which can facilitate functional assignments for poorly annotated genes.

Exploring Effects of Expression Dosage on Molecular Phenotypes

Previous reports have reported that gene dosage can influence the expression of transcriptional signatures [39, 40]. In diploid systems, certain biological phenotypes are observed upon complete removal of both gene copies (eg arising from homozygous gene deletion) while other phenotypes can arise when only one copy is removed (‘haploinsufficiency’) [41]. We explored if dcPerturb-seq might allow us to evaluate how molecular phenotypes might be influenced by the expression dosage of target genes at the single-cell level. Prior to analysis, we defined three potential classes of perturbation targets - Dosage-sensitive (DS), Dosage-Linear (DL), and Dosage Tolerant (DT). DS genes correspond to targets where modest changes in gene expression can already cause transcriptional effects analogous to complete

expression depletion [41]. DL genes correspond to genes where reductions in expression are associated with correspondingly graded alterations in transcriptional phenotype. Lastly, DT genes represent targets showing transcriptional changes only when target expression is severely reduced (**Fig. 2A**).

For each of the 216 target genes, we stratified cells into discrete populations showing either partial or complete depletion of target expression, conducting differential analysis against Wildtype (WT) negative control cells. To ensure data robustness, we applied a stringent single cell-based differential analysis pipeline focusing on strong and significant downstream alterations (**Methods**). We developed a Kullback–Leibler (KL) divergence-based method to classify Perturbation-Phenotype patterns and employed for validation a distinct additional Euclidian distance-based approach (**Fig. 2B**) (**Methods**). Under this analysis framework, two-thirds of dosage-effect patterns fell into the DS category (66.7%), while the remainder belonged to the DL category (33.3%) (**Fig. 2C**). No DT patterns were detected among the 216 perturbed targets potentially due to limited cell numbers or initial gene selection. This finding suggests that for many of the genes selected in this study, modest changes in expression may be sufficient to generate robust molecular phenotypes.

As examples of DS genes, cells with partial *KAT6A* and *NFXL1* depletion (target expression >0) exhibited near-complete global shifts toward profiles seen in completely depleted cells (target expression = 0) on density plots (**Fig. 2D** and **E**). Examining gene signature compositions, we confirmed that cells with partial *KAT6A* and *NFXL1* depletion displayed molecular phenotypes closely resembling those of completely depleted cells (**Fig. 2F** and **G; Supplementary Fig. S3A**). This finding is consistent with previous studies reporting *KAT6A* as a haploinsufficient gene [42]. For DL genes such as *RHOA* and *MASTL*, density plots indicated an intermediate state between WT control and CD cells (**Fig. 2H** and **I**), and expression profiles of partially-depleted cells exhibited intermediate levels of enrichment between WT negative control and completely depleted cells (**Fig. 2J** and **K; Supplementary Fig. S3B**). These results suggest that dcPerturb-seq can provide insights into the sensitivity of perturbation-associated molecular phenotypes to expression dosage. This would be challenging to discern by bulk RNA analysis which aggregates average expression across a large cell population.

dcPerturb-seq Perturbations Modulate Cellular Responses to Epigenetic Therapies

We proceeded to explore how dcPerturb-seq gene perturbations might influence the sensitivity or resistance of cells to therapeutic challenge. For this study, we elected to focus on epigenetic therapies, due to their ability to elicit widespread transcriptional re-programming [43]. After two weeks of library infection, independent pools of GES1 cells were treated with representative drugs covering four distinct epigenetic pathways - histone deacetylase (HDAC) inhibition (SAHA) [44]; DNA methyltransferse (DMNT) inhibition (5AZA) [45]; Bromodomain and Extraterminal Domain Inhibition (BETi) (JQ1) [46]; and Polycomb repressive complex 2 (PRC2) inhibition (GSK126) [47]. Drug exposures were maintained for another two weeks at concentrations causing 40-60% cell death. In total, we generated 121,672 high-quality cells across five experimental conditions: DMSO/Control (28,721 cells), SAHA (41,123 cells), 5AZA (11,432 cells), JQ1 (29,736 cells), and GSK126 (10,660 cells) (**Fig. 3A**).

Cells were first analyzed for CRISPR sgRNA counts ("viability scores") where sgRNA counts for each epigenetic treatment were compared to control cells (ie library-infected cells treated with DMSO for the same duration) (**Fig. 3B**). We found that ~20% and ~15% of the perturbed genes were associated with either heightened resistance or sensitivity to at least one epigenetic therapy [48]. Viability scores based on single-cell measurements correlated highly with pseudo-bulk CRISPR read-based viability measurements (**Supplementary Fig. S4A**). To validate the viability scores, we integrated our findings across multiple independent datasets and performed experimental validation. First, we surveyed the DepMap database containing genome-wide CRISPR/Cas9-based dropout screens across >1,200 cell lines [49]. Focusing on DepMap gastrointestinal lines (n=237), we grouped the lines by their expression of epigenetic drug targets hypothesizing that decreased expression of these targets might function as a surrogate for target inhibition. For example, in dcPerturb-seq data *EP300* perturbed cells exhibited resistance to SAHA treatment. In DepMap, we similarly observed decreased dependency on *EP300* in *HDAC1* low-expressing cells (**Fig. 3C**). Conversely, in dcPerturb-seq data *FOXM1* perturbation imparted increased sensitivity to SAHA. Similarly, in DepMap higher dependencies were observed for *FOXM1* in *HDAC1* low-expressing cells, indicating potential synthetic lethality. Other gene perturbation/therapy response interactions observed in both our data and DepMap included cells with low *EZH2* expression (GSK126 target) exhibiting reduced

ACTL6A dependency, and higher sensitivity to *MET* depletion (**Fig. 3C**). Second, we examined an independent genome-wide CRISPR screening dataset of JQ1-treated MCF7 breast cancer cells [50]. We observed a consistent and significant pattern of enrichment and depletion of perturbed target genes similar to our viability scores ($p<0.05$; **Fig. 3D**). Third, to independently validate the gene perturbation/therapy response interactions, we experimentally perturbed three candidate targets (*MED12*, *KLF4* and *EP300*) in GES1 cells. The perturbed cells all exhibited resistance phenotypes to SAHA treatment, with increased cell viability compared to control (WT) cells after SAHA exposure (**Fig. 3E**).

Unsupervised clustering and correlation analysis revealed broad similarities between HDAC, DMNT, and BET inhibitors, with PRC2 inhibitors showing the least similarity (**Fig. 3F** and **G**). Among the perturbed genes, common enriched targets (ie resistance-associated) after SAHA, 5AZA and JQ1 treatment were primarily involved in DNA repair and metabolic processes, while depleted genes (ie sensitivity-associated) were related to cell cycle related processes (**Supplementary Fig. S4B** and **S4C**). For GSK126 (EZH2 inhibition), we observed a somewhat opposite trend where perturbation of cell cycle related targets conferred resistance to GSK126 treatment whereas histone modification pathways were enriched among the depleted targets (**Supplementary Fig. S4D**). These results provide insights into how drug response phenotypes can be influenced by intrinsic gene-perturbations.

dcPerturb-seq Perturbations Reveal Distinct Molecular Phenotypes Associated with Therapy Resistance

Therapy resistance is a convergent functional cellular phenotype that can arise from different mechanisms [17, 51]. We asked if dcPerturb-seq molecular phenotypes might provide insights into the heterogeneity of pathways leading to epigenetic drug resistance. As an initial use-case, we focused on SAHA and the top 50 resistance-associated gene targets exhibiting post-SAHA viability score enrichment ("enriched targets") (**Fig. 4A**). Unsupervised clustering of post-SAHA molecular phenotypes associated with these enriched targets revealed four distinct clusters (G1-G4) (**Fig. 4B**) (**Methods**). For non-enriched targets, less distinct clustering patterns were observed (**Supplementary Fig. S5A**). We performed GSEA analysis to elucidate pathways associated with the G1-G4 clusters (**Fig. 4C**; see **Methods**). Pathways associated with EMT and WNT signaling were significantly up-regulated in cluster G1,

ribosome biogenesis in cluster G3, and pathways related to ATP synthesis, metabolism, cell cycle, and endoplasmic reticulum cluster were enriched in cluster G4. DNA repair, tumor-promoting inflammation, and NF-kB signaling were also enriched in cluster G4 (**Supplementary Fig. S5B**). Cluster G2 exhibited minimal cluster-specific pathway alterations. Focusing on specific genes within each cluster, we selected *MED12* in G1 and *KLF4* in G4. For *MED12* (Cluster G1), we observed consistent up-regulation of cell-substrate adhesion pathways after *MED12* dcPerturb-seq perturbation and SAHA treatment **(Fig. 4D)**. We independently experimentally confirmed the up-regulation of cell substrate adhesion pathway-related targets at the bulk mRNA level in *MED12* perturbed cells treated with SAHA compared to DMSO (**Fig. 4E; Supplementary Fig. S5C**). For *KLF4* (Cluster G4), we observed an enrichment of cell cycle-related processes in *KLF4* dcPerturb-seq perturbed cells after SAHA treatment (**Supplementary Fig. S5D**). We independently validated the up-regulation of cell cycle-related genes (*ACTL6A*, *MKI67*, *KIF18A*) in SAHA-treated *KLF4*-perturbed cells (**Supplementary Fig. S5E** and **S5F**).

We applied the same analysis framework to 5AZA, JQ1 and GSK126. For 5AZA, we identified two post-Aza transcriptional clusters associated with 35 enriched target genes, with one cluster displaying up-regulation of processes related to cytoplasmic translation, cell cycle and stress response, while the other cluster was enriched for metabolic activity, lysosomal transport and MAPK pathways (**Fig. 4F**). One of these clusters contained *ARID1A* and *ARID1B*, both of which have been reported to exhibit increased expression in cell lines after DNMT inhibitor treatment [52]. For JQ1, we identified three post-JQ1 transcriptional clusters associated with 41 enriched targets (**Supplementary Fig. S6A**). Of these, cluster 3 exhibited up-regulation of histone deubiquitination processes and contained *FBXW7*, the loss of which has been reported to mediate BRAF degradation and resistance to BET inhibitors in leukemia [53]. Lastly, for GSK126, we detected a post-GSK126 resistance signature associated with ARID family (*ARID1B* and *ARID2*) perturbed cells, characterized by up-regulation of processes related to spliceosomal assembly, cytoplasmic translation and ribosome biogenesis (**Supplementary Fig. S6B**). This transcriptional signature was distinct from that observed in JQ1-treated *ARID* perturbed cells, indicating that distinct molecular phenotypes can arise even to the same genetic perturbation, under different extrinsic drug treatments.

We evaluated the extent to which dcPerturb-seq resistance-associated molecular phenotypes identified under one treatment might be similarly enriched across other treatments. Specifically, we compared molecular phenotypes associated with SAHA resistance to the molecular phenotypes of enriched targets under 5AZA, JQ1, and GSK126 treatment (**Supplementary Fig. S6C)**. We observed that molecular phenotypes related to SAHA gene cluster G1 and G4 were also enriched in resistance-associated phenotypes for JQ1 and GSK126 ($p < 0.05$). A minor enrichment of SAHA gene cluster G1-related phenotypes was also observed for 5AZA ($p=0.05$), suggesting the potential existence of conserved resistance mechanisms. Collectively, these results showcase how dcPerturb-seq can elucidate the complexity, diversity and heterogeneity of molecular phenotypes converging onto a common drug-resistant functional phenotype.

Cell Extrinsic Drug Exposures Reveal Additional Gene Perturbation/Molecular Phenotype Network Interactions

To explore how cell extrinsic drug exposures might modulate gene-gene network interactions, we integrated the entire compendium of dcPerturb-seq molecular phenotypes across all five exposure conditions (DMSO control plus four epigenetic drugs), using a common analysis pipeline to quantify global transcriptomic changes between gene-perturbed and control cells under distinct epigenetic treatments (**Fig. 5A**). Surveying ~4.3 million interactions, we observed that transcriptomic shifts associated with the same gene perturbation were often altered, sometimes dramatically, in a drug-dependent manner. For example, in *KDM5A*-perturbed cells, we observed down-regulation of cell growth related processes following SAHA treatment, compared with down-regulation of apoptotic and cell-substrate adhesion pathways after 5AZA treatment (**Supplementary Fig. S7A**). To systematically examine how transcriptional responses to the same genetic perturbation can be modulated by epigenetic drug exposures, we quantified the number and identity of up-regulated genes either shared across multiple therapies or private to specific therapies (**Fig. 5B**). This analysis revealed distinct patterns. Perturbation of *STK11*, a serine-threonine kinase commonly altered in various cancers [54], induced a broadly common set of changes across all five conditions, while *ESRRG* gene perturbation, a member of the estrogen receptor-related family, exhibited both distinct and shared transcriptional alterations across conditions. *PBRM1*, a remodeling complex encoding

gene [55], showed strong drug-specific responses upon perturbation, possibly related to previously described context-specific functions of the SWI/SNF complex [56].

To visualize the landscape of gene perturbation-epigenetic drug interactions from a network perspective, we integrated the dcPerturb-seq compendium with CORUM [38] and constructed gene-gene interaction networks [57]. Distinct networks were observed for each epigenetic treatment, indicating that drug exposures can uncover additional gene-gene interactions that are not overtly discernible in the absence of extrinsic therapeutic challenges (**Fig. 5C**). For example, in DMSO-treated cells, chromatin remodeling genes (*ACTL6A* and *ARID2*) formed a regulatory core correlated with the DNA damage repair genes (*MSH2*, *MSH6*, *BRCA1* and *BRCA2*) (FDR < 0.05). However, under SAHA treatment, *EP300* emerged as a regulatory core together with *CREBBP*. Notably, both *EP300* and *CREBBP* encode acetyltransferases [58], and it is possible that perturbation of these two acetyltransferases may normalize abnormal SAHA-induced histone acetylation resulting in therapy resistance. In 5AZA-treated cells, chromatin complex encoding genes (*ARID1A*, *ARID1B* and *ACTR5*) were enriched, consistent with **Fig. 4F** where these three genes exhibited similar patterns of gene program and GO pathway regulation.

To explore if resistance associated dcPerturb-seq gene perturbations observed in GES1 cells are also present in other gastric cell lines, we performed dcPerturb-seq using the same four epigenetic treatments in two other GC lines, SNU719 (192,778 cells) and NUGC3 (201,478 cells) and another three GC lines treated with SAHA: LMSU (44,409 cells), SNU1750 (41,228 cells), and HGC27 (24,301 cells). Analysis of cell viability profiles revealed significant positive correlations across the lines (**Supplementary Fig. S7B**). Specifically, *MED12*, *EP300,* and *KLF4* perturbations induced resistance to SAHA in multiple lines, a finding we further experimentally and independently validated in SNU719 cells (**Supplementary Fig. S7C** for *MED12* and *EP300,* similar to **Fig. 3E**). Molecular phenotypes associated with gene perturbations were also conserved. For example, *MED12* perturbation in SAHA-treated cells resulted in upregulation of EMT processes across all six cell lines, while *EP300* perturbation led to the upregulation of Notch signaling (**Supplementary Fig. S8A**). Likewise, *KLF4* perturbation led to the upregulation of Myc targets, as well as glucose metabolism and DNA repair pathways in multiple lines. Perturbations of SAHA_G1 genes (*MED12*, *ZFP3* and *NRAS*) were consistently associated with the enrichment of EMT-related signatures, whereas perturbations of SAHA_G4 genes (*ZBTB7C,*

E2F7 and *HOXC10*) were linked to the enrichment of cell cycle-related signatures (**Supplementary Fig. S8B**). Taken collectively, these results reveal conserved gene-perturbation molecular phenotypes associated with cell-extrinsic drug challenges.

Clinical Application of dcPerturb-seq Molecular Signatures

Finally, we explored if *in vitro* derived dcPerturb-seq molecular signatures might inform the clinical behavior of primary GCs *in vivo* (**Fig. 6A**). We compiled all eleven resistance-associated gene clusters (eg G1-G4 for SAHA; G1-G3 for JQ1) and applied them to primary GC expression profiles from the ACRG and TCGA cohorts (**Fig. 6B**) [59, 60]. Primary GCs were stratified into high- and low-expressing groups based on the expression levels of the resistance-associated gene clusters. We found that decreased expression of the SAHA G1 and G4 gene clusters were consistently associated with significantly poorer survival in both the ACRG and TCGA cohorts, with survival associations remaining significant in multivariate analyses (p value < 0.05) (**Fig. 6B** and **C**). Patients with high expression of GO pathways related to EMT and WNT signaling, as observed in G1-perturbed cells *in vitro* (**Fig. 4C**), also exhibited significant poorer survival (**Supplementary Fig. S9A** and **S9B)**. These results highlight the potential clinical utility of molecularly subtyping primary GCs using dcPerturb-seq molecular phenotypes associated with drug resistance.

As GCs are known to exhibit significant intra-tumor heterogeneity [61], we examined if dcPerturb-seq molecular phenotypes could be applied to improve our understanding of GC spatial heterogeneity. Applying the SAHA G1 and G4 gene clusters to an in-house spatial transcriptomic dataset using NanoString GeoMx Digital Spatial Profiler (GeoMx DSP) technology [15], we observed clear spatial distributions of tumor spatial regions of interest (ROIs) exhibiting distinct G1- and G4-low signatures within individual patients (NGC512, Moran's $I = 0.24$, $p = 2.6 \times 10^{-4}$; NGC510, Moran's $I = 0.48$, $p = 6.1 \times 10^{-4}$) (see **Methods**) (**Fig. 6D; Supplementary Fig. S9C**). Deconvolution analysis revealed comparable proportions of tumor and non-tumor cells within G1-low and G4-low ROIs (**Supplementary Fig. S9D**). This finding suggests that the expression of resistance gene signatures are not uniformly distributed across a tumor section.

Our *in vitro* analysis suggests that certain drug resistance signatures are conserved across multiple therapies, potentially highlighting mechanisms causing multi-drug resistance (**Supplementary Fig. S6C**). We explored if *in vitro* derived

resistance gene sets might have relevance for targeted therapy response in primary GCs. Specifically, we analyzed a separate GeoMx DSP transcriptomic dataset of patients with HER2+ metastatic GC treated with trastuzumab combined with chemotherapy as first-line therapy [62]. Comparing pre-treatment biopsies to post-trastuzumab samples collected after the development and onset of clinical resistance, we observed in two patients significant reductions of G1 cluster gene expression after treatment (PT8 and PT5; **Fig. 6E**), while in another two patients (PT6 and PT13) we separately observed G4 gene clusters downregulation in the trastuzumab resistant samples. In conclusion, these results suggest that dcPerturb-seq resistance signatures are associated with patient prognosis, intra-tumor heterogeneity, and acquired drug resistance *in vivo*.

Discussion

Clinical management of GC remains highly challenging, with current therapeutic strategies often limited by the emergence of drug resistance [63]. Frequent late-stage diagnoses further contributes to poor prognosis, with 5-year overall survival rates at 32% and only 6% for patients with advanced disease [64]. While some targeted therapies have demonstrated efficacy (eg trastuzumab for HER2-positive tumors [65], immune checkpoint inhibitors for PDL1-positive and MSI-positive tumors) [66], GC treatment options still remain limited. GCs exhibit substantial genomic heterogeneity, such as genetic alterations in *TP53, PTEN*, and *ARID1A* that can contribute to metastasis and terminal resistance. For example, *ARID1A* loss may promote EMT enhancing metastatic potential [56], while *PTEN* loss may impart resistance to *HER2*-targeted therapies [67]. Notably, compared to the many putative GC targets identified by genomic sequencing and transcriptomic profiling studies [60, 68, 69], comparatively little is known about the functional pathways regulated by these targets. Here, we interrogated 216 genes frequently altered in GC, systematically characterizing their downstream molecular phenotypes in the context of epigenetic therapy response. This data set, which is publicly available, should provide a valuable resource for future research.

Previous studies have highlighted that functional cancer phenotypes are often influenced by gene-gene (GxG) and gene-drug (GxD) interactions. For example, HER2-positive GCs with co-occurring alterations in cell cycle components (eg *CCNE1*) have impaired responses to trastuzumab [67], and in colorectal cancer *KRAS* mutations can act as drivers of acquired resistance to anti-EGFR therapy [70]. Novel molecular phenotypes have also been shown to arise from the cooperative activity on multiple oncogenic alterations, beyond a purely additive model [20]. Besides gene-based alterations, functional cellular phenotypes can also exhibit analogous molecular phenotypes. For example, the hypoxic functional phenotype is closely associated with molecular phenotypes involving the HIF, UPR and mTOR pathways [71]. To date, few studies have systematically explored the vast combinatorial space of GxG and GxD interactions at scale, with most studies limited to single or a limited set of genes and drugs. Technically, the need to explore multiple GxG and GxD combinations poses significant challenges for traditional bulk RNA sequencing (RNA-seq) approaches, which require each perturbation to be performed as a separate experiment. Recently, advancements in Perturb-seq CRISPR screening technologies offer unprecedented

potential to dissect complex GxG and GxD interactions at the single-cell level. Previous Perturb-seq studies have investigated transcriptional networks governing tissue development, immune responses, and cancer progression [24, 72, 73], however these have typically focused on a limited number of genes (<50), primarily to validate gene sets identified through genome-wide bulk CRISPR screens. More recently, genome-scale Perturb-seq screens have demonstrated the capacity to assign gene functions and dissection of complex cellular phenotypes [12, 74], however these studies have largely focused on methodological development. To date, few Perturb-seq studies have considered how therapies can influence gene perturbation linked molecular phenotypes, or the contribution of context- and lineage-specificity across different tumor types.

In this study, we integrated dcPerturb-seq molecular phenotyping with epigenetic drug interventions to investigate combinatorial GxD interactions and their influence on molecular and phenotypic signatures in the context of GC. Providing confidence in the reliability of our platform, we observed high concordance with previously characterized gene-specific molecular phenotypes, such as *FBXW7* in mTOR signaling regulation [75], and *KDM5C* in WNT signaling [76]. By interpreting the >200 molecular phenotypes in the context of a network, we gained a deeper understanding into the potential cellular functions of poorly characterized genes such as *ZBTB41*. Previous studies have posited a relationship between *ZBTB41* expression and the p53 signaling pathway, and *ZBTB41* has been proposed as biomarker for cervical cancer [77]. Besides *ZBTB41*, the dcPerturb-seq data revealed other intriguing molecular phenotypes, such as inflammatory pathways in response to genetic perturbations in *FOXM1*. These findings provide initial inroads into the subsequent study of these frequently altered genes in GC development.

We investigated if single-cell CRISPR molecular phenotyping can be used to dissect gene dosage-dependent effects, in a manner not achievable using bulk RNA-seq. Notably, we observed several dosage-sensitive (DS) genes among our targets, including well-characterized haploinsufficient genes such as *KAT6A*, *APC* and *GATA6*. Specifically, APC haploinsufficiency has been linked to impaired hematopoiesis [78], while GATA6 haploinsufficiency is associated with pancreatic agenesis [79]. Our findings also revealed a separate population of genes exhibiting a dosage-linear (DL) pattern, aligning with previous findings for genes such as *JAK1* and *IRF1*, which demonstrate gradual changes in transcriptional regulation with increasing degrees of

perturbation [80]. While our dataset did not reveal a distinct dosage-tolerant (DT) pattern, we speculate that this may be attributed to limitations in cell number or our selection of target genes. Overall, these observations emphasize the highly nuanced relationship between gene dosage and transcriptional regulation, which single-cell CRISPR screening is uniquely positioned to explore.

By integrating the dcPerturb-seq platform with drug exposures, we identified targets exhibiting resistance and sensitivity to epigenetic treatment. We chose to study epigenetic drugs as more than 70% of our perturbation targets are epigenetic regulators and transcription factors, and it is known that classic oncogenes and tumor suppressors such as *EGFR* and *TP53*, can also induce profound changes in chromatin landscapes [8]. Clinically, epigenetic therapies, such as those targeting DNA methyltransferases (DNMTs) and histone deacetylases [29, 81], are also being explored in clinical trials. Reassuringly, our dcPerturb-seq molecular phenotypes recapitulated several published findings, such as *MED12* loss inducing an EMT-like molecular signature, leading to chemotherapy resistance in colon cancer and resistance to EGFR-inhibitor in lung cancer [82]. We found that SAHA (HDACi) resistance could be caused by *EP300* perturbation. In breast cancer, *EP300* loss has been associated with paclitaxel resistance [83]. Notably, HDAC inhibition has been shown to alter EP300 acetylation and suppress gene expression in an EP300-dependent manner [84], raising the possibility that loss of EP300 may restore the effects of HDAC inhibition As another finding, *KLF4* deficiency in our study was also associated with HDACi resistance, and *KLF4* has been linked to increased tamoxifen resistance in breast cancer [85]. Besides targets where perturbation leads to resistance, our study also revealed targets conferring sensitivity upon perturbation. These may also represent promising avenues for further investigation in the context of synthetic lethality. For example, *FOXM1* suppression have been shown to compromise tumor growth in various cancer types and sensitize cancer cells to cisplatin, tamoxifen and proteasome inhibitors [86].

Our ability to link molecular phenotypes with functional cell viability yielded insights into pathways associated with epigenetic drug resistance. Notably, we found that diverse molecular pathways can converge to produce a common physiological phenotype (drug resistance). This is conceptually akin to melanoma cells exhibiting mesenchymal-like, dedifferentiated, and metabolically reprogrammed "starved-like" states after MAPK inhibition, demonstrating the coexistence of multiple drug-tolerance

transcriptional circuits [87]. In our data, the SAHA_G1 resistance gene cluster exhibited enriched molecular phenotypes associated with phenotypic plasticity, such as epithelial-to-mesenchymal transition (EMT), WNT signalling and tissue development processes. Such plasticity, including epithelial-to-mesenchymal transition (EMT), lineage transdifferentiation, and reversion to stem or progenitor-like states [28], have been shown to strongly correlate with transcriptional plasticity and treatment resistance [17, 88-90]. In contrast to SAHA_G1, the SAHA_G4 resistance gene cluster showed enrichment of molecular signatures associated with cell cycle, metabolic, and stress-related responses, all of which have also been linked to therapy resistance. For example, heterogeneity in cell cycle behaviour has been identified as a critical determinant to chemotherapy resistance [91], and enrichment of carbohydrate and nucleotide metabolism pathways have been observed in PARP-inhibitor-resistant ovarian cancer cells [92]. When queried across multiple epigenetic drugs, we observed a consistent enrichment of SAHA_G1 and G4 related signatures across treatments, suggesting the presence of conserved resistance signatures despite separate genetic perturbations and distinct epigenetic challenges.

The data richness of our dcPerturb-seq molecular phenotyping dataset allowed us to examine the complex interplay between epigenetic drug treatment and genetic perturbations. Specifically, our analysis revealed that certain gene perturbations, such as *STK11*, *PCNA*, *TEAD4*, and *ZNF165*, tended to elicit broadly similar responses across treatment conditions, while other perturbations (e.g., *PBRM1*, *TP53*, *SMARCA4*, and *SMAD4*) elicited more treatment specific molecular phenotypes. Epigenetic challenge also enabled the emergence of distinct gene-gene regulatory networks not observable in the absence of drug exposure. For example, JQ1 treatment highlighted two dominant regulatory cores: genes encoding components of the SWI/SNF complex (*SMARCD1* and *SMARCB1*) and the *CDK2* gene, while GSK126 challenge uncovered an enriched core of genes (*NFKB1*, *MSH2*, *ARID2*, *ACTL6A* and *ARID1B*), associated with downregulation of stress-related pathways, reminiscent of phenotypic inertia [73]. Gene-gene regulatory network analysis further identified a consistent regulatory core involving *EP300* and *CREBBP* in SNU719 cells (**Supplementary Fig. S9E**), mirroring observations in GES1 cells under SAHA treatment. Importantly, our study also demonstrates that molecular signatures derived *in vitro* (SAHA_G1, associated with phenotypic plasticity, and SAHA_G4, linked to cell

cycle and metabolic alterations) can be effectively mapped onto bulk and spatial primary human tumor expression profiles, highlighting their clinical relevance.

This study and its associated platforms can be expanded in several directions. Incorporating additional time points will enable the tracking of longitudinal changes in molecular phenotypes, as drug resistance is a gradual and multifaceted process [93]. Extending the duration of drug treatment may facilitate the detection of induced mutational signatures, as observed in solid tumors receiving epigenetic therapy [94]. The accuracy and generalizability of our findings can be further enhanced by incorporating additional cell lines, increasing the number of replicates, cells per condition, and sequencing depth. A more comprehensive understanding of the role of chromatin modifiers will inevitably require investigation at the chromatin level. Techniques such as SHARE-seq [95] or other single-cell multi-omic approaches will be crucial for elucidating the intricate interplay between chromatin structure and cellular function. Finally, experimental costs can be further reduced by exploring alternative scRNA-seq methods, such as scifi-RNA-seq [96] and SPLiT-seq [97]. Looking ahead, the development of *in vivo* spatial single-cell screening techniques [98] will be crucial for recapitulating the complex cellular and environmental challenges encountered during tumor growth. This approach will allow for the investigation of how drug gradients, spatial heterogeneity, and cell-intrinsic stress responses influence tumor evolution within a more physiologically relevant context.

In conclusion, the dcPerturb-seq platform, applied here in the context of GC and epigenetic drugs, can be readily adapted to assess the impact of genetic perturbations on a broad range of disease-related targets under various conditions, such as pharmaceutical, metabolic, hypoxic, and acidic environments. Information provided through such platforms may augment recent advancements in single-cell foundation models, exemplified by scGPT [99] and GET [100], to predict cellular responses to perturbations and infer gene regulatory networks, culminating in context-specific large-language models for GC disease development and predicting patient outcomes. These findings will contribute significantly to data-driven resources for translational cancer research, paving the way for the development of more effective and personalized therapeutic strategies.

Methods

Cell culture, dcPerturb-seq CRISPR library design and lentiviral production

GES1 cells were a gift from Dr. Alfred Cheng, Chinese University of Hong Kong. HFE145 cells were a gift from Dr Hassan Ashktorab, Howard University. YCC1 cells were a gift from Yonsei University College of Medicine, Seoul. SNU719 and SNU1750 cells were obtained from Korean Cell Line Bank. NUGC3 cells were obtained from the Japan Health Science Research Resource Bank. HGC27 cells were obtained from CLS Cell Lines Service. LMSU cells were obtained from Riken Cell Bank. Cell line identities were confirmed by STR DNA profiling at the Centre for Translational Research and Diagnostics, National University of Singapore. All lines were negative for mycoplasma contamination as assessed by the MycoAlert™ Mycoplasma Detection Kit (Lonza). Culture medium information and latest mycoplasma test dates are summarized in **Supplementary Table S2.**

CRISPR sgRNA libraries were synthesized by Sigma using scaffolds (U6-gRNA-10X:EF1a-puro-2A-Cas9-2A-tGFP) containing Capture Sequences compatible with 10x Genomics Feature barcoding technology (**Supplementary Table S1**). Single guide RNA containing oligos were cloned into lentiCRISPR v2 backbones for independent validation. Lentivirus particles were produced by co-transfecting HEK293T cells with transfer plasmids and standard packaging vectors. For dcPerturb-seq, cells were infected with pools of lentiviral particles containing four independent sgRNAs per target gene, and also non-targeting sgRNAs as controls. Cells were transduced with lentiviral pools at an MOI of 0.2 to attain single sgRNA entries per cell, and selected with puromycin to achieve a final library representation of roughly 2,000 cells per sgRNA. Lentivirus infected HFE145 cells were collected 1 week after infection. GES1, SNU719 and NUGC3 cells were recovered after 2 weeks,and subsequently treated with SAHA (SML0061, Merck), 5AZA (A3656, Merck), JQ1 (11187, Cayman Chemical) and GSK126 (S7061, Selleck Chemicals) for an additional 2 weeks before proceeding to single cell profiling. Similarly, SNU1750, HGC27 and LMSU cells were treated with SAHA (SML0061, Merck) for 2 weeks before undergoing single cell profiling. Drug concentrations for all cell lines used are presented in **Supplementary Table S3**.

dcPerturb-seq library preparation and sequencing

Cells were prepared for single-cell RNA-sequencing in 1×PBS with 0.1% BSA as detailed in the 10x Genomics Single Cell Protocols Cell Preparation Guide (10x Genomics, CG000053 Rev C). Cells were separated into droplet emulsions using the Chromium Controller (10x Genomics) and the 10x Genomics Chromium Single Cell 3' Reagent Kits v3.1 with Feature Barcode Library Kit (10x Genomics, PN-1000121 and PN-1000079), with the goal of recovering 10,000 cells per GEM group before filtering. For preparation of gene expression and sgRNA libraries, samples were processed according to 10x Genomics Chromium Next GEM Single Cell 3' Reagent Kits v3.1 User Guide with Feature Barcode technology for CRISPR Screening (CG000205 Rev D). Sequencing libraries were generated using unique sample indexes, which were quantified using the Kapa library kit. mRNA and sgRNA libraries were pooled at a 4:1 ratio and sequenced on both Illumina Hiseq4000 and NovaSeq X plus sequencing platforms (Illumina) according to the 10x Genomics User Guide.

Cell viability and Quantitative Real-Time PCR assays

For cell viability assays, cells were seeded in 96-well plates, treated with epigenetic drugs, and assessed for viability using the Cell Counting Kit-8 platform (Dojindo) accordingly to the manufacturer's protocol. For quantitative Real-Time PCR assays, cells were seeded in 6-well plates and treated with epigenetic drugs for 72 hours before proceeding to sample collection for downstream analysis. Total RNA was extracted using RNA extraction reagents (Qiagen), and reverse transcribed using iScript Reverse Transcriptase Supermix (Biorad). Quantitative real-time PCR was performed using SYBR Green PCR Master Mix (Life Technologies). Samples were analyzed using the Bio-Rad CFX96 system. Experimental Ct values were normalized to *GAPDH*, and relative mRNA expression was calculated as fold changes compared to *GAPDH* expression (2-ΔCt). Sequences of primers are listed in **Supplementary Table S4**.

dcPerturb-seq data processing

Cell Ranger 6.1.2 software (10x Genomics) was used for alignment of scRNA-seq reads to human transcriptomes, alignment of sgRNA reads to the dcPerturb-seq library, collapsing reads to UMI counts, and cell calling. The 10x Genomics GRCh38 version 2020-A genome build was used as a reference transcriptome. Subsequently, Seurat V4.3.0 [101] was utilized to perform basic quality control (QC) filtering. To filter out low-quality cells and to retain the most informative genes, we filtered out cells with >12% (HFE145), >20% (GES1), >25% (SNU719), >25% (NUGC3), >12% (SNU1750), >12% (HGC27) and >12% (LMSU) mitochondrial UMIs per cell. We then removed cells with RNA counts below the 1st percentile or above the 99th percentile and genes with counts below the 1st percentile. The threshold number of guides was set to 2-10 based on the threshold plot, and cells with only one guide were kept. Each cell was categorized according to its guide identity representing a single genetic perturbation. Only cells bearing a single guide were used for downstream analysis. The data was then normalized using Seurat (Relative Counts normalization, RC-norm). After RC-normalization, genes were further filtered by being highly expressed (> 2 after RC-norm) in no less than five cells. The data was then analysed for highly variable genes (HCG) using the FindVariableFeatures() function, scaling across cell populations using the ScaleData() function, and dimensional reduction analysis using PCA.

To evaluate the overall expression profiles of each perturbed target gene and to remove possible background effects, we first calculated the mean expression of all cells on the whole transcriptome for each perturbed target gene. We then applied Z-transformation on the average gene expression matrix using mean expression and standard deviation values for negative control sgRNA cells as the standard reference. Correlations between any two perturbed target genes was calculated based on these Z-transformed vectors.

Shift distance test and differential gene expression analysis

To assess the impact of gene perturbations on molecular phenotypes, we employed energy-based statistical tests. These tests are well-suited for comparing distributions in high-dimensional data, making them ideal for analyzing gene expression profiles represented as vectors of principal component scores. Specifically, each cell was characterized by its top 20 principal components, and we evaluated whether the distribution of these vectors differed between control cells and those subjected to gene perturbations. Divergences in these distributions indicate that the

perturbation alters the structure or distribution of transcriptional profiles within perturbed cells.

Recognizing the inherent biological variability of gene expression, the incomplete penetrance of certain perturbations, and the heterogeneity of some gene expression programs, we adopted a non-parametric approach. This approach circumvents the need for specific distributional assumptions about gene expression. Specifically, gene expression was z-normalized relative to control cells. For each gene, we tested whether the distribution of normalized expression values was identical between control cells harboring non-targeting sgRNAs and cells subjected to each gene perturbation. The Anderson-Darling test was utilized for this purpose due to its broad sensitivity to differences in distributions, allowing detection of subtle changes in expression profiles.

Minimum-Distortion Embedding (MDE) based 2D-Projection

Minimum-distortion embedding (MDE) has been used in single-cell data to identify connections of different elements using high-dimensional gene profiling data [12]. Using the Z-transformed gene profiling vectors of each target gene, we first used PyMDE (https://github.com/cvxgrp/pymde) to generate a 20 dimension spectral embedding (affinity='nearest_neighbors', n_neighbors=7, eigen_solver='arpack'), and then performed dimension reduction to 2D for visualization (n_neighbors=7, constraint=pymde.Standardized(), repulsive_fraction=5). PyMDE was run multiple times in parallel to find the optimal distortion and residual norm values. The embeddings were plotted as graphs using the pymde.plot() function, and then manually annotated using Gene Ontology and CORUM database annotation.

Permutation test for RPL genes

To evaluate the significance of the essentiality of RPL genes for cell survival and proliferation, we randomly selected 40 negative control sgRNAs, matching the number of sgRNAs targeting RPL genes. We then calculated the number of cells containing the selected negative control sgRNAs and the sgRNAs targeting RPL genes. To determine the significance of enrichment, we performed 10,000 random samplings. An empirical P-value was derived by counting the number of instances in which the number of cells containing the randomly selected negative control sgRNAs was smaller than the number of cells containing sgRNAs targeting RPL genes.

CRISPR viability scores

For each condition (DMSO, SAHA, JQ1, 5AZA, and GSK126), we determined the number of cells assigned to a specific sgRNA-targeted gene. This count was normalized by dividing it by the total number of cells in the respective condition and multiplying the result by 10^5. To mitigate computational errors, a pseudo-cell count was added to the value. The resulting metric represents the number of cells containing a specific sgRNA-targeted gene per 10^5 cells. To calculate CRISPR viability scores for a specific gene perturbation under a drug condition, we computed the $\log_2$ fold change of the resulting metric between the drug and DMSO conditions. Finally, CRISPR viability scores were z-normalized by subtracting the mean and dividing by the standard deviation of the scores across all sgRNA-targeted genes. When calculating CRISPR viability scores using sgRNA counts, we replaced the cell numbers with the counts of sgRNAs targeting a specific gene. This substitution allowed us to use sgRNA numbers as the basis for determining CRISPR viability scores.

Gene dosage effect and cell cycle analysis

For each target gene, we used the RC-normalized gene expression profiles and defined cells with zero expression of the gene as “complete gene removal” (CR), and defined cells that still have intermediate expression of the gene as “partial gene removal” (PR). For each target gene, we merged the cells with negative control cells, then reperformed the normalization step on the merged cells, using Seurat to perform log-normalized raw counts, found variable features (n = 1000), and data scaling to enhance variability. Differential gene expression analysis was performed between negative control cells and fully depleted cells using the Seurat FindAllMarkers() function (min.pct = 0.25, logfc.threshold = 0.25). Identified DEGs with adjusted p-value <=0.05 were retained. The expression patterns of negative control cells, partially depleted cells and fully depleted cells were visualized using the R package Pheatmap. We used two methods to model gene dosage effect, namely the Kullback–Leibler divergence (KL divergence) based method, and the Euclidian distance-based method. For these two methods, we calculated the distances between the “negative control - PR” and the “CR - PR” separately for each target gene using the expression profiles (log-normalized) of DEGs (between negative control and CR) for all cells, and then calculated differences between the distances, forming a distribution of the distances for classifying the dosage effect subtypes.

For the KL Divergence based method:

KL(A || C) = $\sum_i P_A(\mathrm{x_i})\log(\frac{P_A(\mathrm{x_i})}{P_C(\mathrm{x_i})})\,\Delta x$

KL(B || C) = $\sum_i P_B(\mathrm{x_i})\log(\frac{P_B(\mathrm{x_i})}{P_C(\mathrm{x_i})})\,\Delta x$

Differences = KL(A || C) - KL(B || C)

Where $P_A(x)$, $P_B(x)$, $P_C(x)$ denotes estimated probability through Kernel Density Estimation (KDE) on gene expression values of negative control cells (A), completely depleted target gene KO cells (B) and partially deplete target gene KO cells (C). x_i denotes the i-th point on a common grid of values (determined from the KDE).

For the Euclidean Distance based method:

Euclidean Distance$_{AC}$ = $\frac{1}{|A|\cdot|C|}\sum_{i=1}^{|A|}\sum_{j=1}^{|C|}\sqrt{(i-j)^2}$

Euclidean Distance$_{BC}$ = $\frac{1}{|B|\cdot|C|}\sum_{k=1}^{|B|}\sum_{j=1}^{|C|}\sqrt{(k-j)^2}$

Differences = Euclidean Distance$_{AC}$ - Euclidean Distance$_{BC}$

Where A represents gene expression values of negative control cells, B represents gene expression values of completely depleted target gene KO cells (CR), and C represents gene expression values of intermediate-depleted target gene KO cells (PR). $|A|$, $|B|$ and $|C|$ denote the sizes of sets A, B, and C, respectively. I, j and k represent a gene expression value in a single cell in A, B and C, respectively.

Determination of dosage effect categories:

Using differences in either KL Divergence or Euclidean Distance for multiple DEGs, we first estimated the probability through Kernel Density Estimation (KDE) on all DEGs as P(x). Then we determined the category (DS, DL, or DT) of the dosage effect of a target gene by checking the Area Under Curve (AUC) patterns.

$AUC_{a,b} = \int_a^b P(x)dx$

-a = b = (Max(x) – Min(x))*0.05

Where a, b were determined by the differences between maximum and minimum values.
Dosage-Sensitive (DS): $AUC_{b,max(x)} = Max(AUC_{min(x),a}, AUC_{a,b}, AUC_{b,max(x)})$
Dosage-Linear (DL): $AUC_{a,b} = Max(AUC_{min(x),a}, AUC_{a,b}, AUC_{b,max(x)})$
Dosage Tolerant (DT): $AUC_{min(x),a} = Max(AUC_{min(x),a}, AUC_{a,b}, AUC_{b,max(x)})$
Cell circle analysis was performed using the Seurat function CellCycleScoring, which used default s.genes and g2m.genes from references [102]. Cell cycle proportion changes were visualized using the R package ggplot2.

Clustering of gene expression programs and GO biological pathways

For top resistance-associated target genes (n = 50), we first excluded target genes which have less than 35 cells. The Z-transformed gene expression profiles for the remaining target genes were then used for unsupervised clustering using the R package pheatmap (distance method is 'Euclidean', clustering method is 'ward.D2'). The number of clusters was manually decided based on their profiles on the heatmap. We performed pathway enrichment analysis using Gene Set Enrichment Analysis (GSEA) with the GO Biological Process (GO BP) database. To generate the ranked gene list, we calculated a sum statistic defined as sign (mean expression change) $\times -\log^{10}$ (P-value). The mean expression change for a gene was calculated as the difference between the mean expression of cells containing the sgRNA for the target gene and the mean expression of cells containing Negative Control sgRNAs. P-values were determined using the Anderson-Darling test. Cluster-enriched pathways were identified by comparing the normalized enrichment scores (NES) of pathways within the selected gene cluster to the NES of pathways in the other gene clusters. Pathways with a false discovery rate (FDR) of <0.001 and a mean difference >0.1 in NES scores were considered significant.

Survival analysis

300 GC patients in the ACRG cohort and 415 GC patients in the TCGA were collected to study associations between expression levels of dcPerturb-seq resistance molecular phenotypes and patient outcome. The expression levels of selected gene signatures were calculated using ssGSEA. We selected the top 25% of patients with the highest ssGSEA score of the gene signature as the " Expression_high" group, and the bottom 25% of patients with the lowest ssGSEA score of the gene signatures as the " Expression_low" group. We employed Kaplan–Meier survival analysis with the overall survival as the outcome metric. Log-rank tests were used to estimate the significance of Kaplan–Meier curves. We established multivariate Cox proportional-hazards models using the R package "survminer." Gene expression of the ACRG cohort was profiled using Affymetrix Human Genome U133 plus 2.0 Array (GSE62254, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62254) [59]. RNA-seq profiles of 415 primary GCs from the TCGA cohort were obtained from the Broad Institute TCGA Genome Data Analysis Center (GDAC) Firehose (https://gdac.broadinstitute.org/).

Gene-Gene Interaction Network Analysis

Using the gene-gene interaction of all possible pairs of perturbed target genes based on Z-transformed gene expression profile vectors, we employed a published software (https://github.com/robinmeyers/pan-meyers-et-al) [57] to construct gene-gene interaction networks based on the CORUM database (https://mips.helmholtz-muenchen.de/corum/). Target genes that are not in the CORUM database were

excluded from analysis. Default software parameters were used. We selected rank 40 to visualize the network.

Functional similarity analysis

To apply resistance associated molecular phenotypes from SAHA to other drug conditions/ cell lines/ cohorts, we applied single-sample gene set enrichment analysis (ssGSEA) algorithm using the R package GSVA. Gene names of the target genes in each group were used directly as ssGSEA input signatures. To identify enriched GO pathway signatures for each group, we performed ssGSEA for the GO pathways and then calculated average scores according to the identified GO pathway signatures.

DSP analysis

Collection and bioinformatics pre-processing on the DSP GC and HER2-GC dataset was performed according to previous studies [15, 62]. Cell type proportion of the DSP data was estimated by CIBERSORTx using paired scRNA-seq samples. Similar to the method of applying the identified resistance signatures from SAHA to other drug conditions/ cell lines/ cohorts, ssGSEA was used to map signatures to DSP tumor regions of interest (ROI). Spatial autocorrelation of the difference in ssGSEA scores between the G1 and G4 clusters within each primary GC sample was evaluated using Moran's I test, as implemented in the R package spdep (v1.2-8).

Data availability

The raw data generated for this study has been deposited in NCBI SRA PRJNA1219803. The Reviewer link is https://dataview.ncbi.nlm.nih.gov/object/PRJNA1219803?reviewer=jkioao13ghd0ro7hmc955bssl3

References

1. Hanahan, D., *Hallmarks of cancer: new dimensions.* Cancer discovery, 2022. **12**(1): p. 31-46.
2. Lambert, A.W., D.R. Pattabiraman, and R.A. Weinberg, *Emerging biological principles of metastasis.* Cell, 2017. **168**(4): p. 670-691.
3. Golub, T.R., et al., *Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.* science, 1999. **286**(5439): p. 531-537.
4. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proceedings of the National Academy of Sciences, 2005. **102**(43): p. 15545-15550.
5. Qian, Y., et al., *Prognostic cancer gene expression signatures: current status and challenges.* Cells, 2021. **10**(3): p. 648.
6. Guinney, J., et al., *The consensus molecular subtypes of colorectal cancer.* Nature medicine, 2015. **21**(11): p. 1350-1356.
7. Lamb, J., et al., *The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease.* science, 2006. **313**(5795): p. 1929-1935.
8. Yao, X., et al., *Epigenomic consequences of coding and noncoding driver mutations.* Trends in Cancer, 2016. **2**(10): p. 585-605.
9. Yao, X., et al., *VHL deficiency drives enhancer activation of oncogenes in clear cell renal cell carcinoma.* Cancer discovery, 2017. **7**(11): p. 1284-1305.
10. You, J.S. and P.A. Jones, *Cancer genetics and epigenetics: two sides of the same coin?* Cancer cell, 2012. **22**(1): p. 9-20.
11. Shen, H. and P.W. Laird, *Interplay between the cancer genome and epigenome.* Cell, 2013. **153**(1): p. 38-55.
12. Replogle, J.M., et al., *Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq.* Cell, 2022. **185**(14): p. 2559-2575. e28.
13. Gameiro, P.A. and K. Struhl, *Nutrient deprivation elicits a transcriptional and translational inflammatory response coupled to decreased protein synthesis.* Cell reports, 2018. **24**(6): p. 1415-1424.
14. Michealraj, K.A., et al., *Metabolic regulation of the epigenome drives lethal infantile ependymoma.* Cell, 2020. **181**(6): p. 1329-1345. e24.
15. Ma, H., et al., *Spatially Resolved Tumor Ecosystems and Cell States in Gastric Adenocarcinoma Progression and Evolution.* Cancer Discovery, 2025.
16. Psaila, B. and D. Lyden, *The metastatic niche: adapting the foreign soil.* Nature Reviews Cancer, 2009. **9**(4): p. 285-293.
17. Marine, J.-C., S.-J. Dawson, and M.A. Dawson, *Non-genetic mechanisms of therapeutic resistance in cancer.* Nature Reviews Cancer, 2020. **20**(12): p. 743-756.
18. Zou, M., et al., *Transdifferentiation as a mechanism of treatment resistance in a mouse model of castration-resistant prostate cancer.* Cancer discovery, 2017. **7**(7): p. 736-749.
19. Zakowski, M.F., M. Ladanyi, and M.G. Kris, *EGFR mutations in small-cell lung cancers in patients who have never smoked.* New England Journal of Medicine, 2006. **355**(2): p. 213-215.
20. McMurray, H.R., et al., *Synergistic response to oncogenic mutations defines gene class critical to cancer phenotype.* Nature, 2008. **453**(7198): p. 1112-1116.
21. Bock, C., et al., *High-content CRISPR screening.* Nature Reviews Methods Primers, 2022. **2**(1): p. 1-23.
22. Datlinger, P., et al., *Pooled CRISPR screening with single-cell transcriptome readout.* Nature methods, 2017. **14**(3): p. 297-301.
23. Dixit, A., et al., *Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens.* cell, 2016. **167**(7): p. 1853-1866. e17.
24. Genga, R.M., et al., *Single-cell RNA-sequencing-based CRISPRi screening resolves molecular drivers of early human endoderm development.* Cell reports, 2019. **27**(3): p. 708-718. e10.

25. Otto, J.E., et al., *Structural and functional properties of mSWI/SNF chromatin remodeling complexes revealed through single-cell perturbation screens.* Molecular Cell, 2023. **83**(8): p. 1350-1367. e7.
26. Frangieh, C.J., et al., *Multimodal pooled Perturb-CITE-seq screens in patient models define mechanisms of cancer immune evasion.* Nature genetics, 2021. **53**(3): p. 332-341.
27. Sung, H., et al., *Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA: a cancer journal for clinicians, 2021. **71**(3): p. 209-249.
28. Boumahdi, S. and F.J. de Sauvage, *The great escape: tumour cell plasticity in resistance to targeted therapy.* Nature reviews Drug discovery, 2020. **19**(1): p. 39-56.
29. Padmanabhan, N., T. Ushijima, and P. Tan, *How to stomach an epigenetic insult: the gastric cancer epigenome.* Nature reviews Gastroenterology & hepatology, 2017. **14**(8): p. 467.
30. Replogle, J.M., et al., *Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing.* Nature biotechnology, 2020. **38**(8): p. 954-961.
31. Tan, K., D.G. Stupack, and M.F. Wilkinson, *Nonsense-mediated RNA decay: an emerging modulator of malignancy.* Nature Reviews Cancer, 2022. **22**(8): p. 437-451.
32. Ursu, O., et al., *Massively parallel phenotyping of coding variants in cancer with Perturb-seq.* Nature biotechnology, 2022. **40**(6): p. 896-905.
33. Chang, C.-J., et al., *p53 regulates epithelial–mesenchymal transition and stem cell properties through modulating miRNAs.* Nature cell biology, 2011. **13**(3): p. 317-323.
34. Strzalka, W. and A. Ziemienowicz, *Proliferating cell nuclear antigen (PCNA): a key factor in DNA replication and cell cycle regulation.* Annals of botany, 2011. **107**(7): p. 1127-1140.
35. Herceg, Z., et al., *Disruption of Trrap causes early embryonic lethality and defects in cell cycle progression.* Nature genetics, 2001. **29**(2): p. 206-211.
36. Iliopoulos, D., et al., *STAT3 activation of miR-21 and miR-181b-1 via PTEN and CYLD are part of the epigenetic switch linking inflammation to cancer.* Molecular cell, 2010. **39**(4): p. 493-506.
37. Agrawal, A., A. Ali, and S. Boyd, *Minimum-distortion embedding.* Foundations and Trends® in Machine Learning, 2021. **14**(3): p. 211-378.
38. Giurgiu, M., et al., *CORUM: the comprehensive resource of mammalian protein complexes—2019.* Nucleic acids research, 2019. **47**(D1): p. D559-D563.
39. Bhatia, S., et al., *Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia.* The American Journal of Human Genetics, 2013. **93**(6): p. 1126-1134.
40. Bruneau, B.G., et al., *A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease.* Cell, 2001. **106**(6): p. 709-721.
41. Johnson, A.F., H.T. Nguyen, and R.A. Veitia, *Causes and effects of haploinsufficiency.* Biological Reviews, 2019. **94**(5): p. 1774-1785.
42. Tham, E., et al., *Dominant mutations in KAT6A cause intellectual disability with recognizable syndromic features.* The American Journal of Human Genetics, 2015. **96**(3): p. 507-513.
43. Jones, P.A., J.-P.J. Issa, and S. Baylin, *Targeting the cancer epigenome for therapy.* Nature Reviews Genetics, 2016. **17**(10): p. 630-641.
44. Marks, P.A. and R. Breslow, *Dimethyl sulfoxide to vorinostat: development of this histone deacetylase inhibitor as an anticancer drug.* Nature biotechnology, 2007. **25**(1): p. 84-90.
45. Christman, J.K., *5-Azacytidine and 5-aza-2′-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy.* Oncogene, 2002. **21**(35): p. 5483-5495.
46. Filippakopoulos, P., et al., *Selective inhibition of BET bromodomains.* Nature, 2010. **468**(7327): p. 1067-1073.
47. McCabe, M.T., et al., *EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations.* Nature, 2012. **492**(7427): p. 108-112.

48. O'Neil, N.J., M.L. Bailey, and P. Hieter, *Synthetic lethality and cancer.* Nature Reviews Genetics, 2017. **18**(10): p. 613-623.
49. Tsherniak, A., et al., *Defining a cancer dependency map.* Cell, 2017. **170**(3): p. 564-576. e16.
50. Nagarajan, S., et al., *ARID1A influences HDAC1/BRD4 activity, intrinsic proliferative capacity and breast cancer treatment response.* Nature genetics, 2020. **52**(2): p. 187-197.
51. Vasan, N., J. Baselga, and D.M. Hyman, *A view on drug resistance in cancer.* Nature, 2019. **575**(7782): p. 299-309.
52. Aso, T., et al., *Loss of ARID1A, ARID1B, and ARID2 expression during progression of gastric cancer.* Anticancer research, 2015. **35**(12): p. 6819-6827.
53. Yeh, C.-H., et al., *Loss of FBXW7-mediated degradation of BRAF elicits resistance to BET inhibitors in adult T cell leukemia cells.* Molecular cancer, 2020. **19**: p. 1-13.
54. Sanchez-Cespedes, M., et al., *Inactivation of LKB1/STK11 is a common event in adenocarcinomas of the lung.* Cancer research, 2002. **62**(13): p. 3659-3662.
55. Varela, I., et al., *Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma.* Nature, 2011. **469**(7331): p. 539-542.
56. Xu, C., et al., *Comprehensive molecular phenotyping of ARID1A-deficient gastric cancer reveals pervasive epigenomic reprogramming and therapeutic opportunities.* Gut, 2023. **72**(9): p. 1651-1663.
57. Pan, J., et al., *Interrogation of mammalian protein complex structure, function, and membership using genome-scale fitness screens.* Cell systems, 2018. **6**(5): p. 555-568. e7.
58. Attar, N. and S.K. Kurdistani, *Exploitation of EP300 and CREBBP lysine acetyltransferases by cancer.* Cold Spring Harbor perspectives in medicine, 2017. **7**(3): p. a026534.
59. Cristescu, R., et al., *Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes.* Nature medicine, 2015. **21**(5): p. 449-456.
60. Network, C.G.A.R., *Comprehensive molecular characterization of gastric adenocarcinoma.* Nature, 2014. **513**(7517): p. 202-209.
61. Kumar, V., et al., *Single-cell atlas of lineage states, tumor microenvironment, and subtype-specific expression programs in gastric cancer.* Cancer discovery, 2022. **12**(3): p. 670-691.
62. Sheng, T., et al., *Spatial Profiling of Patient-Matched HER2 Positive Gastric Cancer Reveals Resistance Mechanisms to Trastuzumab and Trastuzumab Deruxtecan.* medRxiv, 2024: p. 2024.10. 29.24316248.
63. Alsina, M., et al., *Current developments in gastric cancer: from molecular profiling to treatment strategy.* Nature Reviews Gastroenterology & Hepatology, 2023. **20**(3): p. 155-170.
64. Siegel, R., K. Miller, and R.A. Jemal, *Cancer facts & figures 2021.* 2021.
65. Bang, Y.-J., et al., *Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial.* The Lancet, 2010. **376**(9742): p. 687-697.
66. Kang, Y.-K., et al., *Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial.* The Lancet, 2017. **390**(10111): p. 2461-2471.
67. Kim, J., et al., *Preexisting oncogenic events impact trastuzumab sensitivity in ERBB2-amplified gastroesophageal adenocarcinoma.* The Journal of clinical investigation, 2014. **124**(12): p. 5145-5158.
68. Zang, Z.J., et al., *Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes.* Nature genetics, 2012. **44**(5): p. 570-574.
69. Wang, K., et al., *Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer.* Nature genetics, 2011. **43**(12): p. 1219.

70. Misale, S., et al., *Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer.* Nature, 2012. **486**(7404): p. 532-536.
71. Wilson, W.R. and M.P. Hay, *Targeting hypoxia in cancer therapy.* Nature Reviews Cancer, 2011. **11**(6): p. 393-410.
72. Belk, J.A., et al., *Genome-wide CRISPR screens of T cell exhaustion identify chromatin remodeling factors that limit T cell persistence.* Cancer cell, 2022. **40**(7): p. 768-786. e7.
73. Loukas, I., et al., *Selective advantage of epigenetically disrupted cancer cells via phenotypic inertia.* Cancer Cell, 2023. **41**(1): p. 70-87. e14.
74. Schraivogel, D., et al., *Targeted Perturb-seq enables genome-scale genetic screens in single cells.* Nature methods, 2020. **17**(6): p. 629-635.
75. Mao, J.-H., et al., *FBXW7 targets mTOR for degradation and cooperates with PTEN in tumor suppression.* Science, 2008. **321**(5895): p. 1499-1502.
76. Karwacki-Neisius, V., et al., *WNT signalling control by KDM5C during development affects cognition.* Nature, 2024. **627**(8004): p. 594-603.
77. Wang, M., et al., *TAF1A and ZBTB41 serve as novel key genes in cervical cancer identified by integrated approaches.* Cancer Gene Therapy, 2021. **28**(12): p. 1298-1311.
78. Wang, J., et al., *Haploinsufficiency of Apc leads to ineffective hematopoiesis.* Blood, The Journal of the American Society of Hematology, 2010. **115**(17): p. 3481-3488.
79. Allen, H.L., et al., *GATA6 haploinsufficiency causes pancreatic agenesis in humans.* Nature genetics, 2012. **44**(1): p. 20-22.
80. Jiang, L., et al., *Systematic reconstruction of molecular pathway signatures using scalable single-cell perturbation screens.* Nature Cell Biology, 2025: p. 1-13.
81. Juergens, R.A., et al., *Combination epigenetic therapy has efficacy in patients with refractory advanced non–small cell lung cancer.* Cancer discovery, 2011. **1**(7): p. 598-607.
82. Huang, S., et al., *MED12 controls the response to multiple cancer drugs through regulation of TGF-β receptor signaling.* Cell, 2012. **151**(5): p. 937-950.
83. Asaduzzaman, M., et al., *Tumour suppressor EP300, a modulator of paclitaxel resistance and stemness, is downregulated in metaplastic breast cancer.* Breast cancer research and treatment, 2017. **163**: p. 461-474.
84. Rafehi, H., et al., *Systems approach to the pharmacological actions of HDAC inhibitors reveals EP300 activities and convergent mechanisms of regulation in diabetes.* Epigenetics, 2017. **12**(11): p. 991-1003.
85. Jia, Y., et al., *KLF4 overcomes tamoxifen resistance by suppressing MAPK signaling pathway and predicts good prognosis in breast cancer.* Cellular signalling, 2018. **42**: p. 165-175.
86. Halasi, M. and A.L. Gartel, *Targeting FOXM1 in cancer.* Biochemical pharmacology, 2013. **85**(5): p. 644-652.
87. Rambow, F., et al., *Toward minimal residual disease-directed therapy in melanoma.* Cell, 2018. **174**(4): p. 843-855. e19.
88. Farmer, P., et al., *A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer.* Nature medicine, 2009. **15**(1): p. 68-74.
89. Byers, L.A., et al., *An epithelial–mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance.* Clinical cancer research, 2013. **19**(1): p. 279-290.
90. Brabletz, T., et al., *EMT in cancer.* Nature Reviews Cancer, 2018. **18**(2): p. 128-134.
91. Maleki, E.H., A.R. Bahrami, and M.M. Matin, *Cancer cell cycle heterogeneity as a critical determinant of therapeutic resistance.* Genes & diseases, 2024. **11**(1): p. 189-204.
92. França, G.S., et al., *Cellular adaptation to cancer therapy along a resistance continuum.* Nature, 2024. **631**(8022): p. 876-883.
93. Vander Velde, R., et al., *Resistance to targeted therapies as a multifactorial, gradual adaptation to inhibitor specific selective pressures.* Nature communications, 2020. **11**(1): p. 2393.

94. Yamagishi, M., et al., *Mechanisms of action and resistance in histone methylation-targeted therapy.* Nature, 2024. **627**(8002): p. 221-228.
95. Ma, S., et al., *Chromatin potential identified by shared single-cell profiling of RNA and chromatin.* Cell, 2020. **183**(4): p. 1103-1116. e20.
96. Datlinger, P., et al., *Ultra-high-throughput single-cell RNA sequencing and perturbation screening with combinatorial fluidic indexing.* Nature methods, 2021. **18**(6): p. 635-642.
97. Rosenberg, A.B., et al., *Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding.* Science, 2018. **360**(6385): p. 176-182.
98. Renz, P.F., et al., *In vivo single-cell CRISPR uncovers distinct TNF programmes in tumour evolution.* Nature, 2024. **632**(8024): p. 419-428.
99. Cui, H., et al., *scGPT: toward building a foundation model for single-cell multi-omics using generative AI.* Nature Methods, 2024: p. 1-11.
100. Fu, X., et al., *A foundation model of transcription across human cell types.* Nature, 2025.
101. Hao, Y., et al., *Integrated analysis of multimodal single-cell data.* Cell, 2021. **184**(13): p. 3573-3587. e29.
102. Kowalczyk, M.S., et al., *Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells.* Genome research, 2015. **25**(12): p. 1860-1872.

Figure 1

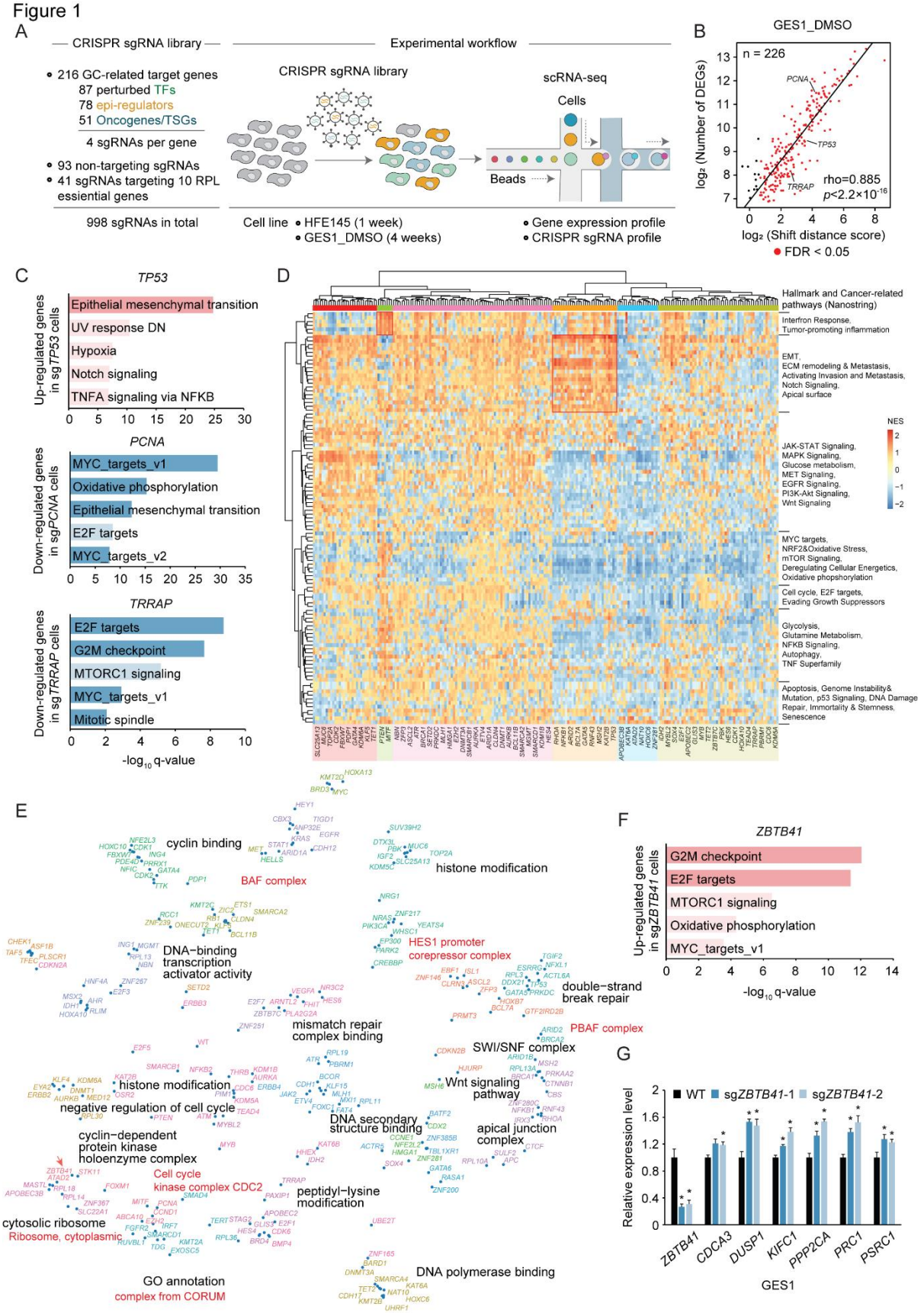

A
CRISPR sgRNA library
Experimental workflow
216 GC-related target genes
87 perturbed TFs
78 epi-regulators
51 Oncogenes/TSGs
4 sgRNAs per gene
93 non-targeting sgRNAs
41 sgRNAs targeting 10 RPL essential genes
998 sgRNAs in total
CRISPR sgRNA library
scRNA-seq
Cells
Beads
Cell line
HFE145 (1 week)
GES1_DMSO (4 weeks)
Gene expression profile
CRISPR sgRNA profile
B
GES1_DMSO
n = 226
PCNA
TP53
TRRAP
rho=0.885
p<2.2×10-16
log2 (Number of DEGs)
log2 (Shift distance score)
FDR < 0.05
C
TP53
Up-regulated genes in sgTP53 cells
Epithelial mesenchymal transition
UV response DN
Hypoxia
Notch signaling
TNFA signaling via NFKB
PCNA
Down-regulated genes in sgPCNA cells
MYC_targets_v1
Oxidative phosphorylation
Epithelial mesenchymal transition
E2F targets
MYC_targets_v2
TRRAP
Down-regulated genes in sgTRRAP cells
E2F targets
G2M checkpoint
MTORC1 signaling
MYC_targets_v1
Mitotic spindle
-log10 q-value
D
Hallmark and Cancer-related pathways (Nanostring)
Interferon Response, Tumor-promoting inflammation
EMT, ECM remodeling & Metastasis, Activating Invasion and Metastasis, Notch Signaling, Apical surface
JAK-STAT Signaling, MAPK Signaling, Glucose metabolism, MET Signaling, EGFR Signaling, PI3K-Akt Signaling, Wnt Signaling
MYC targets, NRF2&Oxidative Stress, mTOR Signaling, Deregulating Cellular Energetics, Oxidative phosphorylation
Cell cycle, E2F targets, Evading Growth Suppressors
Glycolysis, Glutamine Metabolism, NFKB Signaling, Autophagy, TNF Superfamily
Apoptosis, Genome Instability& Mutation, p53 Signaling, DNA Damage Repair, Immortality & Stemness, Senescence
NES
E
cyclin binding
BAF complex
histone modification
DNA-binding transcription activator activity
HES1 promoter corepressor complex
double-strand break repair
mismatch repair complex binding
PBAF complex
SWI/SNF complex
histone modification
Wnt signaling pathway
negative regulation of cell cycle
DNA secondary structure binding
apical junction complex
cyclin-dependent protein kinase holoenzyme complex
Cell cycle kinase complex CDC2
peptidyl-lysine modification
cytosolic ribosome
Ribosome, cytoplasmic
DNA polymerase binding
GO annotation
complex from CORUM
F
ZBTB41
Up-regulated genes in sgZBTB41 cells
G2M checkpoint
E2F targets
MTORC1 signaling
Oxidative phosphorylation
MYC_targets_v1
-log10 q-value
G
WT
sgZBTB41-1
sgZBTB41-2
Relative expression level
ZBTB41
CDCA3
DUSP1
KIFC1
PPP2CA
PRC1
PSRC1
GES1

Figure 1. Functional Gene Annotations from Direct Capture CRISPR Single-Cell Perturbation Sequencing (dcPerturb-seq)

(A) Schematic representation of the dcPerturb-seq experimental design and workflow.

(B) Correlation between shift distance scores relative to wild-type (WT) cells and numbers of differentially expressed genes (DEGs) for each perturbed target gene in GES1 cells. Significance was determined using Pearson correlations.

(C) Enrichment analysis of biological pathways of perturbed genes (*TP53*, *PCNA*, and *TRRAP)* in GES1 cells.

(D) Heatmap showing normalized enrichment scores (NES) for pathway analysis and hierarchical clustering of perturbation-associated molecular phenotypes. Gene names at the bottom represent selected genes from the larger dcPeturb-seq panel.

(E) Minimum distortion embedding plot where each dot represents a gene perturbation. Clusters were annotated using Gene Ontology (GO) pathway enrichments (black) and CORUM database annotations (red).

(F) Biological pathway enrichment analysis of molecular phenotypes associated with *ZBTB41* gene perturbation in GES1 cells.

(G) qRT-PCR results showing relative expression levels of *ZBTB41*, *CDCA3*, *DUSP1*, *KIFC1*, *PPP2CA*, *PRC1* and *PSRC1* genes in WT and *ZBTB41*-perturbed GES1 cells (n=3; mean ± SD) (*, $p < 0.05$, two-sided *t*-test).

Figure 2

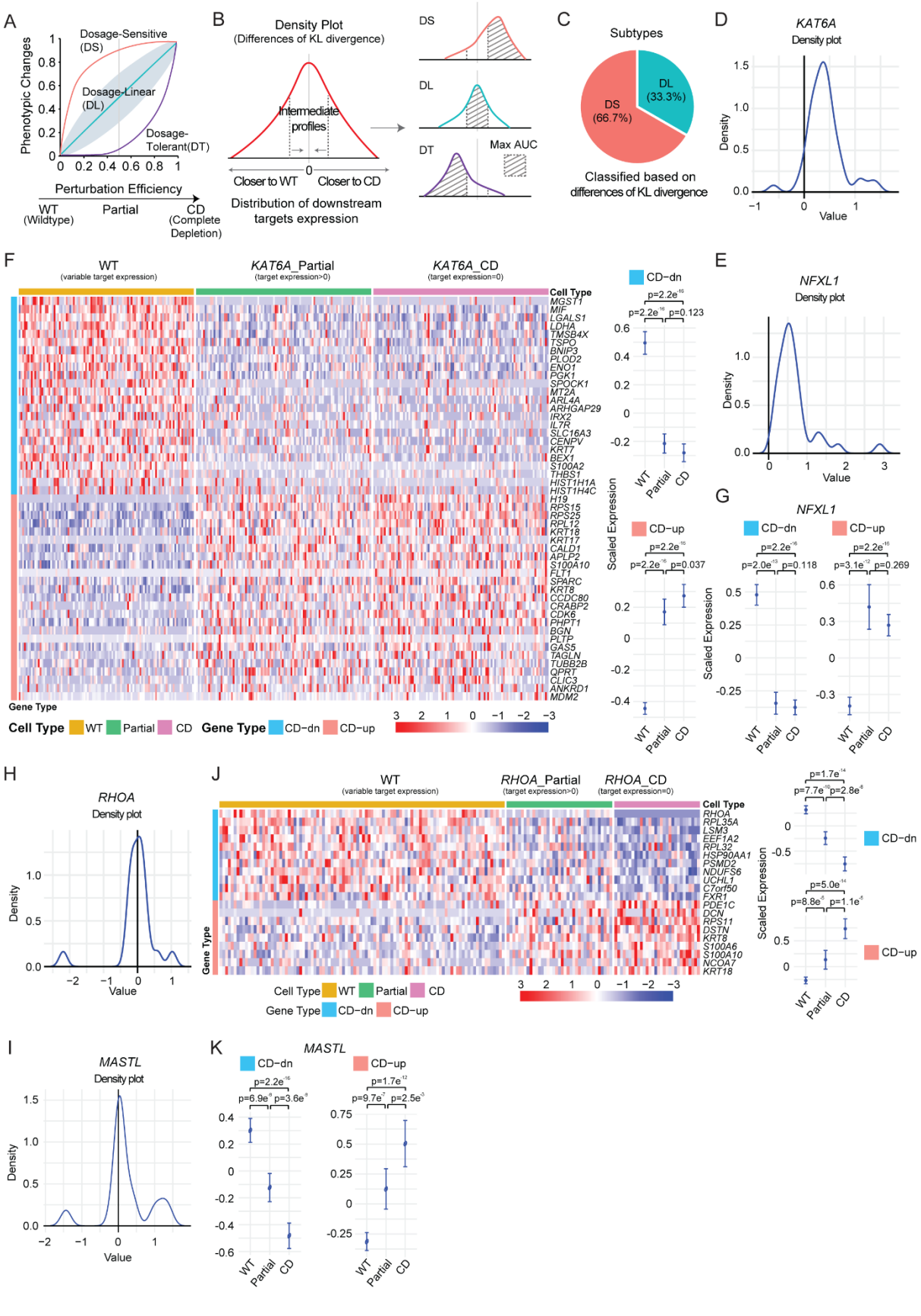

A
Phenotypic Changes
Dosage-Sensitive (DS)
Dosage-Linear (DL)
Dosage-Tolerant(DT)
Perturbation Efficiency
WT (Wildtype)
Partial
CD (Complete Depletion)
B
Density Plot
(Differences of KL divergence)
Intermediate profiles
Closer to WT
Closer to CD
Distribution of downstream targets expression
DS
DL
DT
Max AUC
C
Subtypes
DS (66.7%)
DL (33.3%)
Classified based on differences of KL divergence
D
KAT6A
Density plot
Density
Value
F
WT
(variable target expression)
KAT6A_Partial
(target expression>0)
KAT6A_CD
(target expression=0)
Cell Type
Gene Type
Cell Type WT Partial CD
Gene Type CD−dn CD−up
CD−dn
CD−up
Scaled Expression
E
NFXL1
Density plot
G
NFXL1
H
RHOA
Density plot
J
WT
(variable target expression)
RHOA_Partial
(target expression>0)
RHOA_CD
(target expression=0)
I
MASTL
Density plot
K
MASTL

Figure 2. Effects of Target Gene Expression Dosage and Molecular Phenotypes

(A) Diagram showing alternative classes of transcriptional responses to gene perturbations.

(B) Overview of classification methods used to define categories of transcriptional response associated with **Fig. 2A**.

(C) Distribution of target gene expression dosage against transcriptional response subtypes, determined using Kullback-Leibler (KL) divergence.

(D) Density plot showing shifts in cell profiles with partial *KAT6A* expression depletion compared to profiles observed in cells with complete *KAT6A* expression deletion.

(E) Density plot showing shifts in cell profiles with partial *NFXL1* expression depletion compared to profiles observed in cells with complete *NFXL1* expression deletion.

(F) Heatmap of downstream target gene expression in WT, *KAT6A* partially-depleted and *KAT6A* completely-depleted cells (left), along with a summary of scaled expression (right) (*P* values calculated using Wilcoxon rank sum test).

(G) Summary of scaled downstream target gene expression in WT, *NFXL1* partially-depleted and *NFXL1* completely-depleted cells (*P* values calculated using Wilcoxon rank sum test).

(H) Density plot showing intermediate molecular phenotype expression in cells with partial *RHOA* expression depletion compared to WT cells and cells with complete *RHOA* expression deletion.

(I) Density plot showing intermediate molecular phenotypes expression in of cells with partial *MASTL* expression compared to WT cells and cells with complete *MASTL* expression depletion.

(J) Heatmap of downstream target expression in WT, *RHOA* partially-depleted and *RHOA* completely-depleted cells (left), with a summary of scaled expression (right) (*P* values calculated using Wilcoxon rank sum test).

(K) Summary of scaled downstream target expression in WT, *MASTL* partially-depleted and *MASTL* completely-depleted cells (*P* values calculated using Wilcoxon rank sum test).

Figure 3

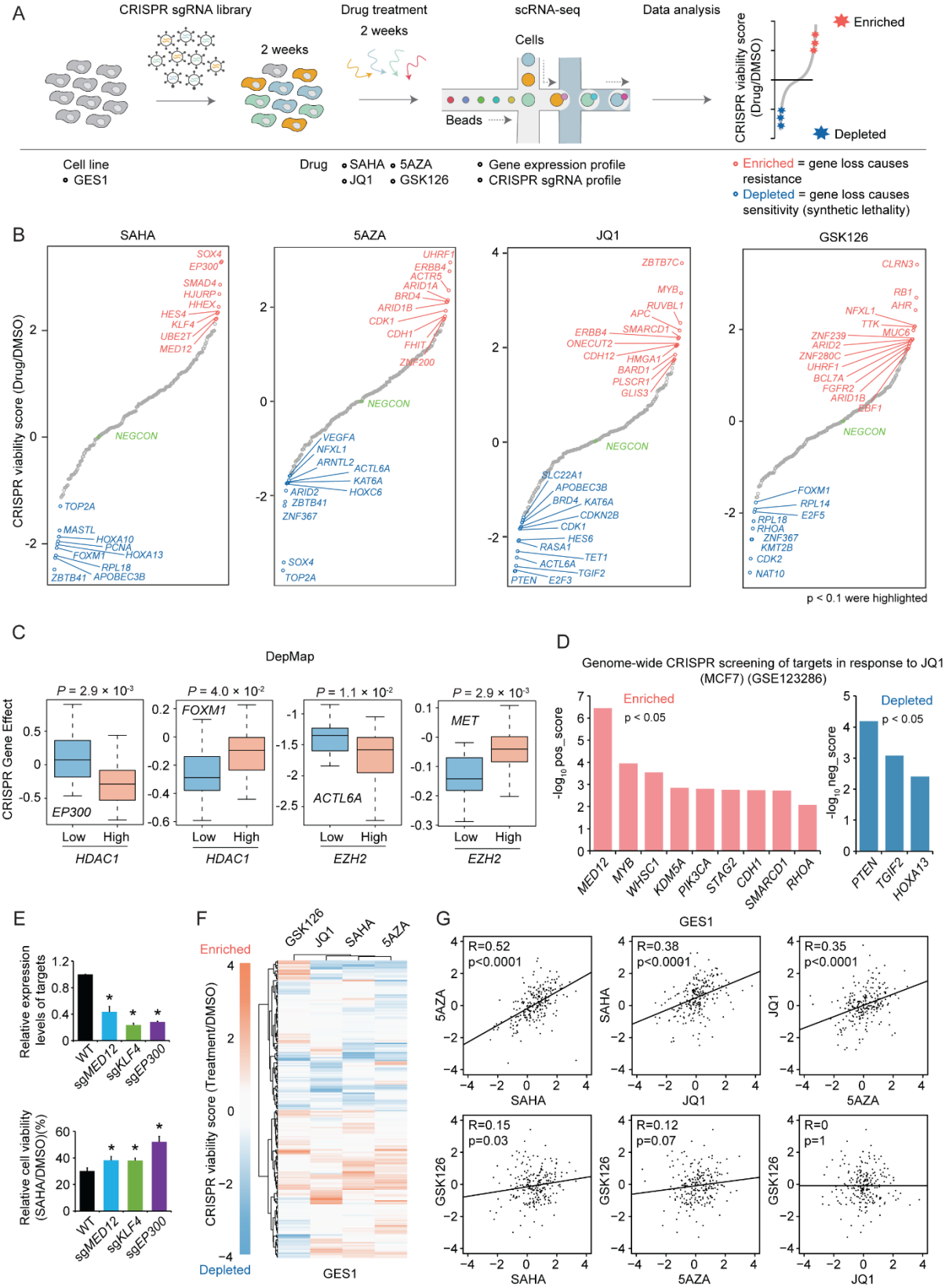

A
CRISPR sgRNA library
2 weeks
Drug treatment
2 weeks
scRNA-seq
Cells
Beads
Data analysis
CRISPR viability score (Drug/DMSO)
Enriched
Depleted
Cell line
GES1
Drug
SAHA
5AZA
JQ1
GSK126
Gene expression profile
CRISPR sgRNA profile
Enriched = gene loss causes resistance
Depleted = gene loss causes sensitivity (synthetic lethality)
B
SAHA
5AZA
JQ1
GSK126
CRISPR viability score (Drug/DMSO)
NEGCON
p < 0.1 were highlighted
C
DepMap
CRISPR Gene Effect
P = 2.9 × 10-3
P = 4.0 × 10-2
P = 1.1 × 10-2
P = 2.9 × 10-3
EP300
FOXM1
ACTL6A
MET
Low
High
HDAC1
HDAC1
EZH2
EZH2
D
Genome-wide CRISPR screening of targets in response to JQ1 (MCF7) (GSE123286)
Enriched
p < 0.05
Depleted
p < 0.05
-log10 pos_score
-log10 neg_score
MED12
MYB
WHSC1
KDM5A
PIK3CA
STAG2
CDH1
SMARCD1
RHOA
PTEN
TGIF2
HOXA13
E
Relative expression levels of targets
Relative cell viability (SAHA/DMSO)(%)
WT
sgMED12
sgKLF4
sgEP300
F
Enriched
Depleted
CRISPR viability score (Treatment/DMSO)
GSK126
JQ1
SAHA
5AZA
GES1
G
GES1
R=0.52
p<0.0001
R=0.38
p<0.0001
R=0.35
p<0.0001
R=0.15
p=0.03
R=0.12
p=0.07
R=0
p=1
5AZA
SAHA
JQ1
GSK126

Figure 3. dcPerturb-seq Gene Perturbations Modulate Cellular Responses to Epigenetic Therapies

(A) Schematic of dcPerturb-seq platform combined with epigenetic drug treatments (SAHA, 5AZA, JQ1 and GSK126). Drugs were added after two weeks of dcPerturb-seq library transfection.

(B) Distribution of normalized CRISPR viability scores comparing drug-treated conditions to DMSO control for all genes targeted by the sgRNA library. Orange dots indicate significantly enriched genes (ie resistance-associated), while blue dots highlight significantly depleted genes (ie sensitivity-associated; $p < 0.1$, Benjamini-Hochberg procedure).

(C) Boxplots comparing DepMap CRISPR gene effect scores for *EP300* and *FOXM1* perturbations in cell lines with low (n=20) and high (n=20) *HDAC1* expression (left), and for *ACTL6A* and *MET* perturbations in cell lines with low (n=20) and high (n=20) *EZH2* expression (right) (*P* values are adjusted p values calculated by Wilcoxon rank sum test).

(D) Representative target genes exhibiting significant enrichment or depletion in response to JQ1 treatment in MCF7 cells identified based on negative or positive effect scores using MAGeCK ($p < 0.05$). MCF7 data is from GSE123286.

(E) qRT-PCR analysis of relative expression levels of *MED12*, *KLF4*, and *EP300* in WT and *MED12*, *KLF4*, and *EP300* perturbed cells (top); Relative cell viability of WT and *MED12*, *KLF4*, and *EP300* perturbed cells treated with SAHA (1.25μM) compared to DMSO is shown (bottom) (GES1) (n=3; mean ± SD) (*, $p < 0.05$, two-sided *t*-test).

(F) Unsupervised clustering of normalized CRISPR viability scores of gene targets under SAHA, 5AZA, JQ1 and GSK126 treatments.

(G) Correlation analysis of normalized CRISPR viability scores across drug treatments. Significance was determined using Pearson correlations.

Figure 4

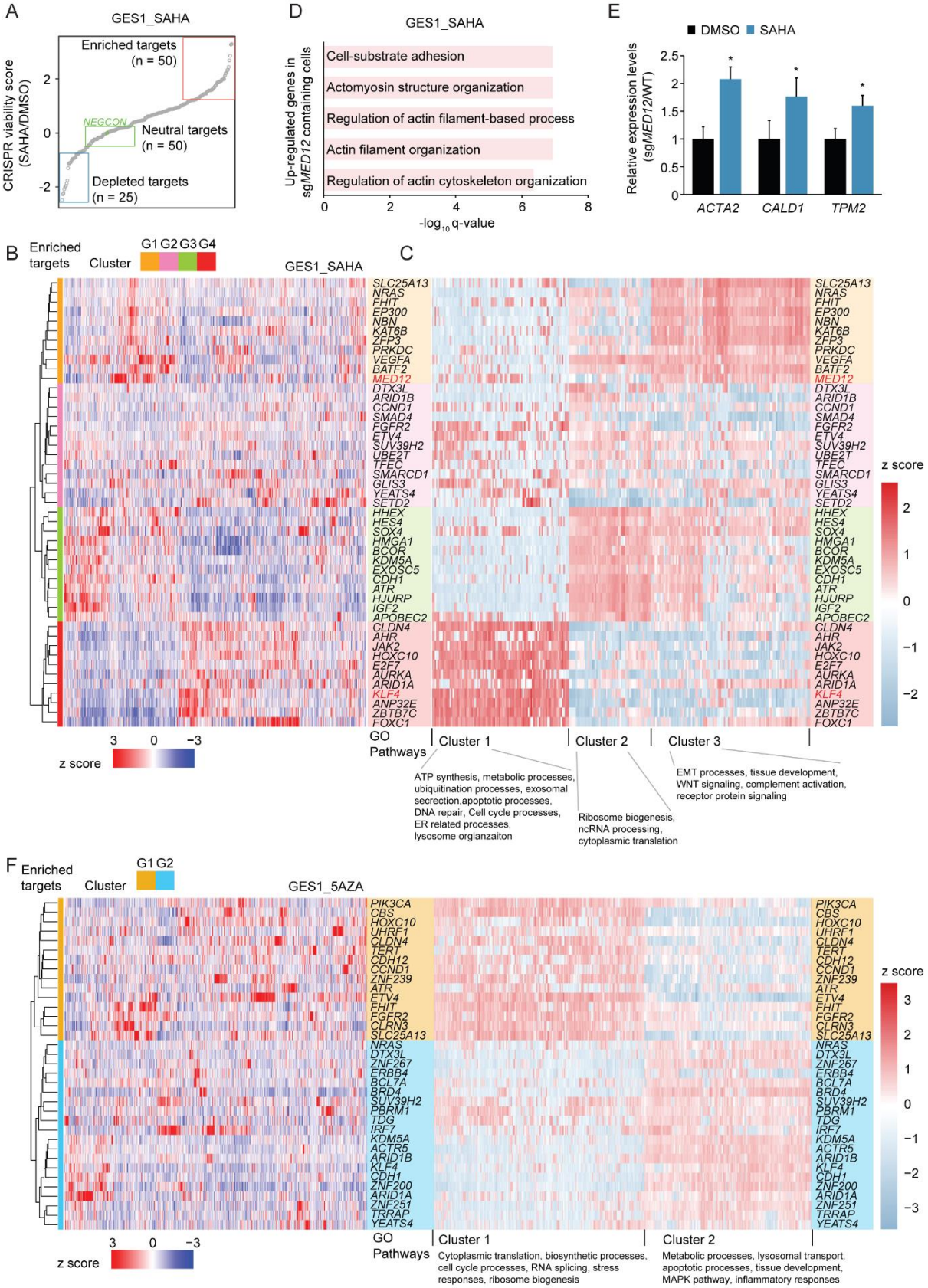

A
GES1_SAHA
Enriched targets (n = 50)
NEGCON
Neutral targets (n = 50)
Depleted targets (n = 25)
CRISPR viability score (SAHA/DMSO)
D
GES1_SAHA
Up-regulated genes in sgMED12 containing cells
Cell-substrate adhesion
Actomyosin structure organization
Regulation of actin filament-based process
Actin filament organization
Regulation of actin cytoskeleton organization
-log10 q-value
E
DMSO
SAHA
Relative expression levels (sgMED12/WT)
ACTA2
CALD1
TPM2
B
Enriched targets
Cluster
G1 G2 G3 G4
GES1_SAHA
C
z score
GO Pathways
Cluster 1
Cluster 2
Cluster 3
ATP synthesis, metabolic processes, ubiquitination processes, exosomal secrection,apoptotic processes, DNA repair, Cell cycle processes, ER related processes, lysosome orgianzaiton
Ribosome biogenesis, ncRNA processing, cytoplasmic translation
EMT processes, tissue development, WNT signaling, complement activation, receptor protein signaling
F
Enriched targets
Cluster
G1 G2
GES1_5AZA
z score
GO Pathways
Cluster 1
Cluster 2
Cytoplasmic translation, biosynthetic processes, cell cycle processes, RNA splicing, stress responses, ribosome biogenesis
Metabolic processes, lysosomal transport, apoptotic processes, tissue development, MAPK pathway, inflammatory responses

Figure 4. dcPerturb-seq Perturbation-Associated Molecular Phenotypes Reveal Distinct Pathways Associated with Therapy Resistance

(A) Diagram illustrating the classification of gene perturbation targets as enriched, neutral, or depleted based on normalized CRISPR viability scores.

(B) Hierarchical clustering of enriched gene perturbation targets in response to SAHA treatment, based on *Z* scored expression profiles.

(C) Heatmap depicting composite scores for Gene Ontology (GO) pathways associated with enriched gene perturbation targets in response to SAHA treatment. Clusters were annotated by their cellular functions.

(D) Biological pathway enrichment analysis of genes perturbed by sgRNAs targeting *MED12* in GES1 cells under SAHA treatment.

(E) qRT-PCR analysis of relative expression levels of *ACTA2*, *CALD1*, and *TPM2* genes in *MED12*-perturbed cells compared to WT cells under DMSO and SAHA treatment (GES1) (n=3; mean ± SD) (*, $p < 0.05$, two-sided *t*-test).

(F) Hierarchical clustering of enriched targets in response to 5AZA treatment based on *Z* scored expression profiles (left); Heatmap showing composite scores for GO pathways associated with enriched target in response to 5AZA treatment, with clusters labeled using function annotations.

Figure 5

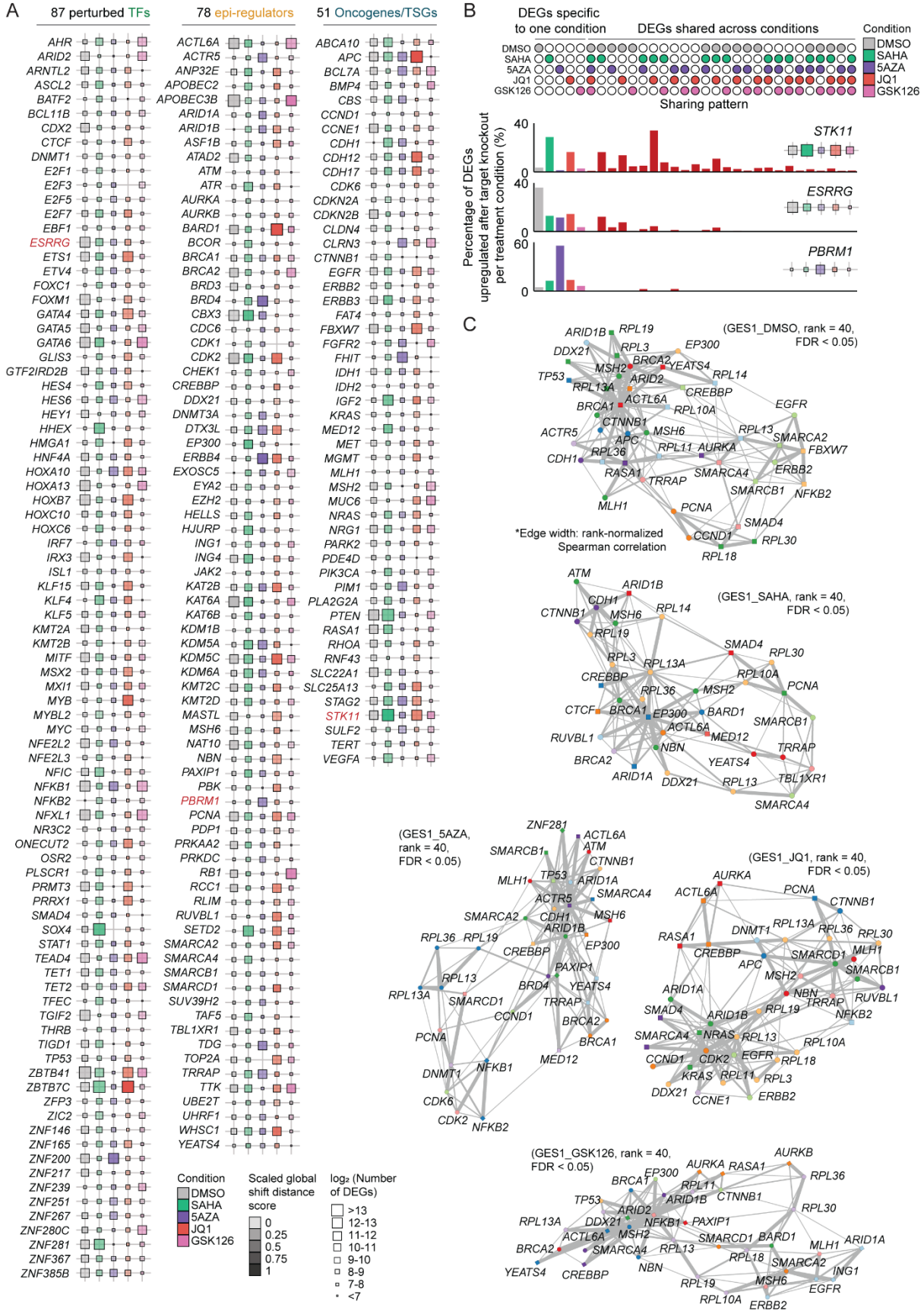

A
87 perturbed TFs
78 epi-regulators
51 Oncogenes/TSGs
B
DEGs specific to one condition
DEGs shared across conditions
Condition
DMSO
SAHA
5AZA
JQ1
GSK126
Sharing pattern
Percentage of DEGs upregulated after target knockout per treatment condition (%)
STK11
ESRRG
PBRM1
C
(GES1_DMSO, rank = 40, FDR < 0.05)
*Edge width: rank-normalized Spearman correlation
(GES1_SAHA, rank = 40, FDR < 0.05)
(GES1_5AZA, rank = 40, FDR < 0.05)
(GES1_JQ1, rank = 40, FDR < 0.05)
(GES1_GSK126, rank = 40, FDR < 0.05)
Condition
DMSO
SAHA
5AZA
JQ1
GSK126
Scaled global shift distance score
0
0.25
0.5
0.75
1
log2 (Number of DEGs)
>13
12-13
11-12
10-11
9-10
8-9
7-8
<7

Figure 5. Epigenetic Drug Exposures Reveal New dcPerturb-seq Gene-Gene Network Interactions

(A) Graphic cartoon representing the magnitude of transcriptomic alterations for each dcPerturb-seq gene perturbation under DMSO and epigenetic drug exposure, relative to negative control cells. Each treatment was analyzed independently and represented by a distinct color.

(B) Percentage of up-regulated DEGs (AD test, FDR < 0.05) following each drug treatment, grouped by sharing patter. (Left) Genes upregulated in only one condition (left) or shared by two or more conditions (right). (Bottom) Examples of distributions associated with *STK11*, *ESRRG* and *PBRM1* perturbations.

(C) Gene-gene network interactions using correlations of mean molecular phenotype profiles from dcPerturb-seq under DMSO and epigenetic drug treatment. Edge width: rank-normalized Spearman correlation for dcPerturb-seq network (rank = 40, FDR < 0.05, permutation test with 10,000 random permutations).

Figure 6

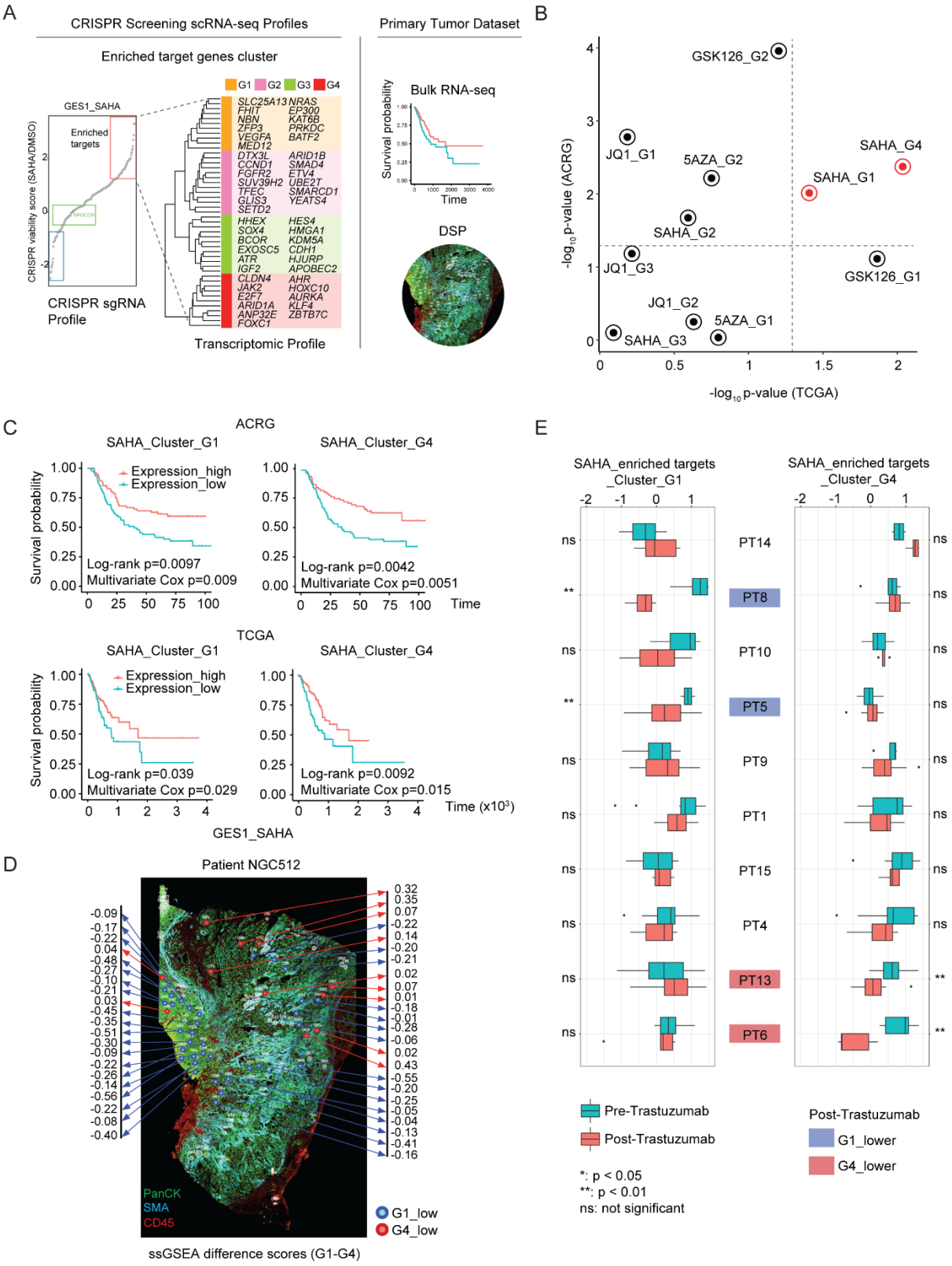

A
CRISPR Screening scRNA-seq Profiles
Primary Tumor Dataset
Enriched target genes cluster
GES1_SAHA
Enriched targets
CRISPR viability score (SAHA/DMSO)
CRISPR sgRNA Profile
G1 G2 G3 G4
SLC25A13 NRAS
FHIT EP300
NBN KAT6B
ZFP3 PRKDC
VEGFA BATF2
MED12
DTX3L ARID1B
CCND1 SMAD4
FGFR2 ETV4
SUV39H2 UBE2T
TFEC SMARCD1
GLIS3 YEATS4
SETD2
HHEX HES4
SOX4 HMGA1
BCOR KDM5A
EXOSC5 CDH1
ATR HJURP
IGF2 APOBEC2
CLDN4 AHR
JAK2 HOXC10
E2F7 AURKA
ARID1A KLF4
ANP32E ZBTB7C
FOXC1
Transcriptomic Profile
Bulk RNA-seq
Survival probability
Time
DSP
B
GSK126_G2
JQ1_G1
5AZA_G2
SAHA_G4
SAHA_G1
SAHA_G2
JQ1_G3
GSK126_G1
JQ1_G2
5AZA_G1
SAHA_G3
-log10 p-value (ACRG)
-log10 p-value (TCGA)
C
ACRG
SAHA_Cluster_G1
SAHA_Cluster_G4
Expression_high
Expression_low
Survival probability
Log-rank p=0.0097
Multivariate Cox p=0.009
Log-rank p=0.0042
Multivariate Cox p=0.0051
Time
TCGA
Log-rank p=0.039
Multivariate Cox p=0.029
Log-rank p=0.0092
Multivariate Cox p=0.015
Time (x10³)
GES1_SAHA
D
Patient NGC512
PanCK
SMA
CD45
G1_low
G4_low
ssGSEA difference scores (G1-G4)
E
SAHA_enriched targets _Cluster_G1
SAHA_enriched targets _Cluster_G4
PT14
PT8
PT10
PT5
PT9
PT1
PT15
PT4
PT13
PT6
Pre-Trastuzumab
Post-Trastuzumab
Post-Trastuzumab
G1_lower
G4_lower
*: p < 0.05
**: p < 0.01
ns: not significant

Figure 6. Application of dcPerturb-seq Molecular Phenotypes in Primary GCs

(A) Diagram illustrating the application of molecular phenotypes associated with enriched target genes (left) to primary tumor expression data and NanoString GeoMx Digital Spatial Profiler (DSP) profiles.

(B) Associations between the expression of molecular phenotype gene clusters and GC patient survival in the ACRG (n=300) and TCGA (n=415) cohorts. SAHA_G1 and SAHA_G4 molecular phenotypes are associated with patient outcome in both cohorts.

(C) Survival analysis of GCs with high and low expression of SAHA-enriched molecular phenotypes clusters G1 and G4 in the ACRG and TCGA cohorts. Significance was determined using the univariate log-rank test and multivariate Cox proportional hazard model.

(D) Stained GeoMx DSP slides annotated with low expression of SAHA-enriched clusters G1 (blue) and G4 (red), based on ssGSEA difference scores (G1-G4), for patient NGC512. Each circle within the stained slide represents a tumor ROI, annotated with ssGSEA difference scores.

(E) Boxplots showing significantly reduced levels of SAHA-enriched cluster G1 in PT5 and PT8 samples (left) and cluster G4 in PT13 and PT6 samples after the develop of trastuzumab resistance, as accessed using DSP profiling. (*, $p < 0.05$, **, $p < 0.01$, ns, not significant, Wilcoxon rank sum test).

Figure 1

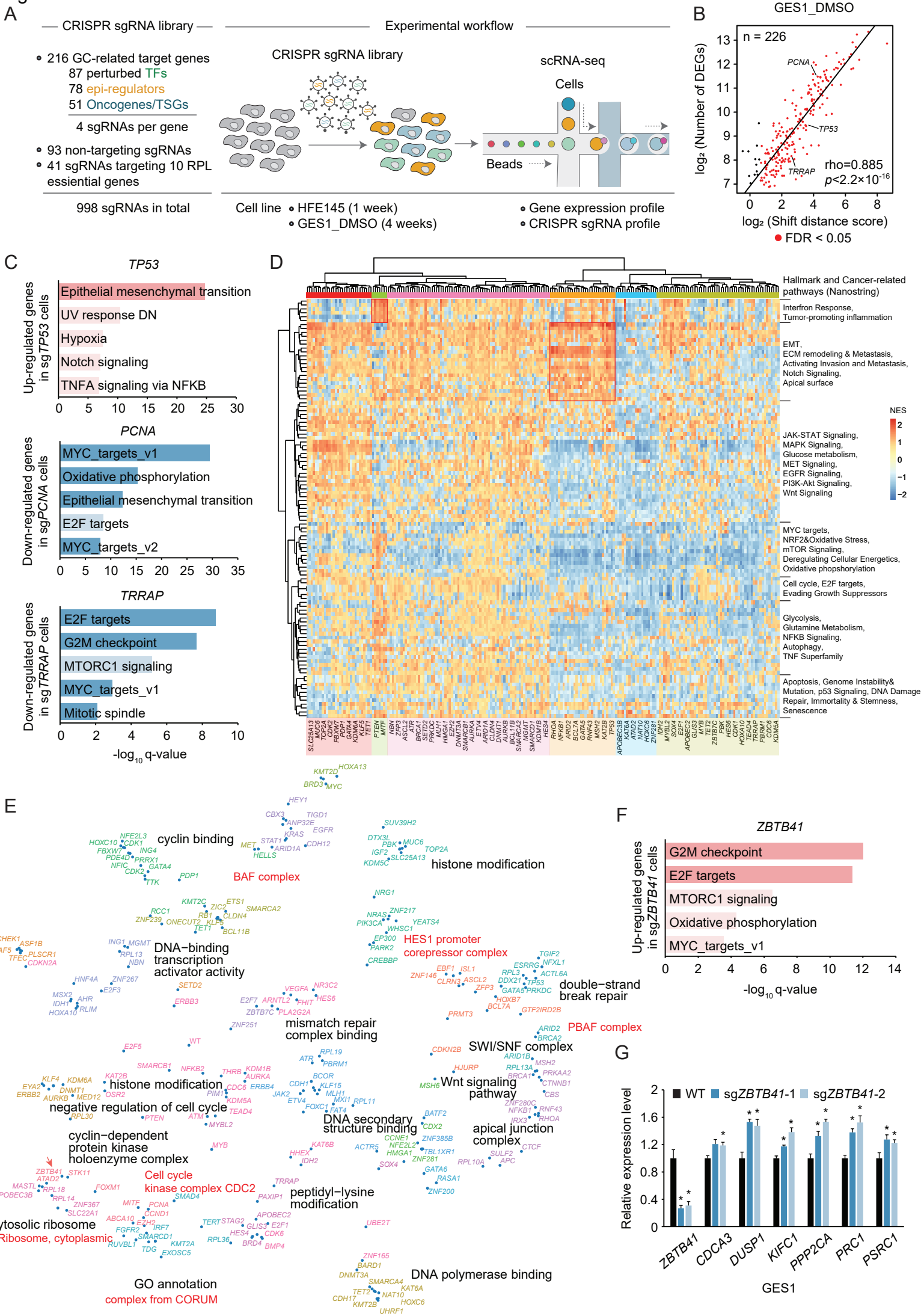

A
CRISPR sgRNA library
Experimental workflow
216 GC-related target genes
87 perturbed TFs
78 epi-regulators
51 Oncogenes/TSGs
4 sgRNAs per gene
93 non-targeting sgRNAs
41 sgRNAs targeting 10 RPL essiential genes
998 sgRNAs in total
CRISPR sgRNA library
scRNA-seq
Cells
Beads
Cell line
HFE145 (1 week)
GES1_DMSO (4 weeks)
Gene expression profile
CRISPR sgRNA profile
B
GES1_DMSO
n = 226
PCNA
TP53
TRRAP
rho=0.885
p<2.2×10-16
log2 (Number of DEGs)
log2 (Shift distance score)
FDR < 0.05
C
TP53
Up-regulated genes in sgTP53 cells
Epithelial mesenchymal transition
UV response DN
Hypoxia
Notch signaling
TNFA signaling via NFKB
PCNA
Down-regulated genes in sgPCNA cells
MYC_targets_v1
Oxidative phosphorylation
Epithelial mesenchymal transition
E2F targets
MYC_targets_v2
TRRAP
Down-regulated genes in sgTRRAP cells
E2F targets
G2M checkpoint
MTORC1 signaling
MYC_targets_v1
Mitotic spindle
-log10 q-value
D
Hallmark and Cancer-related pathways (Nanostring)
Interfron Response, Tumor-promoting inflammation
EMT, ECM remodeling & Metastasis, Activating Invasion and Metastasis, Notch Signaling, Apical surface
JAK-STAT Signaling, MAPK Signaling, Glucose metabolism, MET Signaling, EGFR Signaling, PI3K-Akt Signaling, Wnt Signaling
MYC targets, NRF2&Oxidative Stress, mTOR Signaling, Deregulating Cellular Energetics, Oxidative phopshorylation
Cell cycle, E2F targets, Evading Growth Suppressors
Glycolysis, Glutamine Metabolism, NFKB Signaling, Autophagy, TNF Superfamily
Apoptosis, Genome Instability& Mutation, p53 Signaling, DNA Damage Repair, Immortality & Stemness, Senescence
NES
E
cyclin binding
BAF complex
histone modification
DNA−binding transcription activator activity
HES1 promoter corepressor complex
double−strand break repair
mismatch repair complex binding
PBAF complex
SWI/SNF complex
histone modification
negative regulation of cell cycle
Wnt signaling pathway
DNA secondary structure binding
apical junction complex
cyclin−dependent protein kinase holoenzyme complex
Cell cycle kinase complex CDC2
peptidyl−lysine modification
cytosolic ribosome
Ribosome, cytoplasmic
DNA polymerase binding
GO annotation
complex from CORUM
F
ZBTB41
Up-regulated genes in sgZBTB41 cells
G2M checkpoint
E2F targets
MTORC1 signaling
Oxidative phosphorylation
MYC_targets_v1
-log10 q-value
G
WT
sgZBTB41-1
sgZBTB41-2
Relative expression level
ZBTB41
CDCA3
DUSP1
KIFC1
PPP2CA
PRC1
PSRC1
GES1

Figure 2

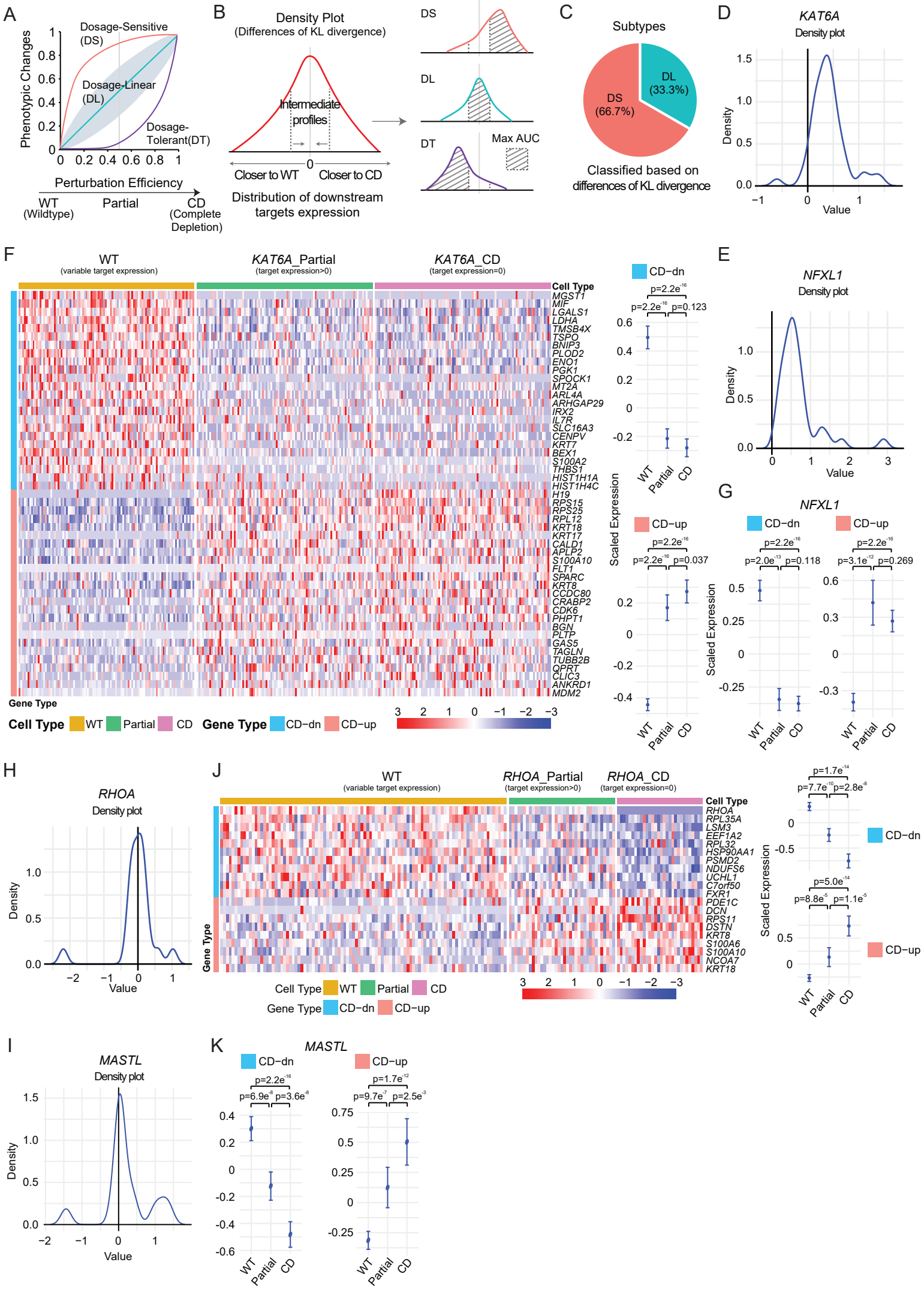

A
Phenotypic Changes
Dosage-Sensitive (DS)
Dosage-Linear (DL)
Dosage-Tolerant(DT)
Perturbation Efficiency
WT (Wildtype)
Partial
CD (Complete Depletion)
B
Density Plot
(Differences of KL divergence)
Intermediate profiles
Closer to WT
Closer to CD
Distribution of downstream targets expression
DS
DL
DT
Max AUC
C
Subtypes
DS (66.7%)
DL (33.3%)
Classified based on differences of KL divergence
D
KAT6A
Density plot
Density
Value
E
NFXL1
Density plot
F
WT
(variable target expression)
KAT6A_Partial
(target expression>0)
KAT6A_CD
(target expression=0)
Cell Type
Gene Type
CD−dn
CD−up
Scaled Expression
G
H
RHOA
Density plot
J
RHOA_Partial
RHOA_CD
I
MASTL
Density plot
K

Figure 3

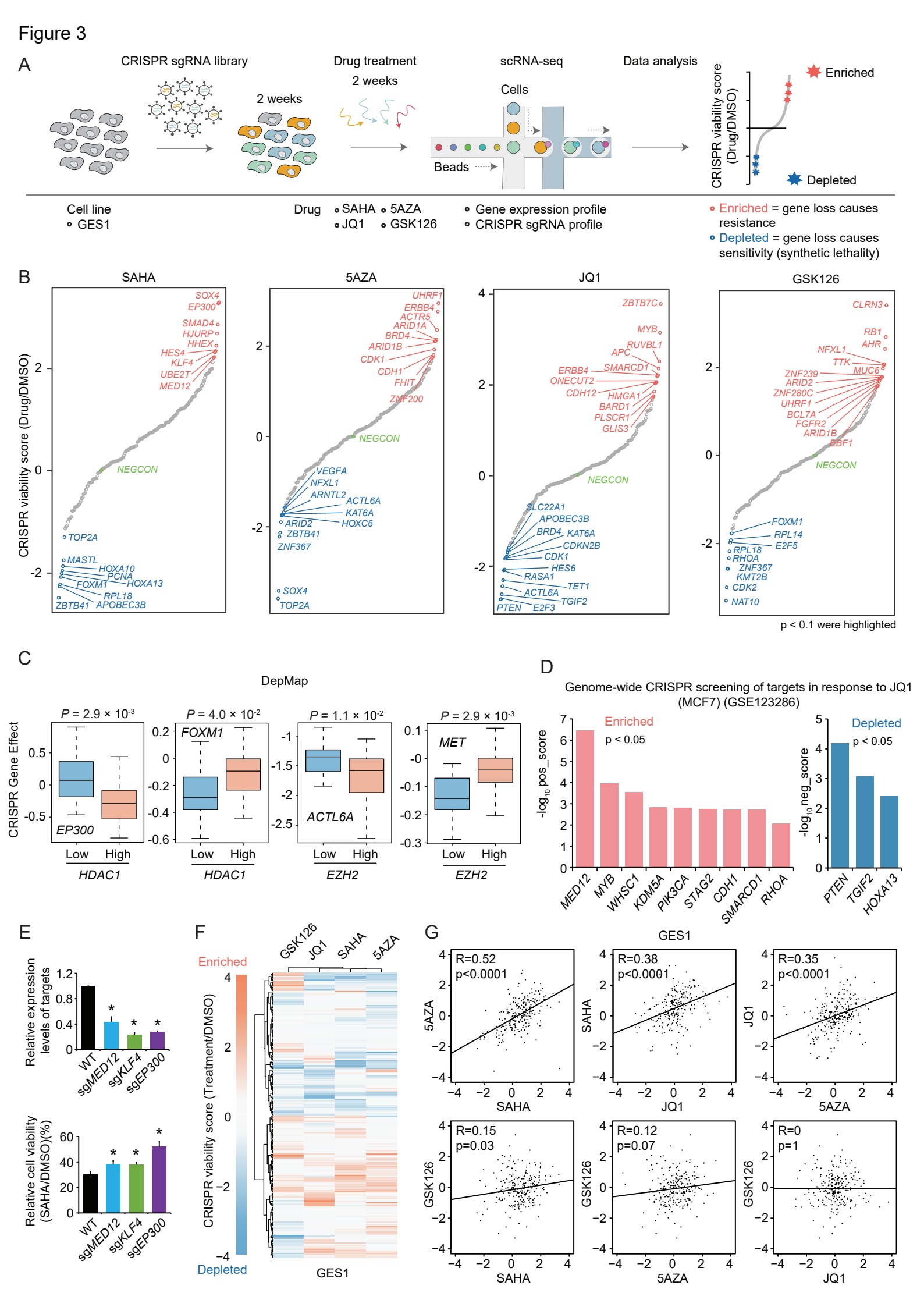

A
CRISPR sgRNA library
2 weeks
Drug treatment
2 weeks
scRNA-seq
Cells
Beads
Data analysis
CRISPR viability score (Drug/DMSO)
Enriched
Depleted
Cell line
GES1
Drug
SAHA
5AZA
JQ1
GSK126
Gene expression profile
CRISPR sgRNA profile
Enriched = gene loss causes resistance
Depleted = gene loss causes sensitivity (synthetic lethality)
B
SAHA
5AZA
JQ1
GSK126
CRISPR viability score (Drug/DMSO)
NEGCON
p < 0.1 were highlighted
C
DepMap
CRISPR Gene Effect
P = 2.9 × 10-3
P = 4.0 × 10-2
P = 1.1 × 10-2
P = 2.9 × 10-3
EP300
FOXM1
ACTL6A
MET
Low
High
HDAC1
EZH2
D
Genome-wide CRISPR screening of targets in response to JQ1 (MCF7) (GSE123286)
Enriched
p < 0.05
Depleted
p < 0.05
-log10 pos_score
-log10 neg_score
MED12
MYB
WHSC1
KDM5A
PIK3CA
STAG2
CDH1
SMARCD1
RHOA
PTEN
TGIF2
HOXA13
E
Relative expression levels of targets
Relative cell viability (SAHA/DMSO)(%)
WT
sgMED12
sgKLF4
sgEP300
F
GSK126
JQ1
SAHA
5AZA
Enriched
Depleted
CRISPR viability score (Treatment/DMSO)
GES1
G
GES1
R=0.52
p<0.0001
R=0.38
p<0.0001
R=0.35
p<0.0001
R=0.15
p=0.03
R=0.12
p=0.07
R=0
p=1
5AZA
SAHA
JQ1
GSK126

Figure 4

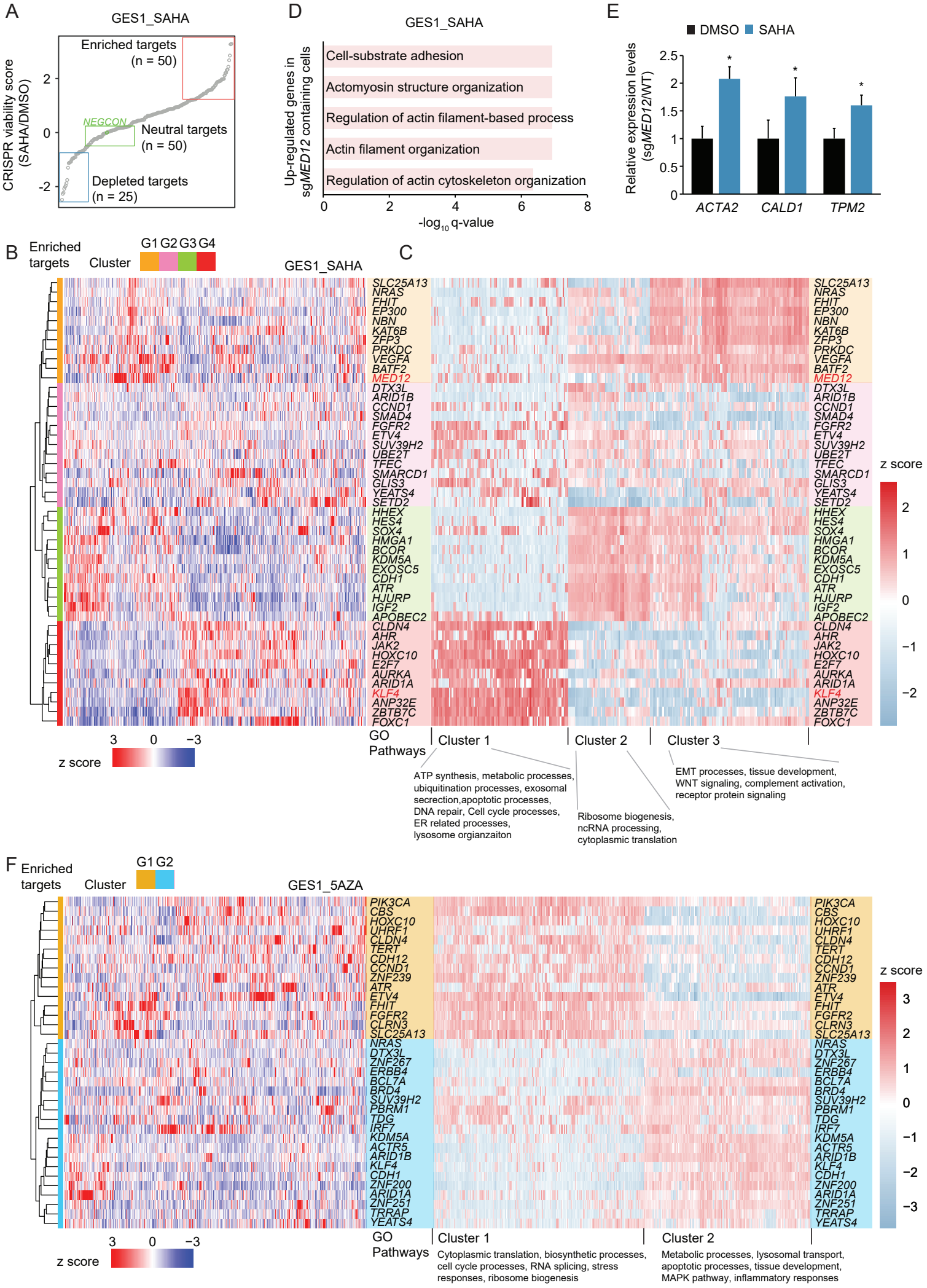

A
GES1_SAHA
Enriched targets (n = 50)
NEGCON
Neutral targets (n = 50)
Depleted targets (n = 25)
CRISPR viability score (SAHA/DMSO)
D
GES1_SAHA
Cell-substrate adhesion
Actomyosin structure organization
Regulation of actin filament-based process
Actin filament organization
Regulation of actin cytoskeleton organization
Up-regulated genes in sgMED12 containing cells
-log10 q-value
E
DMSO
SAHA
Relative expression levels (sgMED12/WT)
ACTA2
CALD1
TPM2
B
Enriched targets
Cluster
G1 G2 G3 G4
GES1_SAHA
z score
C
GO Pathways
Cluster 1
Cluster 2
Cluster 3
ATP synthesis, metabolic processes, ubiquitination processes, exosomal secrection,apoptotic processes, DNA repair, Cell cycle processes, ER related processes, lysosome orgianzaiton
Ribosome biogenesis, ncRNA processing, cytoplasmic translation
EMT processes, tissue development, WNT signaling, complement activation, receptor protein signaling
F
Enriched targets
Cluster
G1 G2
GES1_5AZA
z score
GO Pathways
Cluster 1
Cytoplasmic translation, biosynthetic processes, cell cycle processes, RNA splicing, stress responses, ribosome biogenesis
Cluster 2
Metabolic processes, lysosomal transport, apoptotic processes, tissue development, MAPK pathway, inflammatory responses

Figure 5

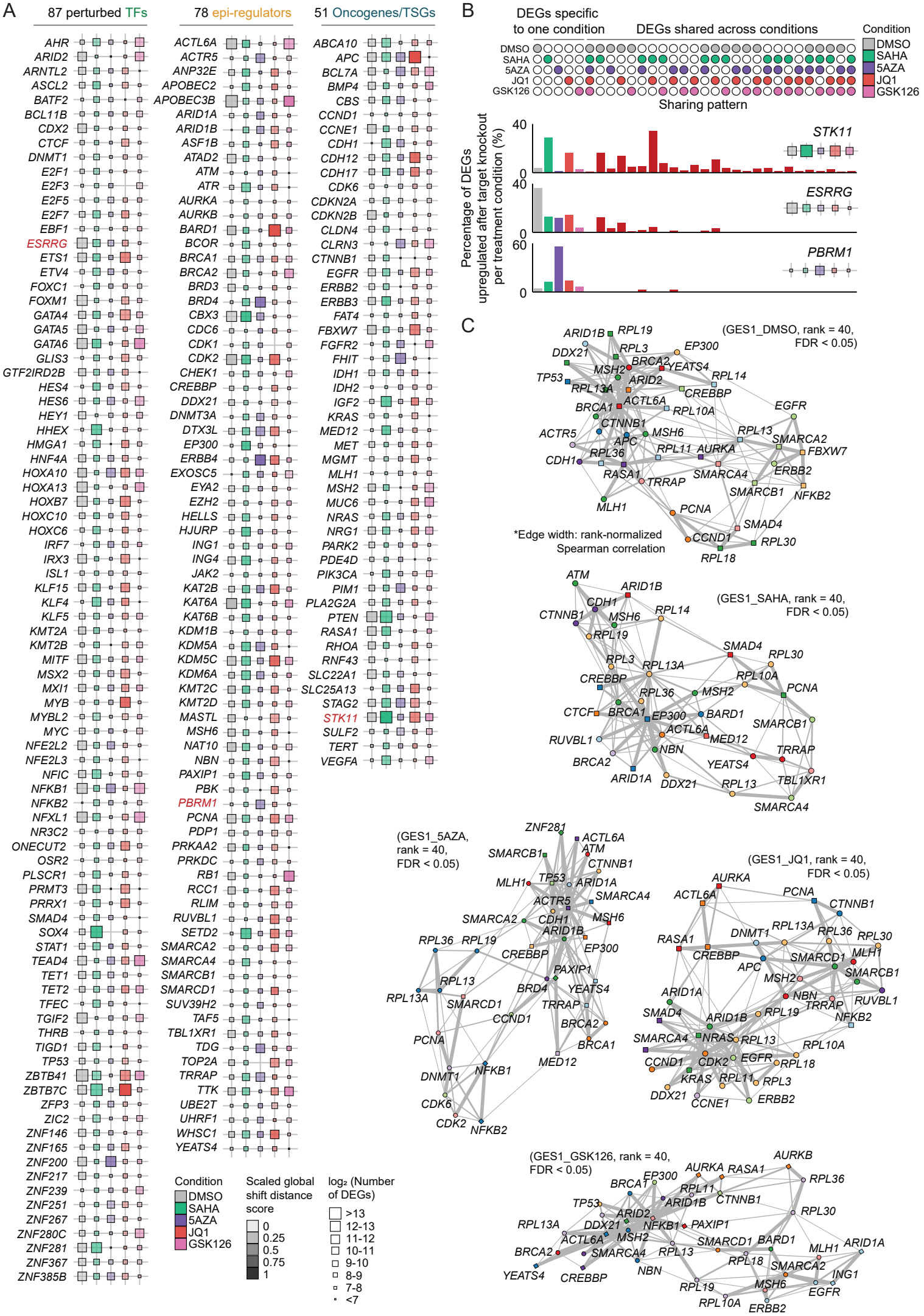
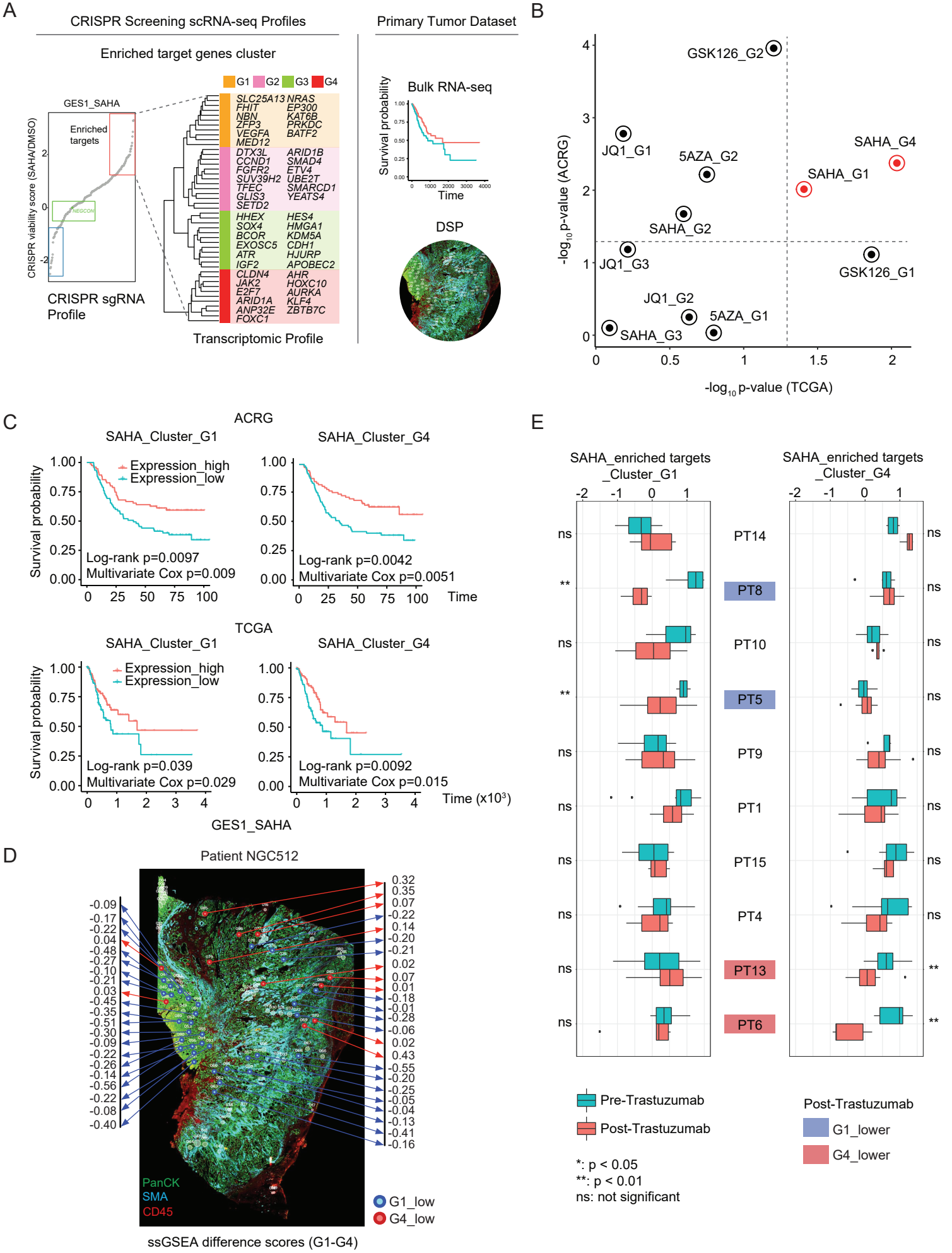

A
87 perturbed TFs
78 epi-regulators
51 Oncogenes/TSGs
B
DEGs specific to one condition
DEGs shared across conditions
Condition
DMSO
SAHA
5AZA
JQ1
GSK126
Sharing pattern
Percentage of DEGs upregulated after target knockout per treatment condition (%)
STK11
ESRRG
PBRM1
C
(GES1_DMSO, rank = 40, FDR < 0.05)
*Edge width: rank-normalized Spearman correlation
(GES1_SAHA, rank = 40, FDR < 0.05)
(GES1_5AZA, rank = 40, FDR < 0.05)
(GES1_JQ1, rank = 40, FDR < 0.05)
(GES1_GSK126, rank = 40, FDR < 0.05)
Condition
DMSO
SAHA
5AZA
JQ1
GSK126
Scaled global shift distance score
0
0.25
0.5
0.75
1
log2 (Number of DEGs)
>13
12-13
11-12
10-11
9-10
8-9
7-8
<7

Figure 6


A
CRISPR Screening scRNA-seq Profiles
Primary Tumor Dataset
Enriched target genes cluster
GES1_SAHA
Enriched targets
NEGCON
CRISPR viability score (SAHA/DMSO)
CRISPR sgRNA Profile
G1 G2 G3 G4
SLC25A13 NRAS FHIT EP300 NBN KAT6B ZFP3 PRKDC VEGFA BATF2 MED12
DTX3L ARID1B CCND1 SMAD4 FGFR2 ETV4 SUV39H2 UBE2T TFEC SMARCD1 GLIS3 YEATS4 SETD2
HHEX HES4 SOX4 HMGA1 BCOR KDM5A EXOSC5 CDH1 ATR HJURP IGF2 APOBEC2
CLDN4 AHR JAK2 HOXC10 E2F7 AURKA ARID1A KLF4 ANP32E ZBTB7C FOXC1
Transcriptomic Profile
Bulk RNA-seq
Survival probability
Time
DSP
B
GSK126_G2
JQ1_G1
5AZA_G2
SAHA_G4
SAHA_G1
SAHA_G2
JQ1_G3
GSK126_G1
JQ1_G2
5AZA_G1
SAHA_G3
-log10 p-value (ACRG)
-log10 p-value (TCGA)
C
ACRG
SAHA_Cluster_G1
SAHA_Cluster_G4
Expression_high
Expression_low
Survival probability
Log-rank p=0.0097
Multivariate Cox p=0.009
Log-rank p=0.0042
Multivariate Cox p=0.0051
Time
TCGA
Log-rank p=0.039
Multivariate Cox p=0.029
Log-rank p=0.0092
Multivariate Cox p=0.015
Time (x10³)
GES1_SAHA
D
Patient NGC512
PanCK
SMA
CD45
G1_low
G4_low
ssGSEA difference scores (G1-G4)
E
SAHA_enriched targets _Cluster_G1
SAHA_enriched targets _Cluster_G4
PT14
PT8
PT10
PT5
PT9
PT1
PT15
PT4
PT13
PT6
Pre-Trastuzumab
Post-Trastuzumab
Post-Trastuzumab
G1_lower
G4_lower
*: p < 0.05
**: p < 0.01
ns: not significant