
Scalable, compressed phenotypic screening using pooled perturbations

In the format provided by the
authors and unedited

Table of Contents

- 1. Key Resource Table**
- 2. Supplementary Table Legends**
- 3. Supplementary Figures**

Supplementary Figure 1: Viability and quality control analyses for the PBMC compressed screen.

Supplementary Figure 2: Perturbation effect coefficients from deconvolution on LPS+control data.

Supplementary Figure 3: Perturbation effect coefficients from deconvolution on IFN β +control data.

- 4. Supplementary Figure Legends**
- 5. Supplementary Methods**

Key Resource Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
See Table S1 for a list of small molecules and ligands screened in U-2 OS, PDAC organoid and PBMC screens		
See Table S8 for a list of organoid culture reagents and recombinant proteins		
90 JUMP-MOA compounds for the PBMC screen	Broad Drug Repurposing Hub: JUMP-Cell Painting Consortium	https://github.com/jump-cellpainting/JUMP-MOA
Critical commercial assays		
Honeycomb scRNA-seq kit	Honeycomb	
Nextseq 500/550 High output v2.5 kit (75 cycles)	Illumina	Cat# 20024906
High Sensitivity D5000 ScreenTape	Agilent	Cat# 5067-5592
Qubit dsDNA High-Sensitivity kit	Thermo Fisher	Cat# Q32854
BD hashing kit	BD Biosciences	
Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Dual Index) with Feature Barcode technology for Cell Surface Protein	10X Genomics	#PN-1000268
TotalSeq-B with 10x Feature Barcoding Technology kit	BioLegend	Cat #633781
Deposited data		
Ground truth and compressed Cell Painting screens	This paper	https://www.ebi.ac.uk/biostudies/bioimages/studies/S-BIAD1319 .

Compressed and validation scRNA-seq data from PDAC organoid screens	This paper	https://singlecell.broadinstitute.org/single_cell/study/SCP2020/efficiently-generating-bio-chemical-perturbation-signatures-in-complex-cellular-systems-by-compressed-screening#study-summary .
Compressed and validation scRNA-seq data from PBMC screens	This paper	https://singlecell.broadinstitute.org/single_cell/study/SCP2622/compressed-phenotypic-screening-empowers-scalable-biological-discovery-pbmc-screens
single-cell transcriptomic data from PDAC patient samples	Raghavan et al. 2021	https://singlecell.broadinstitute.org/single_cell/study/SCP1644 dbGaP:phs002712.v1.p1

Primary PDAC genomic and transcriptomic data (TCGA, Pancreatic Ductal Adenocarcinoma)	Cancer Genome Atlas Research Network, 2017	https://portal.gdc.cancer.gov/projects/TCGA-PAADdbGaP:phs000178.v11.p8
Genomic Data Commons (GDC) TCGA Pancreatic Cancer (PAAD) gene expression RNAseq and phenotype (survival) data	Genomic Data Commons Data Portal	https://xenabrowse.r.net/datapages/?cohort=GDC%20TCGA%20Pancreatic%20Cancer%20(PAAD)&removeHub=https%3A%2F%2Fxcna.treehouse.gi.ucsc.edu%3A443
Experimental models:		
U-2 OS cell line	ATCC	CAT # HTB-96
Human PDAC organoids	Raghavan <i>et al.</i> 2021	N/A
Primary Human PBMCs		N/A
Oligonucleotides		
BD-HC WTA Linker: 5' GTTGTCAAGATGCTACCGTTCAGAG 3'	Integrated DNA technologies	N/A
BD-HC Linker-UPS: 5'GAGGAAGCAGTGGTATCAACGCACCGCA GAGTTGTCAAGATGCTACCGTTCAGAG 3'	Integrated DNA technologies	N/A
Bead-specific WTA: 5' CAGTGGTATCAACGCAGAGT*A*C 3'	Integrated DNA technologies	N/A
Software and algorithms		

GitHub Repository to reproduce the results	This paper	https://github.com/ShalekLab/compressed_screening
Distributed Cell Profiler	Broad Institute	https://github.com/DistributedScience/Distributed-CellProfiler
STARsolo	Kaminow <i>et al.</i> 2021	https://dockstore.org/workflows/github.com/lilab-bcb/cumulus/STARsolo:2.0.0?tab=info
Citeseq-count	https://hoohm.github.io/CITE-seq-Count/	https://portal.firecloud.org/?return=terra#methods/bemead/CITEseq-count/7
cNMF	Kotliar <i>et al.</i> 2019	https://portal.firecloud.org/?return=terra#methods/cNMF_Ivy/cNMF_gene_file/2
CellphoneDB	Garcia-Alonso <i>et al.</i> 2021	https://github.com/ventolab/CellphoneDB
CellRanger	10x Genomics	https://portal.firecloud.org/?return=terra#methods/cumulus/cellranger_workflow/28

Supplementary Table Legends

Supplementary Table S1

1. List of small molecules for U2OS cell line screens.
2. List of protein ligands used for PDAC compressed screens.
3. List of 90 mechanism-of-action (MOA) compounds used in the PBMC compressed screens.

Supplementary Table S2

Metadata on 886 morphological features selected for U2OS CellPainting screen analysis.

Supplementary Table S3

Consensus non-negative matrix factorization (cNMF) gene expression programs (GEPs) from compressed screen and single-ligand validation screen on PDAC.

Supplementary Table S4

Mean expression of cognate receptors for ligands used in the PDAC compressed screen on negative control organoid models.

Supplementary Table S5

Single-ligand PDAC IL-4/IL-13 response GEP and other public IL-4 signatures used in this paper.

Supplementary Table S6

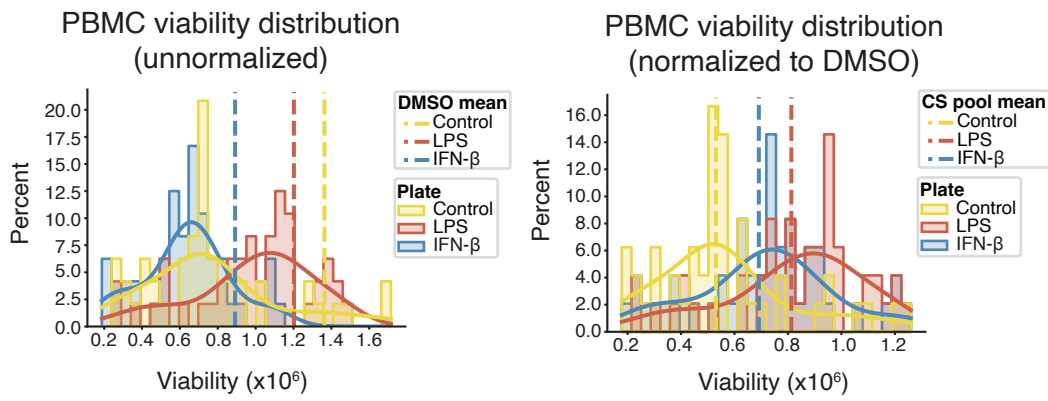
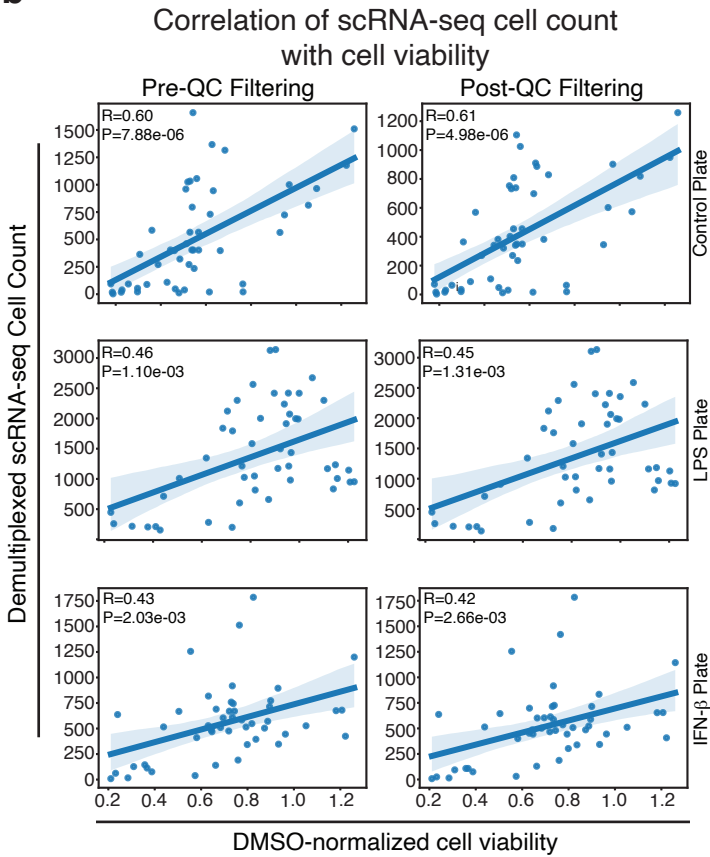
cNMF GEPs from each PBMC cell type in LPS stimulation+control and IFN β stimulation+control screens.

Supplementary Table S7

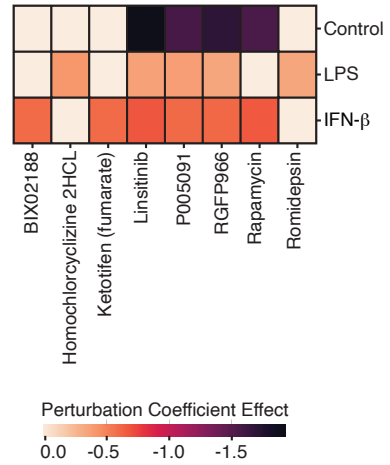
Cost analysis on compressed PBMC screens.

Supplementary Table S8

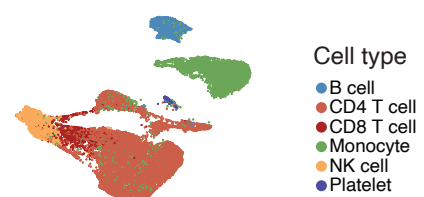
Organoid culture reagents and recombinant proteins used for PDAC organoid culture.

a**b****c**

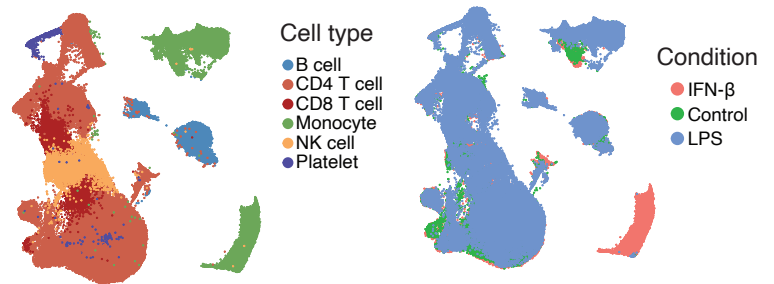
Drugs with significant effect on cell viability

**d**

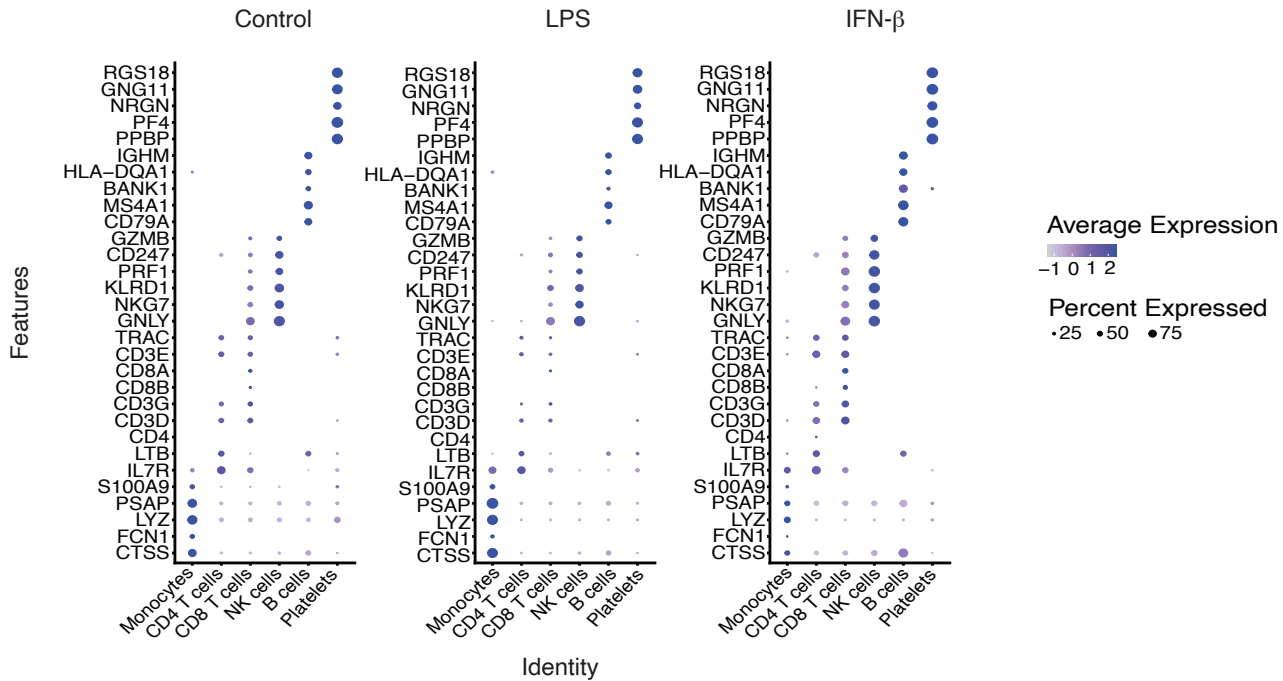
LPS (n = 65,868)

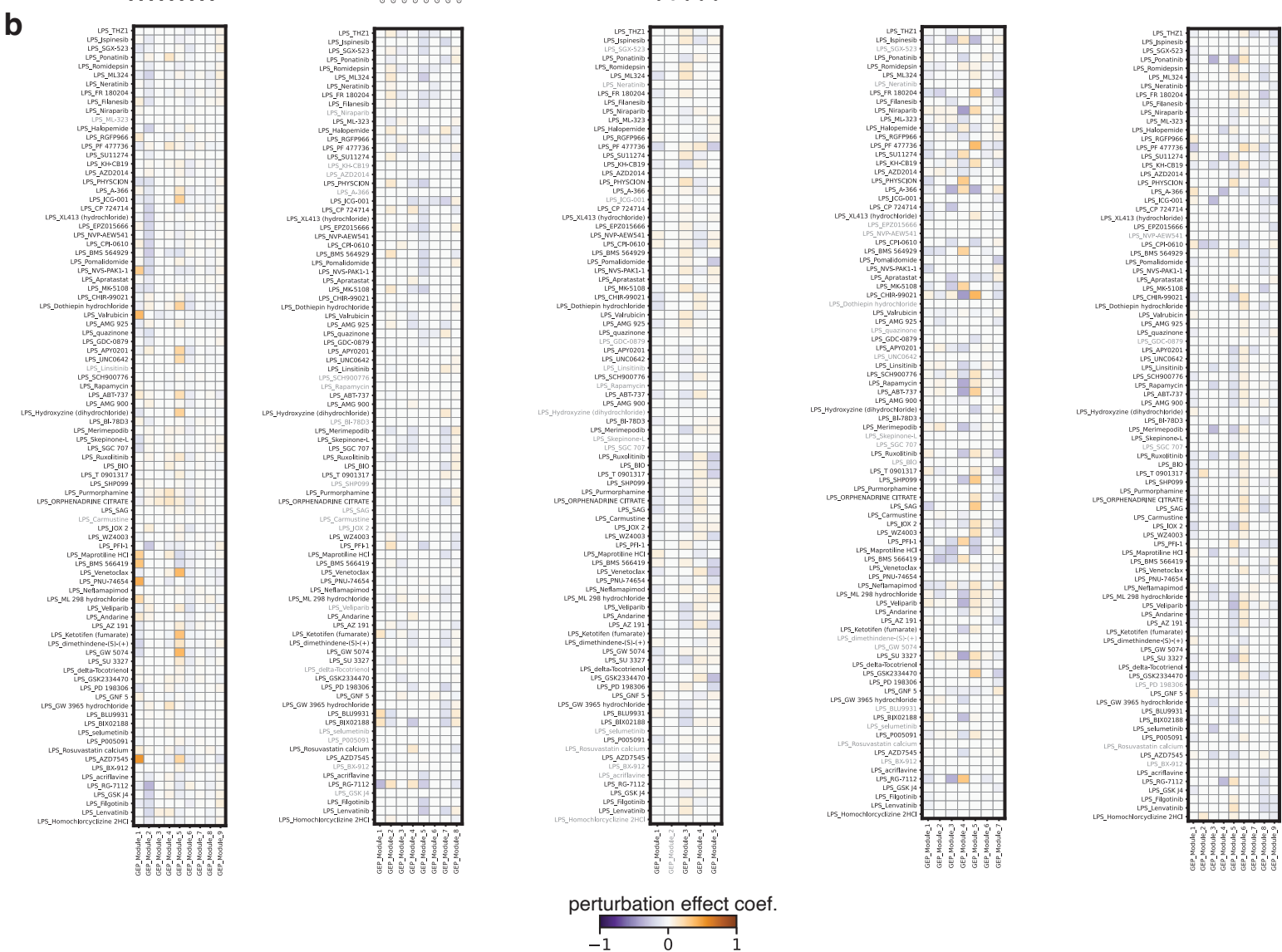
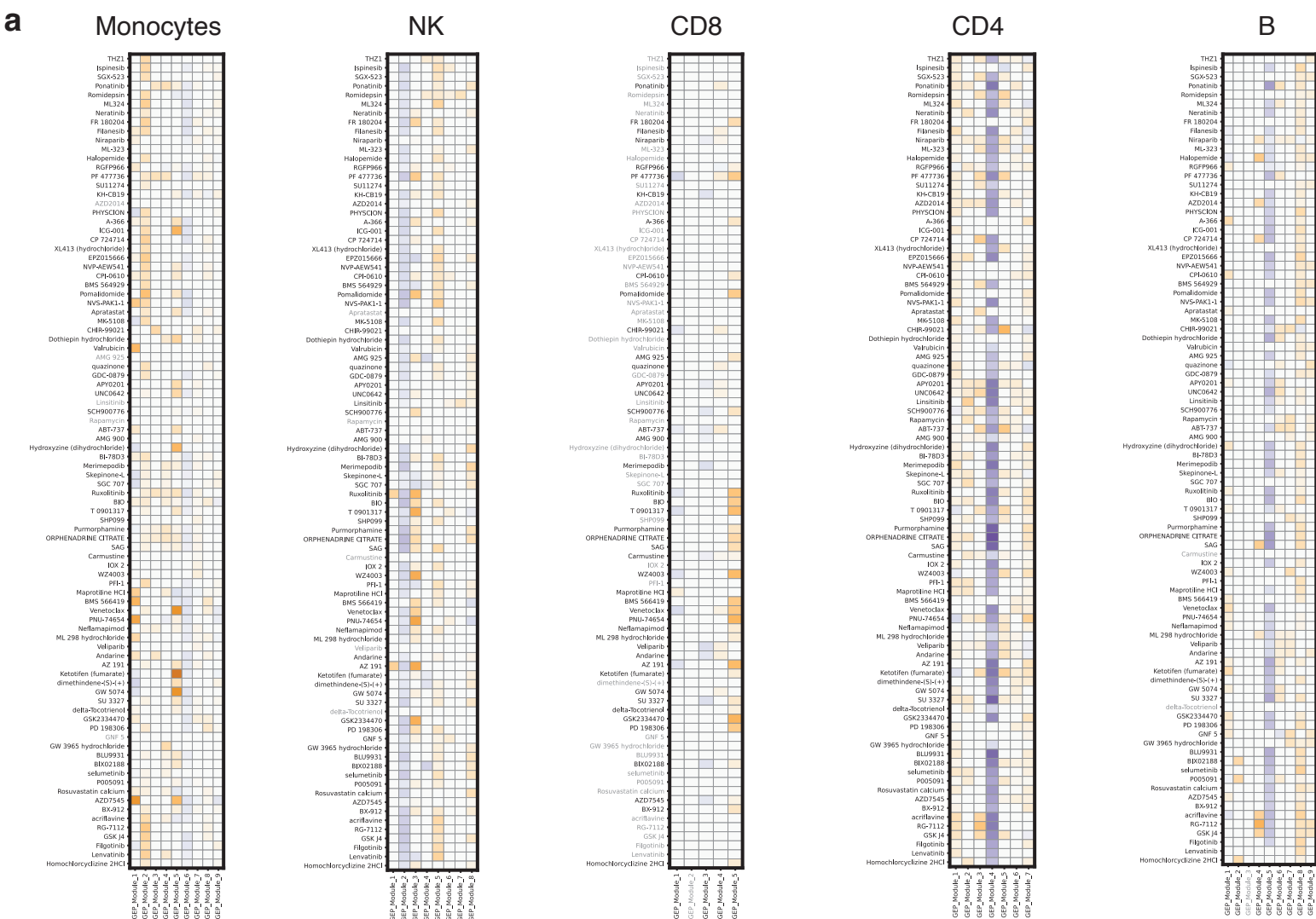
**e**

Integrated UMAP of PBMCs

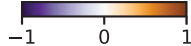
**f**

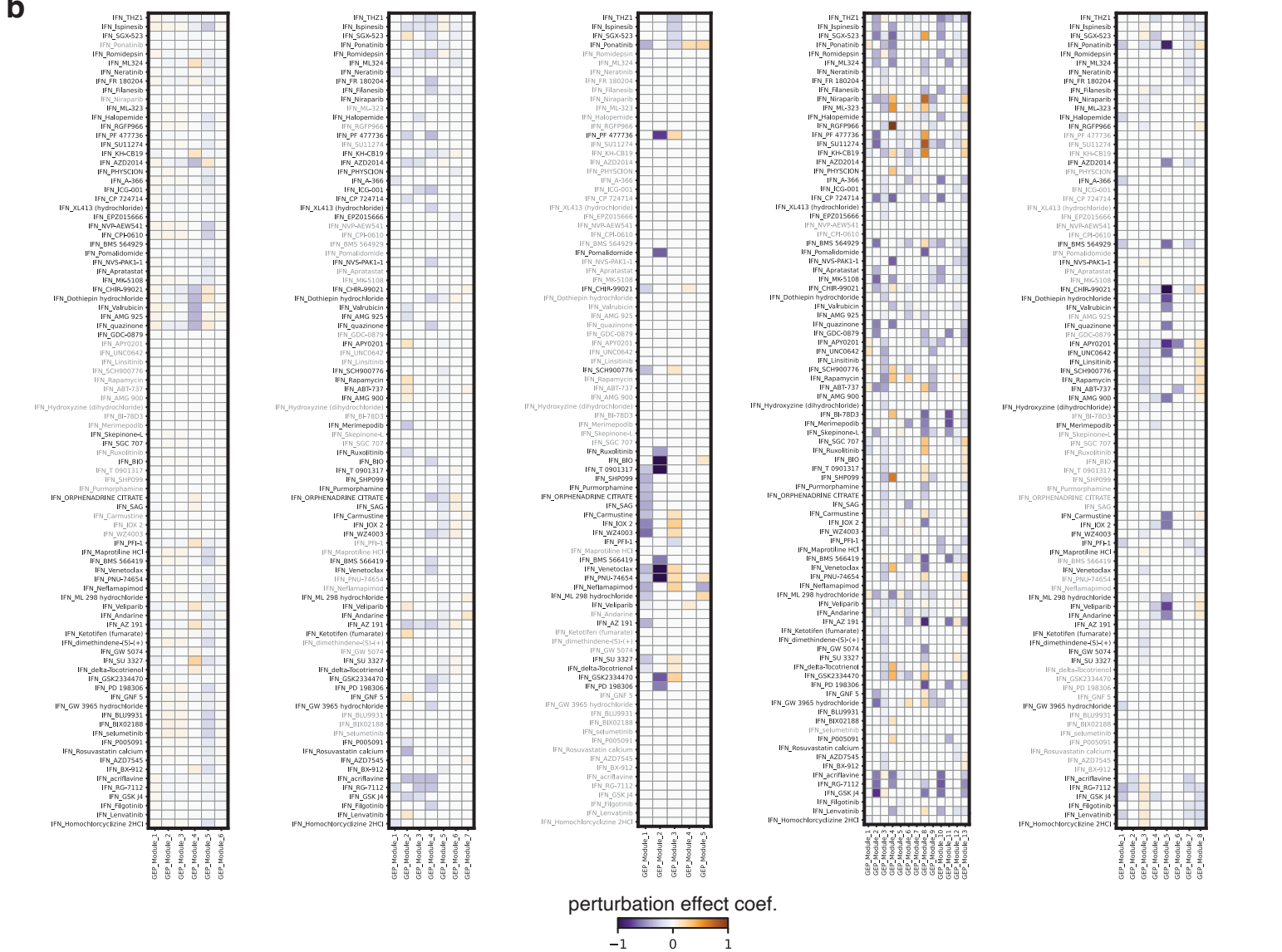
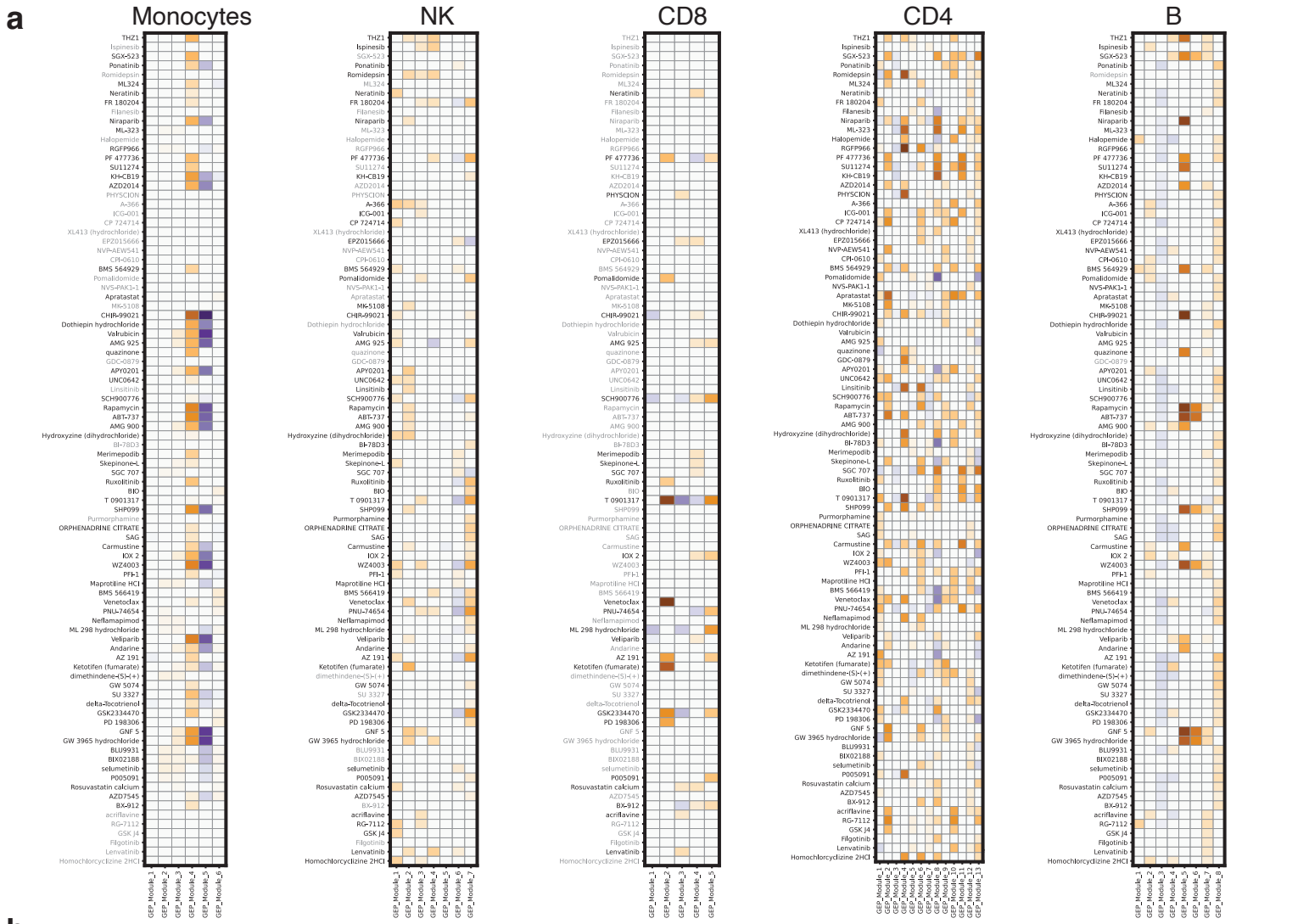
Top 5 Azimuth reference model marker genes per cell type





perturbation effect coef.





Supplementary Figure Legends

Supplementary Figure 1: Viability and quality control analyses for the PBMC compressed screen.

- a. Distributions showing viable cells across the 3 different plate conditions (no stimulation control, LPS stimulation, IFN β stimulation). Vertical dashed lines represent the mean of each group. Left: raw viability; right: normalized to DMSO.
- b. Correlation between demultiplexed scRNA-seq cell counts and the number of viable cells across all stimulation conditions. Left column: before quality control (QC) filtering for low quality cells; Right column: post-QC filtering. Error bar indicates 95% confidence interval. Significance for Pearson correlation was calculated using a two-sided correlation test implemented in SciPy, no multiple hypothesis correction was made.
- c. Heatmap of deconvolution results showing the effect of drug perturbations on cell viability. Highlighted are compounds with significant effects on cell viability across all stimulation conditions.
- d. Low-dimensional embedding (UMAP) of CS experiment where LPS was used as the stimulant.
- e. Integrated low dimensional embedding (UMAP) of scRNA-seq results of all 3 stimulation conditions.
- f. Dot plot of expression level of top 5 marker genes for each cell type from the Azimuth reference model across stimulation conditions.

Supplementary Figure 2: Perturbation effect coefficients from deconvolution on LPS+control data.

- a. Compound effect (β_{drugs}) of all 90 compounds from the mechanism of action (MOA) library in each PBMC cell type. Compounds in gray have no significant, non-zero effect on any of the cNMF GEPs for the given cell type.
- b. Interaction effect ($\beta_{\text{interactions}}$) between compounds from the MOA library and LPS in each PBMC cell type. Interactions in gray have no significant, non-zero effect on any of the cNMF GEPs for the given cell type.

Supplementary Figure 3: Perturbation effect coefficients from deconvolution on IFN β +control data.

- a. Compound effect (β_{drugs}) of all 90 compounds from the mechanism of action (MOA) library in each PBMC cell type. Compounds in gray have no significant, non-zero effect on any of the cNMF GEPs for the given cell type.

- b.** Interaction effect (β interactions) between compounds from the MOA library and IFN β in each PBMC cell type. Interactions in gray have no significant, non-zero effect on any of the cNMF GEPs for the given cell type.

Supplementary Methods

PDAC single-ligand screening

PDAC organoids were cultured and maintained in “OWRNA” complete organoid medium without WNT3A, RSPONDIN-1, NOGGIN, and A-8301 consisting of Advanced DMEM/F12 (Thermo Fisher), 10 mM HEPES (Thermo Fisher), 1x GlutaMAX (Thermo Fisher), 50 ng/mL mEGF (Peprotech), 100 ng/mL hFGF10 (Peprotech), 10 nM hGastrin I (Sigma), 1.25 mM N-ace-tylcysteine (Sigma), 10 mM Nicotinamide (Sigma), 1x B27 supplement (Thermo Fisher), 100 U/mL penicillin/streptomycin (Thermo Fisher), and 1x Primocin (Invivogen). Organoids were dissociated to single cells using 1X TrypLE Express (Thermo Fisher) for 30 minutes, counted and seeded in suspension (10% Matrigel, 90% media) at a density of 40,000 viable cells/well in a 96-well plate at a volume of 50uL/well. Ligands were added after 24 hours at 2x the desired final concentration in 50uL to bring the total volume per well to 100uL at 1x ligand concentration. Organoids were collected after 7 days for single-cell RNA sequencing. Treated organoids were tested in duplicate while control organoids (treated with 0.1% BSA in 1X PBS as vehicle control) were assessed in quadruplicate. Candidate ligands and their concentrations can be found in **Supplementary Table S1**.

PDAC scRNA-seq with cell hashing

PDAC organoids for compressed screens and validation experiments were dissociated to single cells, hashed with antibody-derived oligo tags, pooled and captured for scRNA-seq via Honeycomb Bio HIVEs. Briefly, 200 μ L of RP-10 (RPMI – ThermoFisher #12633-012 + 10% FBS - ThermoFisher #26140) was added to each well of PDAC organoids in the 96-well assay plate (300 μ L total volume) with vigorous mixing. The full volume was then transferred to the upper chamber of a 30-40 μ m 96 well filter plate (Pall #8027) situated on a vacuum manifold (Honeycomb Bio). Media was aspirated under vacuum, with organoids larger than the filter pore size retained in the upper chamber; an additional 300 μ L of RP-10 was added, mixed, and aspirated under vacuum to remove residual Matrigel and cell debris. The bottom of the filter plate was blotted dry and sealed with a plate seal (ThermoFisher #AB0558), and the upper chamber was resuspended with 100 μ L of pre-warmed TrypLE Express (ThermoFisher #12604013), mixed 10-15 times, and incubated for 20 minutes at 37°C with intermittent mixing. Digestion was then quenched with the addition of 200 μ L RP-10 + 50 μ g/mL DNase (Stem Cell Technologies), and the single cell suspension was collected by centrifugation (300g, 5 minutes) into a 96-well u-bottom ultra-low attachment plate

(Corning #7007). Quenching solution was aspirated and 50 μ L / well of cell staining solution (45uL cell staining buffer - BD #554656 + 5uL human multiplexing antibody – BD #633781; resulting in a 1:10 dilution) was added to each well and incubated at room temperature for 20 minutes. Human multiplexing antibodies were distributed column-wise with each column having a unique tag. Following staining, cells were washed in triplicate by adding 300 μ L cell staining buffer, pelleting by centrifugation (400g, 5 minutes), and aspirating the wash solution. Cells were then resuspended in a volume of cell staining buffer such that when samples are collapsed row-wise (containing non-overlapping tags) the total sample volume was 200 μ L. Each row was then treated as a sample for loading onto a Honeycomb HIVE.

Honeycomb HIVE scRNA-seq was performed as described in their October 2021 protocol (v21.10), with modifications to accommodate cell hashing. Briefly, each cell suspension was counted by hemocytometer and normalized to a concentration of 15,000 cells in 1 mL of RPMI. Cells were loaded into the HIVEs by centrifugation at 30g for 3 minutes and stored in cryopreservation solution at -20°C until library preparation (no longer than 2 weeks). HIVEs were thawed, sealed, lysed and hybridized following the standard protocol. Bead recovery was followed by first strand synthesis (reverse transcription), bead clean up (exonuclease treatment), and second strand synthesis following the standard protocol. Whole transcriptome amplification was modified with the addition of 2.5 μ L of 10 μ M BD-HC WTA linker primer per 400 μ L reaction preparation, and PCR was run following the standard protocol. Following PCR, 15 μ L of each WTA reaction per HIVE was pooled size selected with a 0.65X SPRI bead (Beckman Coulter #A63881) addition. The bead bound fraction (cDNA) was washed with a 0.65X SPRI concentrate (Honeycomb Bio), eluted, and used in index PCR per the standard protocol. The supernatant (cell hashing library) had additional SPRI beads added to bring the final ratio to 1.2X. Product was eluted from the beads, and run through a round of PCR (10 cycles of the standard WTA PCR, with 2.5 μ L each of 10 μ M BD-HC linker-UPS and 10 μ M Bead-specific primers per 50uL reaction preparation) to append the Linker-UPS sequence to the cell hashing library. This product was then used as input to the standard index PCR protocol. Following index PCR, cDNA product was cleaned via SPRI beads and quantified per the standard protocol. The cell hashing library was size selected with a reverse 0.65X SPRI selection, followed by a 1.2X SPRI pull down, and a second round of reverse 0.65X SPRI selection and 1.2X SPRI pull down, and quantified per the standard protocol. Quantified libraries were pooled and sequenced on either Illumina NextSeq 500 or NextSeq 2000 with HIVE custom read and custom index primers (Honeycomb Bio) with Read 1: 25 Read 2: 50 Index 1: 8 Index 2:8 bases. BD oligonucleotide sequence can be found in the **Key Resource Table** above.

Ruxolitinib validation experiment

PBMCs were isolated from whole blood as described above and were seeded at a density of 2×10^5 cells in 100 μ L/well in triplicate for each condition (DMSO only, DMSO + LPS, Ruxolitinib only, Ruxolitinib + LPS) in either a canonical v-bottom 96-well plate (Corning Evergreen #222-8031—1V) – to be processed for sequencing – or a white flat-bottomed 96-well plate (PerkinElmer #6005680) – to be processed for cell viability – pre-loaded with drugs to be screened using the Echo 555 Liquid Handler (Beckman Coulter, Serial Number E555-1115). 24 hours later, 100 μ L of media/DMSO or media/Ruxolitinib mix was to bring final concentration of 3 μ M Ruxolitinib in treated wells. Cells were then allowed to rest for 16 hours at 37°C /5% CO₂ before acute stimulation with 5 μ g/mL LPS (ThermoFisher #00-4976-93) or no-stimulant control for 4 hours. 300,000 cells are sufficient to test viability, however, to ensure we obtained an adequate number of cells for sequencing, we pooled the triplicate wells of each condition giving us a total of 600,000 cells/condition.

Cell viability was assessed using Cell Titer Glo as described in the compressed PBMC screen.

PBMC scRNA-seq with cell hashing

For the compressed screen, duplicate plates were first collapsed by discarding the media and washing with 100 μ L fresh media followed by collapse of each identical pair of wells to obtain around 600,000 cells per well. Next, all wells were processed for antibody tagging using BioLegend's TotalSeq-B with 10x Feature Barcoding Technology protocol. 8 antibodies (BioLegend 394631, 394633, 394635, 394637, 394639, 394641, 394643, 394645; at a 1:50 dilution) were used, hence combining 8 conditions into a single pool collapsing our 144 conditions to 18 pools. Before collapse, cell number in each well was determined and this information was used to make pools comprised of 3,750 cells from each condition for a total of 30,000 cells/pool to be loaded onto the 10X Chromium Controller and the protocol was followed according to the manual for Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Dual Index) with Feature Barcode technology for Cell Surface Protein (#PN-1000268, # PN-1000269, # PN-1000262, # PN-1000215, # PN-1000242, # PN-1000120).

Samples for the Ruxolitinib validation were treated similarly: triplicate cells were pooled to obtain 600,000 cells condition. Next, all wells were processed for antibody tagging using BioLegend's TotalSeq-B with 10x Feature Barcoding Technology protocol. 4 antibodies (BioLegend, 394631, 394635, 394637, 394641) were used, combining all 4 conditions into a single pool. Before collapse, cell number in each well was determined and this information was used to make a pool comprised of equivolumes from each condition for a total of 30,000 cells/pool to be loaded onto the 10X Chromium Controller and the protocol was followed according to the manual for Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Dual Index)

with Feature Barcode technology for Cell Surface Protein (#PN-1000268, # PN-1000269, # PN-1000262, # PN-1000215, # PN-1000242, # PN-1000120).

Cell painting data preprocessing

We used the Distributed Cell Profiler software on AWS to process the Cell Painting images (<https://github.com/CellProfiler/Distributed-CellProfiler>). Our pipeline first applied an illumination correction to each image based on per-channel measurements and applied intensity correction on each channel between batches based on differences in the batch-averaged median channel intensity. We next adopted an existing approach for using the isolation forest outlier detection algorithm implemented in sklearn to remove outlier images confounded by artifacts³⁰ based on the whole-image measurement of 97 features. This was performed batch-wise and excluded 4.6% of ground truth batch 1 and 2, 2.0% of compressed batch 1, 2.2% of compressed batch 2, and 1.7% of compressed batch 3 images. We next segmented individual cells, nuclei, and cytoplasm, and measured a total of 3,402 morphological features across all five imaging channels and compartments. Each feature was then median aggregated over all cells within a given well, constituting a single sample. We then removed empty wells and wells with fewer than 50 segmented cells and scaled each feature by plate using the RobustScaler function in sklearn centered on DMSO wells and scaled by all wells within each plate. Finally, we selected robust and variable features within a dataset (either dose-time analysis or our ground truth and compressed screen analysis) by retaining features with low noise and high signal. We accomplished this by first calculating the median absolute deviation (MAD) for each feature within a plate. Low-noise features were all features with a MAD greater than the 10th percentile across greater than 90% of the plates in a dataset – features which consistently showed at least some variability in most plates. Within the dose-time analysis, 1,859 features passed the low-noise criteria, while in the ground truth and compressed screen analysis 2,117 features passed. High-signal features were all features with a MAD greater than the 90th percentile across greater than 10% of the plates in a dataset – features which showed the most variability in at least some of the plates. Within the dose-time analysis, 1,254 features passed the high-signal criteria, while in the ground truth and compressed screen analysis, 1,054 features passed. The final feature set was then taken as the intersection of low-noise and high-signal feature sets. For the dose-time analysis, this resulted in 861 final features; in the ground truth and compressed screen analyses, it was 886 final features. We provide the metadata for these features in **Supplementary Table S2**.

Identifying optimum doses and timepoints for cell painting screens

To determine the dose and timepoint where the compounds in our library had the most activity, we screened all compounds in our library in duplicate at nine dose-timepoint pairwise combinations (0.1 μ M, 1 μ M, 10

μM; 6 hours, 24 hours, 48 hours). We then quantified the effect of each perturbation by calculating the Mahalanobis distance D_p between the mean vector for each perturbation and the negative control (DMSO) samples at the same dose and timepoint. Mahalanobis distance is a multidimensional generalization of a z-score and quantifies how many standard deviations a vector $\vec{x}$ is away from a distribution with mean $\vec{u}$ and covariance S :

$$D_p(\vec{x}) = \sqrt{(\vec{x} - \vec{u})^T S^{-1} (\vec{x} - \vec{u})}.$$

We then calculated the coefficient of variation of the Mahalanobis distance values at each dose/timepoint in order to determine the dose/timepoint combination with the broadest range of effects. This occurred at a dose of 1μM and a 24-hour timepoint. We then screened 4 more replicates of each drug at this dose and timepoint to create a final ground truth dataset consisting of 6 replicates per drug, which we used to recalculate the Mahalanobis distance between each drug and the negative control distribution. We refer to this dataset as our ground truth (GT).

Evaluation of cell viability in cell painting ground truth screen

During the staining procedure for the Cell Painting assay, the vast majority of dead cells are washed away; thus, we used the count of nuclei per well as a proxy for understanding the impact of the compounds on viability. In our pilot experiments to decide the optimal dosage for the screen, we observed a similar distribution of cell counts for our drug and control wells across doses (**Extended Data Fig. 1d**). For the compressed setting, meanwhile, we did not observe a shift in the cell count distribution across different compressions, although there we did detect more cells in some of our GT wells (Mann-Whitney U test p-value<0.05; **Extended Data Fig. 2b**).

Compressed screens scRNA-seq alignment, hash demultiplexing, and initial quality control

For the PDAC screen sequencing runs, raw sequencing reads were converted from bcl files to FASTQs using bcl2fastq (v2.2) based on Honeycomb Bio indices that corresponded to individual samples. Demultiplexed FASTQs were then aligned to the human GRCh38 genome using STARsolo as implemented by Cumulus (v2.0.0) on Terra.bio. We implemented a 2-pass alignment to first generate a pseudo-whitelist (15,000 cell barcodes with the greatest number of unique molecular identifiers - UMIs) and then realign with this pseudo-whitelist for cell barcode collapse. Additionally, in alignment we specified the following parameters: Multi-mappers: EM, UMI Dedup: 1MM_CR, Solo Features: FullGene, UMI Filtering: MultiGeneUMI_CR, Cell Filter: CellRanger2.2 3000 0.99 10. To align the cell hashing libraries, we provided as input to citeseqcount v1.4.3, implemented via a Terra.bio pipeline, the demultiplexed FASTQs, the pseudo-whitelist from the first pass STARsolo alignment and the hash barcodes per experiment.

For the PBMC screen samples processed with 10X Genomics, raw sequencing reads of both cell hashing and GEX libraries from Illumina NextSeq2000 was converted to FASTQs using *cellranger mkfastq* as implemented by Cumulus (v6.1.2) on Terra.bio and those sequenced on NovaSeqX were converted to FASTQs using *bcl2fastq* (v2.2), following 10X Genomics dual indices that correspond to individual samples. Demultiplexed FASTQs were then aligned to the human GRCh38 genome using *cellranger count* by the same Cumulus implementation with default parameters. To align the cell hashing libraries, we provided as input the 8 TotalSeq™-B Hashtag barcodes from BioLegend for each pooled sample.

Following alignment of sample transcriptomes and cell hashing libraries, we performed initial QC filtering and sample demultiplexing via cell hashing. All samples were filtered to retain only cell barcodes with greater than 500 unique genes and 500 UMIs, and less than 75,000 UMIs and 50% mitochondrial transcripts. Cell barcodes were further filtered by hashing libraries such that only barcodes with greater than 10 hash UMIs, and a 1.1 signal-to-noise ratio (maximum ADT UMI count divided by the second highest ADT UMI count) were retained. Next, samples were demultiplexed and labeled by original sample identity by iterating across a range of quantile parameters (0.85 to 1 - (1x10⁻⁶)) in the Seurat HTODemux function (v4) and choosing the quantile parameter maximizing the number of singlets. Retaining only labeled singlets, datasets were filtered to retain genes detected in at least 10 cells, and used for subsequent analysis.

PDAC single-ligand validation screen scRNA-seq analysis

We adopted the same approach as above to run cNMF, process scRNA-seq data, and identify and annotate highly variable GEPs and annotate in the single-ligand data. We then correlated the gene spectra scores in these GEPs to those from the compressed data and focused our analysis on the GEPs that highly correlated with the compressed data. As we observed a large difference in the mean Moffitt Classical score in negative control samples from the three different batches of samples in the single-ligand dataset (indicative of batch effects), we assessed the effect of ligands on each GEP by running a linear regression with batch included as a covariate in order to identify ligand effects independent of batch.

Cell viability calculation for PBMC compressed screens

Cell viability from drug and stimulation treatments were read out through CellTiter-Glo (CTG) viability assay. For a 96-well plate, we took the mean of each row *i*'s empty wells (columns 7-12) as the background measurement ($V_{empty}(i)$) for each row's readout and subtract this value from each non-empty well:

$$V_{adj}(i,j) = V(i,j) - V_{empty}(i)$$

In each treatment plate, we then calculated the mean viability of DMSO-only wells after subtracting the background signal ($\underline{V_{adj,DMSO}}$). All non-empty wells were normalized against that value:

$$V_{norm}(i, j) = V_{adj}(i, j) / \underline{V_{adj, DMSO}}$$

These values were then compared against each other in mean pool values for each compound and run through linear deconvolution to check significant effects by compounds.

PBMC drug effect classification and measurement

To evaluate the effect of each drug on a stimulant, we classified a drug into either a (positive/negative) potentiator, a (positive/negative) interferer, or a non-interactor for p-value cutoffs from 0.001 to 0.05. After filtering by p-value cutoff on the regression coefficient table, we identified these categories as follows:

- Positive potentiator: for a module A that the stimulant *upregulates*, if drug X has a positive interaction term with the stimulant on module A and drug X has a non-negative regression coefficient on module A , then X is deemed a positive potentiator of the stimulant on module A .
- Negative potentiator: for a module A that the stimulant *downregulates*, if drug X has a negative interaction term with the stimulant on module A and drug X has non-positive regression coefficient on module A , then X is deemed a negative potentiator of the stimulant on module A .
- Positive interferer: for a module A that the stimulant *upregulates*, if drug X has a negative interaction term with stimulant on module A and drug X has a non-positive regression coefficient on module A , then X is deemed a positive interferer of the stimulant on module A .
- Negative interferer: for a module A that the stimulant *downregulates*, if drug X has a positive interaction term with stimulant on module A and drug X has non-negative regression coefficient on module A , then X is deemed a negative interferer of the stimulant on module A .
- Non-interactor: not a potentiator or interferer of any stimulant effect module in the cell type.

The strength of a drug's potentiating effect is calculated as follows:

$$I(is\ potentiator?) * abs(\beta_{drugs} + \beta_{stimulants} + \beta_{interactions}) / abs(\beta_{stimulants}),$$

and the strength of a drug's interfering effect is calculated as follows:

$$I(is\ interferer?) * abs(\beta_{drugs} + \beta_{interactions}) / abs(\beta_{stimulants})$$

where I is the indicator function and β is the coefficient from deconvolution after filtering by permutation test p-value cutoff. For the heatmap visualization in **Fig. 4h**, the strength of negative potentiators or interferers are multiplied by -1 for ease of visualization. For visualization, we used seaborn function

clustermmap with default parameters to run hierarchical clustering on the effect strength of all drugs. Similar analysis was run for **Extended Data Figure 6** without clustering on all cNMF GEPs.

For demonstrating the effect of stimulants and drugs in the compressed screen (e.g., **Fig. 4f-g**), we took the normalized cell usage score output from cNMF runs of each program. For **Fig. 4g, Extended Data Fig. 7b, 9b,&10d**, we normalized the usage scores for each GEP against the median of the DMSO+IFN β /LPS condition in each comparison group for better illustration of the directional changes. For evaluating the compressed screen module in our single-compound validation experiment (**Fig. 4j**), we took the selected GEP from the compressed screen and found the top genes in the module by running KneeLocator in python on the sorted gene spectra scores to identify the top genes contributing to each GEP. We then used `score_genes` function in scanpy to calculate a module score for the selected module on the log-normalization validation data. Mann-Whitney U test was used to ascertain statistical significance between usage score differences.

To interpret GEPs in each cell type, we used a similar approach to what was used in the PDAC screen. We ran decoupler and the “`score_genes`” function in scanpy to generate module scores for PROGENy and MsigDB genesets and for cell cycle and scRNA-seq quality gene signatures, then correlated all of these module scores with the GEP usages across all cells.