Cancer-Associated Fibroblasts Regulate Patient-Derived Organoid Drug Responses

María Ramos Zapatero[1], Alexander Tong[2,3], Jahangir Sufi[1], Petra Vlckova[1], Ferran Cardoso Rodriguez[1], Callum Nattress[1], Xiao Qin[1], Daniel Hochhauser[4], Smita Krishnaswamy[2]*, Christopher J. Tape[1]*

[1]Cell Communication Lab, Department of Oncology, University College London Cancer Institute, London, UK.
[2]Department of Computer Science, Yale University, New Haven, CT, USA.
[3]Department of Computer Science and Operations Research, Université de Montréal, Mila — The Quebec AI Institute, Montreal, Canada.
[4]Drug-DNA Interactions Group, Department of Oncology, University College London Cancer Institute, London, UK.
*Correspondence: c.tape@ucl.ac.uk, smita.krishnaswamy@yale.edu

Abstract

Patient-derived organoids (PDOs) model personalized cancer therapy responses. However, existing bulk PDO screening technologies cannot reveal drug response mechanisms or model how cells of the tumor microenvironment alter therapy performance. To address this, we developed a highly-multiplexed thiol-reactive organoid barcoding *in situ* (TOB*is*) mass cytometry platform to perform single-cell post translational modification (PTM) signaling analysis of colorectal cancer (CRC) PDOs and cancer-associated fibroblasts (CAFs) in response to clinical therapies. To compare patient- and microenvironment-specific treatment effects in thousands of single-cell PTM datasets, we developed *Trellis* — a highly-scalable, hierarchical tree-based treatment effect analysis method. Trellis analysis of >2,500 single-cell PTM PDO-CAF organoid cultures revealed that on-target cell-cycle blockage and DNA-damage drug effects are common, even in chemorefractory PDOs. However, drug-induced apoptosis is patient-specific. We found drug-induced apoptosis does not correlate with genotype or clinical staging, but does align with cell-intrinsic PTM signaling in PDOs. We observe that CAFs protect chemosensitive PDOs by shifting cancer cells into a slow-cycling cell-state in a patient-specific manner and show that CAF chemoprotection can be reversed by inhibiting YAP. These results reveal that PTM signaling flux is a major determinant of chemosensitivity and demonstrate CAFs regulate patient-specific drug responses by altering cancer cell-state.

1 Introduction

Tumors are heterogeneous cellular systems comprising cancer cells, stromal fibroblasts, and various immune cells. Tumor phenotypes are regulated by cell-intrinsic mutations within cancer cells and cell-extrinsic cues from the tumor microenvironment (TME) [1]. Colorectal cancer (CRC) kills >0.9 million

people per year worldwide [2] and is characterized by a high inter-patient genetic heterogeneity and patient-specific responses to therapy [3]. Cancer associated fibroblasts (CAFs) are one of the most abundant cell-types in the CRC TME [4]. CAF abundance correlates with poor overall survival [5], and influences response to both targeted therapies [6] and radiotherapy [7]. However, there is a lack of understanding on how CAFs regulate cancer cell response to therapy and to what extent stromal regulation is patient-specific.

Patient-derived organoids (PDOs) are personalized cancer models [8] that can mimic their parent tumors' response to chemotherapies [9] — with several studies proposing PDOs as personalized avatars of drug response [10]. However, epithelial PDO monocultures do not recapitulate the heterocellular TME that therapies must navigate *in vivo* and therefore cannot model the influence of stromal cells on therapy response. PDOs can be co-cultured with stromal and immune cells to recapitulate elements of the TME [11], but how this alters PDO phenotypes and personalized drug response is unknown. Moreover, PDO drug sensitivity is typically measured using bulk live/dead viability assays [12] that can not resolve cell-type-specific data from TME co-cultures and provide no mechanistic insight into drug responses [13].

To overcome these limitations, we developed a highly-multiplexed Thiol-reactive Organoid Barcoding *in situ* (TOB*is*) mass cytometry [14, 15] platform to study how anti-cancer therapies regulate the cell-state, DNA-damage response, and post-translational modification (PTM) signaling of CRC PDOs in the presence or absence of CAFs at single-cell resolution across >2,500 PDO-CAF cultures. To compare single-cell drug responses from thousands of cell-type-specific datasets, we developed *Trellis*, a hierarchical tree-based treatment effect analysis method that derives generalized optimal transport distances between samples after normalizing by their own controls. TOB*is* mass cytometry and Trellis revealed PDO-specific drug response PTM signaling and demonstrated CAFs protect CRC cells from chemotherapy by shifting epithelial cell-state. CAF chemoprotection could be rationally reversed using insights from single-cell PTM data, demonstrating the utility of PTM-focused drug screening for overcoming therapy resistance. These results illustrate the functional intertumoral heterogeneity of patient-specific drug response mechanisms and suggest TME cells should be included in future PDO models.

2 Results

Patient- and Microenvironment-Specific Single-Cell PTM PDO-CAF Drug Analysis

To study how CAFs regulate patient-specific drug response signaling, we established a high-throughput 3D organoid co-culture system comprising 10 CRC PDOs [12] (Table S1) cultured either alone or with CRC CAFs [16, 17]. Organoid cultures were treated in triplicate with either vehicle control, or titrated combinations of clinical therapies fluoropyrimidine 5-fluorouracil (5-FU), SN-38 (active metabolite of Irinotecan), Oxaliplatin, and Cetuximab (EGFR inhibitor). LGK974 (PORCN inhibitor) and Berzosertib (VX-970) (ATR inhibitor) were also studied as novel CRC therapies (Figure 1a) (Table S2). Each culture was fixed, stained with thiol-reactive monoisotopic TOB*is* barcodes *in situ* [15], pooled, dissociated into single-cells, stained with a panel of 44 rare-earth metal antibodies (spanning cell-type, cell-state, DNA-damage, and PTM signaling markers (Table S3)), and analyzed by mass cytometry (Figure 1b). Following multiplexed debarcoding [18] and cell-type-specific gating, we obtained >10 million cancer cells and >15 million CAFs from 2,520 3D cultures (3,360 cell-type-specific single-cell PTM signaling datasets).

Trellis: Hierarchical Tree-Based Single-Cell Treatment Effect Analysis

Highly-multiplexed single-cell cytometry screening data presents several major challenges. First, prior work on large single-cell data is designed with the manifold structure of transcriptomic technologies, where

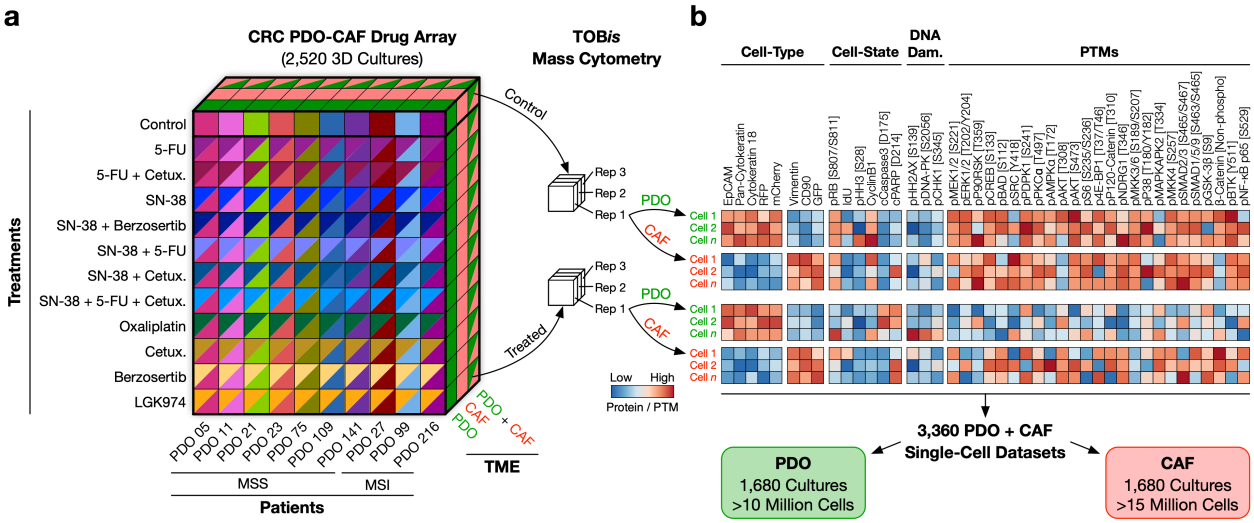

Figure 1: TOB*is* MC Single-Cell PTM PDO-CAF Drug Responses. a) Multidimensional array of 10 CRC PDO (7 microsatellite stable (MSS), 3 microsatellite instable (MSI)) treated with 11 titrated drug combinations either alone or in co-culture with CRC CAFs in triplicate (2,520 3D cultures). **b)** PDO-CAFs were barcoded *in situ* using TOB*is*, stained with 44 rare-earth metal antibodies spanning cell-type identification, cell-state, DNA-damage response, and PTM signaling, and analyzed by MC (3,360 single-cell PTM datasets).

cell distances are locally Euclidean [19–23]. In cytometry data, antibody panels are designed based on prior biological knowledge, and analyzed using gating strategies that follow a hierarchical structure, which are better described by tree distances rather than a single smooth manifold. Second, our PDO-CAF PTM screening data contains >2,250 conditions with >25 million cells. Existing state of the art is to compare cluster proportions between single-cell samples [24–26]. Emerging methods can compare distributions using earth mover's distance (EMD), but only at course granularity [21], or by using graph diffusion which does not account for the hierarchical tree structure of cytometry data [22]. As highly-multiplexed cytometry datasets are becoming increasingly common, there is a need for tools that can efficiently compare thousands of single-cell cytometry conditions. Finally, large scale screening datasets compare independent systems (e.g. patients, microenvironments, and/or technical batches) perturbed by constant treatments. For this, internal controls need to be leveraged, such that multiple controls and treatments can be directly compared in a common computational space. To solve these problems, we developed Trellis.

Comparisons between single-cell datasets typically treat all markers equally — irrespective of prior biological knowledge. While equal weighting may be appropriate for unbiased single-cell methods such as scRNA-Seq, Trellis leverages the experimental design of cytometry data using a 'branch' tree hierarchy of well-defined biological processes (e.g. cell-type hierarchies or cell states) that supervenes upon randomized 'leaves' of latent biological significance (via four levels of four k-means clusters) (Figure 2a). This enables an automated assessment of cell populations that mimics human intuition in the design of the experiment, and subsequently its interpretation (Figure S1a). Trellis can leverage any gating strategy that returns a single hierarchy or multiple hierarchies (Algorithm 1 line 3).

Once one or multiple hierarchies are defined, Trellis then embeds each single-cell distribution into a vector such that for two distributions, the L^1 distance between embeddings is equivalent to the EMD between the two distributions along the defined tree or forest (Algorithm 1 lines 4-5). By reducing a complicated

Algorithm 1: High-level Trellis algorithm for comparing single-cell treatment effects

1: **Input:** Dataset containing controls and variables
2: **Output:** Distances between treatment effects to their relative controls
3: Build tree or trees incorporating prior knowledge on markers followed by random construction with edge weights w for each node.
4: Embed each distribution to a vector with each element as fraction of dataset in that node forming abundance matrix $\boldsymbol{A}$.
5: Multiply elementwise $w \otimes \boldsymbol{A}$ to calculate Trellis embeddings $\boldsymbol{E}$
6: (optionally) subtract relevant control vectors for paired Trellis embeddings $\tilde{\boldsymbol{E}}$
7: **return** Relevant L^1 distances between embeddings.

and inefficient distance calculation to a vector distance, Trellis can scale to larger datasets by leveraging existing work in high-dimensional distance computation (Figure S2b). For instance, if we only need to find the nearest neighbor treatments for non-linear embedding [20, 27, 28], we leverage fast nearest neighbor algorithms such as KD-trees as used in PHATE [20], Annoy [29] used in UMAP [27] and Scanpy [30], or locally sensitive hashing [31, 32].

As single-cell screens increase in size and complexity, the use of internal controls enables the comparison of independent variables in parallel. Existing distribution comparison methods cannot easily incorporate this, indeed EMD is not even defined for the difference of distributions. To solve this, Trellis can easily be extended to 'paired' Trellis (Algorithm 1 line 6), where paired controls are subtracted from treatment samples to directly compare treatment effects. We prove this is equivalent to a Kantorovich-Rubenstein (KR) norm with tree ground distance (Prop. 2). This KR norm cannot be computed with standard Wasserstein distance methods (even for small problems [33, 34]) but can be calculated by Trellis. Paired Trellis therefore enables thousands of variables to be compared to their internal controls in a common computational space — enabling clear distinction of individual treatment effects in paralleled high-dimensional single-cell screening data (Figure S1a).

In summary, Trellis uses a prior-driven tree domain to compute the generalized Wasserstein distance between thousands of single-cell samples. Pairing treatments to controls enables paralleled visualization of treatment effects (Figure S1a) and reduces batch effects in serially acquired datasets (Figure S1b). Prior-driven branches help resolve biological treatment effects (Figure S1c) and the tree domain structure enables thousands of single-cell datasets to be analyzed (Figure S1d) (Figure S2a, b). Further detail on Trellis' scalability, theoretical soundness, and robustness can be found in Methods.

Trellis-Based Single-Cell Analysis of PDO Cell-State and PTM Signaling

Anti-cancer drugs typically induce major shifts in cell-cycle and apoptosis that can be detected by mass cytometry. For example, SN-38 inhibits topoisomerase 1 [35], resulting in S-phase blockage, cell-cycle exit, and induction of apoptosis (Figure 2b). Similarly, 5-FU blocks nucleotide biosynthesis by inhibiting thymidylate synthase [36] which subsequently stalls S-phase entry, whereas oxaliplatin induces ribosome biogenesis stress to block mitotic progression [37]. Capturing shifts in cell-state is therefore crucial for understanding on-target drug responses.

In mass cytometry, cell-state is identified by hierarchical gating of pRB, IdU, pHH3, Cyclin B1, and cPARP/cCaspase3 [38, 39] and is therefore well suited for Trellis. For cell-type-specific analysis of PDO-CAF co-cultures we designed a Trellis hierarchy using cell-state-driven branches that supervene upon ran-

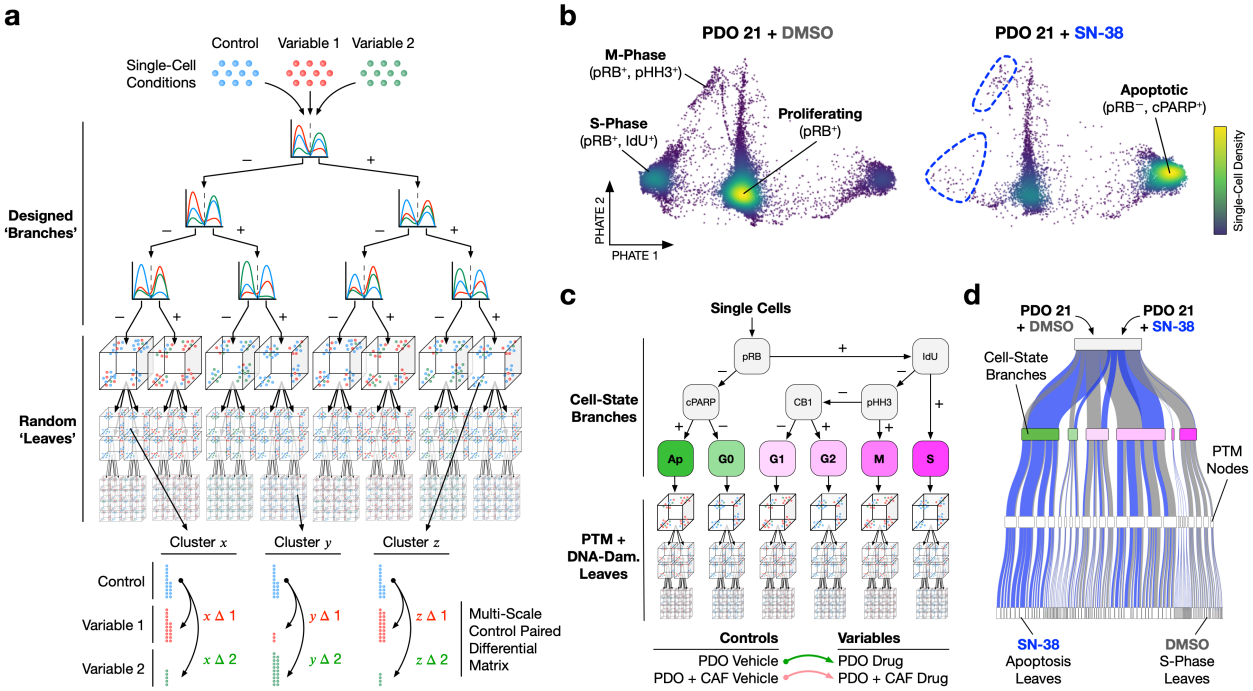

Figure 2: Trellis. a) Single cells from control and variable conditions are distributed through a tree comprising designed binary 'branches' that supervene upon randomized k-means clustering nodes, ending in 'leaves'. Branches weigh hierarchical gating strategies while nodes and leaves leverage latent parameters. In each node of the tree, variables are subtracted from paired controls to create a multi-scaled differential matrix (based on a Kantorovich-Rubinstein norm) that scales to thousands of conditions. **b)** Single-cell density PHATEs of PDO 21 treated with DMSO or SN-38 (irinotecan). SN-38 results in cell-cycle exit (IdU^{-}, pHH3^{-}, and pRB^{-}) and induction of apoptosis (cPARP^{+}). **c)** Trellis hierarchy for single-cell PDO on-target drug responses leveraging cell-state branches and randomized PTM and DNA-damage parameters. Trellis scores are calculated per PDO by comparing untreated controls to drugs for both mono-cultures and co-cultures. CB1, Cyclin B1. **d)** Sankey diagram showing data from b) distributing through the Trellis layout in c) (terminal leaves not shown).

domized DNA-damage and PTM signaling leaves (Figure 2c) (Figure S3a). This topology sensitizes Trellis scores to canonical on-target shifts in cell-cycle and apoptosis that occur following drug treatments while also leveraging latent changes in DNA-damage and PTM signaling (Figure 2d) (Figure S3b-e).

TOB*is* Mass Cytometry Trellis Analysis of Cell-Type-Specific PDO-CAF Drug Responses

We used Trellis to analyze 3,360 (1,680 PDO, 1,680 CAF) single-cell PTM profiles (Figure 3a) in order to explore drug-, patient-, and microenvironment-specific therapy responses for both PDOs (Figure 3b-d) and CAFs (Figure S4). Since Trellis performs pairwise normalization to internal controls, all controls group on the left side of the graph (Figure 3b) (Figure S1a), and each treatment embeds relative to their controls, depending on their distribution through the Trellis tree. This enables therapeutic effects to be visualized across PHATE 1, and mechanistic response in PHATE 2 (Figure S5). Trellis revealed PDO treatment profiles are dominated not by drug type, but by patient-specific PTM signaling responses.

We observed four grouped responses to 5-FU, SN-38, and oxaliplatin chemotherapies: 1) broadly chemosensitive with high apoptosis (patients 21 and 75), 2) broadly chemosensitive with apoptosis and a strong DNA-damage response (patients 23 and 27), 3) anecdotally chemosensitive (i.e. only apoptotic with a

specific drug) (patients 99 and 109), and 4) chemorefractory with minimal apoptosis and low DNA-damage response (patients 05, 11, 141, and 216). Cetuximab, Berzosertib, and LGK974 generally had modest effects on PDO cell-state and PTMs relative to chemotherapies (Figure 3b) (Figure S5). While PDOs demonstrate clear patient- and microenvironment-specific drug responses, CAF signaling does not cluster by patient or drug (Figure S4), suggesting chemotherapies mainly alter the cell-state, DNA-damage, and PTM profiles of PDOs, not CAFs. Trellis further revealed CAFs protect some PDOs from chemotherapies (Figure 3d).

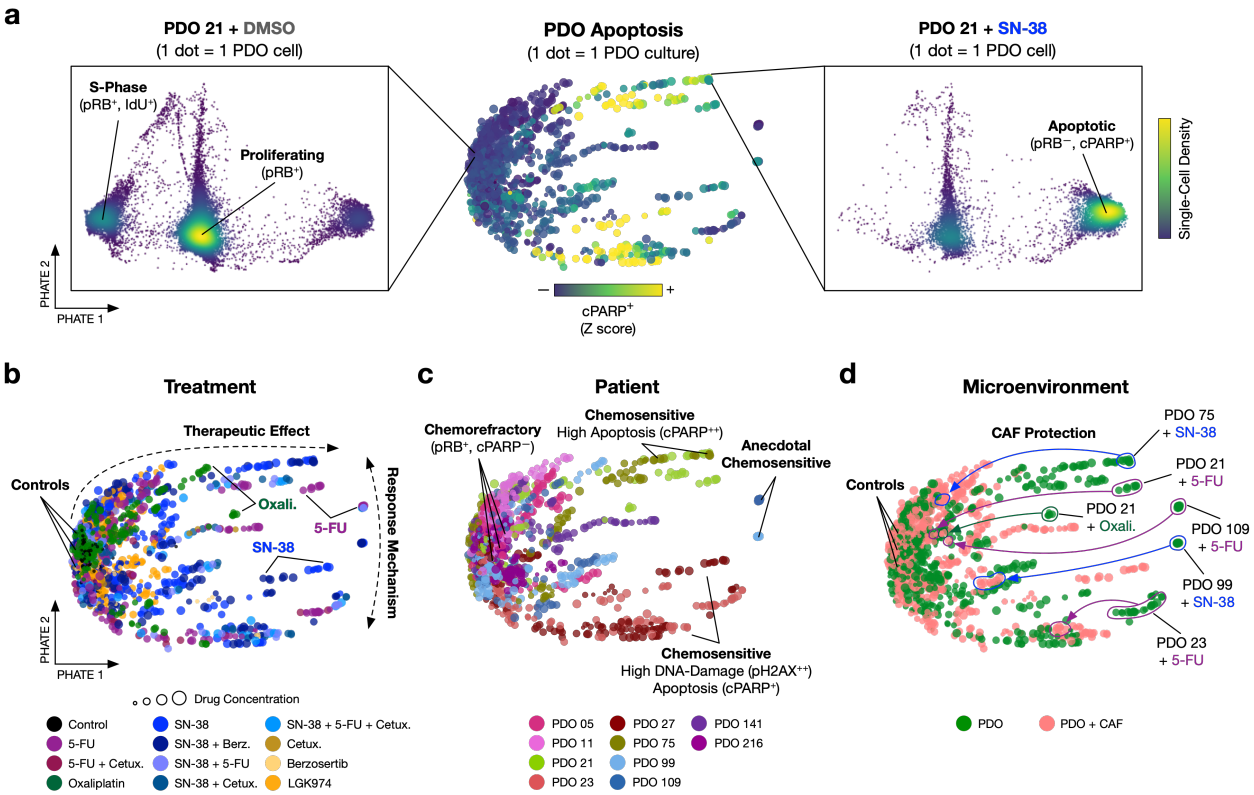

Figure 3: Trellis Analysis of Single-Cell PDO-CAF Drug Responses. a) Trellis-PHATE of 1,680 PDO single-cell PTM profiles (1 dot = 1 organoid culture comprising >5,000 single-cells) colored by apoptosis with representative single-cell density embeddings of PDO 21 + DMSO or SN-38. **b)** Drug treatment-specific responses. Controls group on the left, with treatment effects spreading across PHATE 1 and response mechanisms resolving across PHATE 2. **c)** Patient-specific drug responses illustrate different chemosensitive mechanisms and chemorefractory patients. **d)** CAFs provide patient-specific chemoprotection from 5-FU, SN-38, and oxaliplatin.

PDO Drug Signaling Responses Are Patient-Specific

PDOs have been proposed as personalized avatars of drug response [10], but how clinical treatments mechanistically alter patient-specific PDO biology is not well understood. To explore patient-specific drug response signaling, we altered the designed portion of the Trellis tree by combining cell-state parameters with a pHH2AX [S139] detection layer to enrich on-target DNA double-strand breaks and analyzed each patient drug response separately (Figure S6a-d). Patients continue to display either broad (PDOs 21, 23, 27, and 75) or anecdotal (PDOs 99 and 109) chemotherapeutic sensitivity, and multiple examples of drug

insensitivity (Figure 4a).

Unlike univariate live/dead metrics used in traditional drug screens, TOB*is* mass cytometry can detect on-target treatment effects that do not result in cell death. For example, SN-38 induces on-target S-phase blockage and double-strand breaks in both PDO 21 and PDO 05, yet only PDO 21 translates genotoxic stress into apoptosis (Figure 4b). Similarly, in PDOs 23 and 99, 5-FU and SN-38 result in a large DNA-damage response and stalled mitosis respectively, but no apoptosis (Figure S6e). 5-FU and SN-38 can clearly induce double-strand breaks and cell-cycle arrest in these PDOs, but they do not translate genotoxic replication stress into cell death. In fact, across nearly all PDOs tested, SN-38 (Figure S6f), oxaliplatin (Figure S6g), and 5-FU (Figure S6h) display on-target mitotic arrest (83%), but only a subset of patient and treatment combinations trigger apoptosis (40%). This suggests on-target drug responses are common in CRC PDOs, but often insufficient to induce cell death.

The specific chemosensitivity demonstrated by several PDOs reinforces the notion that PDOs could be used to identify drugs uniquely potent to an individual's cancer. For example, in PDO 99, 5-FU blocks mitosis and SN-38 causes a large DNA-damage response – yet neither chemotherapy induces substantial apoptosis. However, when treated with oxaliplatin, PDO 99 cells exit the cell-cycle and enter apoptosis (Figure S6e). Unlike 5-FU and SN-38, oxaliplatin does not kill cells directly through blocking S-phase, but via inducing ribosome biogenesis stress [37]. PDO 99 appears refractory to cytostatic stress but hypersensitive to ribosome biogenesis stress. Similarly, unlike all other patients, PDO 27 is highly sensitive to the ATR inhibitor Berzosertib. ATR inhibitors block single-stranded DNA-damage response and typically synergize with DNA-damage inducing drugs [40]. PDO 27 may experience high baseline DNA-damage as a result of its microsatellite instability (MSI). However, alternative MSI PDOs 99 and 216 are insensitive to Berzosertib, suggesting PDO genotype alone is insufficient to predict treatment response.

Chemosensitive PDOs Have Distinct Cell-Intrinsic PTM Signaling Profiles

We next sought to understand features common to chemosensitive and chemorefractory PDOs. Therapeutic response does not correlate with MSI/MSS status, clinical staging, anatomical location, *KRAS*, or *APC* genotypes (Figure S7a) (Table S1). However, baseline PDO cell-state and PTM signaling profiles are patient-specific and correlate with chemosensitivity (Figure 4c) (Figure S7b-d). Chemosensitive PDOs 21, 23, 27, and 75 are highly proliferative at baseline and experience canonical S-phase blockage, increased DNA-damage, and apoptosis when treated with both 5-FU and SN-38. In contrast, chemorefractory PDOs generally have lower cell-intrinsic mitotic PTM signaling activity than chemosensitive PDOs (Figure 4c). When treated with 5-FU, SN-38, and oxaliplatin, chemorefractory PDOs experience a reduction in S-phase and blocked M-phase consistent with on-target drug responses, but generally elicit a lower double-strand break response compared to chemosensitive patients and do not activate PARP or Caspase3 (Figure 4d) (Figure S6e). This suggests that even chemorefractory PDOs experience on-target drug responses, but their slow mitotic signaling flux at point of treatment means drug-induced cytostatic stress is insufficient to trigger widespread DNA-damage and apoptosis. In summary, TOB*is* mass cytometry and Trellis reveal on-target drug performance is common in CRC PDOs (even in chemorefractory PDOs) but cytotoxicity is patient-specific and correlates with cell-intrinsic PDO PTM signaling.

CAFs Chemoprotect PDOs by Altering PDO Cell-State

CAFs have both pro- and anti-cancer roles across a variety of solid tumors [4]. Trellis analysis of all PDOs revealed CRC CAFs can chemoprotect some chemosensitive PDOs (Figure 3d). To functionally explore the role of CAFs in patient-specific CRC PDO drug response, we performed paralleled analysis

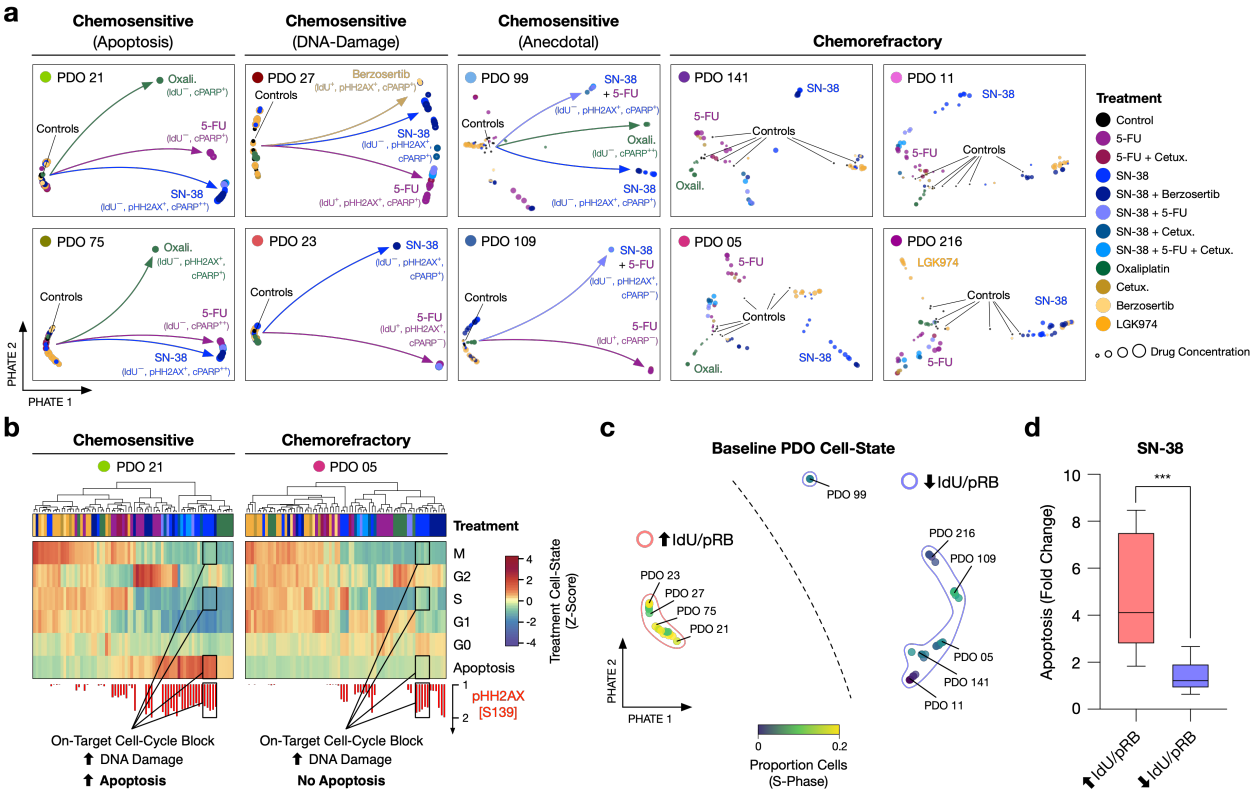

Figure 4: PDO Drug Response Mechanisms Are Patient-Specific and Align With Baseline Cell-State. **a)** Trellis-PHATE patient-specific PDO drug responses (840 single-cell PTM datasets). **b)** Patient-specific distribution of cells within Trellis branches reveals on-target cell-state shifts upon drug treatments. Treatment cell-state quantifies the fold change of the proportion of cells/cell state over the controls for each treatment (Z-score). DNA damage is quantified by the fold change of the proportion pHH2AX^{+} cells over the controls. **c)** Trellis-PHATE baseline PDO cell-state and PTM signaling resolves high IdU/pRB (red outline) and low IdU/pRB (blue outline) PDO groups (colored by proportion of cells in S-phase). **d)** SN-38-induced apoptosis in low IdU/pRB and high IdU/pRB PDOs. Unpaired *t*-test, *** <0.001.

of PDO monocultures and PDO-CAF co-cultures (Figure 5a). Trellis revealed CAFs provide varying degrees of chemoprotection in a patient- and drug-specific manner. For example, CAFs completely protect chemosensitive PDOs 21 and 75 from SN-38, 5-FU, and oxaliplatin-induced apoptosis, whereas PDOs 23, 27, and 99 only experience partial chemoprotection (Figure S8). Chemorefractory PDOs 05, 11, and 141 are largely unaffected by CAFs. This dichotomy suggests CAFs deregulate cancer cells in a patient-specific manner.

We next sought to understand why CAFs have such different patient-specific regulation of PDO drug response. Chemosensitive PDOs 21 and 75 are highly proliferative in monoculture but reduce cell-cycle activity when co-cultured with CAFs (Figure 5b) (Figure S8a). CAFs that protect PDOs also have a distinct PTM signaling profile in co-culture (Figure 5c), suggesting patient-specific reciprocal signaling between PDOs and CAFs occurs during chemoprotection. CAFs do not cause PDOs 21 and 75 to exit the cell-cycle, but instead shift PDOs towards a pRB^{+}, IdU^{low} slow-cycling state. This may explain why chemotherapies targeting DNA-replication such as SN-38 and 5-FU are less potent when PDOs 21 and 75 are cultured with CAFs.

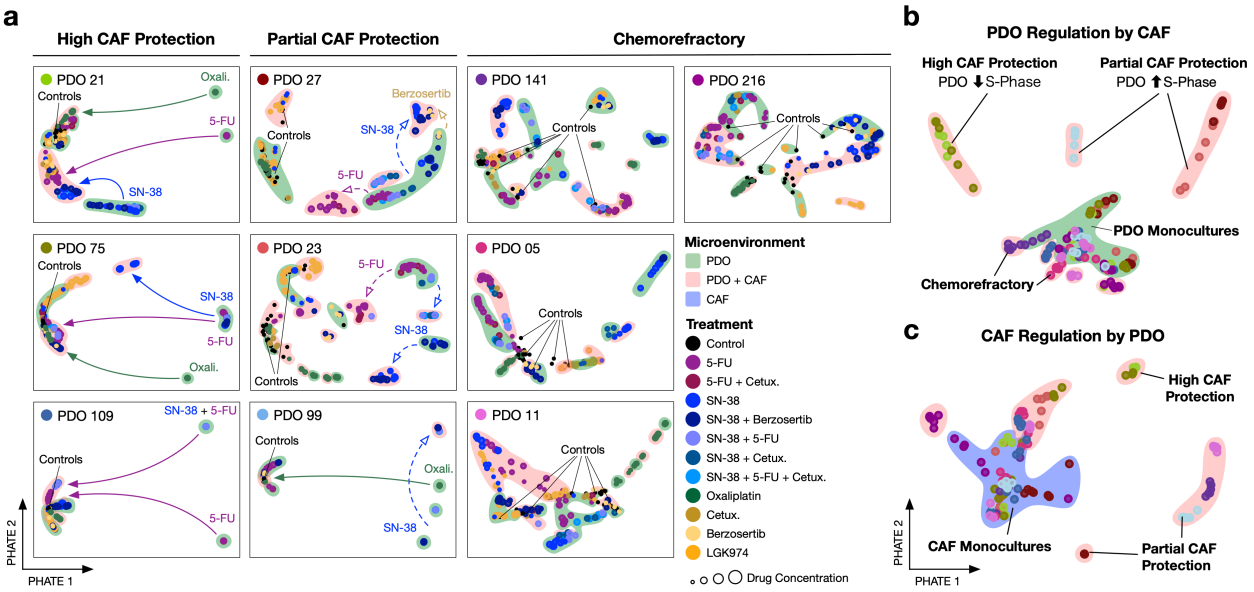

Figure 5: CAFs Chemoprotect PDOs by Altering PDO Cell-State. a) Trellis-PHATE of patient-specific PDO PTM drug responses with or without CAFs illustrates CAFs can protect PDOs from therapy (1,680 PDO-CAF cultures). Dots colored by treatment, shapes colored by microenvironment. Solid arrows refer to full protection, dashed arrows refer to partial protection by CAFs. **b)** Alterations of PDO cell-state and PTM signaling by CAFs correlates with chemoprotection. Dots correspond to 6 replicates colored by PDO, shapes colored by microenvironment. **c)** Baseline CAF cell-state and PTM signaling when co-cultured with PDOs correlates with chemosensitivity protection. Dots correspond to 6 replicates colored by PDO, shapes colored by microenvironment.

Re-sensitizing CAF-Protected PDOs

Mechanistic understanding of PDO drug responses by single-cell PTM analysis could identify opportunities to rationally re-sensitize refractory PDOs. For example, Trellis revealed CAFs protect chemosensitive PDOs from SN-38 — not by reducing on-target S-phase blockage or DNA-damage — but by shifting cancer cells towards a slow-cycling cell-state (Figure 6a-e). Moreover, we observed CAFs activate TGF-β, JNK, and NF-κB signaling and suppress mitotic MAPK and PI3K pathways in PDO 21 (Figure S8d). It has recently been shown that CRC cells can escape chemotherapy by differentiating towards a slow-cycling revival stem cell (revSC) cell-fate [41]. Intestinal revSCs are induced by fibroblast-derived TGF-β and WNT ligands in a YAP-dependent manner [42].

Using the PTM signaling and cell-state insights provided by Trellis, we hypothesized that CAFs chemoprotect cancer cells by shifting PDOs towards a YAP-driven revSC-like state. To test this, we treated PDO 21-CAF cultures +/- Verteporfin (YAP-TEAD complex inhibitor) +/- SN-38 and measured PTM and cell-state responses using TOB*is* mass cytometry (Figure 6f). YAP inhibition did not induce apoptosis in PDO 21 either alone or in co-culture with CAFs. However, when combined with SN-38, Verteporfin re-sensitizes CAF-protected PDOs to chemotherapy-induced apoptosis. Crucially, YAP inhibition does not alter on-target SN-38 DNA-damage in PDO 21, and does not regulate CAF cell-cycle progression or apoptosis (Figure S8e-g) — indicating YAP re-sensitization functions by targeting the unique PDO-CAF cell-state. These results demonstrate that single-cell PTM analysis can be used to re-sensitize CAF-protected PDOs and underscores the value of mechanism-focused drug screening.

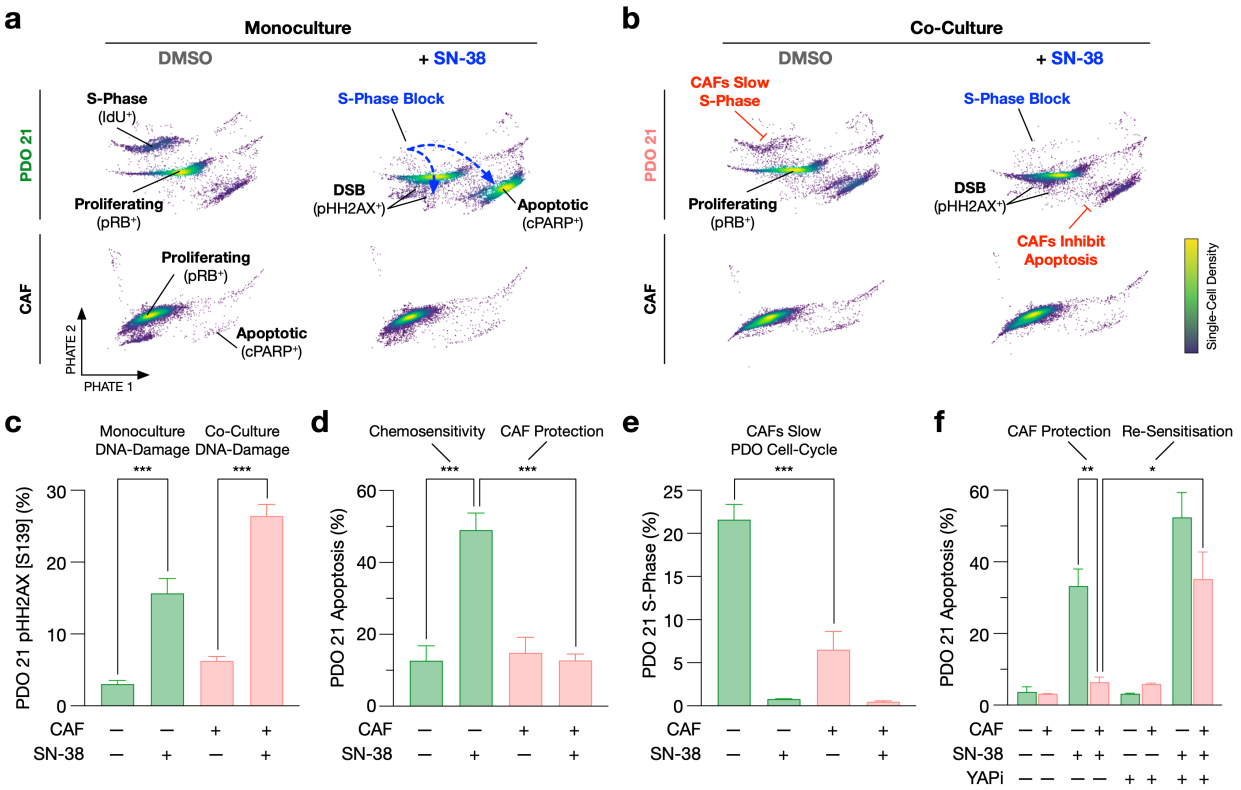

Figure 6: Re-sensitizing CAF Protected PDOs. a-b) Single-cell density PHATEs of PDO 21 and CAFs during SN-38 treatment illustrates cell-state shifts by drug and co-culture (5,000 cells). **c)** SN-38 induces on-target DNA-double strand breaks (DSB) (pHH2AX^{+}) in PDOs irrespective of CAFs, yet **d)** CAFs protect PDOs from SN-38-induced apoptosis. **e)** CAFs slow PDO S-phase entry and PDOs experience on-target S-phase blockage by SN-38 irrespective of CAFs. **f)** Verteporfin (YAPi) completely re-sensitizes CAF-protected PDOs to SN-38-induced apoptosis. Unpaired *t*-test, *** <0.0001, ** <0.001, * <0.01.

3 Discussion

PDOs have been widely proposed as personalized avatars of patient-specific drug responses [43]. However, bulk screening technologies have limited previous studies to PDO monocultures alone and provide no mechanistic insight into PDO drug response [13]. Using highly-multiplexed single-cell PTM profiling by TOB*is* mass cytometry and hierarchical treatment effect analysis by Trellis, we demonstrate PDO drug response signaling is patient-specific and reveal CAFs regulate PDO chemosensitivity by altering PDO cell-state. PDO-CAF interactions are also patient-specific, with CAFs both stimulating and repressing PTM signaling and cell-cycle activity in a patient-specific manner. Crucially, we demonstrate mechanistic profiling of patient-specific drug responses can be used to re-sensitize CAF-protected PDOs.

Unlike static diagnostic metrics (e.g. pharmacogenomics) that have failed to substantially advance precision oncology [43], PDOs are functional biopsies that can be experimentally tested to reveal patient-specific drug responses alongside clinical care in real-time [44–46]. However, recent studies have suggested PDOs alone are not sufficient to biomimetically predict drug response. For example, only 20% of monoculture drug combination 'hits' could be validated in *ex vivo* organotypic CRC tumours containing a TME [44], and cell-extrinsic regulation of PDO cell-state can change pancreatic ductal adenocarcinoma (PDAC) organoid

drug responses [47]. Our results confirm PDO-CAF interactions are a source of functional inter-tumor heterogeneity and should not be generalized. Given that cell-extrinsic signals can have dramatic effects on drug performance, we propose TME cells should be considered in future studies evaluating PDOs as personalized functional biopsies.

Phenotypic plasticity is an emerging hallmark of cancer [48] and therapeutic targeting of cancer-specific cell-states is a growing area of cancer research [49, 50]. As stem cell-driven model systems, PDOs are capable of high differentiation plasticity [8] and are therefore well-suited to studying drug- or TME-induced cancer cell plasticity. We observed that PTM cell-state (not MSI/MSS status, tumor stage, anatomical location, or genotype) aligned with patient-specific drug response (Figure 4c-d) (Figure S7) and found CAFs can transition PDOs into a refractory cell-state to protect PDOs from specific therapies. Previous work in PDAC has demonstrated PDO transcriptional profiles correlate with drug response [51] and recent studies of oncogenic [52] and kinase [53] activity suggest signaling flux predicts patient survival better than genotype. Taken with our observations, mounting evidence suggests metrics that more closely describe cancer cell-state such as organoid transcription and PTM signaling may more accurately predict patient-specific drug responses than genomic profiles or clinical staging. The combination of PDO plasticity with mechanism-focused single-cell screening technologies will enable characterization of functional cell-state plasticity and therapy-induced canalization in cancer in future studies.

In contrast to traditional live/dead drug screens, TOB*is* mass cytometry reveals molecular insights into PDO drug responses. We observed PDOs frequently experience on-target drug responses (83%), but only a subset of PDOs enter drug-induced apoptosis (40%). This suggests chemorefractory PDOs do not translate cytostatic and genotoxic stress into apoptosis. Single-cell PTM profiling further revealed CAFs chemoprotect PDOs by shifting cancer cells into a slow-cycling revSC-like cell-state. We used this mechanistic insight to re-sensitize PDOs by blocking revSC activation via YAP. Given that drug synergy is rare when using unbiased screens [54], mechanism-focused screening could be used to rapidly identify rational drug synergies to re-sensitize refractory cancers.

The advent of high-dimensional single-cell technologies such as mass cytometry and scRNA-Seq provides new opportunities to study heterogeneous drug response mechanisms beyond simple viability scores [13]. However, high-dimensional drug screening data is challenging to interpret — with existing tools designed to analyze dozens, not thousands of samples. Trellis overcomes this scalability bottleneck by distributing single-cell data across a tree domain structure, enabling the KR norm between thousands of single-cell samples to be computed rapidly. While we use cell-state branches to sensitize Trellis results towards canonical on-target anti-cancer drug responses, alternative branching structures could in theory be designed to enrich for cell-type hierarchies (e.g. in immune profiling) or PTM signaling hierarchies (e.g. for kinase inhibitor screens). Trellis' scalability is independent of supervening branches and is therefore a flexible platform for future single-cell screening applications.

Although this study has focused on PDOs and CAFs, single-cell technologies also enable mechanistic analysis of organoid-leukocyte co-culture models [13]. In addition to studying the role of leukocytes in regulating chemical drug responses, single-cell PTM analysis could be a powerful approach to study preclinical co-culture organoid models of anti-solid tumor cellular biotherapeutics (e.g. CAR-T cells) where understanding the biology of the drug (engineered T-cell) is as important as the PDO target-cell killing. This study demonstrates high-throughput single-cell screening of PDO cell-cell drug interactions is feasible and we expect the technology will be rapidly adapted to study biological therapies.

In summary, we demonstrate highly-multiplexed single-cell PTM profiling by TOB*is* mass cytometry and hierarchical treatment effect analysis by Trellis can reveal patient-specific drug responses in thousands of PDO-CAF cultures. CAFs regulate PDO drug response by altering PDO cell-state in a patient-specific

manner and PTM signaling data can be used to overcome CAF protection. We propose single-cell PTM analysis as a powerful alternative to traditional bulk viability analysis of PDOs and suggest TME cells should be considered in future precision medicine models.

4 Methods

4.1 CRC PDO and CRC CAF Culture

CRC PDOs were obtained from the Human Cancer Models Initiative (Sanger Institute, Cambridge, UK) [12] and expanded in 12-well plates (Helena Biosciences 92412T) in x3 25 μL droplets of Growth Factor Reduced Matrigel (Corning 354230) per well with 1 mL of Advanced DMEM F/12 (Thermo 12634010) containing 2 mM L-glutamine (Thermo 25030081), 1 mM N-acetyl-L-cysteine (Sigma A9165), 10 mM HEPES (Sigma H3375), 500 nM A83-01 (Generon 04-0014), 10 μM SB202190 (Avantor CAYM10010399-10), and 1X B-27 Supplement (Thermo 17504044), 1X N-2 Supplement (Thermo 17502048), 50 ng ml^{-1} EGF (Thermo PMG8041), 10 nM Gastrin I (Sigma SCP0152), 10 mM Nicotinamide (Sigma N0636), and 1X HyClone Penicillin-Streptomycin Solution (Fisher SV30010), and conditioned media produced as described in [55] at 5 % CO_2, 37 °C. PDOs were dissociated into single cells with 1X TripLE Express Enzyme (Gibco 12604013) (incubated at 37 °C for 20 minutes) and passaged every 10 days. L-cells for conditioned media production were obtained from Shintaro Sato (Research Institute of Microbial Diseases, Osaka University, Osaka, Japan). To aid cell-type-specific visualization and gating, CRC PDO were transfected with H2B-RFP (Addgene 26001). CRC CAFs (+GFP) were a kind gift from Prof. Olivier De Wever (University of Gent) [16, 17] and cultured in DMEM (Thermo 11965092) enriched with 10 % FBS (Gibco 10082147), and 1X HyClone Penicillin-Streptomycin Solution (Fisher SV30010) at 5% CO_2, 37 °C.

4.2 PDO-CAF Drug Treatments

PDOs were dissociated into single cells on day 0, and expanded in 12-well plates in Growth Factor Reduced Matrigel (Corning 354230) with Advanced DMEM F/12 (Thermo 12634010) containing 2 mM L-glutamine (Thermo 25030081), 1 mM N-acetyl-L-cysteine (Sigma A9165), 10 mM HEPES (Sigma H3375), 1X B-27 Supplement (Thermo 17504044), 1X N-2 Supplement (Thermo 17502048), 50 ng ml^{-1} EGF (Thermo PMG8041), 10 nM Gastrin I (Sigma SCP0152), 10 mM Nicotinamide (Sigma N0636), 500 nM A83-01 (Generon 04-0014), 10 μM SB202190 (Avantor CAYM10010399-10) and 1X HyClone Penicillin-Streptomycin Solution (Fisher SV30010) at 5% CO_2, 37 °C for 4 days. On day 5, PDOs were starved in Reduced media (containing only 2 mM L-glutamine, 1 mM N-acetyl-L-cysteine, 10 mM HEPES, 1X B-27 Supplement, 1X N-2 Supplement, 10 mM Nicotinamide, and 1X HyClone-Penicillin Streptomycin Solution) at 5 % CO_2, 37 °C. In parallel, CAFs were starved in 2 % FBS DMEM with 1X Hyclone-Penincillin Streptomycin Solution. PDOs and CAFs were seeded on day 6 in 96-well plates (Helena Biosciences 92696T) in 50 μL Matrigel stacks with 300 μL of reduced media. PDO monocultures are seeded at a density of $\sim$ 1.5 x10^3 organoids/well, and CAFs at 2.5 x 10^5 cells/well, co-cultures were mixed in Matrigel on ice at the densities described, and seeded together on the plates for polymerization. On day 7, media was replaced with titrated concentrations of SN-38 (Sigma H0165), 5-FU (Merck F6627), Oxaliplatin (Merck O9512), Cetuximab (MedChem Express HY-P9905), VX-970 (Stratech), and LGK-974 (Peprotech 1241454) (Table S2) diluted in Reduced media. On day 8, media was replaced with the corresponding treatments (same as on day 7). After 72 hours of co-culture, and 48-hours of treatment (day 9), cultures were processed for TOB*is* mass cytometry (see below). Verteporfin (Cambridge Bioscience CAY17334) was used at 100 nM as above.

4.3 PDO-CAF TOB*is* Mass Cytometry

PDO-CAF co-cultures were analyzed using the TOB*is* mass cytometry protocol outlined in detail by Sufi and Qin *et al.*, Nature Protocols, 2021 [15]. Briefly, following drug treatment, PDO-CAF cultures

were incubated with 25 μM (5-Iodo-2'-deoxyuridine) (^{127}IdU) (Fluidigm 201127) at 37 °C for 30 minutes, and 5 minutes before the end of this incubation, 1X Protease Inhibitor Cocktail (Sigma, P8340) and 1 XPhosSTOP (Sigma 4906845001) are added into the media. After the incubation with ^{127}IdU, protease inhibitors and PhosSTOP, each well is fixed in 4 % PFA/PBS (Thermo J19943K2) for 1 hour at 37 °C. PDO-CAFs were washed with PBS, dead cells were stained using 0.25 μM ^{194}Cisplatin (Fluidigm 201194), and PDO-CAFs were barcoded *in situ* with 126-plex (*9*-choose-*4*) TOB*is* overnight at 4 °C. Unbound barcodes were quenched in 2 mM GSH and all PDO-CAFs were pooled. PDO-CAFs were dissociated into single-cells using 1 mg ml^{-1} Dispase II (Thermo 17105041), 0.2 mg ml^{-1} Collagenase IV (Thermo 17104019), and 0.2 mg ml^{-1} DNase I (Sigma DN25) in C-Tubes (Miltenyi 130-096-334) via gentleMACS™ Octo Dissociator with Heaters (Miltenyi 130-096-427). Single PDO and CAF cells were washed in cell staining buffer (CSB) (Fluidigm 201068) and stained with extracellular rare-earth metal conjugated antibodies (Table S3) for 30 minutes at room temperature. PDO-CAFs were then permeabilized in 0.1 % (vol/vol) Triton X-100/PBS (Sigma T8787), 50 % methanol/PBS (Fisher 10675112), and stained with intracellular rare-earth metal conjugated antibodies for 30 minutes at room temperature. PDO-CAFs were then washed in CSB and antibodies were cross-linked to cells using 1.6 % (vol/vol) FA/PBS for 10 minutes. PDO-CAFs were incubated in 125 nM ^{191}Ir/^{193}Ir DNA intercalator (Fluidigm 201192A) overnight at 4 °C. PDO-CAFs were washed, resuspended in 2 mM EDTA (Sigma 03690) in water (Fluidigm 201069), and analyzed using a Helios Mass Cytometer (Fluidigm) fitted with a 'Super Sampler' (Victorian Airships) or CyTOF XT (Fluidigm) at 200-400 events s^{-1}.

4.4 TOB*is* Mass Cytometry Data Preprocessing

Multiplexed FCS files were debarcoded into separate experimental conditions by using the Zunder Lab Single Cell Debarcoder (https://github.com/zunderlab/single-cell-debarcoder) [18]. Debarcoded FCS files were uploaded to Cytobank (Beckman Coulter), gated for Gaussian parameters, and DNA (^{191}Ir/^{193}Ir). Epithelial cells were gated on PCK$^+$ and EpCAM$^+$, and CAFs were gated on Vimentin$^+$ and GFP$^+$. Arcsinh transformed values were mean centered across batches before Trellis analysis.

4.5 Trellis Computational Background

Single-cell data are being collected in experiments with ever more numerous conditions in order to characterize libraries of treatments [56] including small-molecules [57] and gene-perturbations [58]. One method that directly generalizes bulk measurements to single-cell samples is through the theory of optimal transport and more specifically, the Wasserstein distance [21–23].

Optimal transport is well suited to the formulation of distances between collections of points, as it generalizes the notion of distances between points to distances between distributions. Intuitively, the distance between distributions should be the minimum total work to move a pile of dirt located at a source distribution to a target distribution. This framework yields a natural definition of similarity between experimental conditions, namely two conditions are similar when their collections of cells are not far from each other.

These distances aim at answering a deeper question: *Which treatments have similar and different effects on the system?* To answer this question we need a metric between changes to densities. We assume that for each treated condition X we have access to an associated control condition X_c. When all treated conditions are measured relative to a single X_c we show approaches based on the Wasserstein distance are a valid metric between changes in densities. However, in larger experiments it is impossible to measure all treated conditions within a single batch, and thus treated conditions may have different controls. In this case, we show that Wasserstein-based approaches fail, and show that a generalization to an approach based on the

Kantorovich and Rubinstein norm gives a valid metric between changes in densities in this more general multi-control case.

4.5.1 Integral Probability Metrics

Integral probability metrics (IPMs) are metrics over probability measures μ, ν some common space $\mathcal{X}$ that can be expressed as

$$\text{IPM}_{\mathcal{F}}(\mu, \nu) = \sup_{f \in \mathcal{F}} \left| \int_{\mathcal{X}} f d\mu - \int_{\mathcal{X}} f d\nu \right| \tag{1}$$

where $\mathcal{F}$ is a family of real-valued bounded measurable functions on $\mathcal{X}$. For specific choices of $\mathcal{F}$ the Dudley metric, Total variation distance, Kologorov distance, maximum mean discrepancy, and Wasserstein distance can all be expressed as IPMs.

IPMs are often useful when we only have samples $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$, $\hat{\nu} = \frac{1}{m}\sum_{j=1}^{m} \delta_{y_j}$ drawn from probability measures μ, ν [59]. In this case it is possible to directly estimate IPMs, unlike for the class of ϕ-divergences which either do not converge or are $+\infty$. The Wasserstein metric is of particular interest as it has an interpretable primal formulation as the transport of mass between distributions.

4.5.2 The Wasserstein Metric as a Norm

Let μ, ν be two probability distributions on a measurable space $\mathcal{X}$ with metric d, let $\Pi(\mu, \nu)$ be the set of joint probability distributions π on the space $\mathcal{X}^2$ where for any subset $\omega \subset \mathcal{X}$, $\pi(\omega \times \mathcal{X}) = \mu(\omega)$ and $\pi(\mathcal{X} \times \omega) = \nu(\omega)$. The α-Wasserstein distance is defined as:

$$W_d^{\alpha}(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}^2} d(x, y)^{\alpha} \pi(dx, dy) \right)^{1/\alpha}. \tag{2}$$

The *Kantorovich–Rubinstein dual* for the Wasserstein distance on arbitrary measures is

$$\sup_{(f,g) \in \mathcal{C}(\mathcal{X})^2} \int_{\mathcal{X}} f(x) d\mu(x) + \int_{\mathcal{X}} g(y) d\nu(y) \tag{3}$$

subject to $f(x) + g(y) \leq d(x, y)^{\alpha}$ for all $(x, y) \in \mathcal{X}^2$. Most work applying the Wasserstein distance focuses on $\alpha = 2$ [60] or more general convex costs with $\alpha > 1$ [61], due to the provable regularity of the transport map. We instead focus on the case where $0 < \alpha \leq 1$. Here the transportation map loses regularity but admits a simplification of the dual as when $0 < \alpha \leq 1$, it can be shown that Eq. 3 achieves optimality when $g = -f$ [62, Prop. 6.1] and so simplifies to:

$$W_d^{\alpha} = \sup_{f} \left\{ \int_{\mathcal{X}} f(x)(d\mu(x) - d\nu(x)) : \mathcal{H}_d^{\alpha}(f) \leq 1 \right\} \tag{4}$$

where

$$\mathcal{H}_d^{\alpha}(f) := \sup_{(x,y) \in \mathcal{X}^2} \left\{ \frac{|f(x) - f(y)|}{d(x, y)^{\alpha}} : x \neq y \right\}. \tag{5}$$

When $0 < \alpha \leq 1$, Eq. 4 shows that W_d^{α} is the dual of the α-Hölder functions $\{f : \mathcal{H}_{\alpha}(f) \leq 1\}$ and is a norm, namely

$$W_d^{\alpha}(\mu, \nu) = \|\mu - \nu\|_{W_d^{\alpha}}, \tag{6}$$

and is valid for any measures μ, ν such that $\int_{\mathcal{X}} \mu = \int_{\mathcal{X}} \nu$. Of particular interest is that W_d^α is still a norm even for non-positive measures. This generalization to non-positive measures will form the basis for our Trellis metric between datasets and is known as the *Kantorovich–Rubinstein norm* [63] when applied to differences of non-positive measures.

Definition 1 ([63])**.** *The Kantorovich-Rubinstein (KR) distance between measures μ, ν such that $\int_{\mathcal{X}} \mu = \int_{\mathcal{X}} \nu$ with respect to ground distance d as*

$$\mathrm{KR}_d^\alpha(\mu, \nu) := \sup_f \left\{ \int_{\mathcal{X}} f(x)(d\mu(x) - d\nu(x)) : \mathcal{H}_d^\alpha(f) \leq 1 \right\} = \|\mu - \nu\|_{\mathrm{KR}_d^\alpha}. \tag{7}$$

For simplicity we will drop the α term and assume $\alpha = 1$, but all statements apply to $0 < \alpha \leq 1$ unless otherwise specified. Trellis can be thought of as an efficient implementation of the KR norm over a tree ground distance.

4.5.3 The Wasserstein Distance with Tree Ground Distance

Consider discrete distributions $\mu = \sum_{i=1}^n \mu_i \delta_i$ and $\nu = \sum_{i=1}^n \nu_i \delta_i$ where δ is the dirac function in $\mathbb{R}^d$ and $\sum_{i=1}^n \mu_i - \nu_i = 0$. Then for general costs, the Wasserstein distances between μ and ν can be computed exactly in $\tilde{O}(n^3)$ using the Hungarian algorithm [33], and approximated using a slightly modified entropy regularized problem in $\tilde{O}(n^2)$ with the Sinkhorn algorithm [34].

However, for some classes of the ground distance, there exist more efficient algorithms (See Table 1). For example, if d is the Euclidean distance in $\mathbb{R}$, then the Wasserstein distance can be computed in $O(n \log n)$ time and is equivalent to sorting [62, 64]. This special case is exploited in sliced-Wasserstein metrics [65, 66] to compute approximate Wasserstein distances in higher dimensions.

Another more general class of ground distances where there exist efficient algorithms is the class of tree metrics. Let $\mathcal{T}$ be a rooted tree with non-negative edge lengths, and let $d_{\mathcal{T}}$ be a *tree metric* on $\mathcal{T}$. Then for two measures μ, ν over $\mathcal{T}$, the Wasserstein distance with respect to $d_{\mathcal{T}}$, $W_{d_{\mathcal{T}}}(\mu, \nu)$, can be computed in $O(n)$ time by exploiting the fact that there is a single path between any pair of masses [31, 67, 68]. In this case the 1-Wasserstein distance, also known as the Earth Mover's Distance (EMD) can be expressed as

$$W_{d_{\mathcal{T}}}^1 = \left(\sum_{x \in \mathcal{T}} w_x |\mu(\Gamma(x)) - \nu(\Gamma(x))| \right)^{1/\alpha} \tag{8}$$

where w_x is the weight / distance to the parent node of x and $\Gamma(x)$ represents the set of nodes in the subtree of x. Let $P(x, y)$ be the unique path between x and y, then $\Gamma(x) = \{y \in \mathcal{T} \mid x \in P(r, y)\}$. This alternative formulation can be embedded in l_1:

$$W_{d^{\mathcal{T}}} = \|v(\mu) - v(\nu)\|_1^{1/\alpha} \tag{9}$$

where $v : \mu(\mathcal{T}) \to \mathbb{R}^n$ is a function such that $v(\mu)_x = w_x \mu(\Gamma(x))$.

Approximating the Euclidean distance with a tree distance can be done probabilistically with $O(d \log \Delta)$ distortion in expectation where Δ is a resolution parameter [69]. Following the result of Charikar [70], this implies that the α-Wasserstein distance with tree ground distance has the same order distortion. One simple tree construction that achieves this distortion is known as "Quadtree", where each node has four children in $\mathbb{R}^2$ and 2^d children in $\mathbb{R}^d$ [31]. We introduce a new tree construction based on k-means clustering, which we show is a generalization of the Quadtree construction but can be applied to higher dimensions.

Algorithm 2: $\text{Trellis}(X, \boldsymbol{\mu}, k, l, \mathcal{T}_m, \boldsymbol{c})$

Input: $n \times f$ data matrix X, $n \times m$ distributions $\boldsymbol{\mu}$, # of clusters k, and # of levels l, manual tree $\mathcal{T}_m$, and (optional) control mapping $\boldsymbol{c}$ specifying control distribution set for each distribution.
Output: $m \times |\mathcal{T}|$ distribution embeddings $\boldsymbol{v}$
$\mathcal{T} \leftarrow \text{BuildTree}(X, k, l, \mathcal{T}_m)$
for Node $\mathcal{T}_i$ with parent edge weight w_i in $\mathcal{T}$ **do**
  $\boldsymbol{v}[:, i] \leftarrow w_i \boldsymbol{\mu}(\Gamma(\mathcal{T}_i))$
end for
if $\boldsymbol{c}$ is $null$ **then**
  return $\boldsymbol{v}$
end if
for control distribution set μ_c for each distribution μ in $\boldsymbol{c}$ **do**
  $\boldsymbol{v}[\mu] \leftarrow \boldsymbol{v}[\mu] - \text{mean}_{\mu_c}(\boldsymbol{v}[\mu_c])$
end for
return $\boldsymbol{v}$

4.6 Unpaired and Paired Trellis

We start with a more detailed overview of the Trellis algorithm for comparing the effects of drugs on different experimental conditions. The Trellis algorithm is summarized in Algorithm 2. At a high level Trellis consists of four steps:

1. Construct a hierarchical tree partitioning of the data $\mathcal{T}$
2. Embed each distribution μ^i over $\mathcal{T}$ to a vector $v(\mu^i)$ such that $\text{Trellis}(\mu^i, \mu^j) = \|v(\mu^i) - v(\mu^j)\|_1^{1/\alpha}$
3. Optional: subtract a control distribution embedding $v(\mu_c^i)$ from each $v(\mu^i)$ for paired Trellis
4. Compute nearest Trellis neighbor distributions exploiting L^1 geometry

We discuss potential methods of constructing $\mathcal{T}$ in section 4.6.1, how to embed an empirical distribution to a vector and its equivalence to the Wasserstein distance in section 4.6.2, the effect of subtracting a control distribution embedding in section 4.6.3, and finally how to construct a Trellis-metric nearest neighbor graph for subsequent visualization with a non-linear embedding algorithm such as PHATE [20], UMAP [27], or t-SNE [28] in section 4.6.4.

4.6.1 Constructing Trees on Single-Cell Mass Cytometry Data

Trellis gives a distance between measures or differences in measures over a tree metric space. Often the data is not associated with an explicit tree metric, but is naturally hierarchical such as in the case of single-cell cytometry data. Previous methods have used manual gating, automatic gating, or a combination of the two to hierarchically cluster single-cell mass cytometry data [71]. These methods build trees, but are missing the 'metric' component, which can be encoded as the edge weights between parent and child clusters. We use a simple tree metric where each edge weight for node x is the Euclidean distance between the cluster center $\text{mean}(x)$ and the center of its parent $\text{mean}(Pa(x))$.

$$w_x = \|\text{mean}(x) - \text{mean}(Pa(x))\|_2. \tag{10}$$

Algorithm 3: $\text{BuildTree}(X, k, l, \mathcal{T}_m)$

Input: $n \times f$ data matrix X, # of clusters k, # of levels l and (optional) manual base tree $\mathcal{T}_m$.
Output: Weighted hierarchical clustering tree $\mathcal{T}$.
if $l = 0$ **then**
  return $null$
end if
if $\mathcal{T}_m$ is not $null$ **then**
  for leaf node n_i in $\mathcal{T}_m$ **do**
    $\mathcal{T}_i \leftarrow \text{BuildTree}(X[n], k, l, null)$
  end for
  // Where TreeJoin replaces each of the leaves with the respective subtree
  return $\mathcal{T} \leftarrow \text{TreeJoin}(\mathcal{T}_m, [\mathcal{T}_i])$
end if
$labels \leftarrow \text{Kmeans}(X)$
for $i = 1$ to k **do**
  $\mathcal{T}_i \leftarrow \text{BuildTree}(X[labels = i], k, l - 1)$
end for
return $\mathcal{T} \leftarrow [\mathcal{T}_i]_{i=1}^k$

The tree metric between two nodes $u, v \in \mathcal{T}$ is the sum of the path lengths along the unique path geodesic between u and v in $\mathcal{T}$ denoted by $P_{\mathcal{T}}(u, v)$ then

$$d_{\mathcal{T}}(x, x) = \sum_{v \in P_{\mathcal{T}}(x,y)} w_v. \tag{11}$$

Trellis applies to any clustering, we demonstrate the Trellis framework using a simple combination of manual gating for non-Euclidean features and automatic gating to approximate Euclidean distances among sub populations. This strategy allows us to leverage manual gating when appropriate due to prior biological knowledge, or automatic gating using repeated *k*-means clustering with no prior on the biological splits. This clustering method is particularly interesting because for specific settings we can show that the Trellis metric is topologically equivalent to an α-Wasserstein distance with Euclidean ground distance in $\mathbb{R}^d$.

Given a number of clusters at each level k and a depth h construct a divisive hierarchical clustering of the data as described in Algorithm 3. Where Kmeans is the *k*-means algorithm with some fixed set of parameters. Interestingly, with a specific setting of *k*-means we show Trellis is topologically equivalent to the α-Wasserstein distance with Euclidean ground distance. This is formalized in the following proposition.

Proposition 1. *Let $k = 2^d$, max_iter $= 0$, data X be normalized such that $X \in [-1, 1]^d$ with precision Δ and initialize the kth cluster at level l with parent center p as $p + 2^{1-l}(\textit{Binary}(k) - 1/2)$. Then there exists constants c, C such that*

$$c \cdot W_{\|\cdot\|_2}(\mu, \nu) \leq \mathbb{E}[\text{Trellis}(\mu, \nu)] \leq C \log \Delta \cdot W_{\|\cdot\|_2}(\mu, \nu). \tag{12}$$

This can be seen by first noting that this initialization is equivalent to a QuadTree construction in the topological sense. If two points are clustered together in our construction at some level then they are also clustered together in QuadTree at the equivalent level. In addition, the edge weights are equivalent up to a constant with the edge weights decaying by $1/2$ at every level in both constructions. Once these two

properties are verified, then we cal leveraging existing results on QuadTree constructions from [31] and [70] to show that the inequalities hold. We also note that there exist results on the approximate nearest neighbors of this construction in [68].

While these parameters for kmeans-clustering work well in low dimensions, the number of clusters scales exponentially with dimension. In practice we use four levels of four clusters. This expectation holds over a randomly selected initialization of the zero'th level cluster. In practice, we take the expectation over k-means initializations, building ten parallel trees with different initializations.

Trellis can be applied to any tree metric or ensemble of tree metrics. We have presented a method that allows for combining manual and automatic gating, as well as an automatic gating method that in expectation is similar to a Euclidean distance. Many other choices for partitioning CyTof data have been explored in the automatic gating literature [71–74]. These automatic gating methods are generally used for partitioning the data not building a tree metric. However, it is simple to convert them into tree metrics by assigning edge weights based on cluster means. This strategy can be applied to a precomputed hierarchical clustering of the data with no knowledge of how those clusters were chosen. This allows for adaptation of Trellis to different systems where either manual or automatic gating is preferred.

4.6.2 Trellis Given a Metric Tree

Given a general metric tree $\mathcal{T}$, we first define the embedding function $v : \mu(\mathcal{T}) \to \mathbb{R}^{|\mathcal{T}|}$ which takes distributions defined over the tree and embeds them in a vectorspace where the L^1 between vectors is equivalent to the Wasserstein distance with tree ground distance. Given edge weights w_x and denoting the subtree at node x as $\Gamma(x) = \{y \in \mathcal{T} | x \in P(r, y)\}$, then v is defined element-wise as

$$v(\mu) = [w_x \mu(\Gamma(x))]_{x \in \mathcal{T}}. \tag{13}$$

Intuitively, this can be thought of computing the sum of the mass below each node times the edge weight at each node. The difference between $v(\mu)_x - v(\nu)_x$ for a given node $x \in \mathcal{T}$ can be thought of as the amount of work needed to move μ to ν. If this difference is positive, then this means that mass of μ is greater in the subtree $\Gamma(x)$ than the mass of ν. This means that the transport map must move exactly $\mu(\Gamma(x)) - \nu(\Gamma(x))$ mass upwards from x at cost w_x. Adding up these aggregate movements over all nodes gives the total work needed and is equivalent to the work required by the Wasserstein distance.

For our tree construction in Section 4.6.1 with the additional manual tree step, we define the unpaired Trellis distance ($\mathrm{uTrellis}$) as

$$\mathrm{uTrellis}(\mu, \nu) = \|v(\mu) - v(\nu)\|_1. \tag{14}$$

without the manual tree step considering only the k-means construction we define this as the TreEMD distance which is similar to previous Tree-based Wasserstein distance constructions for high dimensions [67, 68].

These two unpaired distances are comparable to existing methods for computing the Wasserstein distance between distributions. We discuss related methods for computing or approximating the Wasserstein distance in Section 4.6.5. However, these distances do not take into account control, treatment, batch, and replicate information. Given information on which samples were taken under similar conditions, we are able to improve the distances with *Paired Trellis*.

4.6.3 Paired Trellis

To examine the effects of a drug across many conditions it is useful to measure the differences of the treated condition relative to the control.

For each sample μ and ν, let the associated control distributions be μ_c and ν_c respectively, and v be defined as above. Then we define the Paired Trellis metric between changes in distributions as:

$$\text{pTrellis}(\mu, \nu) := \|v(\mu) - v(\mu_c) - v(\nu) - v(\nu_c)\|_1.$$

Intuitively, the Paired Trellis distance measures the difference in the change in density between treated conditions from their respective controls. This allows us to control for unmeasured confounders that are implicit in the treated cell population μ and ν respectively.

Proposition 2. *For two distributions μ, ν with their respective controls μ_c, ν_c, the Paired Trellis is equivalent to a Kantorovich-Rubenstein distance with tree ground distance as in Eq. 8*

$$\textit{pTrellis}(\mu, \nu) = \mathrm{KR}_{d_\mathcal{T}}(\mu - \mu_c, \nu - \nu_c). \tag{15}$$

Proof. The equivalence of paired Trellis to a Kantorovich-Rubenstein distance can be verified through algebraic manipulation following [75]. We start with the definition of the Kantorovich-Rubenstein distance and show that this is equivalent to pTrellis for an arbitrary tree domain $\mathcal{T}$ with ground distance $d_\mathcal{T}$. Denote the family of Hölder functions under $d_\mathcal{T}$ as $\mathcal{F} = \{f : \mathcal{H}^\alpha_{d_\mathcal{T}}(f) \leq 1 \& f(r) = 0\}$ and let λ be the (unique) length measure on $\mathcal{T}$ such that $d_\mathcal{T}(x, y) = \lambda(P(x, y))$. Then there exists a unique function $g : \mathcal{T} \to [-1, 1]$ such that $f(x) = \int_{P(r,x)} g(z)\lambda(dz) = \int_\mathcal{T} \mathbf{1}_{z \in P(r,x)} g(z)\lambda(dz)$.

$$\int_\mathcal{T} f(x)d\mu(x) = \int_\mathcal{T}\int_\mathcal{T} \mathbf{1}_{z\in P(r,x)} g(z)\lambda(dz)d\mu(x) = \int_\mathcal{T} g(z)\mu(\Gamma(z))\lambda(dz). \tag{16}$$

For the optimal witness function f^*, we have

$$g(z) = \begin{cases} 1 & \text{if } \mu(\Gamma(z)) > \nu(\Gamma(z)) \\ -1 & \text{else} \end{cases}. \tag{17}$$

Plugging this equivalence into Eq. 7 we have:

$$KR_{d_\mathcal{T}}(\mu, \nu) = \sup_f \left\{ \int_\mathcal{T} f(x)(d\mu(x) - d\nu(x)) : \mathcal{H}^\alpha_{d_\mathcal{T}}(f) \leq 1 \right\} = \int_\mathcal{T} |\mu(\Gamma(z)) - \nu(\Gamma(z))|\lambda(dz). \tag{18}$$

Therefore, for two measures a, b over $\mathcal{T}$ such that $\int_\mathcal{T} a(x)dx = \int_\mathcal{T} b(x)dx = c$ we have that $a(\Gamma(r)) = b(\Gamma(r)) = c$ and for $v : \mathcal{T} \to \mathbb{R}^+$ as defined in Eq. 13 we have

$$KR_{d_\mathcal{T}}(a, b) = \sum_{x\in\mathcal{T}} w_x |a(\Gamma(x)) - b(\Gamma(x))| = \|v(a) - v(b)\|_1. \tag{19}$$

substituting $a = \mu - \mu_c$ and $b = \nu - \nu_c$ yields the proposition since $\int_\mathcal{T} a = \int_\mathcal{T} b = 0$ for any distributions μ, μ_c, ν, and ν_c. □

4.6.4 Nearest Trellis Neighbors

Fast nearest neighbor calculation is useful in graph-based methods which use the k-nearest neighbor graph for down stream tasks such as clustering [76], classification [68], or visualization [19, 20, 27]. For nearest neighbors in normed spaces such as the L^2 norm, the geometry of the space can be utilized for fast

exact or approximate nearest neighbor calculation in time scaling logarithmically with the number of points. For more general distances between objects, these algorithms may not apply.

For instance, to compute the k-nearest neighbor distributions in terms of the Wasserstein distance for m distributions, there is no faster algorithm than computing the Wasserstein distance to all other distributions then computing the k closest ones in $O(m)$ time. However, the Unpaired and Paired Trellis versions of the Wasserstein distance for finite data can be expressed as norms in a finite dimensional space, this allows us to apply fast nearest neighbor algorithms which exploit the induced geometry between distributions. In this case, to find nearest neighbor distributions we can apply tree-based algorithms such as KD-Trees, or Ball-Trees as used in PHATE [20] and scikit-learn [77], locality sensitive hashing in $O(T \log m)$ time for m distributions on trees of size T.

Method	Exact	KR-control	Ground cost	knn-Time
Exact EMD [33]	Yes	No	Any	$O(m^2n^3)$
Sinkhorn EMD [34]	No	No	Any	$O(m^2n^2)$
PhEMD [21]	No	No	$d_{\mathcal{M}}$	$O(m^2T^3 + n^3)$
Mean	No	Yes	Any	$\tilde{O}(kmn)$
Diffusion EMD [22]	No	Yes	$d_{\mathcal{M}}$	$\tilde{O}(kmn)$
Trellis	Yes	Yes	$d_{\mathcal{T}}$	$\tilde{O}(kmT + n)$

Table 1: Comparison of Computational Earth Mover's Distance Methods separated into super-linear (top), and log-linear methods (bottom) based on time-complexity of computing k-Wasserstein-nearest-neighbors. Assumes a dataset of m distributions over n points with (optionally) a tree of size $T < n$.

4.6.5 Related Work and Time Complexity

There are many methods for computing or approximating the Wasserstein distance. In Table 1 we present methods for computing the nearest neighbor distributions according to the Wasserstein distance split into two groups. Here we consider the time it takes for the method to compute the k-Wasserstein-nearest-neighbors on a dataset with m distributions over n points with access to a precomputed tree over the data of size $T < n$. The first three methods are widely used, but do not scale well to large datasets with a large number of distributions or a large number of points. For the first three methods, the Hungarian algorithm [33], the Sinkhorn algorithm [34], and PhEMD [21], to find the k-nearest-neighbors for a distribution it is necessary to compute the distance to all m other distributions. This implies that they scale poorly with the number of distributions as illustrated in Figure S2b. PhEMD saves significant time by only computing the distances between a small set of clusters, however, eventually this is dominated by an increasing number of distributions.

Trellis / TreEMD scales log linearly in the number of points, distributions, and the size of the precomputed tree $\mathcal{T}$. Constructing the tree partitioning for Trellis takes $\tilde{O}(n)$ time. Embedding the distributions takes $O(mT)$ time. Subtracting the control distribution embedding for paired Trellis takes $O(T)$ time. finally, computing the k-nearest neighbors of the Trellis distance takes $\tilde{O}(kmT)$ time. In total both unpaired and paired Trellis take $\tilde{O}(kmT + n)$ time to compute the k nearest neighbor distributions.

When $T \ll n$ as in our case, we can see substantial increases in speed in line with simply taking the Euclidean distance between means of clusters. As T achieves its upperbound of $2n - 1$, Trellis has the same complexity as computing the nearest distribution means and of DiffusionEMD [22].

5 Data Availability

All mass cytometry files are available on Cytobank at: https://community.cytobank.org/cytobank/projects/1461. Compiled TOB*is* mass cytometry PDO-CAF dataframe is available at: Ramos, Maria (2022), "Ramos Zapatero et. al (Cancer-Associated Fibroblasts Regulate Patient-Derived Organoid Drug Responses)", Mendeley Data, V1, doi: 10.17632/hc8gxwks3p.1 (with a key in Table S2).

6 Code Availability

Trellis code is available at: https://github.com/KrishnaswamyLab/Trellis. Code to reproduce all PHATE embeddings in this paper is available at: https://github.com/TAPE-Lab/Ramos-et-al-Trellis.

7 Acknowledgments

We are extremely grateful to M. Garnett and H. Francies for sharing CRC PDOs and O. De Wever for providing CRC CAFs. We thank Y. Guo, K. Boustani, and G. Morrow from the UCL CI Flow-Core for CyTOF support. This work was supported by Cancer Research UK (C60693 / A23783), the Cancer Research UK City of London Centre (C7893 / A26233), the UCLH Biomedical Research Centre (BRC422), the Rosetrees Trust (M872 / A2292), the Yale-UCL Collaborative Student Exchange Programme, NIH grants (R01GM135929 / R01GM130847), NSF Career grant (2047856), the Chan-Zuckerberg Initiative grants (CZF2019-182702 / CZF2019-002440), and the Sloan Fellowship (FG-2021-15883).

8 Author Contributions

M.R.Z. designed the study, performed all PDO-CAF TOB*is* mass cytometry experiments, analyzed the data, and wrote the paper. A.T. conceived and developed Trellis, analyzed the data, and wrote the paper. J.S. conjugated mass cytometry antibodies and developed TOB*is* barcodes. P.V., C.N., and X.Q. provided PDO and CAF support. F.C.R. analyzed the data. D.H. provided chemotherapies and oversaw the project. S.K. conceived and oversaw Trellis. C.J.T. designed the study, analyzed the data, and wrote the paper.

9 Competing Interests

S.K. is on the scientific advisory board of KovaDx and AI Therapeutics.

References

[1] C. J. Tape. The heterocellular emergence of colorectal cancer. *Trends Cancer*, 3(2):79–88, 2017. 1

[2] Y. Xi and P. Xu. Global colorectal cancer burden in 2020 and projections to 2040. *Transl Oncol*, 14 (10):101174, 2021. 2

[3] W. M. Grady and J. M. Carethers. Genomic and epigenetic instability in colorectal cancer pathogenesis. *Gastroenterology*, 135(4):1079–99, 2008. 2

[4] E. Sahai, I. Astsaturov, E. Cukierman, D. G. DeNardo, M. Egeblad, R. M. Evans, D. Fearon, F. R. Greten, S. R. Hingorani, T. Hunter, R. O. Hynes, R. K. Jain, T. Janowitz, C. Jorgensen, A. C. Kimmelman, M. G. Kolonin, R. G. Maki, R. S. Powers, E. Pure, D. C. Ramirez, R. Scherz-Shouval, M. H. Sherman, S. Stewart, T. D. Tlsty, D. A. Tuveson, F. M. Watt, V. Weaver, A. T. Weeraratna, and Z. Werb. A framework for advancing our understanding of cancer-associated fibroblasts. *Nat Rev Cancer*, 20 (3):174–186, 2020. 2, 7

[5] A. Calon, E. Lonardo, A. Berenguer-Llergo, E. Espinet, X. Hernando-Momblona, M. Iglesias, M. Sevillano, S. Palomo-Ponce, D. V. Tauriello, D. Byrom, C. Cortina, C. Morral, C. Barcelo, S. Tosi, A. Riera, C. S. Attolini, D. Rossell, E. Sancho, and E. Batlle. Stromal gene expression defines poor-prognosis subtypes in colorectal cancer. *Nat Genet*, 47(4):320–9, 2015. 2

[6] A. Woolston, K. Khan, G. Spain, L. J. Barber, B. Griffiths, R. Gonzalez-Exposito, L. Hornsteiner, M. Punta, Y. Patil, A. Newey, S. Mansukhani, M. N. Davies, A. Furness, F. Sclafani, C. Peckitt, M. Jimenez, K. Kouvelakis, R. Ranftl, R. Begum, I. Rana, J. Thomas, A. Bryant, S. Quezada, A. Wotherspoon, N. Khan, N. Fotiadis, T. Marafioti, T. Powles, S. Lise, F. Calvo, S. Guettler, K. von Loga, S. Rao, D. Watkins, N. Starling, I. Chau, A. Sadanandam, D. Cunningham, and M. Gerlinger. Genomic and transcriptomic determinants of therapy resistance and immune landscape evolution during anti-egfr treatment in colorectal cancer. *Cancer Cell*, 36(1):35–50 e9, 2019. 2

[7] A. M. Nicolas, M. Pesic, E. Engel, P. K. Ziegler, M. Diefenhardt, K. B. Kennel, F. Buettner, C. Conche, V. Petrocelli, E. Elwakeel, A. Weigert, A. Zinoveva, M. Fleischmann, B. Haupl, C. Karakutuk, H. Bohnenberger, M. H. Mosa, L. Kaderali, J. Gaedcke, M. Ghadimi, F. Rodel, M. C. Arkan, T. Oellerich, C. Rodel, E. Fokas, and F. R. Greten. Inflammatory fibroblasts mediate resistance to neoadjuvant therapy in rectal cancer. *Cancer Cell*, 40(2):168–184 e13, 2022. 2

[8] H. Clevers. Modeling development and disease with organoids. *Cell*, 165(7):1586–1597, 2016. 2, 11

[9] G. Vlachogiannis, S. Hedayat, A. Vatsiou, Y. Jamin, J. Fernandez-Mateos, K. Khan, A. Lampis, K. Eason, I. Huntingford, R. Burke, M. Rata, D. M. Koh, N. Tunariu, D. Collins, S. Hulkki-Wilson, C. Ragulan, I. Spiteri, S. Y. Moorcraft, I. Chau, S. Rao, D. Watkins, N. Fotiadis, M. Bali, M. Darvish-Damavandi, H. Lote, Z. Eltahir, E. C. Smyth, R. Begum, P. A. Clarke, J. C. Hahne, M. Dowsett, J. de Bono, P. Workman, A. Sadanandam, M. Fassan, O. J. Sansom, S. Eccles, N. Starling, C. Braconi, A. Sottoriva, S. P. Robinson, D. Cunningham, and N. Valeri. Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science*, 359(6378):920–926, 2018. 2

[10] A. Letai, P. Bhola, and A. L. Welm. Functional precision oncology: Testing tumors with drugs to identify vulnerabilities and novel combinations. *Cancer Cell*, 40(1):26–35, 2017. 2, 6

[11] K. Yuki, N. Cheng, M. Nakano, and C. J. Kuo. Organoid models of tumor immunology. *Trends Immunol*, 41(8):652–664, 2020. 2

[12] M. van de Wetering, H. E. Francies, J. M. Francis, G. Bounova, F. Iorio, A. Pronk, W. van Houdt, J. van Gorp, A. Taylor-Weiner, L. Kester, A. McLaren-Douglas, J. Blokker, S. Jaksani, S. Bartfeld, R. Volckman, P. van Sluis, V. S. Li, S. Seepo, C. Sekhar Pedamallu, K. Cibulskis, S. L. Carter, A. McKenna, M. S. Lawrence, L. Lichtenstein, C. Stewart, J. Koster, R. Versteeg, A. van Oudenaarden, J. Saez-Rodriguez, R. G. Vries, G. Getz, L. Wessels, M. R. Stratton, U. McDermott, M. Meyerson, M. J. Garnett, and H. Clevers. Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell*, 161(4):933–45, 2015. 2, 13

[13] X. Qin and C. J. Tape. Deciphering organoids: High-dimensional analysis of biomimetic cultures. *Trends Biotechnol*, 39(8):774–787, 2021. 2, 10, 11

[14] X. Qin, J. Sufi, P. Vlckova, P. Kyriakidou, S. E. Acton, V. S. W. Li, M. Nitz, and C. J. Tape. Cell-type-specific signaling networks in heterocellular organoids. *Nat Methods*, 17(3):335–342, 2020. 2

[15] J. Sufi, X. Qin, F. C. Rodriguez, Y. J. Bu, P. Vlckova, M. R. Zapatero, M. Nitz, and C. J. Tape. Multiplexed single-cell analysis of organoid signaling networks. *Nat Protoc*, 16(10):4897–4918, 2021. 2, 13

[16] E. De Vlieghere, F. Gremonprez, L. Verset, L. Marien, C. J. Jones, B. De Craene, G. Berx, B. Descamps, C. Vanhove, J. P. Remon, W. Ceelen, P. Demetter, M. Bracke, B. G. De Geest, and O. De Wever. Tumor-environment biomimetics delay peritoneal metastasis formation by deceiving and redirecting disseminated cancer cells. *Biomaterials*, 54:148–57, 2015. 2, 13

[17] E. De Jaeghere, E. De Vlieghere, J. Van Hoorick, S. Van Vlierberghe, G. Wagemans, L. Pieters, E. Melsens, M. Praet, J. Van Dorpe, M. N. Boone, R. Ghobeira, N. De Geyter, M. Bracke, C. Vanhove, S. Neyt, G. Berx, B. G. De Geest, P. Dubruel, H. Declercq, W. Ceelen, and O. De Wever. Heterocellular 3d scaffolds as biomimetic to recapitulate the tumor microenvironment of peritoneal metastases in vitro and in vivo. *Biomaterials*, 158:95–105, 2018. 2, 13

[18] E. R. Zunder, R. Finck, G. K. Behbehani, A. D. Amir el, S. Krishnaswamy, V. D. Gonzalez, C. G. Lorang, Z. Bjornson, M. H. Spitzer, B. Bodenmiller, W. J. Fantl, D. Pe'er, and G. P. Nolan. Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm. *Nat Protoc*, 10(2):316–33, 2015. 2, 14

[19] Kevin R. Moon, Jay S. Stanley, Daniel Burkhardt, David van Dijk, Guy Wolf, and Smita Krishnaswamy. Manifold learning-based methods for analyzing single-cell RNA-sequencing data. *Current Opinion in Systems Biology*, 7:36–46, February 2018. ISSN 24523100. doi: 10.1016/j.coisb.2017.12.008. 3, 20

[20] Kevin R. Moon, David van Dijk, Zheng Wang, Scott Gigante, Daniel B. Burkhardt, William S. Chen, Kristina Yim, Antonia van den Elzen, Matthew J. Hirn, Ronald R. Coifman, Natalia B. Ivanova, Guy Wolf, and Smita Krishnaswamy. Visualizing structure and transitions in high-dimensional biological data. *Nat Biotechnol*, 37(12):1482–1492, 2019. 4, 17, 20, 21

[21] William S. Chen, Nevena Zivanovic, David van Dijk, Guy Wolf, Bernd Bodenmiller, and Smita Krishnaswamy. Uncovering axes of variation among single-cell cancer specimens. *Nat Methods*, 17(3): 302–310, March 2020. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-019-0689-z. 3, 14, 21

[22] Alexander Tong, Guillaume Huguet, Amine Natik, Kincaid MacDonald, Manik Kuchroo, Ronald Coifman, Guy Wolf, and Smita Krishnaswamy. Diffusion Earth Mover's Distance and Distribution Embeddings. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 10336–10346. PMLR, 2021. 3, 21

[23] Alexander Tong, Guillaume Huguet, Dennis Shung, Amine Natik, Manik Kuchroo, Guillaume Lajoie, Guy Wolf, and Smita Krishnaswamy. Embedding Signals on Knowledge Graphs with Unbalanced Diffusion Earth Mover's Distance. In *ICASSP*, 2022. 3, 14

[24] Aaron T L Lun, Arianne C Richard, and John C Marioni. Testing for differential abundance in mass cytometry data. *Nat Methods*, 14(7):707–709, July 2017. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth.4295. 3

[25] Chamith Y. Fonseka, Deepak A. Rao, Nikola C. Teslovich, Ilya Korsunsky, Susan K. Hannes, Kamil Slowikowski, Michael F. Gurish, Laura T. Donlin, James A. Lederer, Michael E. Weinblatt, Elena M. Massarotti, Jonathan S. Coblyn, Simon M. Helfgott, Derrick J. Todd, Vivian P. Bykerk, Elizabeth W. Karlson, Joerg Ermann, Yvonne C. Lee, Michael B. Brenner, and Soumya Raychaudhuri. Mixed-effects association of single cells identifies an expanded effector CD4+ T cell subset in rheumatoid arthritis. *Science Translational Medicine*, 10(463):eaaq0305, October 2018. ISSN 1946-6242. doi: 10.1126/scitranslmed.aaq0305.

[26] Belinda Phipson, Choon Boon Sim, Enzo R Porrello, Alex W Hewitt, Joseph Powell, and Alicia Oshlack. propeller: testing for differences in cell type proportions in single cell data. *Bioinformatics*, 38(20):4720–4726, October 2022. ISSN 1367-4803. doi: 10.1093/bioinformatics/btac582. URL `https://doi.org/10.1093/bioinformatics/btac582`. 3

[27] Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, September 2020. 4, 17, 20

[28] Laurens van der Maaten and Geoffrey E Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 2008. 4, 17

[29] ANNOY library. `https://github.com/spotify/annoy`. Accessed: 2017-08-01. 4

[30] F. Alexander Wolf, Philipp Angerer, and Fabian J. Theis. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol*, 19(1):15, December 2018. ISSN 1474-760X. doi: 10.1186/s13059-017-1382-0. 4

[31] Piotr Indyk and Nitin Thaper. Fast image retrieval via embeddings. In *3rd International Workshop on Statistical and Computational Theories of Vision*, 2003. 4, 16, 19

[32] Mayur Datar, Nicole Immorlica, Piotr Indyk, and Vahab S. Mirrokni. Locality-sensitive hashing scheme based on p-stable distributions. In *Proceedings of the Twentieth Annual Symposium on Computational Geometry - SCG '04*, page 253, Brooklyn, New York, USA, 2004. ACM Press. ISBN 978-1-58113-885-6. doi: 10.1145/997817.997857. 4

[33] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics*, 2(1-2): 83–97, March 1955. ISSN 00281441, 19319193. doi: 10.1002/nav.3800020109. 4, 16, 21

[34] Marco Cuturi. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In *Advances in Neural Information Processing Systems 26*, pages 2292–2300, 2013. 4, 16, 21

[35] Y. Pommier. Drugging topoisomerases: lessons and challenges. *ACS Chem Biol*, 8(1):82–95, 2013. 4

[36] D. V. Santi, C. S. McHenry, and H. Sommer. Mechanism of interaction of thymidylate synthetase with 5-fluorodeoxyuridylate. *Biochemistry*, 13(3):471–81, 1974. 4

[37] P. M. Bruno, Y. Liu, G. Y. Park, J. Murai, C. E. Koch, T. J. Eisen, J. R. Pritchard, Y. Pommier, S. J. Lippard, and M. T. Hemann. A subset of platinum-containing chemotherapeutic agents kills cells by inducing ribosome biogenesis stress. *Nat Med*, 23(4):461–471, 2017. 4, 7

[38] G. K. Behbehani, S. C. Bendall, M. R. Clutter, W. J. Fantl, and G. P. Nolan. Single-cell mass cytometry adapted to measurements of the cell cycle. *Cytometry A*, 81(7):552–66, 2012. 4

[39] M. A. Rapsomaniki, X. K. Lun, S. Woerner, M. Laumanns, B. Bodenmiller, and M. R. Martinez. Cellcycletracer accounts for cell cycle and volume in mass cytometry data. *Nat Commun*, 9(1):632, 2018. 4

[40] M. R. Middleton, E. Dean, T. R. J. Evans, G. I. Shapiro, J. Pollard, B. S. Hendriks, M. Falk, I. Diaz-Padilla, and R. Plummer. Phase 1 study of the atr inhibitor berzosertib (formerly m6620, vx-970) combined with gemcitabine +/- cisplatin in patients with advanced solid tumours. *Br J Cancer*, 125 (4):510–519, 2021. 7

[41] A. Alvarez-Varela, L. Novellasdemunt, F. M. Barriga, X. Hernando-Momblona, A. Canellas-Socias, S. Cano-Crespo, M. Sevillano, C. Cortina, D. Stork, C. Morral, G. Turon, F. Slebe, L. Jimenez-Gracia, G. Caratu, P. Jung, G. Stassi, H. Heyn, D. V. F. Tauriello, L. Mateo, S. Tejpar, E. Sancho, C. Stephan-Otto Attolini, and E. Batlle. Mex3a marks drug-tolerant persister colorectal cancer cells that mediate relapse after chemotherapy. *Nat Cancer*, 2022. 9

[42] S. Yui, L. Azzolin, M. Maimets, M. T. Pedersen, R. P. Fordham, S. L. Hansen, H. L. Larsen, J. Guiu, M. R. P. Alves, C. F. Rundsten, J. V. Johansen, Y. Li, C. D. Madsen, T. Nakamura, M. Watanabe, O. H. Nielsen, P. J. Schweiger, S. Piccolo, and K. B. Jensen. Yap/taz-dependent reprogramming of colonic epithelium links ecm remodeling to tissue regeneration. *Cell Stem Cell*, 22(1):35–49 e7, 2018. 9

[43] V. Veninga and E. E. Voest. Tumor organoids: Opportunities and challenges to guide precision medicine. *Cancer Cell*, 39(9):1190–1201, 2021. 10

[44] N. Gavert, Y. Zwang, R. Weiser, O. Greenberg, S. Halperin, O. Jacobi, G. Mallel, O. Sandler, A. J. Berger, E. Stossel, D. Rotin, A. Grinshpun, I. Kamer, J. Bar, G. Pines, D. Saidian, I. Bar, S. Golan, E. Rosenbaum, A. Nadu, E. Ben-Ami, R. Weitzen, H. Nechushtan, T. Golan, B. Brenner, A. Nissan, O. Margalit, D. Hershkovitz, G. Lahat, and R. Straussman. Ex vivo organotypic cultures for synergistic therapy prioritization identify patient-specific responses to combined mek and src inhibition in colorectal cancer. *Nat Cancer*, 2022. 10

[45] C. Pauli, B. D. Hopkins, D. Prandi, R. Shaw, T. Fedrizzi, A. Sboner, V. Sailer, M. Augello, L. Puca, R. Rosati, T. J. McNary, Y. Churakova, C. Cheung, J. Triscott, D. Pisapia, R. Rao, J. M. Mosquera, B. Robinson, B. M. Faltas, B. E. Emerling, V. K. Gadi, B. Bernard, O. Elemento, H. Beltran, F. Demichelis, C. J. Kemp, C. Grandori, L. C. Cantley, and M. A. Rubin. Personalized in vitro and in vivo cancer models to guide precision medicine. *Cancer Discov*, 7(5):462–477, 2017.

[46] K. P. Guillen, M. Fujita, A. J. Butterfield, S. D. Scherer, M. H. Bailey, Z. Chu, Y. S. DeRose, L. Zhao, E. Cortes-Sanchez, C. H. Yang, J. Toner, G. Wang, Y. Qiao, X. Huang, J. A. Greenland, J. M. Vahrenkamp, D. H. Lum, R. E. Factor, E. W. Nelson, C. B. Matsen, J. M. Poretta, R. Rosenthal, A. C. Beck, S. S. Buys, C. Vaklavas, J. H. Ward, R. L. Jensen, K. B. Jones, Z. Li, S. Oesterreich, L. E. Dobrolecki, S. S. Pathi, X. Y. Woo, K. C. Berrett, M. E. Wadsworth, J. H. Chuang, M. T. Lewis, G. T. Marth, J. Gertz, K. E. Varley, B. E. Welm, and A. L. Welm. A human breast cancer-derived xenograft and organoid platform for drug discovery and precision oncology. *Nat Cancer*, 3(2):232–250, 2022. 10

[47] S. Raghavan, P. S. Winter, A. W. Navia, H. L. Williams, A. DenAdel, K. E. Lowder, J. Galvez-Reyes, R. L. Kalekar, N. Mulugeta, K. S. Kapner, M. S. Raghavan, A. A. Borah, N. Liu, S. A. Vayrynen, A. D. Costa, R. W. S. Ng, J. Wang, E. K. Hill, D. Y. Ragon, L. K. Brais, A. M. Jaeger, L. F. Spurr, Y. Y. Li, A. D. Cherniack, M. A. Booker, E. F. Cohen, M. Y. Tolstorukov, I. Wakiro, A. Rotem, B. E. Johnson, J. M. McFarland, E. T. Sicinska, T. E. Jacks, R. J. Sullivan, G. I. Shapiro, T. E. Clancy, K. Perez, D. A. Rubinson, K. Ng, J. M. Cleary, L. Crawford, S. R. Manalis, J. A. Nowak, B. M. Wolpin, W. C. Hahn, A. J. Aguirre, and A. K. Shalek. Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. *Cell*, 184(25):6119–6137 e26, 2021. 11

[48] D. Hanahan. Hallmarks of cancer: New dimensions. *Cancer Discov*, 12(1):31–46, 2022. 11

[49] D. B. Burkhardt, B. P. San Juan, J. G. Lock, S. Krishnaswamy, and C. L. Chaffer. Mapping phenotypic plasticity upon the cancer cell state landscape using manifold learning. *Cancer Discov*, pages OF1–OF13, 2022. 11

[50] J. M. Chan, S. Zaidi, J. R. Love, J. L. Zhao, M. Setty, K. M. Wadosky, A. Gopalan, Z. N. Choo, S. Persad, J. Choi, J. LaClair, K. E. Lawrence, O. Chaudhary, T. Xu, I. Masilionis, I. Linkov, S. Wang, C. Lee, A. Barlas, M. J. Morris, L. Mazutis, R. Chaligne, Y. Chen, D. W. Goodrich, W. R. Karthaus, D. Pe'er, and C. L. Sawyers. Lineage plasticity in prostate cancer depends on jak/stat inflammatory signaling. *Science*, page eabn0478, 2022. 11

[51] H. Tiriac, P. Belleau, D. D. Engle, D. Plenker, A. Deschenes, T. D. D. Somerville, F. E. M. Froeling, R. A. Burkhart, R. E. Denroche, G. H. Jang, K. Miyabayashi, C. M. Young, H. Patel, M. Ma, J. F. LaComb, R. L. D. Palmaira, A. A. Javed, J. C. Huynh, M. Johnson, K. Arora, N. Robine, M. Shah, R. Sanghvi, A. B. Goetz, C. Y. Lowder, L. Martello, E. Driehuis, N. LeComte, G. Askan, C. A. Iacobuzio-Donahue, H. Clevers, L. D. Wood, R. H. Hruban, E. Thompson, A. J. Aguirre, B. M. Wolpin, A. Sasson, J. Kim, M. Wu, J. C. Bucobo, P. Allen, D. V. Sejpal, W. Nealon, J. D. Sullivan, J. M. Winter, P. A. Gimotty, J. L. Grem, D. J. DiMaio, J. M. Buscaglia, P. M. Grandgenett, J. R. Brody, M. A. Hollingsworth, G. M. O'Kane, F. Notta, E. Kim, J. M. Crawford, C. Devoe, A. Ocean, C. L. Wolfgang, K. H. Yu, E. Li, C. R. Vakoc, B. Hubert, S. E. Fischer, J. M. Wilson, R. Moffitt, J. Knox, A. Krasnitz, S. Gallinger, and D. A. Tuveson. Organoid profiling identifies common responders to chemotherapy in pancreatic cancer. *Cancer Discov*, 8(9):1112–1129, 2018. 11

[52] P. East, G. P. Kelly, D. Biswas, M. Marani, D. C. Hancock, T. Creasy, K. Sachsenmeier, C. Swanton, T. RACERx consortium, J. Downward, and S. de Carne Trecesson. Ras oncogenic activity predicts response to chemotherapy and outcome in lung adenocarcinoma. *Nat Commun*, 13(1):5632, 2022. 11

[53] Abel Sousa, Aurelien Dugourd, Danish Memon, Borgthor Petursson, Evangelia Petsalaki, Julio Saez-Rodriguez, and Pedro Beltrao. Pan-cancer landscape of protein activities identifies drivers of signalling dysregulation and patient survival. *bioRxiv*, 2021. doi: 10.1101/2021.06.09.447741. 11

[54] P. Jaaks, E. A. Coker, D. J. Vis, O. Edwards, E. F. Carpenter, S. M. Leto, L. Dwane, F. Sassi, H. Lightfoot, S. Barthorpe, D. van der Meer, W. Yang, A. Beck, T. Mironenko, C. Hall, J. Hall, I. Mali, L. Richardson, C. Tolley, J. Morris, F. Thomas, E. Lleshi, N. Aben, C. H. Benes, A. Bertotti, L. Trusolino, L. Wessels, and M. J. Garnett. Effective drug combinations in breast, colon and pancreatic cancer cells. *Nature*, 603(7899):166–173, 2022. 11

[55] Yu Takahashi, Shintaro Sato, Yosuke Kurashima, Tomohisa Yamamoto, Shiho Kurokawa, Yoshikazu Yuki, Naoki Takemura, Satoshi Uematsu, Chen-Yi Lai, Makoto Otsu, Hiroshi Matsuno, Hideki Osawa, Tsunekazu Mizushima, Junichi Nishimura, Mikio Hayashi, Takayuki Yamaguchi, and Hiroshi Kiyono. A Refined Culture System for Human Induced Pluripotent Stem Cell-Derived Intestinal Epithelial Organoids. *Stem Cell Reports*, 10(1):314–328, January 2018. ISSN 2213-6711. doi: 10.1016/j.stemcr.2017.11.004. URL `https://www.sciencedirect.com/science/article/pii/S2213671117304903`. 13

[56] Yuge Ji, Mohammad Lotfollahi, F. Alexander Wolf, and Fabian J. Theis. Machine learning for perturbational single-cell omics. *Cell Systems*, 12(6):522–537, 2021. ISSN 24054712. doi: 10.1016/j.cels.2021.05.016. 14

[57] Dongju Shin, Wookjae Lee, Ji Hyun Lee, and Duhee Bang. Multiplexed single-cell RNA-seq via transient barcoding for simultaneous expression profiling of various drug perturbations. *Sci. Adv.*, 5 (5):eaav2249, May 2019. ISSN 2375-2548. doi: 10.1126/sciadv.aav2249. 14

[58] Atray Dixit, Oren Parnas, Biyu Li, Jenny Chen, Charles P. Fulco, Livnat Jerby-Arnon, Nemanja D. Marjanovic, Danielle Dionne, Tyler Burks, Raktima Raychowdhury, Britt Adamson, Thomas M. Norman, Eric S. Lander, Jonathan S. Weissman, Nir Friedman, and Aviv Regev. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell*, 167(7): 1853–1866.e17, December 2016. ISSN 00928674. doi: 10.1016/j.cell.2016.11.038. 14

[59] Bharath K. Sriperumbudur, Kenji Fukumizu, Arthur Gretton, Bernhard Schölkopf, and Gert R. G. Lanckriet. On the empirical estimation of integral probability metrics. *Electron. J. Statist.*, 6(none), January 2012. ISSN 1935-7524. doi: 10.1214/12-EJS722. 15

[60] Shibing Chen and Alessio Figalli. Partial W2, regularity for optimal transport maps. *Journal of Functional Analysis*, 272(11):4588–4605, June 2017. ISSN 00221236. doi: 10.1016/j.jfa.2017.02.025. 15

[61] Luis A Caffarelli. The Regularity of Mappings with Convex Potential. *J. Amer. Math. Soc.*, 5(1), 1992. 15

[62] Gabriel Peyré and Marco Cuturi. *Computational Optimal Transport*. arXiv:1803.00567, 2019. 15, 16

[63] Leonid G Hanin. Kantorovich-Rubinstein Norm and Its Application in the Theory of Lipschitz Spaces. *Proc. AMS*, 115(2):345–352, 1992. 16

[64] Sheida Nabavi, Daniel Schmolze, Mayinuer Maitituoheti, Sadhika Malladi, and Andrew H. Beck. EMDomics: A robust and powerful method for the identification of genes differentially expressed between heterogeneous classes. *Bioinformatics*, 32(4):533–541, February 2016. ISSN 1367-4803. doi: 10.1093/bioinformatics/btv634. 16

[65] Soheil Kolouri, Yang Zou, and Gustavo K. Rohde. Sliced Wasserstein Kernels for Probability Distributions. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5258–5267, Las Vegas, NV, USA, June 2016. IEEE. ISBN 978-1-4673-8851-1. doi: 10.1109/CVPR.2016.568. 16

[66] Soheil Kolouri, Kimia Nadjahi, Umut Simsekli, Roland Badeau, and Gustavo Rohde. Generalized Sliced Wasserstein Distances. In *Advances in Neural Information Processing Systems 32*, pages 261–272, 2019. 16

[67] Tam Le, Makoto Yamada, Kenji Fukumizu, and Marco Cuturi. Tree-Sliced Variants of Wasserstein Distances. In *Advances in Neural Information Processing Systems 33*, 2019. 16, 19

[68] Arturs Backurs, Yihe Dong, Piotr Indyk, Ilya Razenshteyn, and Tal Wagner. Scalable Nearest Neighbor Search for Optimal Transport. *ICML*, 2020. 16, 19, 20

[69] Yair Bartal. Probabilistic Approximation of Metric Spaces and its Algorithmic Applications. In *In 37th Annual Symposium on Foundations of Computer Science*, pages 184–193, 1996. 16

[70] Moses S Charikar. Similarity Estimation Techniques from Rounding Algorithms. In *STOC*, 2002. 16, 19

[71] The FlowCAP Consortium, The DREAM Consortium, Nima Aghaeepour, Greg Finak, Holger Hoos, Tim R Mosmann, Ryan Brinkman, Raphael Gottardo, and Richard H Scheuermann. Critical assessment of automated flow cytometry data analysis techniques. *Nat Methods*, 10(3):228–238, March 2013. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth.2365. 17, 19

[72] Chris P. Verschoor, Alina Lelic, Jonathan L. Bramson, and Dawn M. E. Bowdish. An Introduction to Automated Flow Cytometry Gating Tools and Their Implementation. *Front. Immunol.*, 6, July 2015. ISSN 1664-3224. doi: 10.3389/fimmu.2015.00380.

[73] Florian Mair, Felix J. Hartmann, Dunja Mrdjen, Vinko Tosevski, Carsten Krieg, and Burkhard Becher. The end of gating? An introduction to automated analysis of high dimensional cytometry data: Highlights. *Eur. J. Immunol.*, 46(1):34–43, January 2016. ISSN 00142980. doi: 10.1002/eji.201545774.

[74] Andrea Cossarizza, Hyun-Dong Chang, Andreas Radbruch, Andreas Acs, Dieter Adam, Sabine Adam-Klages, William W. Agace, Nima Aghaeepour, Mübeccel Akdis, Matthieu Allez, Larissa Nogueira Almeida, Giorgia Alvisi, Graham Anderson, Immanuel Andrä, Francesco Annunziato, Achille Anselmo, Petra Bacher, Cosima T. Baldari, Sudipto Bari, Vincenzo Barnaba, Joana Barros-Martins, Luca Battistini, Wolfgang Bauer, Sabine Baumgart, Nicole Baumgarth, Dirk Baumjohann, Bianka Baying, Mary Bebawy, Burkhard Becher, Wolfgang Beisker, Vladimir Benes, Rudi Beyaert, Alfonso Blanco, Dominic A. Boardman, Christian Bogdan, Jessica G. Borger, Giovanna Borsellino, Philip E. Boulais, Jolene A. Bradford, Dirk Brenner, Ryan R. Brinkman, Anna E. S. Brooks, Dirk H. Busch, Martin Büscher, Timothy P. Bushnell, Federica Calzetti, Garth Cameron, Ilenia Cammarata, Xuetao Cao, Susanna L. Cardell, Stefano Casola, Marco A. Cassatella, Andrea Cavani, Antonio Celada, Lucienne Chatenoud, Pratip K. Chattopadhyay, Sue Chow, Eleni Christakou, Luka Čičin-Šain, Mario Clerici, Federico S. Colombo, Laura Cook, Anne Cooke, Andrea M. Cooper, Alexandra J. Corbett, Antonio Cosma, Lorenzo Cosmi, Pierre G. Coulie, Ana Cumano, Ljiljana Cvetkovic, Van Duc Dang, Chantip Dang-Heine, Martin S. Davey, Derek Davies, Sara De Biasi, Genny Del Zotto, Gelo Victoriano Dela Cruz, Michael Delacher, Silvia Della Bella, Paolo Dellabona, Günnur Deniz, Mark Dessing,

James P. Di Santo, Andreas Diefenbach, Francesco Dieli, Andreas Dolf, Thomas Dörner, Regine J. Dress, Diana Dudziak, Michael Dustin, Charles-Antoine Dutertre, Friederike Ebner, Sidonia B. G. Eckle, Matthias Edinger, Pascale Eede, Götz R.A. Ehrhardt, Marcus Eich, Pablo Engel, Britta Engelhardt, Anna Erdei, Charlotte Esser, Bart Everts, Maximilien Evrard, Christine S. Falk, Todd A. Fehniger, Mar Felipo-Benavent, Helen Ferry, Markus Feuerer, Andrew Filby, Kata Filkor, Simon Fillatreau, Marie Follo, Irmgard Förster, John Foster, Gemma A. Foulds, Britta Frehse, Paul S. Frenette, Stefan Frischbutter, Wolfgang Fritzsche, David W. Galbraith, Anastasia Gangaev, Natalio Garbi, Brice Gaudilliere, Ricardo T. Gazzinelli, Jens Geginat, Wilhelm Gerner, Nicholas A. Gherardin, Kamran Ghoreschi, Lara Gibellini, Florent Ginhoux, Keisuke Goda, Dale I. Godfrey, Christoph Goettlinger, Jose M. González-Navajas, Carl S. Goodyear, Andrea Gori, Jane L. Grogan, Daryl Grummitt, Andreas Grützkau, Claudia Haftmann, Jonas Hahn, Hamida Hammad, Günter Hämmerling, Leo Hansmann, Goran Hansson, Christopher M. Harpur, Susanne Hartmann, Andrea Hauser, Anja E. Hauser, David L. Haviland, David Hedley, Daniela C. Hernández, Guadalupe Herrera, Martin Herrmann, Christoph Hess, Thomas Höfer, Petra Hoffmann, Kristin Hogquist, Tristan Holland, Thomas Höllt, Rikard Holmdahl, Pleun Hombrink, Jessica P. Houston, Bimba F. Hoyer, Bo Huang, Fang-Ping Huang, Johanna E. Huber, Jochen Huehn, Michael Hundemer, Christopher A. Hunter, William Y. K. Hwang, Anna Iannone, Florian Ingelfinger, Sabine M Ivison, Hans-Martin Jäck, Peter K. Jani, Beatriz Jávega, Stipan Jonjic, Toralf Kaiser, Tomas Kalina, Thomas Kamradt, Stefan H. E. Kaufmann, Baerbel Keller, Steven L. C. Ketelaars, Ahad Khalilnezhad, Srijit Khan, Jan Kisielow, Paul Klenerman, Jasmin Knopf, Hui-Fern Koay, Katja Kobow, Jay K. Kolls, Wan Ting Kong, Manfred Kopf, Thomas Korn, Katharina Kriegsmann, Hendy Kristyanto, Thomas Kroneis, Andreas Krueger, Jenny Kühne, Christian Kukat, Désirée Kunkel, Heike Kunze-Schumacher, Tomohiro Kurosaki, Christian Kurts, Pia Kvistborg, Immanuel Kwok, Jonathan Landry, Olivier Lantz, Paola Lanuti, Francesca LaRosa, Agnès Lehuen, Salomé LeibundGut-Landmann, Michael D. Leipold, Leslie Y.T. Leung, Megan K. Levings, Andreia C. Lino, Francesco Liotta, Virginia Litwin, Yanling Liu, Hans-Gustaf Ljunggren, Michael Lohoff, Giovanna Lombardi, Lilly Lopez, Miguel López-Botet, Amy E. Lovett-Racke, Erik Lubberts, Herve Luche, Burkhard Ludewig, Enrico Lugli, Sebastian Lunemann, Holden T. Maecker, Laura Maggi, Orla Maguire, Florian Mair, Kerstin H. Mair, Alberto Mantovani, Rudolf A. Manz, Aaron J. Marshall, Alicia Martínez-Romero, Glòria Martrus, Ivana Marventano, Wlodzimierz Maslinski, Giuseppe Matarese, Anna Vittoria Mattioli, Christian Maueröder, Alessio Mazzoni, James McCluskey, Mairi McGrath, Helen M. McGuire, Iain B. McInnes, Henrik E. Mei, Fritz Melchers, Susanne Melzer, Dirk Mielenz, Stephen D. Miller, Kingston H.G. Mills, Hans Minderman, Jenny Mjösberg, Jonni Moore, Barry Moran, Lorenzo Moretta, Tim R. Mosmann, Susann Müller, Gabriele Multhoff, Luis Enrique Muñoz, Christian Münz, Toshinori Nakayama, Milena Nasi, Katrin Neumann, Lai Guan Ng, Antonia Niedobitek, Sussan Nourshargh, Gabriel Núñez, José-Enrique O'Connor, Aaron Ochel, Anna Oja, Diana Ordonez, Alberto Orfao, Eva Orlowski-Oliver, Wenjun Ouyang, Annette Oxenius, Raghavendra Palankar, Isabel Panse, Kovit Pattanapanyasat, Malte Paulsen, Dinko Pavlinic, Livius Penter, Pärt Peterson, Christian Peth, Jordi Petriz, Federica Piancone, Winfried F. Pickl, Silvia Piconese, Marcello Pinti, A. Graham Pockley, Malgorzata Justyna Podolska, Zhiyong Poon, Katharina Pracht, Immo Prinz, Carlo E. M. Pucillo, Sally A. Quataert, Linda Quatrini, Kylie M. Quinn, Helena Radbruch, Tim R. D. J. Radstake, Susann Rahmig, Hans-Peter Rahn, Bartek Rajwa, Gevitha Ravichandran, Yotam Raz, Jonathan A. Rebhahn, Diether Recktenwald, Dorothea Reimer, Caetano Reis e Sousa, Ester B.M. Remmerswaal, Lisa Richter, Laura G. Rico, Andy Riddell, Aja M. Rieger, J. Paul Robinson, Chiara Romagnani, Anna Rubartelli, Jürgen Ruland, Armin Saalmüller, Yvan Saeys, Takashi Saito, Shimon Sakaguchi, Francisco Sala-de Oyanguren, Yvonne Samstag, Sharon Sanderson, Inga Sandrock, Angela

Santoni, Ramon Bellmàs Sanz, Marina Saresella, Catherine Sautes-Fridman, Birgit Sawitzki, Linda Schadt, Alexander Scheffold, Hans U. Scherer, Matthias Schiemann, Frank A. Schildberg, Esther Schimisky, Andreas Schlitzer, Josephine Schlosser, Stephan Schmid, Steffen Schmitt, Kilian Schober, Daniel Schraivogel, Wolfgang Schuh, Thomas Schüler, Reiner Schulte, Axel Ronald Schulz, Sebastian R. Schulz, Cristiano Scottá, Daniel Scott-Algara, David P. Sester, T. Vincent Shankey, Bruno Silva-Santos, Anna Katharina Simon, Katarzyna M. Sitnik, Silvano Sozzani, Daniel E. Speiser, Josef Spidlen, Anders Stahlberg, Alan M. Stall, Natalie Stanley, Regina Stark, Christina Stehle, Tobit Steinmetz, Hannes Stockinger, Yousuke Takahama, Kiyoshi Takeda, Leonard Tan, Attila Tárnok, Gisa Tiegs, Gergely Toldi, Julia Tornack, Elisabetta Traggiai, Mohamed Trebak, Timothy I.M. Tree, Joe Trotter, John Trowsdale, Maria Tsoumakidou, Henning Ulrich, Sophia Urbanczyk, Willem Veen, Maries Broek, Edwin Pol, Sofie Van Gassen, Gert Van Isterdael, René A.W. Lier, Marc Veldhoen, Salvador Vento-Asturias, Paulo Vieira, David Voehringer, Hans-Dieter Volk, Anouk Borstel, Konrad Volkmann, Ari Waisman, Rachael V. Walker, Paul K. Wallace, Sa A. Wang, Xin M. Wang, Michael D. Ward, Kirsten A Ward-Hartstonge, Klaus Warnatz, Gary Warnes, Sarah Warth, Claudia Waskow, James V. Watson, Carsten Watzl, Leonie Wegener, Thomas Weisenburger, Annika Wiedemann, Jürgen Wienands, Anneke Wilharm, Robert John Wilkinson, Gerald Willimsky, James B. Wing, Rieke Winkelmann, Thomas H. Winkler, Oliver F. Wirz, Alicia Wong, Peter Wurst, Jennie H. M. Yang, Juhao Yang, Maria Yazdanbakhsh, Liping Yu, Alice Yue, Hanlin Zhang, Yi Zhao, Susanne Maria Ziegler, Christina Zielinski, Jakob Zimmermann, and Arturo Zychlinsky. Guidelines for the use of flow cytometry and cell sorting in immunological studies (second edition). *Eur. J. Immunol.*, 49(10):1457–1973, October 2019. ISSN 0014-2980, 1521-4141. doi: 10.1002/eji.201970107. 19

[75] Steven N. Evans and Frederick A. Matsen. The phylogenetic Kantorovich-Rubinstein metric for environmental sequence samples. *J R Stat Soc Series B Stat Methodol.*, 2012. 20

[76] Ulrike von Luxburg. A tutorial on spectral clustering. *Stat Comput*, 17(4):395–416, December 2007. ISSN 0960-3174, 1573-1375. doi: 10.1007/s11222-007-9033-z. 20

[77] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, and David Cournapeau. Scikit-learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON*, 2011. 21

Table S1: PDO mutations and clinical metadata.

Patient	Mutations	Mutated Genes	MSI	Stage	Location
PDO 05	30	*APC, KRAS, PIK3CA, B2M, TCF7L2, PDGFRA*	No	III	Rectum
PDO 11	41	*APC, KRAS*	No	I	Transverse colon
PDO 21	49	*KRAS, SMAD4, ARID1A, PIK3RI, CTNNB1, NOTCH2*	No	II	Rectum
PDO 23	40	*APC, TP53, SUFU*	No	I	Sigmoid colon
PDO 75	27	*APC, TP53, PTEN, PIK3R1, ALK, ZNF292*	No	I	Rectum
PDO 109	38	*APC, KRAS, SOX9, TP53*	No	III	Sigmoid colon
PDO 141	20	*APC, KRAS, PIK3CA, TP53*	No	III	Sigmoid colon
PDO 27	397	*APC, TP53, B2M, RNF43, ACVR2A, KMT2C, EP300, CEBBP, CCND1, FANCE, FAS, GRIN2A, HDLBP, HNF1A, MSH3, PI3KCB, POLE, SYNE1, TP53BP1, USP9X, ZNF292*	Yes	III	Ascending colon
PDO 99	393	*PIK3CA, SOX9, BRAF, BMPR2, RNF43, MLH1, ACVR2A, AXIN1, CASP8, FAT1, KMT2C, TAG2, TBX3, GPS2, SPEN, AXIN2, BCL9L, FANCA, MSH3, TRAF7, UBR5, ZNF292*	Yes	II	Ascending colon
PDO 216	352	*TP53, PIK3CA, FBXW7, BRAF, ARID1A, ACVR2A, B2M, BMPR2, CD58, RNF43, ZNRF3, KMT2B, KMT2D, PBRM1, BRCA2, NCOR1, KDM6A, GATA3, ASXL1, PTCH1, RASA1, CASP8, TGFBR2, RBM10, BRD7, RPL22, CDKN1B, PPM1D, CUX1, CLI1, DROSHA, FAS, FAT3, FLT4, HNF1A, HNRNPA2B1, HSPG2, LRIG3, MSH6, NAB2, PIM1, QKI, SH2B3, SUFU, XPA*	Yes	III	Transverse colon

Table S2: Drug treatment key for PDO-CAF TOB*is* mass cytometry master dataframe (Ramos, Maria (2022), "Ramos Zapatero et. al (Cancer-Associated Fibroblasts Regulate Patient-Derived Organoid Drug Responses)", Mendeley Data, V1, doi: 10.17632/hc8gxwks3p.1).

Treatment	Label	Conc. Label	Conc. (1)	Conc. (2)	Conc. (3)
SN-38	S	1	1 nM	—	—
	S	2	10 nM	—	—
	S	3	50 nM	—	—
	S	4	100 nM	—	—
5-FU	F	1	200 nM	—	—
	F	2	2 µM	—	—
	F	3	20 µM	—	—
	F	4	200 µM	—	—
Oxaliplatin	O	1	2 nM	—	—
	O	2	20 nM	—	—
	O	3	200 nM	—	—
LGK-974	L	1	1 nM	—	—
	L	2	10 nM	—	—
	L	3	50 nM	—	—
	L	4	5000 nM	—	—
SN-38 (1) + VX-970 (2)	V	1	—	250 nM	—
	VS	2	1 nM	250 nM	—
	VS	3	10 nM	250 nM	—
	VS	4	100 nM	250 nM	—
SN-38 (1) + 5-FU (2) + Cetux. (3)	C	1	—	—	5 nM
	CS	2	10 nM	—	5 nM
	CF	3	—	2 µM	5 nM
	SF	4	10 nM	2 µM	—
	CSF	5	10 nM	2 µM	5 nM
DMSO	DMSO	0	—	—	—
NH_4OH	AH	0	—	—	—
H_2O	H2O	0	—	—	—

Table S3: Mass cytometry antibody panel used in all TOB*is* PDO-CAF experiments.

Extracellular Antigen	**Metal**	**Clone**	**Supplier**
CD326 (EpCAM)	^{113}In	9C4	BioLegend
CHGA	^{170}Er	C-12	Insight Biotechnology
CD90 (THY1)	^{173}Yb	5E10	CST
Intracellular Antigen	**Metal**	**Clone**	**Supplier**
pHistone H3 [S28]	^{89}Y	HTA28	BioLegend
RFP	^{106}Cd	8E5.G7	eBiosciences
mCHERRY	^{110}Cd	16D7	Thermofisher
Vimentin	^{111}Cd	RV202	BD Biosciences
CK18	^{114}Cd	C-04	Abcam
Pan-CK	^{115}In	AE1/AE3	Biolegend UK
GFP	^{116}Cd	5F12.4	eBiosciences
pPDK1 [S241]	^{141}Pr	J66-653.44.22	BD Biosciences
cCaspase 3 [D175]	^{142}Nd	D3E9	CST
Geminin	^{143}Nd	Polyclonal	Santa Cruz
pMEK 1/2 [S221]	^{144}Nd	166F8	CST
pNDRG1 [T346]	^{145}Nd	D98G11	CST
pMKK4 SEK1 [S257]	^{146}Nd	C36C11	CST
pBTK [Y511]	^{147}Sm	24a/BTK	BD Biosciences
pSRC [Y418]	^{148}Nd	SC1T2M3	BD Biosciences
p4E-BP1 [T37/46]	^{149}Sm	236B4	CST
pRB [S807/811]	^{150}Nd	J1112-906	BD Biosciences
pPKCα [T497]	^{151}Eu	K14-984	BD Biosciences
pAKT [T308]	^{152}Sm	J1-223.371	BD Biosciences
pCREB [S133]	^{153}Eu	87G3	CST
pSMAD1/5/9 [S463/465] / [S463/465] / [S465/467]	^{154}Sm	D5B10	CST
pAKT [S473]	^{155}Gd	D9E	CST
pNK-κB p65 [S529]	^{156}Gd	K10-895.12.50	BD Biosciences
pMKK3/pMKK6 [S189/207]	^{157}Gd	D8E9	CST
pP38 [T180/Y182]	^{158}Gd	D3F9	CST
pMAPKAPK2 [T334]	^{159}Tb	27B7	Abcam
pAMPKα [T172]	^{160}Gd	40H9	CST
pBAD [S112]	^{161}Dy	40A9	CST
pHistone H2A.X [S139]	^{162}Dy	D7T2V	CST
pP90RSK [T359]	^{163}Dy	D1E9	CST
p120 Catenin [T310]	^{164}Dy	22/p120 (pT310)	BD Biosciences
β-Catenin [Active]	^{165}Ho	D13A1	CST
pGSK-3β [S9]	^{166}Er	D85E12	CST
pERK1/2 [T202/Y204]	^{167}Er	20A	BD Biosciences
pSMAD2/3 [S465/467] / [S423/425]	^{168}Er	D27F4	CST
PLK1	^{169}Tm	35-206	Thermofisher
pDNAPK [S2056]	^{171}Yb	EPR5670	Abcam
pS6 [S235/236]	^{172}Yb	D57.2.2E	CST
cPARP [D214]	^{174}Yb	D64E10	CST
pCHK1 [S345]	^{175}Lu	133D3	CST
Cyclin B1	^{176}Yb	GNS-11	BD Biosciences

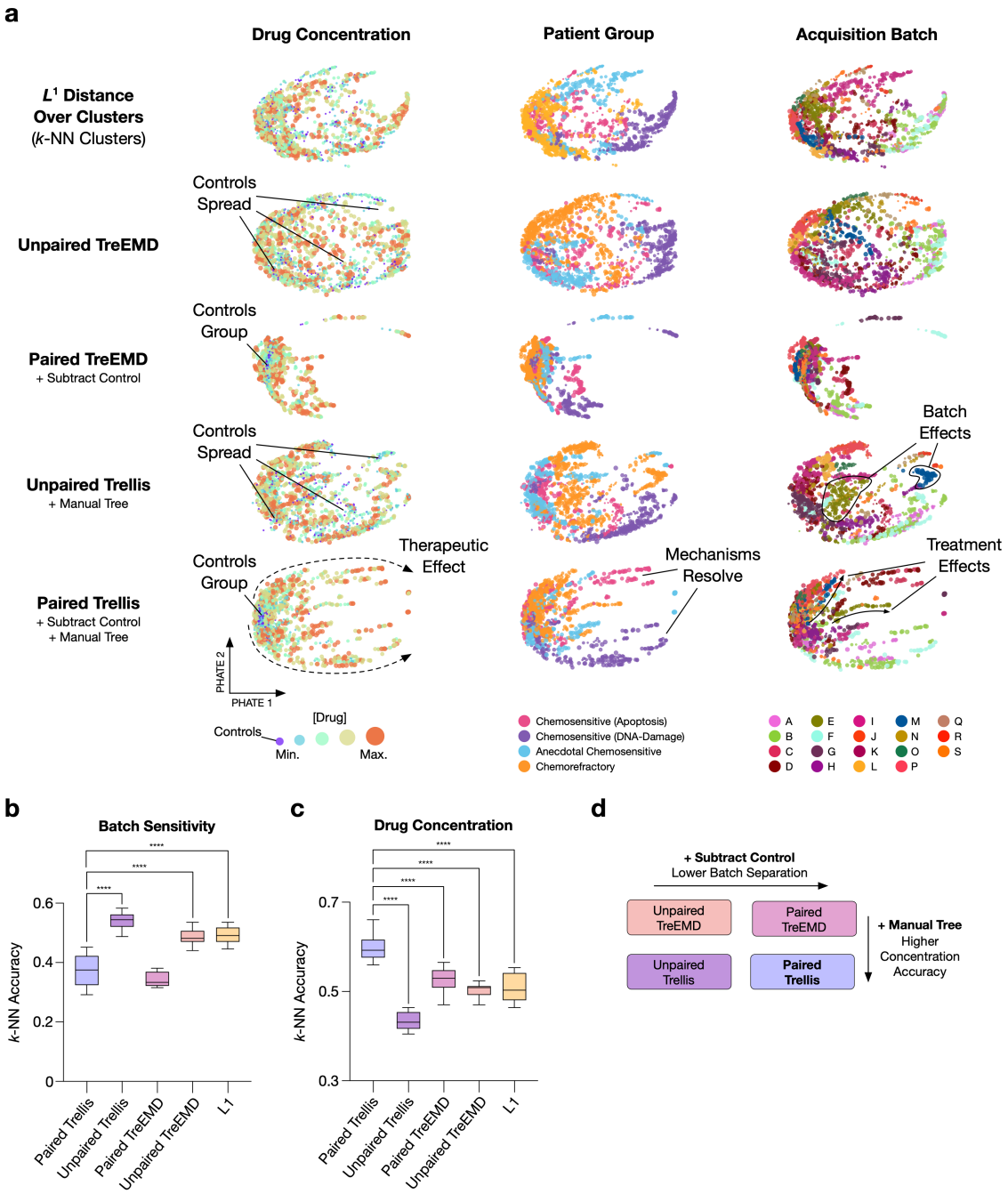

Figure S1: Trellis Ablation Test. a) Comparison of Trellis' ablated algorithm into: L^1 distance over k-NN clusters, Wassertein distance over automatic gating (Unpaired TreEMD), Kantorovich-Rubenstein (KR) norm over automatic gating (Paired TreEMD), Wassertein distance over hierarchical tree partitions of the data by cell-state (Unpaired Trellis), and KR norm hierarchical tree partitions of the data by cell-state (Paired Trellis). **b)** k-NN accuracy score on acquisition batches. A higher k-NN accuracy infers a higher batch separation effect by the method. **c)** k-NN accuracy score on drug concentrations vs controls. Paired Trellis improves drug treatment effect detection. **d)** Schematic representation of the comparison across methods. One-way ANOVA, **** = <0.0001 (n=10).

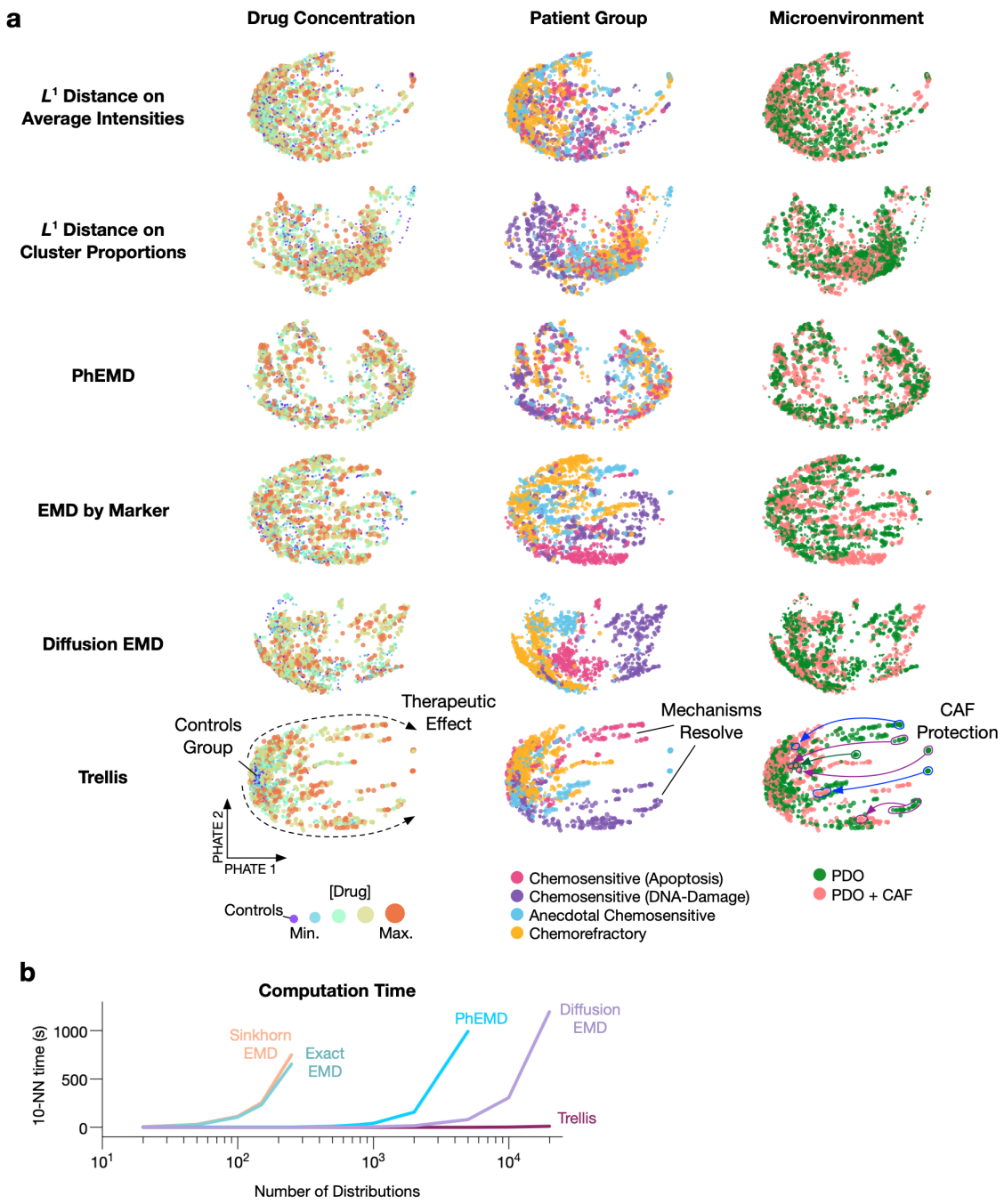

Figure S2: Comparison of Trellis to Alternative Methods. a) Trellis performance compared to existing methods such as L^1 distance of average intensity of the markers, L^1 distance of differential abundance of cells in clusters, PhEMD, EMDs between samples on marker intensities, and Diffusion EMD. Alternative methods fail to capture therapeutic effects and cannot identify CAF protection. **b)** Trellis speed and scalability relative to alternative EMD methods.

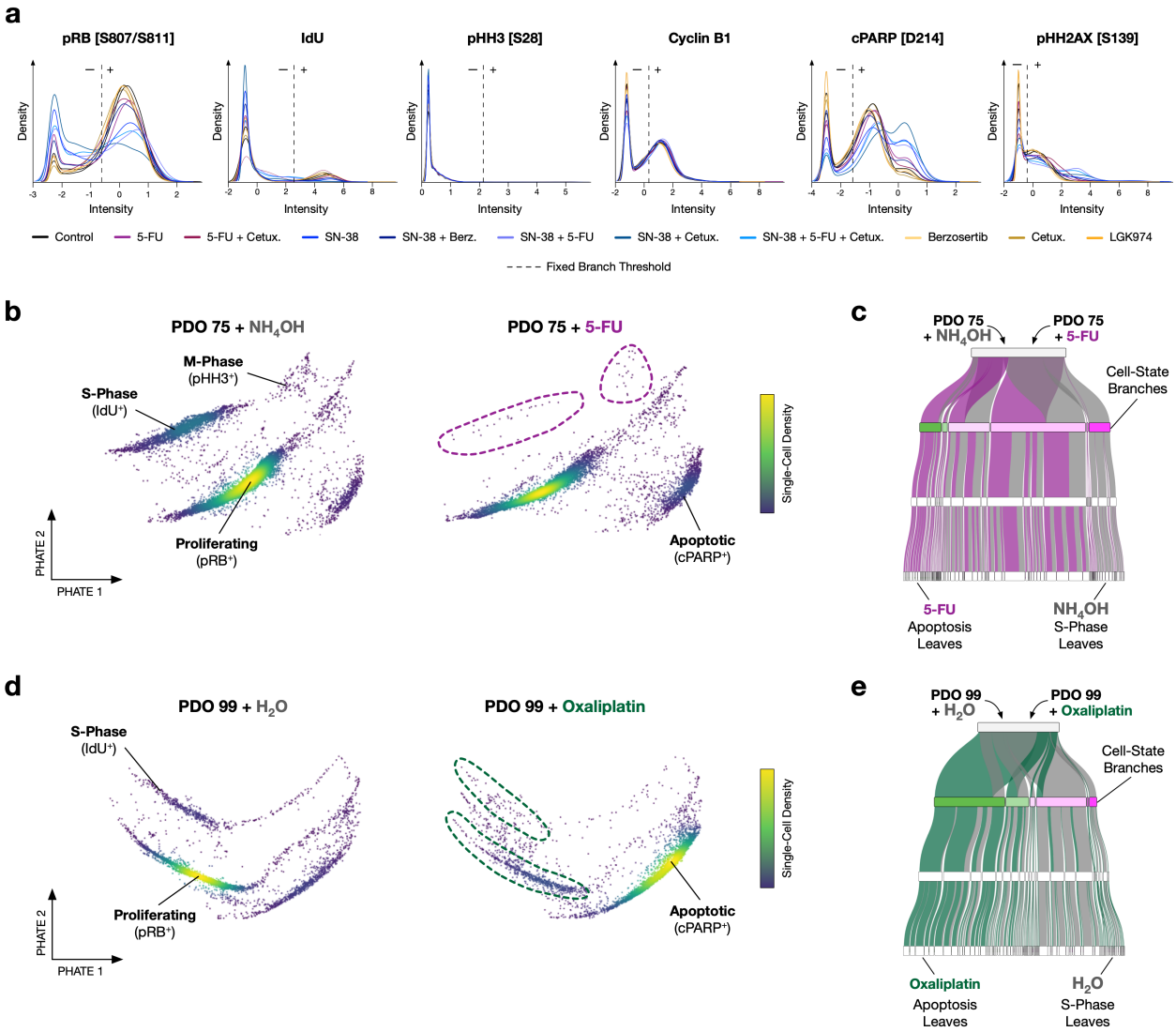

Figure S3: Trellis Detection of PDO Cell-State Drug Responses. a) Trellis cell-state branch thresholds for PDO 21 (batch-mean centered and arcsinh transformed intensities). **b)** Single-cell density PHATEs of PDO 75 treated with NH_4OH vehicle control or 5-FU. **c)** Sankey diagram showing data from b) distributing through the cell-state Trellis layout in Fig. 2 (terminal leaves not shown). **d)** PDO 99 treated with H_2O vehicle control or Oxaliplatin. **e)** Sankey diagram showing data from d) distributing through the cell-state Trellis layout in Fig. 2 (terminal leaves not shown).

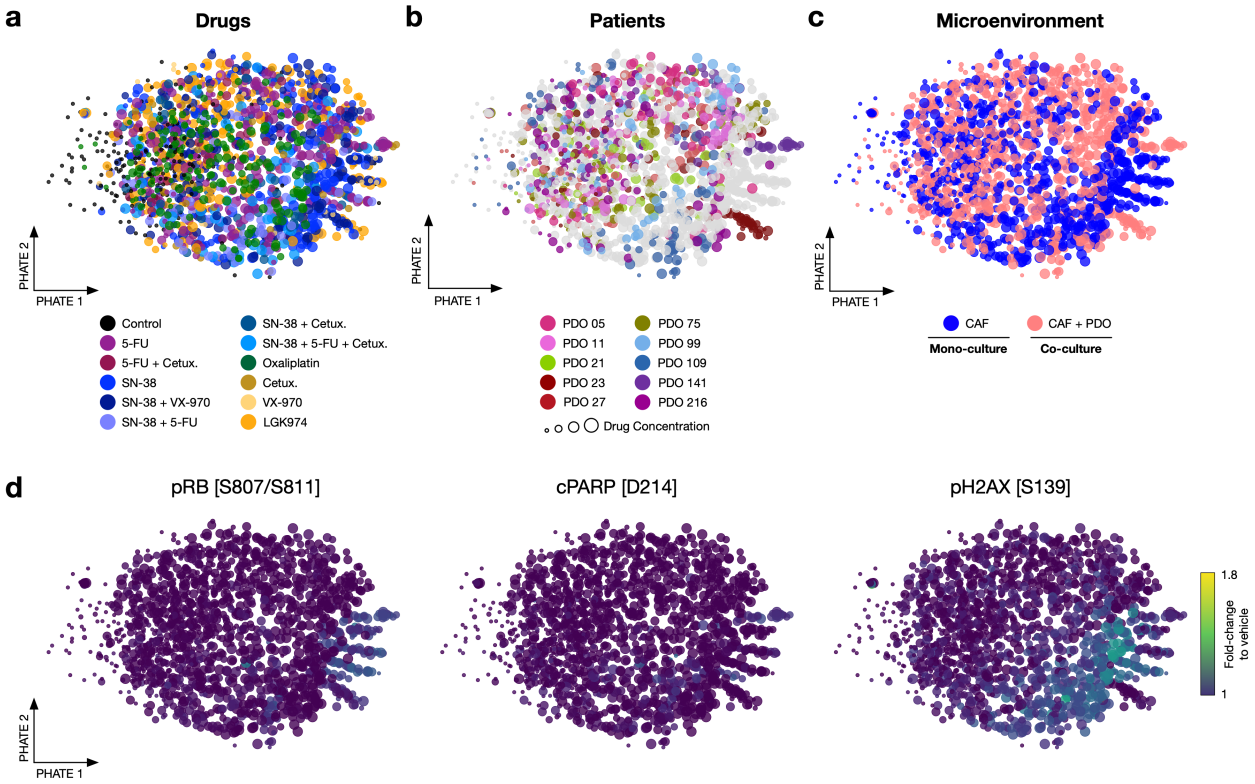

Figure S4: Single-Cell PTM CAF Drug Responses. a) Trellis-PHATE of PTM profiles from PDO-CAF cultures fails to identify drug-specific CAF responses, **b)** patient-specific CAF drug responses, or **c)** microenvironment-specific CAF drug responses. **d)** Fold-change to vehicle of pRB [S807/S811], cPARP [D214], and pHH2AX [S139] fail to resolve drug- or patient-specific shifts in cell-state.

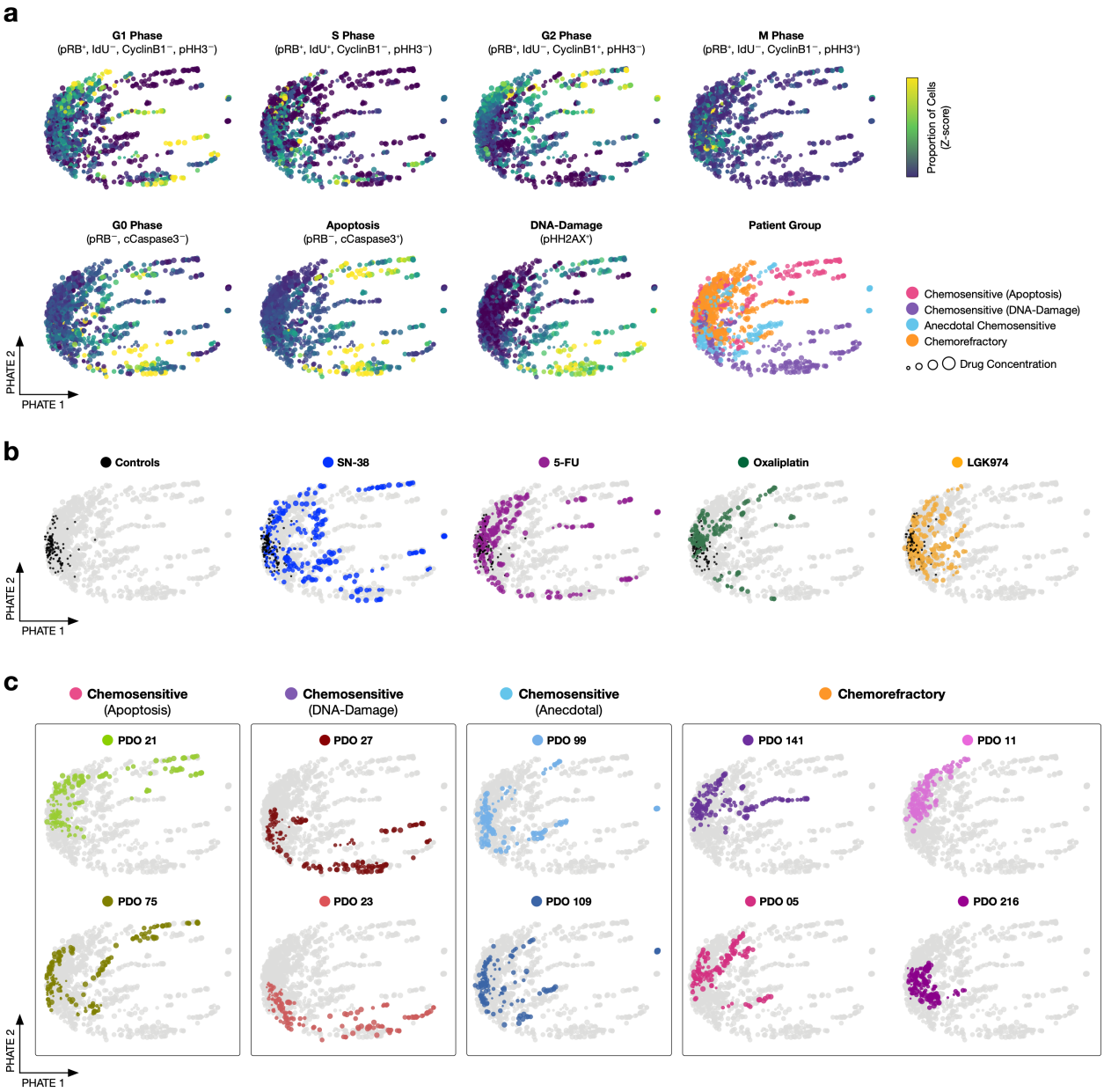

Figure S5: PDO Trellis-PHATE Cell-State Distributions. a) Treatment cell-state (z-score) across 1,680 single-cell PDO cultures reveal mechanistic drug treatment effects. **b)** Control, SN-38, 5-FU, Oxaliplatin, and LGK974 distributions. **c)** Individual patient distributions.

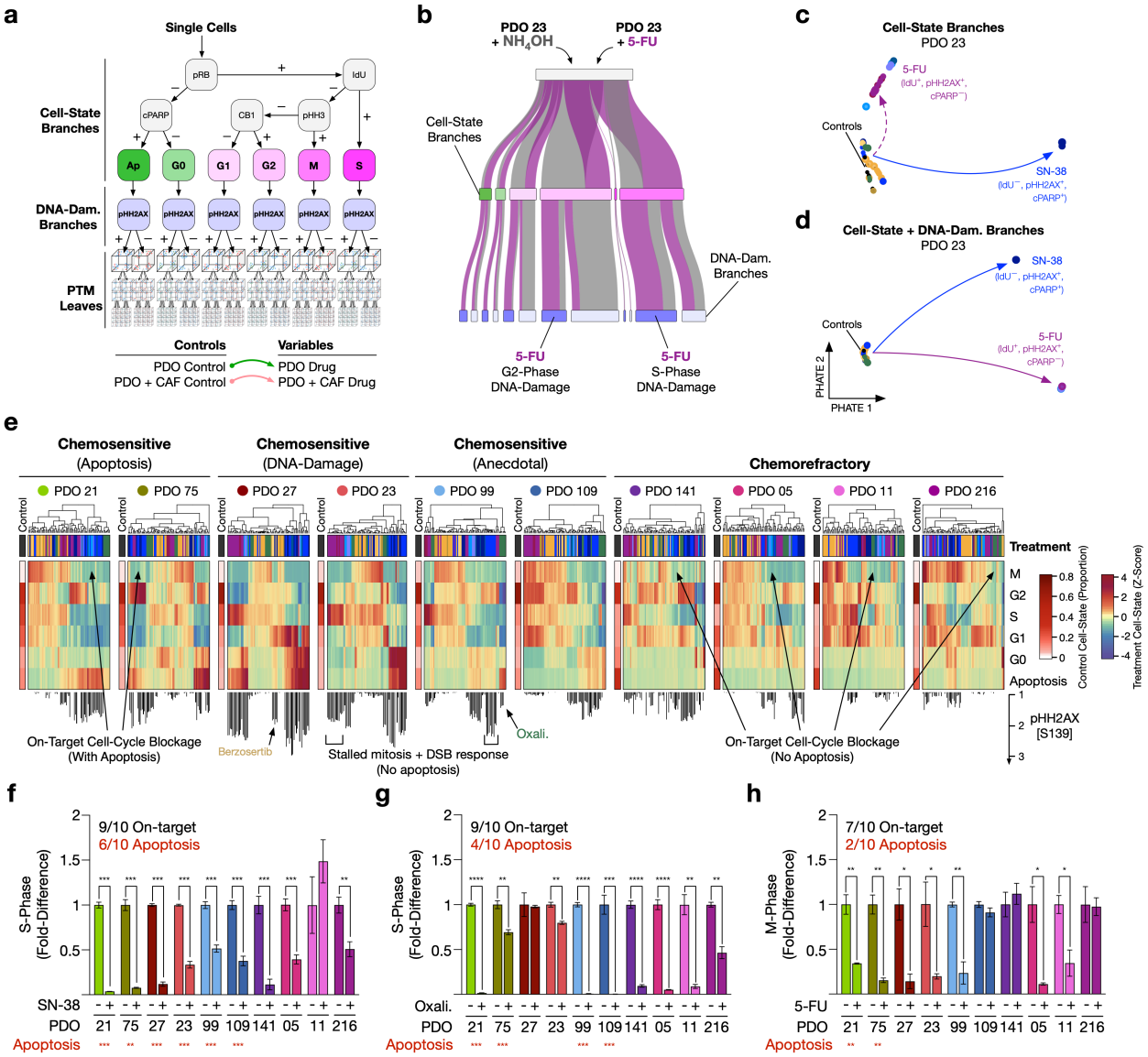

Figure S6: Patient-Specific Regulation of Cell-State and DNA-Damage. **a)** Trellis hierarchy containing cell-state branches with a pHH2AX [S139] DNA double-strand break detection layer. **b)** Sankey diagram showing NH_4OH vehicle control and 5-FU treatment of PDO 23 distributing through the cell-state Trellis branches in a) (leaves not shown). **c)** Trellis-PHATE of PDO 23 treatments analyzed using cell-state branches alone or **d)** cell-state branches and pHH2AX [S139] detection layer. The DNA-damage detection layer improves resolution of 5-FU on-target treatment effect. Solid arrows refer to strong treatment effect, dashed arrows refer to partial treatment effect. **e)** Patient-specific distribution of cells within Trellis branches reveals mechanistic cell-state shifts and DNA-damage upon drug treatments. Treatment cell-state quantifies the fold change of the proportion of cells/cell state over the controls for each treatment (Z-score). DNA damage is quantified by the fold change of the proportion $pHH2AX^+$ cells over the controls. **f)** PDO cells in S-phase following 100 nM SN-38. **g)** PDO cells in S-phase following 200 nM Oxaliplatin. **h)** PDO cells in M-phase following 200 nM 5-FU. PDOs with a significant >1.5 fold increase in apoptosis are indicated in red. Unpaired *t*-test, *** <0.0001, ** <0.001, * <0.01.

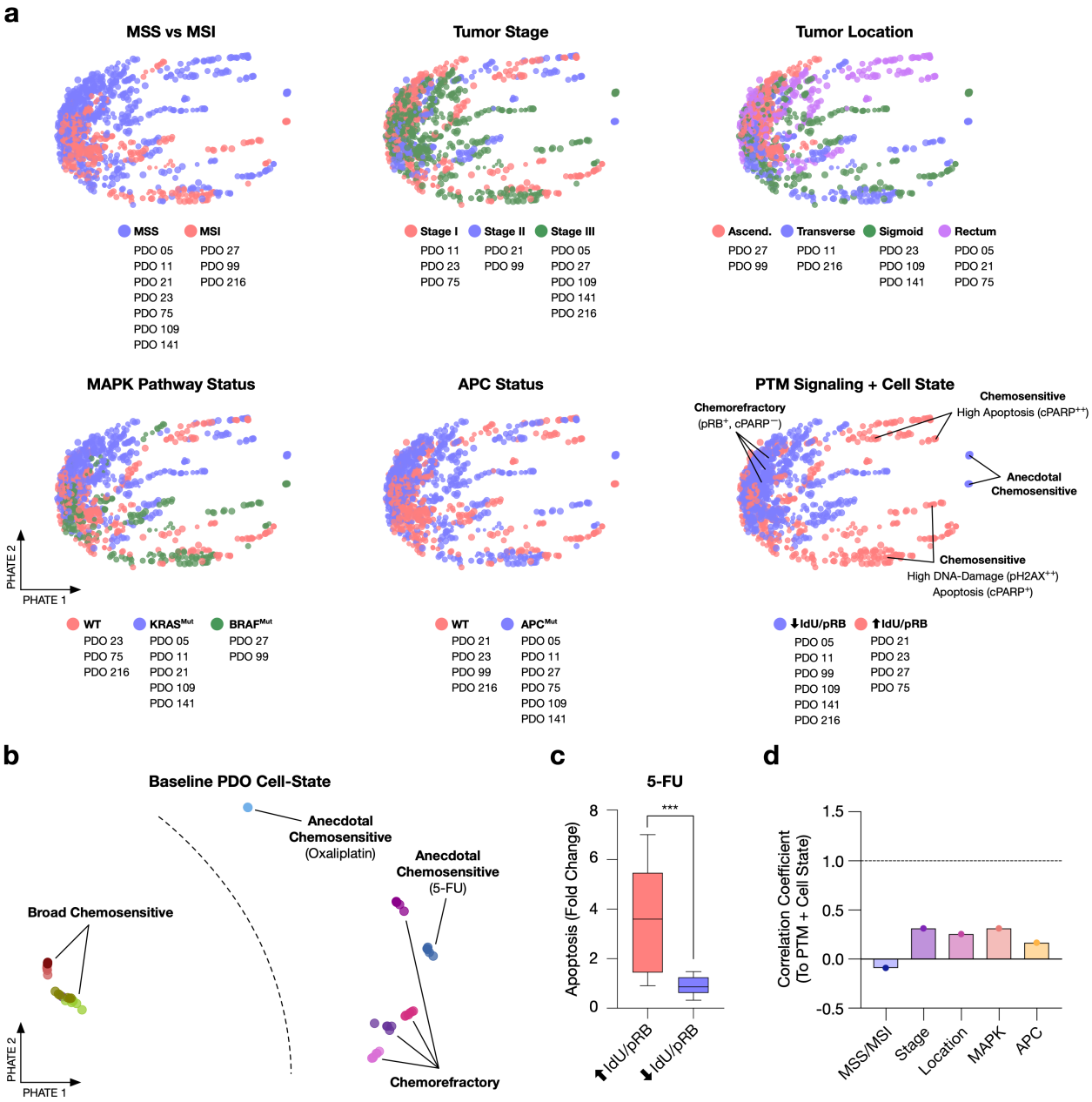

Figure S7: PDO Metadata. **a)** Trellis-PHATE plots of patient metadata. Patient-specific treatment effects do not align with MSS/MSI, tumor stage, tumor location, MAPK pathway mutations, or APC mutations. High baseline cell-cycle activity correlates with broad chemosensitivity. **b)** Trellis-PHATE of baseline PDOs annotated by patient and chemosensitivity. **c)** Quantification of 5-FU chemocytotoxicity in low and high cycling PDOs. **d)** Quantification of the correlation between PDO metadata information and PDO cell-state. Unpaired *t*-test, *** <0.0001.

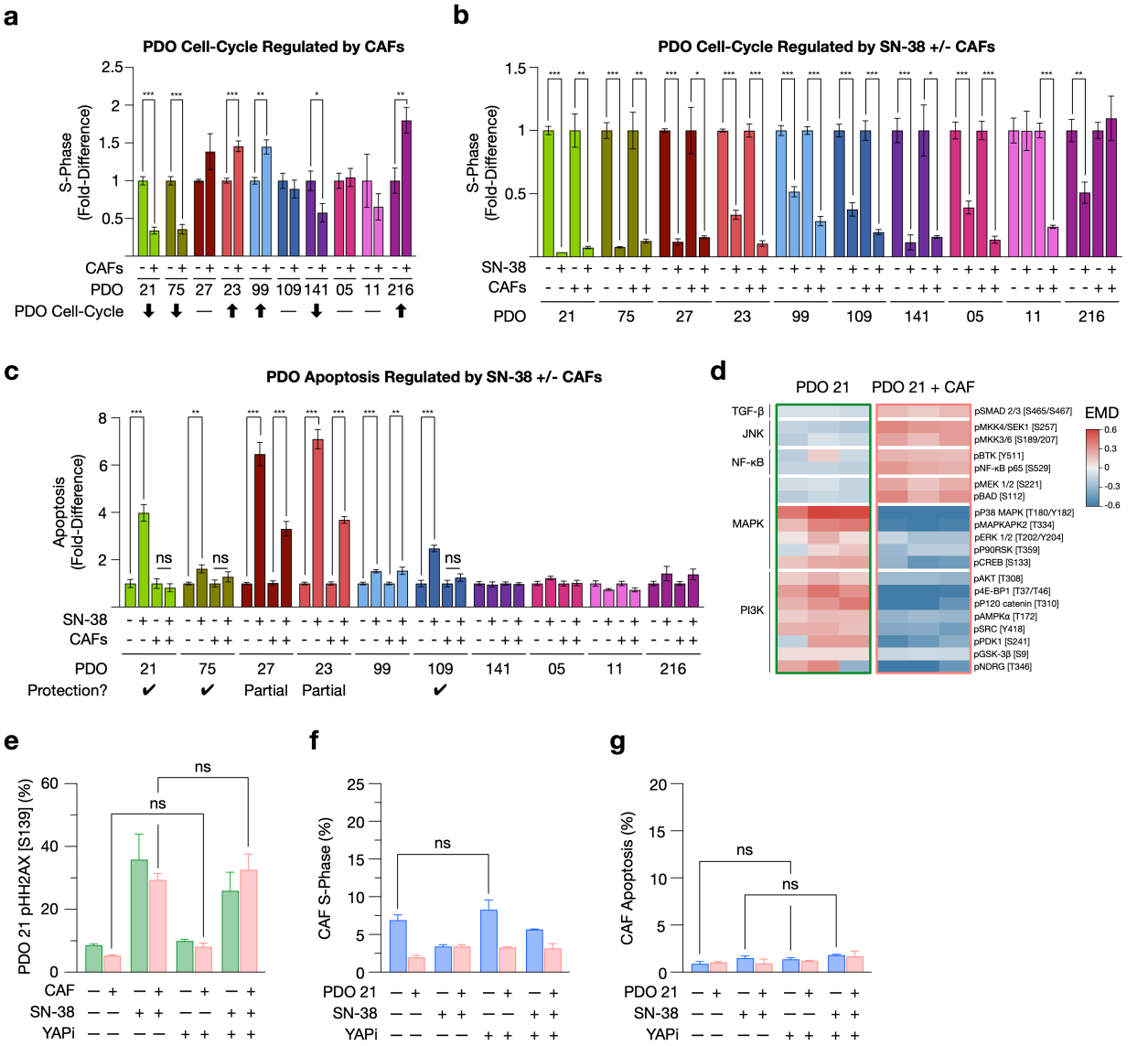

Figure S8: CAF-Induced PDO Cell-State Shifts. a) Fold difference to monoculture of PDO cells in S-phase when co-cultured with CAFs. CAFs both decrease and increase PDO S-phase in a patient-specific manner. **b)** Fold difference to vehicle controls of PDO cells in S-phase when treated with SN-38 either as PDOs alone or in co-culture with CAFs. CAFs do not alter SN-38 on-target S-phase blockage. **c)** PDO SN-38 induced apoptosis +/- CAFs. Partial CAF-protection is defined as a reduction drug-induced apoptosis in co-culture relative to monoculture, yet still apoptosis is still >1.5-fold over control and statistically significant. **d)** EMD heatmap of PTMs in PDO 21 +/- CAFs. **e)** YAP inhibition does not alter SN-38 induces on-target DNA-double strand breaks (pHH2AX^{+}) in PDOs. **f)** YAP inhibition does not alter S-phase or **g)** apoptosis in CAFs. Unpaired *t*-test, *** <0.0001, ** <0.001, * <0.01. ns, not significant.