bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discovery of small molecule pathway regulators by image profile matching

Mohammad H. Rohban, Ashley M. Fuller, Ceryl Tan, Jonathan T. Goldstein, Deepsing Syangtan, Amos Gutnick, Madhura P. Nijsure, Megan Rigby, Joshua R Sacher, Steven M. Corsello, Grace B. Peppler, Marta Bogaczynska, Gabrielle E. Ciotti, Ann DeVine, Minh Doan, Jennifer P. Gale, Rik Derynck, Thomas Turbyville, Joel D. Boerckel, Shantanu Singh, Laura L. Kiessling, Thomas L. Schwarz, Xaralabos Varelas, Florence F. Wagner, Ran Kafri, T.S. Karin Eisinger-Mathason, Anne E. Carpenter*

*to whom correspondence should be addressed

Abstract *(200 word limit, currently 200)*

Identifying chemical regulators of biological pathways is currently a time-consuming bottleneck in developing therapeutics and small-molecule research tools. Typically, thousands to millions of candidate small molecules are tested in target-based biochemical screens or phenotypic cell-based screens, both expensive experiments customized to a disease of interest. Here, we instead use a broad, virtual screening approach that matches compounds to pathways based on phenotypic information in public data. Our computational strategy efficiently uncovered small molecule regulators of three pathways, containing p38α (MAPK14), YAP1, or PPARGC1A (PGC-1α). We first selected genes whose overexpression yielded distinct image-based profiles in the Cell Painting assay, a microscopy assay involving six stains that label eight cellular organelles/components. To identify small molecule regulators of pathways involving those genes, we used publicly available Cell Painting profiles of 30,616 small molecules to identify compounds that yield morphological effects either positively or negatively correlated with image-based profiles for specific genes. Subsequent assays validated compounds that impacted the predicted pathway activities. This image profile-based drug discovery approach could transform both basic research and drug discovery by identifying useful compounds that modify pathways of biological and therapeutic interest, thus using a computational query to replace certain customized labor- and resource-intensive screens.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

The pace of defining new diseases based on genome sequencing is rapidly accelerating[1]. The cost and time required to develop novel therapeutics has also increased dramatically[2], creating huge unmet need. The dominant drug-discovery strategies in the pharmaceutical industry and academia are target-based (biochemical) and phenotypic (cell-based) drug discovery. Both require significant setup time, are tailored to a specific target, pathway, or phenotype, and involve physically screening thousands to millions of candidate compounds at great expense[3]. Computational approaches that allow virtual discovery of small molecule modulators of pathways using the published literature or existing experimental data are beginning to emerge to meet the need for more efficient routes to drug discovery[4,5].

Here we develop a distinct computational approach that uses image profile-based analysis to facilitate drug discovery. We use the complex morphological responses of cells to a genetic perturbation to identify small molecules (i.e., chemical compounds) that produce the same (or opposite) response. Morphological responses are assessed using existing public image-based profiles from the microscopy assay, Cell Painting[6,7]. Conceptually similar to transcriptional profiling[8], Cell Painting is cheaper and has substantial predictive power[9–11].

Recent decades have given rise to an appealing, reductive ideal in the pharmaceutical industry: one drug that targets one protein to target one disease[12]. However, diseases often involve many interacting proteins and successful drugs often impact multiple targets[13,14]. An emerging concept is that target deconvolution—identifying the precise molecular target of a drug—is valuable but not a deciding factor[15], because it is often inconclusive, incomplete, or incorrect[14]. There is therefore a renewed appreciation for identifying small molecules that can modulate *pathways* in living cell systems to yield a desired phenotypic effect, focusing on the network level rather than the individual protein level[12]. Because genes in a pathway often show similar morphology[16] and compounds often show similar morphology based on their mechanism of action[17], we examined image profile-based drug discovery as a promising but untested route to capturing perturbations at the pathway level and accelerating the discovery of useful therapeutics and research tool compounds.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

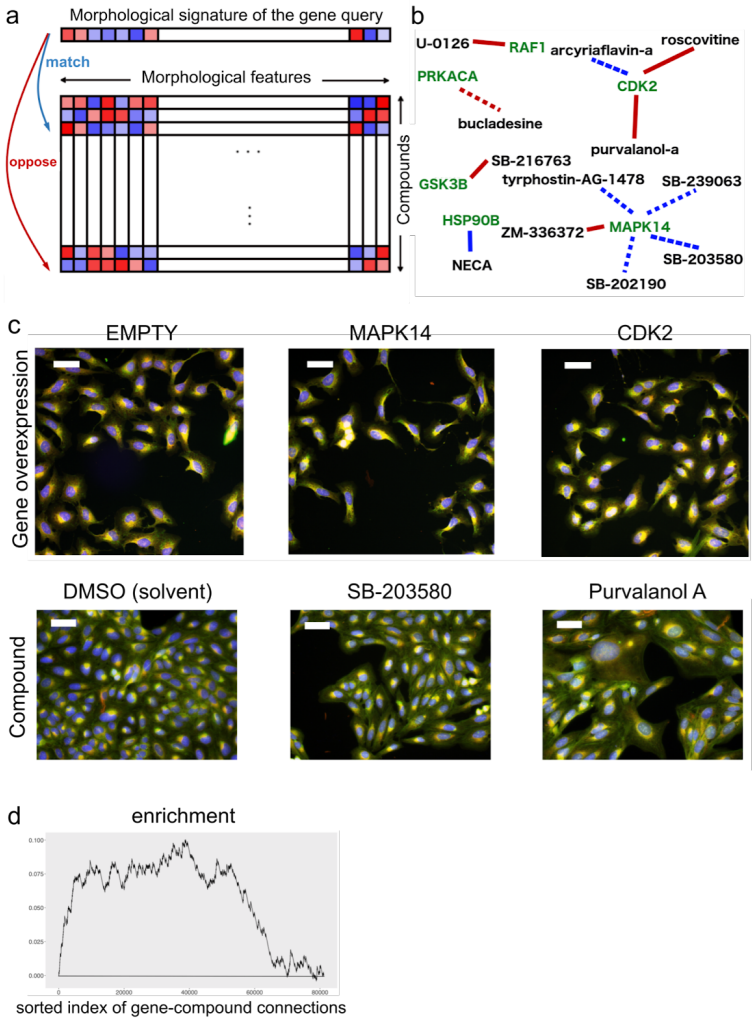


Figure 1: Image profile-based drug discovery offers efficient, virtual discovery of pathway modulators. *If an overexpressed gene changes the morphology of cells, its image-based profile can be used as a query to identify matches in a database of small molecule profiles, looking for those that match (positively correlate) or oppose (negatively correlate). b) Of the 63 genes that have a bioactive compound annotated as targeting it in the dataset, six genes (green text) strongly matched or opposed the correct compound(s) (black text). The lines represent positive (blue) and negative (red) morphological correlations to compounds. They also show whether the morphological correlation is the expected (solid) or unexpected directionality (dotted) based on previously described positive or negative impacts on gene function. c) Cell Painting images for two positive control gene-compound matches that yield observable morphological phenotypes. EMPTY and DMSO are the negative controls in the gene overexpression and compound experiments, respectively; they differ in their confluency and image acquisition conditions. The phenotype of p38α (MAPK14) overexpression matches (correlates to) that of SB-203580, a known p38 inhibitor; in both, elongated/triangular cells and mitotic cells are over-represented. The phenotype of CDK2 overexpression (small cells) negatively correlates to that of purvalanol-a, a known CDK inhibitor, which induces an opposite phenotype (huge cells). Scale bars= 60 μm. d) Enrichment plot of all gene-compound connections sorted based on their absolute profile correlation. Starting from the left, the curve rises a unit if the gene is annotated to interact with a known target of the compound (or a pathway member), and goes down a unit otherwise. The units are normalized to the number of possible relevant pairs, so the maximum height is one and ends in zero. A steep initial increase of the curve indicates enrichment of correct connections towards the top of the rank-ordered list of pairs.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Image-based gene-compound matching: validation

We began with 69 unique genes whose overexpression yields a distinctive morphological phenotype by Cell Painting, from our prior study in U2OS cells[16]. We matched their image-based profiles to our published library of Cell Painting profiles of 30,616 small molecules[18], which includes 747 compounds annotated with the gene(s) they target (Figure 1a). We restricted analysis to the 15,863 tested compounds (52%) whose profiles are distinguishable from negative controls, and confirmed that the profiles show variety rather than a single uniform toxic phenotype (Extended Data Figures 1 and 2).

We first verified that image profiles allow compounds to be matched with other compounds that share the same mechanism of action, for the subset that is annotated. Consistent with past work[17], top-matching compound pairs share a common annotated mechanism-of-action four times more often than for the remainder of pairs (p-value $< 2.2 \times 10^{-16}$, one-sided Fisher's exact test, Supplementary Table 1).

We next attempted gene-compound matching. We did not expect a given compound to produce a profile that matches that of its annotated gene target in all cases, nor even the majority. Expecting simple gene-compound matching takes a reductionist view that may not reflect the complexity of drug action (see Introduction). We therefore included genes annotated as pathway members as a correct match, given our goal of identifying compounds with the same functional impact in the cell. In addition, existing annotations are imperfect, particularly given the prevalence of under-annotation, mis-annotation, off-target effects, and polypharmacology, where small molecules modulate protein functions beyond the intended primary target[13]. Finally, technical reasons can also confound matching. The genetic and compound experiments were conducted years apart and by different laboratory personnel, yielding batch effects. They were performed in U2OS cells which may not be relevant for observing the annotated gene-compound interaction. In addition, the negative controls in a gene overexpression experiment (untreated cells), and a small molecule experiment (treated with the solvent control DMSO), do not produce identical profiles (left column, Figure 1c), and must therefore be normalized to align the negative controls in the feature space (see "Feature set alignment" in Methods). Despite these concerns, we persisted because even if the strategy worked in only a small fraction of cases, a virtual screening approach could be very powerful given millions of dollars saved per screening campaign.

63 of the 69 genes were annotated as targeted by a compound in the set; we used these as positive controls. These positive controls were 2.5-fold overrepresented among the strongest gene-compound pairings (correlation ≥0.35) (p-value = 0.007; Figure 1b, Supplementary Tables 2 and 3); for some matches, we could visually confirm that gene overexpression phenocopies or pheno-opposes the matching/opposing compound (Figure 1c). Looking across the whole spectrum of matches, rather than those above our threshold, we confirmed consistent enrichment in the correct connections (Figure 1d).

In a more practical version of this analysis, we took a gene-centric view and examined the top positively or negatively correlated compounds for each gene (rather than examining all gene-compound matches at once). For 19% of genes, spanning diverse biological pathways (Supplementary Table 4), that list is significantly enriched with the correct compound (12 genes out of 63 genes that had a morphological phenotype and at least one relevant compound in the experiment; adjusted p-value 0.05; see "Enrichment p-value estimation" in Methods).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Image-based gene-compound matching: discovery

We next searched virtually for novel small molecule regulators of pathways. Throughout this study, we looked for compounds that both match (positively correlate) and oppose (negatively correlate) each overexpressed gene profile for two reasons: inhibitors and activators of a given pathway may both be of interest and we previously found that negative correlations among profiles can be biologically meaningful[16]. In addition, overexpression may not increase activity of a given gene product in the cell; it could be neutral or even decrease it via a dominant-negative or feedback loop effect. Finally, the impact of a gene or compound could be cell-type specific. In our validation set, for example, we found that the directionality of correct matches is sometimes the opposite of what is expected; three gene-compound matches showed the expected directionality, one showed the opposite, and two showed mixed results (Figure 1b).

For each of the 69 genes, we created a rank-ordered list of compounds (from the 15,863 impactful compounds of the 30,616 set) based on the absolute value of correlation to that gene (Supplementary Table 5). We then found seven experts studying pathways with strong hits who were willing to conduct exploratory experiments; researchers chose the most relevant biological systems for validation, rather than simply attempting to validate the original finding.

Two cases yielded no confirmation (data not shown): RAS and SMAD3. We selected 236 compounds based on their positive or negative correlations to the wild-type RAS or oncogenic HRAS G12V differential profile (see Methods). The compounds failed to elicit a RAS-specific response in a 72-hour proliferation assay using isogenic mouse embryonic fibroblast (MEF) cell lines driven by human KRAS4b G12D, HRAS WT, or BRAF V600E alleles but otherwise devoid of RAS isoforms[19]. Nine compounds matching or opposing the SMAD3 overexpression profile failed to yield activity in a transcription reporter assay involving tandem Smad binding elements, with and without Transforming growth factor beta 1 (TGF-β1). We cannot distinguish whether the compounds were inactive due to differences in the cell types or readouts, or whether these represent a failure of morphological profiling to accurately identify modulators of the pathway of interest.

Two cases yielded promising initial results but the novel compounds failed to confirm using an orthogonal assay or following compound resynthesis. We tested 17 compounds that negatively correlated with CSNK1E overexpression in a biochemical assay for the closely related kinase CSNK1A1. We found that three (SB 203580, SB 239063, and SKF-86002) had inhibitory IC_{50} concentrations in the nanomolar range at K_m ATP. Inhibition of CSNK1 family members by these compounds is supported by published kinase profiling studies[20–22]. A fourth compound, BRD-K65952656, weakly inhibited CSNK1A1 (IC_{50} 12 uM) but failed to bind any native kinases in a full KINOMEscan panel, suggesting it acts against another molecular target. In the other case, 16 compounds that positively correlated and 17 compounds that negatively correlated to GSK3B were tested for impact on GSK3α and GSK3β (which generally overlap in function) in a non-cell-based, biochemical assay. This yielded four hits with GSK3α IC50s ≤ 10 μM; the two most potent failed to show activity following resynthesis and hit expansion (testing of similarly-structured compounds) (Supplementary Table 6), suggesting the original activity was not due to the expected compound, perhaps due to breakdown. If truly negative, we again cannot distinguish whether their failure reflects our choice of biochemical binding and specific kinase assays (rather than a cell-based functional pathway readout) or whether they represent a failure of the morphological matching method.

We did not pursue these cases further in light of the success for the three other cases, described next.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discovery of small molecules modulating the p38α (MAPK14) pathway

p38α (MAPK14) inhibitors are sought for a wide variety of disorders, including cancers, dementia, asthma, and COVID-19[23,24]. We chose 20 compounds whose Cell Painting profile matched (9) or opposed (11) that of p38α overexpression in U2OS cells. In a single-cell p38 activity reporter assay in retinal pigment epithelial (RPE1) cells[25,26], we identified several inhibiting compounds, including a known p38 MAPK inhibitor, SB202190 (Figure 2), and confirmed activity at 10 μM (Extended Data Figure 3). We also found many activating compounds, which are less interesting given that the p38 pathway is activated by many stressors but rarely inhibited. We conclude that our computational image-based matching method can identify novel compounds impacting the p38 pathway using public Cell Painting data rather than a specific screen designed to measure p38 activity.

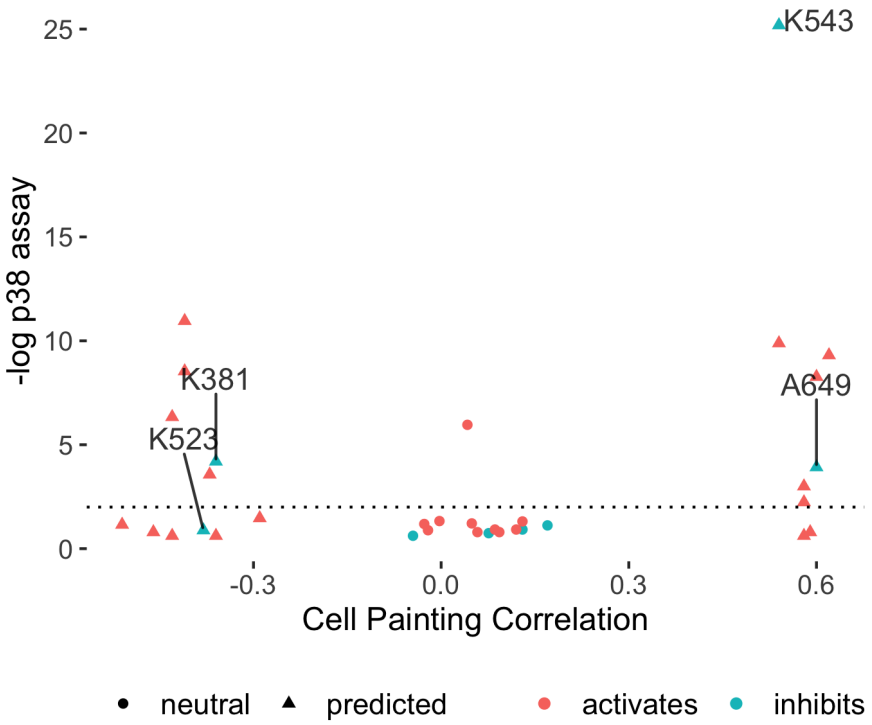


Figure 2: Cell Painting profiles identify compounds impacting the p38 pathway. *Compounds predicted to perturb p38 activity (triangles) and a set of 14 neutral compounds (Cell Painting profile correlations to p38α between -0.2 to 0.2; circles) were tested for their influence on p38 activity at 1 μM using a two-sided t-test on the single cell distributions of a p38 activity reporter[27] (FDR-adjusted -log$_{10}$ p-values shown).Two potential inhibitors were found (BRD-K38197229 <K381> and BRD-A64933752 <A649>); an additional one (BRD-K52394958 <K523>) was identified via an alternative statistical test (Extended Data Figure 3a, h-i). K543 (BRD-K54330070) denotes SB202190, a known p38 inhibitor found as a match.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discovery of small molecules impacting PPARGC1A (PGC-1α) overexpression phenotypes

We next identified compounds with strong morphological correlation to overexpression of peroxisome proliferator-activated receptor gamma coactivator 1-alpha (PGC1α, encoded by the PPARGC1A gene). We found that these compounds tend to be hits in a published, targeted screen for PGC1α activity (p=7.7e-06, Fisher's exact test)[28], validating our image profile-based matching approach. The dominant matching phenotype is mitochondrial blobbiness, which can be quantified as the high standard deviation of the MitoTracker staining at the edge of the cell without major changes to cell proliferation, size, or overall protein content (Figure 3a,b). Cell subpopulations that are large, multi-nucleate, and contain fragmented mitochondria are over-represented when PGC-1α is overexpressed while subpopulations whose organelles are asymmetric are under-represented (Extended Data Figure 4). More symmetric organelle morphology is associated with reduced motility and PGC-1α overexpression[29]. The role of PGC-1α in mitochondrial biogenesis is well-appreciated[30]. The phenotype uncovered here using image profile matching is consistent with other recently discovered mitochondrial phenotypes associated with this gene[31].

We chose 24 compounds whose Cell Painting profiles correlated positively or negatively with PGC-1α overexpression in U2OS cells; one is a known direct ligand for PPAR gamma, GW-9662 (BRD-K9325869). PGC-1α is a transcriptional coactivator of several nuclear receptors including PPAR gamma and ERR alpha[32]. We therefore tested compounds in a reporter assay representing FABP4, a prototypical target gene of the nuclear receptor, PPARG[33], in a bladder cancer cell line (Figure 3c). Three of the five most active compounds leading to reporter activation were structurally related and included two annotated SRC inhibitors, PP1 and PP2, which have a known link to PGC1α[34], as well as a novel analog thereof. CCT018159 (BRD-K65503129) and Phorbol 12-myristate 13-acetate (BRD-K68552125) inhibited reporter activity. Many of the same compounds also showed activity in a ERRalpha reporter assay in 293T cells, albeit with differing effects (Extended Data Figure 5*)*.

Encouraged by these results, we tested the impact of the compounds on mitochondrial motility, given the mitochondrial phenotype we observed and the role of PGC1α in mitochondrial phenotypes and neurodegenerative disorders[35]. In an automated imaging assay of rat cortical neurons[36], we found several compounds decreased mitochondrial motility; none increased motility (Extended Data Figure 6). Although the latter is preferred due to therapeutic potential, this result suggests that the virtual screening strategy, applied to a larger set of compounds, might identify novel motility-promoting compounds. We found 3 of the 23 compounds suppress motility but do not decrease mitochondrial membrane potential; this is a much higher hit rate (13.0%) than in our prior screen of 3,280 bioactive compounds, which yielded two such compounds (0.06%)[36].

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

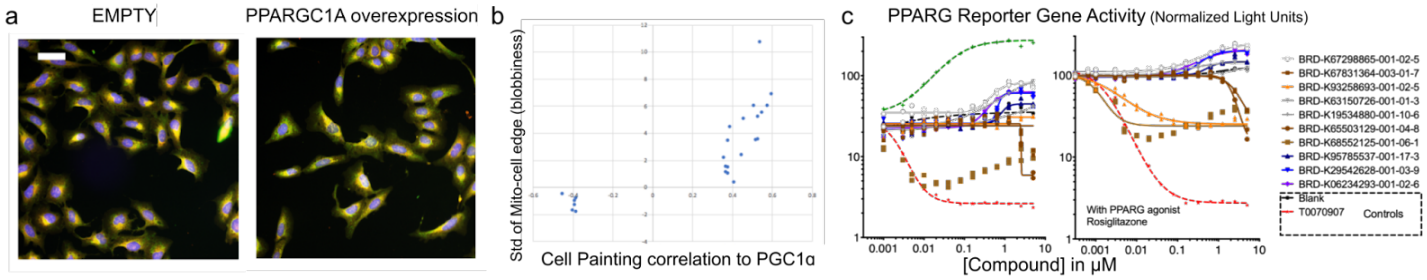


Figure 3: Cell Painting profiles identify compounds impacting PPARGC1A (PGC-1α) overexpression phenotypes. *a) Cell Painting images for PPARGC1A (PGC-1α) overexpression compared to negative control (EMPTY, same image as in Figure 1a). Scale bar = 60 μm. b) Correlation of compounds to PGC-1α overexpression is dominated by one feature, the standard deviation of the MitoTracker staining intensity at the edge of the cell, which we term blobbiness. Compounds with high or low correlations of their Cell Painting profiles to PGC-1α overexpression were chosen for further study (hence all samples are below ~ -0.35 or above ~0.35 on the X axis). The samples with high correlation show generally high blobbiness, as plotted on the Y axis as number of standard deviations (normalized to the negative controls). c) PPARG reporter gene assay dose-response curves in the absence (left) or presence (right) of added PPARG agonist, Rosiglitazone. Representative data of the ten most active compounds is shown and reported as normalized light units. Compounds highlighted in blue/purple are structurally related pyrazolo-pyrimidines.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discovery of small molecules modulating the Hippo pathway

The Hippo pathway plays a key role in development, organ size regulation, and tissue regeneration. Small molecules that alter its activity are highly sought for basic research and as potential therapeutics for cancer and other diseases[37]. We tested 30 compounds whose Cell Painting profile matched (25 compounds) or opposed (5 compounds) the overexpression of the Hippo pathway effector Yes-associated protein 1 (YAP1), which we previously explored[16] (Supplementary Table 7, Extended Data Figure 7). One hit, fipronil, has a known tie to the Hippo pathway: its impact on mRNA profiles matches that of another calcium channel blocker, ivermectin, a potential YAP1 inhibitor[38] (99.9 connectivity score in the Connectivity Map[8]). After identifying 5 promising compounds in a cell proliferation assay in KP230 cells (described later), we focused on the three strongest in various assays and cell contexts, as follows.

N-Benzylquinazolin-4-amine (NB4A, BRD-K43796186) is annotated as an EGFR inhibitor and shares structural similarity with kinase inhibitors. NB4A showed activity in 30 of 606 assays recorded in PubChem, one of which detected inhibitors of TEAD-YAP interaction in HEK-TIYL cells. Its morphological profile positively correlated with that of YAP1 overexpression (0.46) and, consistently, negatively correlated with overexpression of STK3/MST2 (-0.49), a known negative regulator of YAP1.

Because the Hippo pathway can regulate the pluripotency and differentiation of human pluripotent stem cells (hPSCs)[39,40], we investigated the effect of NB4A in H9 hPSCs. NB4A did not affect *YAP1* expression but increased the expression of YAP1 target genes (*CTGF* and *CYR61*) in a dose-dependent manner (Figure 4a), confirming it impacts the Hippo pathway. Accordingly, NB4A increased YAP1 nuclear localization (Figure 4b). While decreasing total YAP1 levels, NB4A also reduced YAP1 S127 phosphorylation (Figure 4c and Extended Data Figure 8a), which promotes YAP1 cytoplasmic sequestration[41].

Effects of NB4A on YAP1 mRNA expression were not universal across cell types, consistent with the Hippo pathway's known context-specific functions. In most cell types represented in the Connectivity Map, *YAP1* mRNA is unaffected, but in HT29 cells, *YAP1* mRNA is up-regulated after six hours of NB4A treatment (z-score = 3.16; also z-score = 2.04 for TAZ) and in A375 cells, *YAP1* mRNA is slightly down-regulated (at 6 and 24 hours; z-score ~ -0.7)[8]. NB4A had no effect in a YAP1-responsive reporter assay following 48h of YAP overexpression in HEK-293 cells (Extended Data Figure 8b).

Compounds influencing the Hippo pathway might be therapeutic for undifferentiated pleomorphic sarcoma (UPS), an aggressive mesenchymal tumor that lacks targeted treatments[42]. In UPS, YAP1 promotes tumorigenesis and is inversely correlated with patient survival[42]. To assess the impact of NB4A on the Hippo pathway, we treated KP230 cells, derived from a mouse model of UPS[42]. In these cells, NB4A did not regulate transcription of *Yap1*, its sarcoma target genes (*Foxm1*, *Ccl2*, *Hbegf*, *Birc5*, and *Rela*), nor Yap1's negative regulator, angiomotin (*Amot*) (data not shown). Instead, pathways such as interferon alpha and gamma responses were up-regulated, whereas pathways such as the epithelial-mesenchymal transition, angiogenesis, and glycolysis were down-regulated, according to RNA sequencing and gene set enrichment analysis (Figure 4d; Supplementary Table 8). Nevertheless, we identified impact on the Hippo pathway: Yap1 protein levels were reduced after 72 hours of treatment (Figure 4e-f, h). NB4A also significantly attenuated Yap1 nuclear localization (Figure 4g-h), which is known to reduce its ability to impact transcription.

Genetic and pharmacologic inhibition of Yap1 suppresses UPS cell proliferation *in vitro* and tumor initiation and progression *in vivo*[42]. Consistent with being a Hippo pathway regulator, NB4A inhibited the proliferation of two YAP1-dependent cell lines: KP230 cells and TC32 human Ewing's family sarcoma cells[43] (Figure 4i). NB4A did

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

not affect the proliferation of two other YAP1-dependent lines, STS-109 human UPS cells (Extended Data Figure 9a) and HT-1080 fibrosarcoma cells (Extended Data Figure 9b)[42,44], nor YAP1-independent HCT-116 colon cancer cells (Extended Data Figure 9c-e). Interestingly, NB4A treatment did not exhibit overt toxicity by trypan blue staining in any of these (not shown), suggesting it inhibits cell proliferation by a mechanism other than eliciting cell death.

Finally, we investigated two structurally similar compounds (BRD-K28862419 and BRD-K34692511, distinct from NB4A's structure) whose Cell Painting profiles negatively correlated with YAP1's overexpression profile (-0.43 for BRD-K28862419 and -0.45 for BRD-K34692511) and positively correlated with TRAF2 overexpression (0.41 for BRD-K28862419 and 0.29 for BRD-K34692511) (Extended Data Figure 7). These compounds are not commercially available, limiting our experiments and past literature.

We assessed their impact on the Hippo pathway using mesenchymal lineage periosteal cells isolated from 4-day old femoral fracture callus from mice with DOX-inducible YAP-S127A. BRD-K34692511 substantially upregulated mRNA levels of relevant Hippo components including *Yap1* and *Cyr61* after 48 hours of treatment, but not at 1 and 4 hours (Extended Data Figure 8c-f). By contrast, the compounds had no effect on *YAP1* or its target genes in H9 hPSCs (Extended Data Figure 8g), nor in a 48 h YAP-responsive reporter assay following YAP overexpression in HEK-293 cells (Extended Data Figure 8b).

Like NB4A, the effects of these compounds on proliferation varied across cell types. In the U2OS Cell Painting images, BRD-K28862419 reduced proliferation (-2.0 st dev). Per PubChem, it inhibits cell proliferation in HEK293, HepG2, A549 cells (AC50 5-18 μM) and it inhibits PAX8, which is known to influence TEAD/YAP signaling[45]. BRD-K34692511 had none of these impacts.

Interestingly, both compounds inhibited KP230 cell proliferation (Extended Data Figure 9f). Also noteworthy, BRD-K28862419 modestly yet significantly reduced KP230 cell viability (Extended Data Figure 9g), indicating its mechanism of action and/or therapeutic index may differ from that of NB4A and BRD-K34692511.

In summary, although deconvoluting the targets and behaviors of these compounds in various cell contexts remains to be further ascertained, we conclude that the strategy identified compounds that modulate the Hippo pathway. This demonstrates that, although the directionality and cell specificity will typically require further study, image-based pathway profiling can identify modulators of a given pathway.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

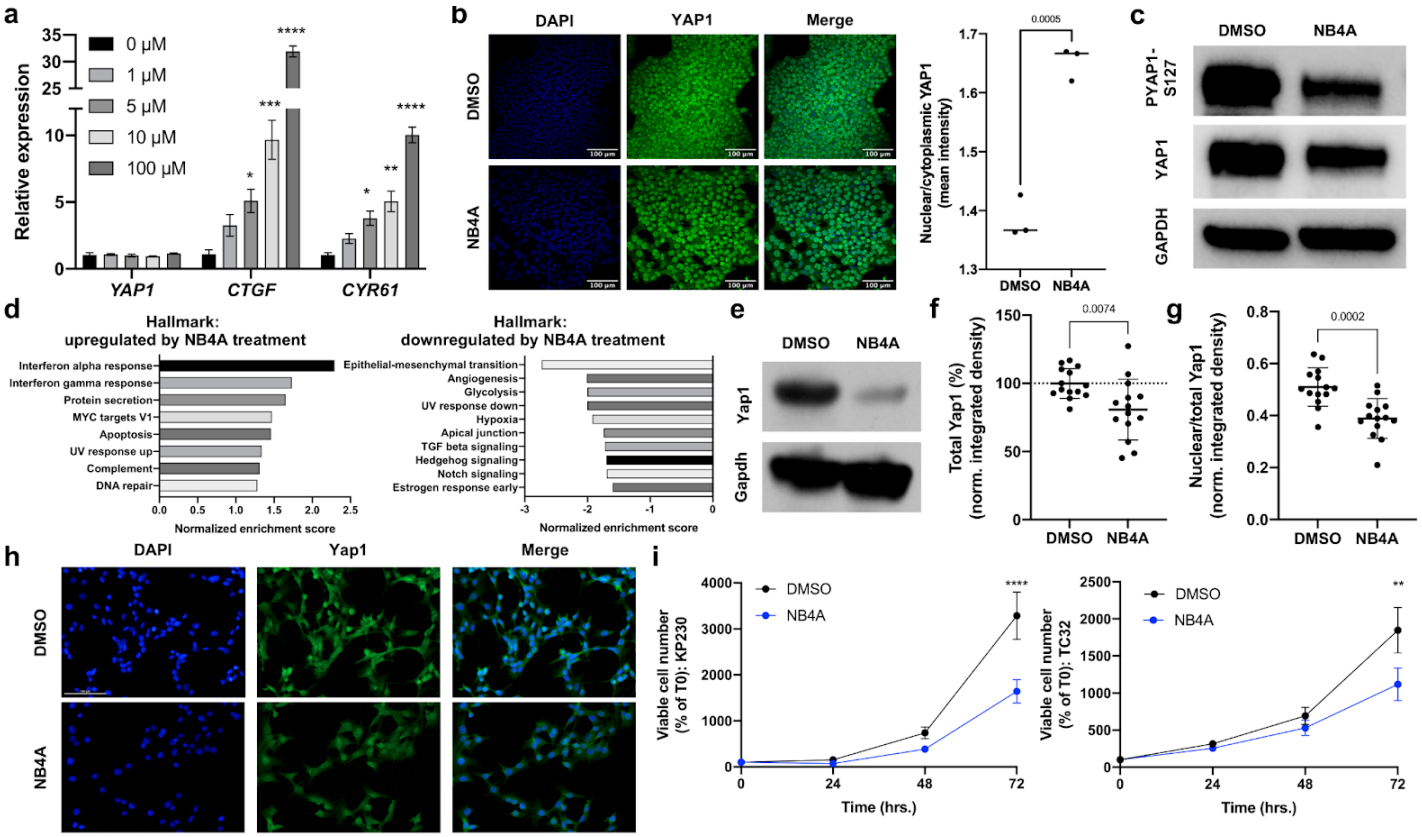


Figure 4: Cell Painting profiles identify compounds impacting the Hippo pathway. *a) Relative transcript levels of YAP1, CTGF, and CYR61 in H9 human pluripotent stem cells treated with NB4A or DMSO control for 24 hrs. *P<0.05; **P<0.01; ***P<0.001; ****P<0.0001 (one-way ANOVA with Dunnett's multiple comparisons test). Mean ± SEM. n = 3. b) Representative images of YAP1 immunofluorescence (left) and quantification of nuclear/cytoplasmic YAP1 mean intensity (right) in H9 cells after treatment with 10 μM NB4A or DMSO control for 24 hours. Two-tailed student's t-test; note the split y axis. n = 3; an average of mean intensities from 3 fields of each biological replicate is calculated. c) Representative western blot analysis of phospho-YAP1 (S127) and total YAP1 from H9 cells treated with DMSO or 10 μM NB4A for 24 hrs, with GAPDH as loading control (quantified in Extended Data Figure 8a). d) Normalized enrichment scores of GSEA show up to 10 of the most significant Hallmark pathways up- and down-regulated in NB4A-treated vs. control KP230 cells (FDR-adjusted P<0.25). n = 3. e) Representative western blot for Yap1 in NB4A-treated and control KP230 cells. f) Immunofluorescence-based analysis of total Yap1 in NB4A-treated and control KP230 cells. Two-tailed student's t-test. Mean ± SEM. n = 3. g) Immunofluorescence-based analysis of nuclear Yap1 in NB4A-treated and control KP230 cells (normalized to total Yap1). Two-tailed student's t-test. Mean ± SEM. n = 3. For f and g, the Y axis is integrated density normalized to cell number and representative images are shown in (h), out of 5 fields acquired per condition. Scale bar (top left panel) = 100 μM. i) Growth curves of NB4A-treated and control KP230 and TC32 sarcoma cells. **P<0.01; ****P<0.0001 DMSO vs. NB4A (72 hrs.; 2-way ANOVA with Sidak's multiple comparisons test). Mean ± SEM. n = 3. For panels d-i, cells were treated with 10 μM NB4A daily for 72 hours.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discussion

We found that small molecule regulators of pathways of interest can be efficiently discovered by virtual matching of genes and compounds using Cell Painting profiles, which we term image profile-based drug discovery. As with all screening approaches, further testing is necessary to confirm activity and directionality in a relevant cell type or model system and to develop hits into useful therapeutics. However, the strategy of computationally matching the phenotypic effect of compounds to that of gene manipulation will in many cases enable rapid and inexpensive identification of compounds with phenotypic impacts at scale. This approach may also be extended to identify which pathways are targeted by novel small molecules of unknown mechanism of action, another significant bottleneck in the drug discovery process[46].

Large-scale data production efforts are underway that will increase the potential for matching profiles: the Library of Integrated Network-Based Cellular Signatures (LINCS)[47] now contains Cell Painting data, the JUMP-Cell Painting Consortium is producing a public dataset of 140,000 chemical and genetic perturbations, and some pharmaceutical and biotechnology companies have even larger proprietary datasets[48]. Expansion to other staining sets or more complex biological models, such as co-cultures, primary cells, or organoids could further increase the probability of success. More advanced methods are also on the horizon, from feature extraction[49] to machine learning on new benchmark datasets of gene-compound pairs[50]. We anticipate that image profile-based drug discovery provides a new, broad, and unbiased route toward meeting the pressing need for novel therapeutics.

Materials and Methods

Data availability

The large-scale Cell Painting datasets used in this paper are publicly available and their details and locations are described in publications (gene overexpression dataset[16] and compound dataset[18]). RNA-sequencing data have been deposited into the NCBI Gene Expression Omnibus (GEO; accession number pending).

Code availability

The code used in this study is available at https://github.com/broadinstitute/GeneCompoundImaging. It is available for use under the BSD 3-clause license, a permissive open-source license.

Cell line and DNA construct availability

Cell lines and DNA constructs are available from the laboratories that performed the experiments using them, or where restricted by licensing, from commercial sources.

Research animals

Boerckel lab: Mouse experiments were conducted in compliance with all relevant regulations. All animal experiments were performed at the University of Pennsylvania under IACUC review and in compliance with IACUC protocol #806482.

Feature set alignment

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

As each experiment was analyzed by a slightly different CellProfiler pipeline, and also the phenotype of the negative controls are quite different (Figure 1c), an extra data preprocessing step is needed to make the feature sets comparable. To achieve this, we first took the intersection of features in the two datasets, which resulted in 605 features (1399 features in the genetic screen, without any feature selection; and 729 features in the compound screen, obtained using the findCorrelation with threshold of 0.90 on the original 1,783 dimensional feature set). In order to compare values of the corresponding features across experiments, each feature is standardized (mean-centered and scaled by standard deviation) with respect to the negative control. This was done platewise based on the mean and standard deviation of the controls at profile level for the compound dataset. The normalization parameters are slightly different for the genetic screen, where median and median absolute deviation (MAD) are used instead, to remove the outlier effects[16]. The code repository for all the analyses are publicly available as described in Code Availability.

Scoring gene-compound connections

We use Pearson correlation on aligned profiles of a gene and compound to score their connection. The profiles are obtained by averaging the replicate profiles feature-wise. We empirically found that an absolute score value greater than 0.35 indicates similar/opposite phenotypes in the gene and compound and used this for validation experiments. For the follow-up experiments of a gene, unless otherwise noted, we used a more stringent filter of 0.40 and picked the top 15 bioactive compounds that are positively correlated to the gene profile, and also 15 most negatively correlated ones. For the diversity-oriented-synthesis compounds in the set which are much less studied, we do the same, except that the top 30 in both directions are picked.

Compound annotations

Compound MOAs and target annotations were mainly acquired from the “Repurposing hub”[51] and then curated to include missing annotations from other sources, such as DrugBank[52]. The protein interaction data, which was used to assess relevance of a protein to compound targets, was collected from BioGRID[53].

Enrichment p-value estimation

We estimate the p-values of candidate compound list enrichment empirically, by counting the number of valid connections in the list, and ranking it against a null distribution. The null distribution is defined as the same count for random lists of the same size as the original list, and is sampled many times. The p-value estimation is repeated many times and the final estimation is obtained by averaging the individual estimates.

SMAD3 experiments

For SMAD3 compounds, given a limit of 10 compounds to study, we chose the top five positive matches and the top two negative matches (which were somewhat cytotoxic based on cell count in the Cell Painting assay), along with three additional negative matches (among the top 15) which were less cytotoxic. One was unavailable.

A549 lung carcinoma cells were transfected with the luciferase reporter plasmids 4xSBE-Luc to measure TGF-b/Smad3-activated transcription[54] and pRL-TK (low expressing, constitutively active Renilla luciferase under the HSV-thymidine kinase promoter) (Promega cat# E224A) to normalize for the 4xSBE Firefly luciferase values. The transfected cell lysates were processed for luciferase assays as described[56] and per manufacturer’s protocol (Promega). In brief, the cells were seeded in 24-well plates at 80% confluency and, after adhering, the media was changed to growth or starvation media (RPMI-1640 with 10% or 2% FBS

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

respectively) for 6 hours. The cells were then transfected with 200 ng 4xSBE-Luc and 100ng RI-Tk-Luc reporter plasmids per well using Lipofectamine 2000 per manufacturer recommendations (Thermo Fisher cat# 11668019). 12 hours after transfection cells were treated for 24 hours with 5 ng/ml TGF-β1 or 5 μM SB431542 to inhibit TGF-b-induced Smad activation, and either of 9 compounds at 10μM in triplicate. All cells were harvested with 200 μl of passive lysis buffer (Promega). Luciferase assays were performed using a Dual−Luciferase assay kit (Promega), and luciferase activities were quantified with a SpectraMax M5 plate luminometer (Molecular Devices) and normalized to the internal Renilla luciferase control and DMSO control.

Ras experiments

Isogenic RAS-less mouse embryonic fibroblast cell lines driven by human KRAS4b G12D, HRAS WT, or BRAF V600E alleles were plated in 384-well plates and dosed with compound or DMSO 18 hours later using an Echo acoustic liquid handler in a 10 point, 2-fold dilution in 0.2% DMSO, with 10μM as the top concentration. After 72 hours, Promega CellTiter-Glo® reagent was added, and the signal was read using Envision software. Values were normalized using day zero and DMSO control readings. Hits were determined by a one log difference in IC50 values between BRAF V600E and RAS-driven cell line responses.

Casein-kinase 1 alpha experiments

CSNK1A1 enzymatic assays were performed by mobility shift assay using the Labchip EZ Reader II (Perkin Elmer). GST-tagged human CSNK1A1 (Carna Biosciences) protein was incubated with ATP, substrate, and assay buffer (20 mM Hepes - pH 7.5, 5 mM MgCl2, and 0.01% Triton X-100). The assay reaction was initiated with 5 μM ATP, 2 mM DTT, and 1 μM Profiler Pro FL-Peptide 16 substrate (Perkin Elmer). Curve fitting and determination of AC50 values for phosphorylation inhibition were performed using Genedata.

GSK3B experiments

The compounds with a Cell Painting profile matching or opposing GSK3 overexpression were tested against GSK3α and GSK3β as previously reported[55]. Purified GSK3β or GSK3α was incubated with tested compounds in the presence of 4.3 μM of ATP (at or just below Km to study competitive inhibitors) and 1.5 μM peptide substrate (Peptide 15, Caliper) for 60 minutes at room temperature in 384-well plates (Seahorse Bioscience) in assay buffer that contained 100 mM HEPES (pH 7.5), 10 mM MgCl2, 2.5 mM DTT, 0.004% Tween-20, and 0.003% Brij-35. Reactions were terminated with the addition of 10 mM ethylenediaminetetraacetic acid (EDTA). Substrate and product were separated electrophoretically, and fluorescence intensity of the substrate and product was determined by Labchip EZ Reader II (Perkin Elmer). The kinase activity was measured as percent conversion to product. The reactions were performed in duplicate for each sample. The positive control, CHIR99021, was included in each plate and used to scale the data in conjunction with “in-plate” DMSO controls. The results were analyzed by Genedata Assay Analyzer. The percent inhibition was plotted against the compound concentration, and the IC50 value was determined from the logistic dose-response curve fitting. Values are the average of at least three experiments. Compounds were tested using a 12-point dose curve with 3-fold serial dilution starting from 33 μM. The two most active compounds were resynthesized for validation and tested along with closely related analogs (Supplemental Methods).

p38 experiments

Cell Painting profiles for two wild-type variants of p38α (MAPK14) were averaged to create a p38α Cell Painting profile. 20 compounds whose Cell Painting profile correlated positively or negatively to that of p38α overexpression were selected; we also chose 14 "non-correlated" compounds (i.e. absolute value of

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

correlation <0.2) as negative/neutral controls. The compounds were tested for their influence on p38 activity using the RPE1-p38 kinase translocation reporter (KTR) line that was previously generated[26]. This cell line has been tested and confirmed to be negative for mycoplasma contamination, but not authenticated. p38 activity is measured by phosphorylation of its substrate, MEF2C, which is preferentially phosphorylated by p38α, while p38β and p38δ contribute less[56]. RPE1-p38KTR cells were cultured in DMEM/F12 medium supplemented with 10% Fetal Bovine Serum at 37C in a humidified atmosphere with 5% CO2. 1000 cells were plated per well in 96-well plates and treated with 1μM and 10μM of each compound (n=4 well per concentration per compound, no replicates) for 48 hours. Only the middle 60 wells were used to prevent potential confounds from the edge effect. Cells were then fixed in 4% paraformaldehyde for 10min, followed by permeabilization in cold methanol at -20C for 5min. Cells were stained with 0.4 μg/mL Alexa Fluor 647 carboxylic acid, succinimidyl ester for 2hr at RT, followed by 1μg/mL DAPI for 10min at RT to facilitate the segmentation of individual cells. p38 activity in single cells was calculated using the ratio of the median intensity of the p38-KTR in a 5-pixel-wide cytoplasmic ring around the nucleus to the median intensity of the p38-KTR in the nucleus. p38 activity measurements were normalized to DMSO within the same plate and column. The Student's t-test or Kolomogorov-Smirnov (KS) test was used to assess the significance of changes in the single cell distributions of p38 activity for each compound relative to control; we note that even for the positive control known inhibitor the effect sizes are small. When reporting hits from the assay, KS test and t-test p-values were adjusted to control the false discovery rate using the Benjamini-Hochberg method, using the p.adjust(method='BH") method in R.

PPARGC1A (PGC-1α) experiments

Reporter assays: To measure PGC-1α activity related to PPARG, RT112/84 cells were obtained from the Cancer Cell Line Encyclopedia (Broad Institute, Cambridge, MA), which obtained them from the original source and performed cell line authentication. The cell line was engineered with the NanoLuc gene cloned into the 3' UTR of the FABP4 (previously described[33]) followed by stable expression of nuclear GFP (pTagGFP2-H2B, Evrogen) and tested negative for mycoplasma (MycoAlert, Lonza). Cells were plated in 384-well plates at ~10,000 cells/well and dosed with indicated compounds in the absence or presence of EC50 of PPARG agonist, rosiglitazone, using an HP D300 digital dispenser. The following day nuclei were counted for normalization (IncuCyte S3, Essen Bioscience) and the reporter activity was evaluated using the NanoGlo Luciferase Assay System (Promega). Normalized data is reported as NanoGlo arbitrary light units divided by cell number. PPARG agonist, rosiglitazone, and inhibitor, T0070907, were obtained from Tocris and included as controls.

To measure effects on PGC1a/ERRalpha, HEK293T cells purchased from ATCC were co-transfected with Gal4-ERRalpha, with and without PGC1a (pCDNA3.1-Flag-HA-PGC-1alpha[57]), kind gifts from Pere Puigserver, in combination with the Gal4 UAS reporter construct, pGL4.35 [luc2P/9XGAL4UAS/Hygro] (Promega) modified by subcloning the HSV-TK promoter into the unique HindIII site that is downstream of the 9xGal4 UAS sites, in addition to a Renilla luciferase expression vector pRL (Promega) for normalization. Cells were dosed with compounds and 24 hours later, plates were analyzed using Dual-Glo Luciferase Assay System (Promega). Normalized light units are reported as Firefly luciferase divided by Renilla luciferase. ERRalpha modulators XCT790, Daidzein, and Biochanin A (Cayman Chemical) were included as controls. 293T cells were not authenticated nor tested for mycoplasma.

High content mitochondrial motility screen: We used our previously published assay to assess mitochondrial motility[36]. Briefly, we plated E18 rat cortical neurons in the middle 60 wells of 96 well plates (Greiner) – 40,000 cells per well in 150 μl enriched Neurobasal media. Neurons were transfected with mito-DsRed at DIV7 using Lipofectamine2000 (Life Technologies). Plating and transfection were all done using an Integra VIAFLO 96/384

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

automated liquid handler. At DIV9, test compounds were added into wells to achieve a final concentration of 10 µM each (4 wells per compound), as well at 10 µM calcimycin for neg. control, and DMSO only for mock treatment. Following a 1-2 hour incubation, plates were imaged on a ArrayScan XTI (Thermo Fisher). Mitochondrial motility data was extrapolated from imaging data using a MATLAB and CellProfiler based computational pipeline. Compounds A01-A12 were tested on one plate; B01-B11 were tested separately on another plate on the same day. The experiment was repeated twice in different weeks. In the second week, TMRE was added to all wells after imaging was completed (20min, then 2 washes) and imaged to measure mitochondrial membrane potential in order to determine mitochondrial and cell health.

YAP1-related compounds

For the initial experiments, quality control of the compounds revealed that purity was 88% for A15 (BRD-K34692511-001-01-9), 81% for A05 (BRD-K28862419-001-01-9), and > 99% for E07 (BRD-K43796186-001-01-1). For subsequent experiments in the Eisinger lab, BRD-K43796186 (NB4A) was ordered from MuseChem (cat. #M189943) and for the Kiessling lab, from Ambinter (Cat # Amb2554311).

YAP1 cell culture and treatments

Eisinger lab: Murine KP230 cells, a Yap1-dependent cancer cell line, were derived from a tumor from the KP mouse model (*Kras*$^{G12/D}$; *Trp53*$^{fl/fl}$*),* as described in[44]. STS-109 UPS cells were derived from a human UPS tumor and validated by Rebecca Gladdy, MD (Sinai Health System, Toronto, Ontario, Canada). TC32 cells were a gift from Patrick Grohar, MD, PhD (Children's Hospital of Philadelphia). HT-1080, HCT-116, and HEK293T cells were purchased from ATCC. KP230, HT-1080, and HEK-293T cells were grown in DMEM with 10% FBS, 1% L-glutamine, and 1% penicillin/streptomycin (P/S). STS-109 cells were cultured in DMEM with 20% FBS, 1% L-glutamine, and 1% P/S. TC32 cells were grown in RPMI with 10% FBS, 1% L-glutamine, and 1% P/S. HCT-116 cells were cultured in McCoy's 5A medium with 10% FBS and 1% P/S. All cells were confirmed to be negative for mycoplasma contamination and were maintained in an incubator at 37C with 5% CO_2. For experimental purposes, cells were cultured for up to 20 passages before being discarded, and were grown to approximately 50% confluence to circumvent the effects of high cell density on Yap1 expression and activity. All cell lines in the Eisinger laboratory were treated with 10 µM of each inhibitor or an equivalent volume of DMSO every 24 hours for 3 days, except for STS-109 cells, which were treated daily for 8 days.

Kiessling lab: H9 hPSCs (WiCell) were maintained on vitronectin (Thermo Fisher)-coated plates in Essential 8 (E8) medium. The cells were routinely passaged using 0.5mM EDTA and treated with 5µM Y-27632 dihydrochloride (Tocris) on the first day. For testing the effects of the small molecules, H9 hPSCs were seeded at 50K cells/cm^2 on vitronectin-coated plates in E8 medium supplemented with 5µM Y-27632 dihydrochloride (day 0). On day 1, the medium was switched to E8 medium. On day 2, the medium was switched to E8 medium supplemented with the small molecules. Following overnight incubation, the cells were collected for subsequent analysis on day 3. The cells were regularly checked for Mycoplasma contaminations (Sigma Aldrich - Lookout Mycoplasma PCR Detection Kit) but were not authenticated.

Boerckel lab: Murine periosteal cells were isolated from a transgenic mouse model (CMV-Cre;R26R-rtTA^{fl}; tetO-YAP^{S127A}; C57Bl/6 strain/background) in which YAP1 can be activated in a doxycycline inducible manner (Camargo 2011). This mouse model expresses a mutant form of YAP1 (YAP^{S127A}) that escapes degradation. Cells were isolated from 3 female mice (age 15 weeks) from a 4-day old femoral fracture callus. Cells were cultured in a-MEM with 15% Fetal Bovine serum (S11550, R&D Systems), 1% GlutaMAX-I (Gibco, 35050-061) and 1% Penicillin/Streptomycin (Gibco, 15140-122).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

YAP1-related lentiviral production

Knockdown of *YAP1* in HCT-116 cells was performed with shRNAs (TRC clone IDs: TRCN0000107266 and TRCN0000107267); a scrambled shRNA was used as a negative control. shRNA plasmids (Dharmacon) were packaged using the third-generation lentiviral vector system (pVSV-G, pMDLG, and pRSV-REV; Addgene) and expressed in HEK-293T cells using Fugene 6 transfection reagent (Promega). Virus-containing supernatants were collected 24 and 48 hours after transfection and concentrated 40-fold by centrifugation with polyethylene glycol 8000.

YAP1-related proliferation assays

NB4A treatment: Cells were treated with 10 μM of each inhibitor or an equivalent volume of DMSO every 24 hours for 3-8 days, and counted with a hemocytometer with trypan blue exclusion daily (KP230, HT-1080, TC32, HCT-116), or every 2 days (STS-109).

shRNA-mediated *YAP1* knockdown: HCT-116 cells were infected with *YAP1* shRNA-encoding lentiviruses in the presence of 8 μg/mL polybrene (Sigma). Antibiotic selection (3 μg/mL puromycin) was performed after 48 hours, after which cells were cultured for an additional 48 hours. Cells were then trypsinized, seeded under puromycin-selection conditions, and counted with a hemocytometer with trypan blue exclusion on days 7, 8, and 9 post-infection.

YAP1-related qRT-PCR

For the Eisinger lab, total RNA from cultured cells was isolated with the QIAGEN RNeasy mini kit, and cDNA was synthesized with the High-Capacity RNA-to-cDNA kit (Life Technologies). qRT-PCR analysis was performed with TaqMan “best coverage” probes on a ViiA7 instrument. Hypoxanthine phosphoribosyltransferase (*HPRT*) and succinate dehydrogenase subunit A (*SDHA*) were used as endogenous controls. Relative expression was calculated using the ddCt method.

For the Kiessling lab, the RNA was extracted using TRIzol (Life Technologies) and Direct-zol™ RNA MiniPrep kit (Zymo Research) as per manufacturer instructions. The RNA was reverse transcribed using iScript cDNA synthesis kit (Bio-Rad). The qPCR was performed on CFX Connect (Bio-Rad) using iTaq Universal SYBR Green Supermix (Bio-Rad). GAPDH was used as a reference gene for normalization. The relative gene expression levels were determined using the ddCt method. The primer sequences used are listed in Supplementary Table 9.

For the Boerckel lab, to induce YAP^{S127A}, 1μM doxycycline was added to the cell culture medium for 48 hours. This was used as a positive control to compare YAP1 mRNA expression. Cells were also treated with BRD-K34692511-001-01-9 at 5μM. mRNA was isolated from cells (n=3/group/time point) at 1, 4 or 48 hours after treatment using Qiagen RNeasy Mini kit (Qiagen, 74106). cDNA was prepared as per the manufacturer’s protocol using the High-Capacity Reverse Transcription kit (Thermofisher scientific, 4368814). qPCR analysis was performed using the QuantStudio 6 Pro Real-Time PCR System.

YAP1-related reporter assay

Varelas lab: HEK293T cells purchased from ATCC were co-transfected using Lipofectamine 3000 (Thermo Fisher) with a TEAD luciferase reporter construct, 8xGTIIC-luciferase (gift from Stefano Piccolo, Addgene plasmid # 34615), a plasmid expressing Renilla Luciferase from a CMV promoter as a transfection control,

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

along with a plasmid expressing 3xFlag-tagged wild-type YAP1 from a CMV promoter (pCMV5 backbone). Following transfection the cells were immediately treated with 0.2% DMSO, 10μM NB4A, BRD-K34692511 or BRD-K28862419 and then lysed 48 hours later. Lysates were examined using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's protocol and measured using a SpectraMax iD3 plate reader (Molecular Devices). Firefly Luciferase activity from the TEAD reporter was normalized to Renilla Luciferase activity and then plotted as relative values. Mycoplasma tests are routinely performed, but cells were not recently authenticated.

YAP1-related RNA-sequencing and data analysis

Total RNA from cultured cells was isolated with the QIAGEN RNeasy Mini Kit with on-column DNase digestion. RNA quality checks were performed with an Agilent 2100 Bioanalyzer (Eukaryotic Total RNA Nano kit). Library preparation (500 ng input RNA) was performed with the NEBNext Poly(A) mRNA Magnetic Isolation Module (#E7490) with SPRIselect Beads (Beckman Coulter), the NEBNext Ultra II Single-End RNA Library Prep kit (#7775S), and the NEBNext Multiplex Oligos for Illumina (Index Primers Set 1) according to the manufacturer's instructions. Library size was confirmed with an Agilent 2100 Bioanalyzer (DNA1000 chip). Pooled libraries were diluted to 1.8 pM (concentrations checked with the Qubit Fluorometer high-sensitivity assay, Thermo Fisher), and sequenced on an Illumina NexSeq 500 instrument with the NexSeq 500 75-cycle high-output kit.

For data analysis, FASTQ files were generated with the *bcl2fastq* command line program (Illumina). Transcript alignment was performed with Salmon[58]. Differential expression analysis (NB4A- vs. DMSO-treated cells) was performed with the DESeq2 R package. DESeq2 "stat" values for each gene were used as inputs to pre-ranked GSEA, where enrichment was tested against the Hallmark gene sets from the Molecular Signatures Database (MSigDB). Access to sequencing data is discussed in the data availability section.

YAP1-related Western blotting

For the Kiessling lab, the cells were lysed in RIPA buffer (Pierce) supplemented with Halt Protease inhibitor cocktail and Halt Phosphatase inhibitor cocktail (Thermo Fisher). The Eisinger lab lysed cells in hot Tris-SDS buffer (pH 7.6) and boiled for 5 minutes at 95ºC. The protein concentration of each sample was quantified using the Pierce BCA protein assay (Thermo Fisher). The proteins were resolved by SDS-PAGE and transferred to PVDF membranes using the Trans-Blot Turbo Transfer system (Bio-Rad). The membranes were blocked in 5% non-fat milk in TBS-T for up to 1 hour at room temperature and incubated with primary antibodies in 5% bovine serum albumin in TBS-T overnight at 4ºC. Then, the membranes were incubated with HRP-conjugated anti-rabbit IgG secondary antibodies at 1:10000 (Kiessling lab; Jackson ImmunoResearch Laboratories, #111-035-003) or 1:2500 (Eisinger lab; Cell Signaling Technology [CST] #7074) for 1 hour at RT and developed in the ChemiDoc MP Imaging system (Kiessling lab) or on autoradiography film (Eisinger lab) using ECL Prime reagent (Amersham). The band intensities in immunoblots were quantified with Image Lab software. The primary antibodies and dilutions used are: anti-YAP1 (CST 4912S and CST 14074 [clone D8H1X]) at 1:1000, anti-phospho-YAP1-S127 (CST 4911S) at 1:1000, and anti-GAPDH (CST 5174 and CST 2118 [clone 14C10]) at 1:15000 and 1:1000, respectively. Primary antibodies were validated commercially in cells both wild-type and deficient (e.g., knockout) for the gene/protein of interest. YAP1-related immunofluorescence and image analysis

For the Eisinger lab, cells grown on poly-L-lysine-coated chamber slides were fixed in 4% PFA (15 minutes at room temperature), permeabilized with 0.5% Triton-X100/PBS (15 minutes at room temperature), and blocked with 5% goat serum (Vector Laboratories S-1000; 1 hour at room temperature). Cells were then incubated with anti-Yap1 primary antibodies (CST #14074 [clone D8H1X]; 1:1000) diluted in blocking buffer overnight at 4ºC.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Subsequently, cells were incubated with Alexa Fluor 488-conjugated secondary antibodies (4 ug/mL in blocking buffer; Thermo Fisher Scientific #A-11008) for 1 hour at room temperature. Coverslip mounting was performed with ProLong Gold Antifade reagent with DAPI. Images (5 fields per condition for each of 3 independent experiments) were acquired with a Nikon Eclipse Ni microscope and Nikon NES Elements software. Image analysis was performed with Fiji as follows: For nuclear staining intensity, watershed analysis of DAPI channel images (8-bit) was performed to “separate” nuclei that appeared to be touching. Nuclei were then converted to regions of interest (ROIs) that were “applied” to the corresponding GFP channel image (8-bit format). Analysis of staining intensity in these nuclear ROIs was then performed, excluding objects smaller than 100 pixels^2 (integrated density normalized to number of nuclei). A similar process was followed to determine whole-cell staining intensity: using 8-bit GFP channel images, cells were distinguished from background via thresholding, and converted to ROIs that were applied back to the 8-bit GFP channel images. Analysis of staining intensity (integrated density normalized to number of nuclei) was then performed in these ROIs, excluding objects smaller than 500 pixels^2. The ratio of nuclear to total Yap1 expression was determined after subtracting out background GFP signal from no-primary antibody controls.

For the Kiessling lab, the cells were fixed with 4% formaldehyde for 15 mins at room temperature. The cells were permeabilized and blocked with PBS containing 2% BSA and 0.1% Triton-X100. The cells were incubated with a primary antibody against YAP1 (Santa Cruz Biotechnology, sc-101199) at 1:200 dilution in a blocking buffer overnight at 4°C. Then, the cells were incubated with a goat anti-mouse Alexa Fluor 488 conjugated secondary antibody (Thermo Fisher, #A11001) at 1:1000 dilution for 1 hour at room temperature. The nuclei were counterstained with DAPI dilactate (Molecular Probes). Images were collected with Olympus FV1200 microscope and analyzed with CellProfiler. Briefly, nuclei and cell bodies were segmented using DAPI and YAP channels respectively. The cell cytoplasm was determined as the region outside nuclei but within the cell bodies. Then, the ratio of mean intensity of YAP in the nucleus to cytoplasm was calculated to determine YAP translocation.

Acknowledgements

The authors thank the researchers who originally helped produce the published data used in this analysis, including the Broad Institute LINCS team and Cancer Program. We appreciate helpful discussions with our colleagues, including Pere Puigserver, Evan Rosen, and Amit Majithia.

The Carpenter-Singh lab team was supported by the National Institutes of Health (NIH R35 GM122547 to AEC). R.K. is supported by the Canadian Institutes of Health Research (343437) and the Natural Sciences and Engineering Research Council of Canada (RGPIN-2015-05805). C.T. is supported by a University of Toronto Open Fellowship. The Eisinger lab team was supported by the National Institutes of Health (R01CA229688 to TSKE and T32-HL007971 to AMF), and is grateful to the University of Pennsylvania High-Performance Computing Facility for providing computational capacity and data archiving services. The Kiessling lab team was supported by the National Institutes of Health (U01CA231079 to LK). The Boerckel lab team was supported by the National Institutes of Health (R01AR073809 to JDB) and the National Science Foundation (CMMI: 15-48571 to JDB). S.M.C. is supported by the National Institutes of Health (K08CA230220). Turbyville and Rigby are supported with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Competing interests

AEC has ownership interest in Recursion, a publicly-traded biotech company using images for drug discovery. JTG reports receiving a commercial research grant from Bayer AG. SMC reports receiving research funding from Bayer and Calico Life Sciences.

References

1. Roessler, H. I., Knoers, N. V. A. M., van Haelst, M. M. & van Haaften, G. Drug Repurposing for Rare Diseases. *Trends Pharmacol. Sci.* **42**, 255–267 (2021).
2. Wouters, O. J., McKee, M. & Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA* **323**, 844–853 (2020).
3. Moffat, J. G., Vincent, F., Lee, J. A., Eder, J. & Prunotto, M. Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nat. Rev. Drug Discov.* **16**, 531–543 (2017).
4. Schneider, P. *et al.* Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug Discov.* **19**, 353–364 (2020).
5. Vamathevan, J. *et al.* Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* **18**, 463–477 (2019).
6. Bray, M.-A. *et al.* Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* **11**, 1757–1774 (2016).
7. Gustafsdottir, S. M. *et al.* Multiplex cytological profiling assay to measure diverse cellular states. *PLoS One* **8**, e80999 (2013).
8. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437–1452.e17 (2017).
9. Wawer, M. J. *et al.* Toward performance-diverse small-molecule libraries for cell-based phenotypic screening using multiplexed high-dimensional profiling. *Proc. Natl. Acad. Sci. U. S. A.* (2014) doi:10.1073/pnas.1410933111.
10. Lapins, M. & Spjuth, O. Evaluation of Gene Expression and Phenotypic Profiling Data as Quantitative

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Descriptors for Predicting Drug Targets and Mechanisms of Action. *Cold Spring Harbor Laboratory* 580654 (2019) doi:10.1101/580654.

11. Becker, T. *et al.* Predicting compound activity from phenotypic profiles and chemical structures. *Cold Spring Harbor Laboratory* 2020.12.15.422887 (2020) doi:10.1101/2020.12.15.422887.
12. Hughes, R. E., Elliott, R. J. R., Dawson, J. C. & Carragher, N. O. High-content phenotypic and pathway profiling to advance drug discovery in diseases of unmet need. *Cell Chem Biol* **28**, 338–355 (2021).
13. Proschak, E., Stark, H. & Merk, D. Polypharmacology by Design: A Medicinal Chemist's Perspective on Multitargeting Compounds. *J. Med. Chem.* **62**, 420–444 (2019).
14. Lin, A. *et al.* Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Sci. Transl. Med.* **11**, (2019).
15. Bunnage, M. E., Chekler, E. L. P. & Jones, L. H. Target validation using chemical probes. *Nat. Chem. Biol.* **9**, 195–199 (2013).
16. Rohban, M. H. *et al.* Systematic morphological profiling of human gene and allele function via Cell Painting. *Elife* **6**, (2017).
17. Chandrasekaran, S. N., Ceulemans, H., Boyd, J. D. & Carpenter, A. E. Image-based profiling for drug discovery: due for a machine-learning upgrade? *Nat. Rev. Drug Discov.* 1–15 (2020).
18. Bray, M.-A. *et al.* A dataset of images and morphological profiles of 30 000 small-molecule treatments using the Cell Painting assay. *Gigascience* **6**, 1–5 (2017).
19. Drosten, M. *et al.* Genetic analysis of Ras signalling pathways in cell proliferation, migration and survival. *EMBO J.* **29**, 1091–1104 (2010).
20. Davis, M. I. *et al.* Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **29**, 1046–1051 (2011).
21. Klaeger, S. *et al.* The target landscape of clinical kinase drugs. *Science* **358**, (2017).
22. Shanware, N. P., Williams, L. M., Bowler, M. J. & Tibbetts, R. S. Non-specific in vivo inhibition of CK1 by the pyridinyl imidazole p38 inhibitors SB 203580 and SB 202190. *BMB Rep.* **42**, (2009).
23. MAPK14 - ClinicalTrials.Gov.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

https://clinicaltrials.gov/ct2/results?term=MAPK14&Search=Apply&age_v=&gndr=&type=&rslt=.

24. Martínez-Limón, A., Joaquin, M., Caballero, M., Posas, F. & de Nadal, E. The p38 Pathway: From Biology to Cancer Therapy. *Int. J. Mol. Sci.* **21**, (2020).
25. Regot, S., Hughey, J. J., Bajar, B. T., Carrasco, S. & Covert, M. W. High-sensitivity measurements of multiple kinase activities in live single cells. *Cell* **157**, 1724–1734 (2014).
26. Liu, S. *et al.* Size uniformity of animal cells is actively maintained by a p38 MAPK-dependent regulation of G1-length. *Elife* **7**, (2018).
27. Kaufman, T. *et al.* Visual barcodes for multiplexing live microscopy-based assays. (2020) doi:10.21203/rs.3.rs-67883/v1.
28. National Center for Biotechnology Information. PubChem Bioassay Record for AID 651723, Source: Broad Institute. https://pubchem.ncbi.nlm.nih.gov/bioassay/651723.
29. Lee, H.-J., Su, Y., Yin, P.-H., Lee, H.-C. & Chi, C.-W. PPAR(gamma)/PGC-1(alpha) pathway in E-cadherin expression and motility of HepG2 cells. *Anticancer Res.* **29**, 5057–5063 (2009).
30. Luo, C., Widlund, H. R. & Puigserver, P. PGC-1 Coactivators: Shepherding the Mitochondrial Biogenesis of Tumors. *Trends Cancer Res.* **2**, 619–631 (2016).
31. Halling, J. F. & Pilegaard, H. PGC-1α-mediated regulation of mitochondrial function and physiological implications. *Appl. Physiol. Nutr. Metab.* **45**, 927–936 (2020).
32. Handschin, C. & Spiegelman, B. M. Peroxisome proliferator-activated receptor gamma coactivator 1 coactivators, energy homeostasis, and metabolism. *Endocr. Rev.* **27**, 728–735 (2006).
33. Goldstein, J. T. *et al.* Genomic Activation of PPARG Reveals a Candidate Therapeutic Axis in Bladder Cancer. *Cancer Res.* **77**, 6987–6998 (2017).
34. Usui, M., Uno, M. & Nishida, E. Src family kinases suppress differentiation of brown adipocytes and browning of white adipocytes. *Genes Cells* **21**, 302–310 (2016).
35. Nierenberg, A. A. *et al.* Peroxisome Proliferator-Activated Receptor Gamma Coactivator-1 Alpha as a Novel Target for Bipolar Disorder and Other Neuropsychiatric Disorders. *Biol. Psychiatry* **83**, 761–769 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

36. Shlevkov, E. *et al.* A High-Content Screen Identifies TPP1 and Aurora B as Regulators of Axonal Mitochondrial Transport. *Cell Rep.* **28**, 3224–3237.e5 (2019).

37. Dey, A., Varelas, X. & Guan, K.-L. Targeting the Hippo pathway in cancer, fibrosis, wound healing and regenerative medicine. *Nat. Rev. Drug Discov.* **19**, 480–494 (2020).

38. Nishio, M. *et al.* Dysregulated YAP1/TAZ and TGF-β signaling mediate hepatocarcinogenesis in Mob1a/1b-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E71–80 (2016).

39. Zaltsman, Y., Masuko, S., Bensen, J. J. & Kiessling, L. L. Angiomotin Regulates YAP Localization during Neural Differentiation of Human Pluripotent Stem Cells. *Stem Cell Reports* **12**, 869–877 (2019).

40. Musah, S. *et al.* Substratum-induced differentiation of human pluripotent stem cells reveals the coactivator YAP is a potent regulator of neuronal specification. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13805–13810 (2014).

41. Zhao, B. *et al.* Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. *Genes Dev.* **21**, 2747–2761 (2007).

42. Ye, S. *et al.* YAP1-Mediated Suppression of USP31 Enhances NFκB Activity to Promote Sarcomagenesis. *Cancer Res.* **78**, 2705–2720 (2018).

43. Hsu, J. H. & Lawlor, E. R. BMI-1 suppresses contact inhibition and stabilizes YAP in Ewing sarcoma. *Oncogene* **30**, 2077–2085 (2011).

44. Eisinger-Mathason, T. S. K. *et al.* Deregulation of the Hippo pathway in soft-tissue sarcoma promotes FOXM1 expression and tumorigenesis. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3402–11 (2015).

45. Elias, K. M. *et al.* Epigenetic remodeling regulates transcriptional changes between ovarian cancer and benign precursors. *JCI Insight* **1**, (2016).

46. Ha, J., Park, H., Park, J. & Park, S. B. Recent advances in identifying protein targets in drug discovery. *Cell Chem Biol* **28**, 394–423 (2021).

47. Keenan, A. B. *et al.* The Library of Integrated Network-Based Cellular Signatures NIH Program: System-Level Cataloging of Human Cells Response to Perturbations. *Cell Syst* **6**, 13–24 (2018).

48. Mullard, A. Machine learning brings cell imaging promises into focus. *Nat. Rev. Drug Discov.* **18**, 653–655

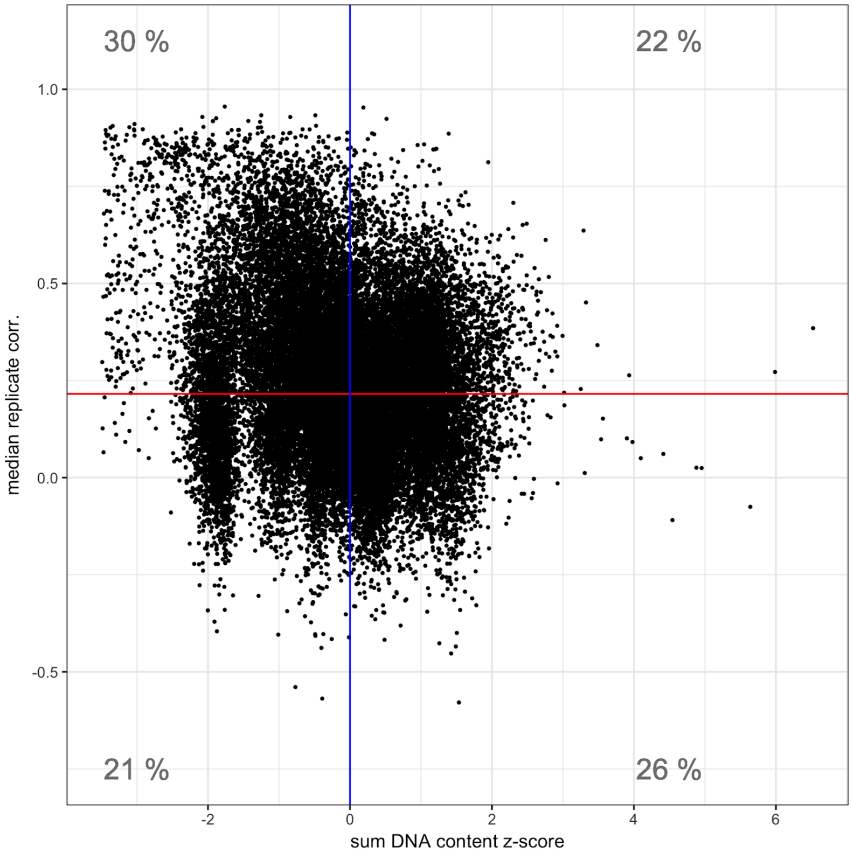
bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(2019).

49. Pratapa, A., Doron, M. & Caicedo, J. C. Image-based cell phenotyping with deep learning. *Curr. Opin. Chem. Biol.* **65**, 9–17 (2021).
50. Chandrasekaran, S. N. *et al.* Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. (2021).
51. REPURPOSING. https://clue.io/repurposing-app.
52. DrugBank Online. https://www.drugbank.ca.
53. Lab, M. T. BioGRID Interaction Database. https://thebiogrid.org.
54. Feng, X. H., Zhang, Y., Wu, R. Y. & Derynck, R. The tumor suppressor Smad4/DPC4 and transcriptional adaptor CBP/p300 are coactivators for smad3 in TGF-beta-induced transcriptional activation. *Genes Dev.* **12**, 2153–2163 (1998).
55. Wagner, F. F. *et al.* Exploiting an Asp-Glu ‘switch’ in glycogen synthase kinase 3 to design paralog-selective inhibitors for use in acute myeloid leukemia. *Sci. Transl. Med.* **10**, (2018).
56. Zetser, A., Gredinger, E. & Bengal, E. p38 mitogen-activated protein kinase pathway promotes skeletal muscle differentiation. Participation of the Mef2c transcription factor. *J. Biol. Chem.* **274**, 5193–5200 (1999).
57. Rodgers, J. T. *et al.* Nutrient control of glucose homeostasis through a complex of PGC-1alpha and SIRT1. *Nature* **434**, 113–118 (2005).
58. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
59. Wang, X. & Schwarz, T. L. Chapter 18 Imaging Axonal Transport of Mitochondria. in *Methods in Enzymology* vol. 457 319–333 (Academic Press, 2009).

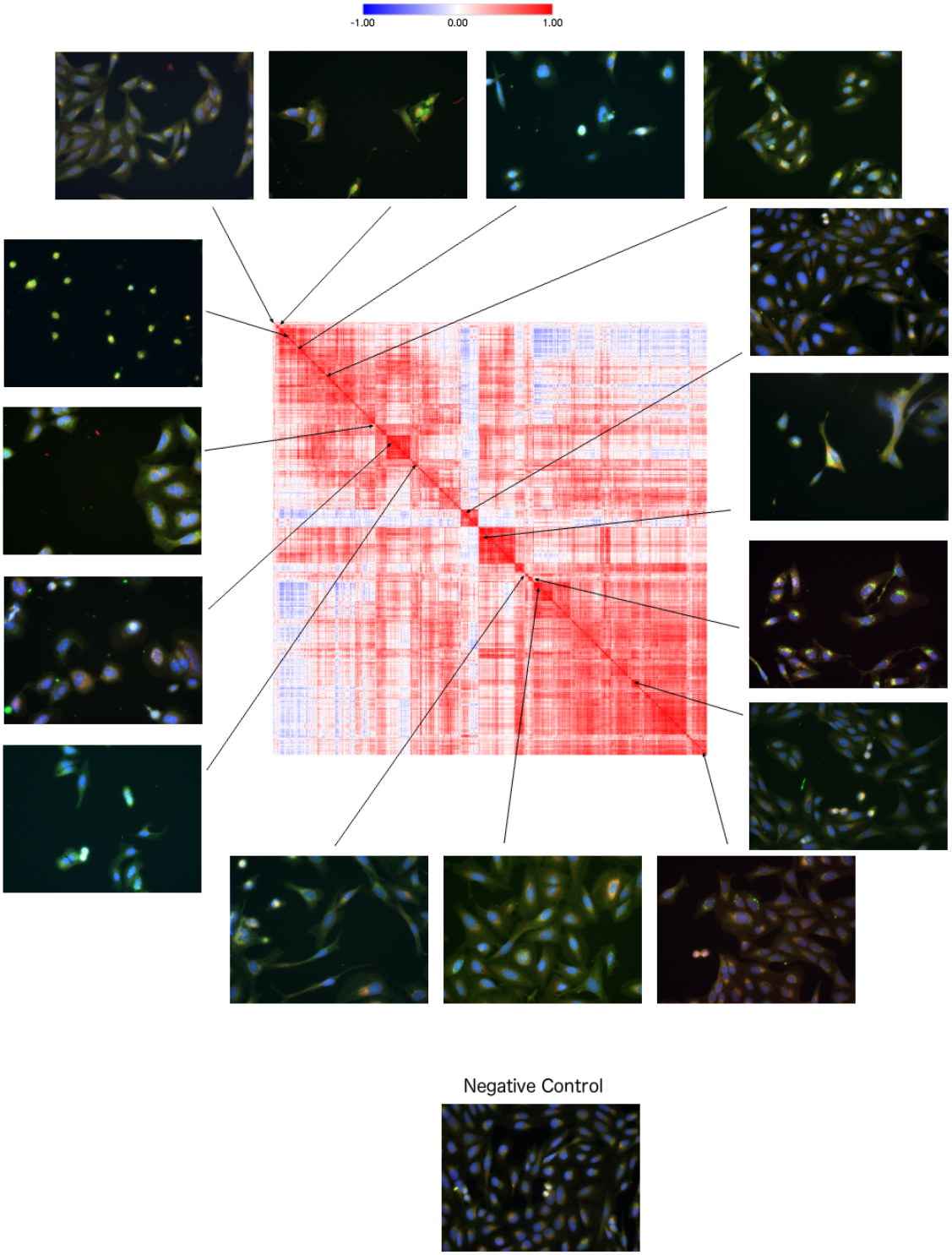
bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Extended Data Figures



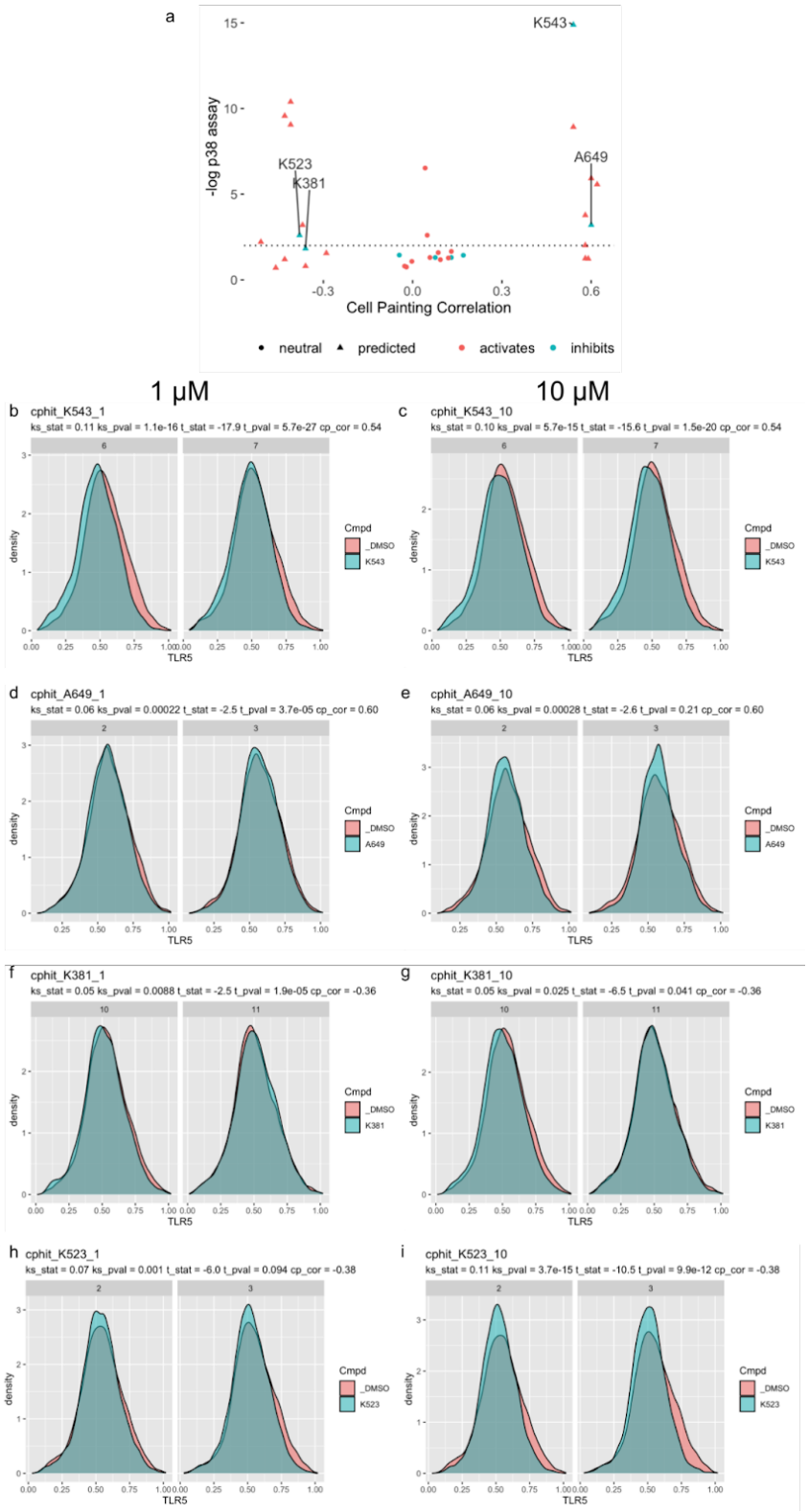
Extended Data Figure 1: Relationship between detectable Cell Painting profiles and cell proliferation rules out toxicity being a single, dominant phenotype. *The Y axis shows the replicate correlation, which is high for compounds that produce detectable morphological phenotypes in the Cell Painting assay. 52% of the compounds have a replicate correlation higher than the 95th percentile of non-replicate correlations (red line) and thus are considered to have a detectable phenotype. The X axis shows the z-score for the sum of DNA content, where higher values represent higher cell proliferation. Although the ratio of low-proliferation samples (left of blue line) with a detectable phenotype (30% vs. 21%) is higher than for high-proliferation samples (right of blue line) (22% vs. 26%), it is clear that impact on cell proliferation does not explain the majority of detectable morphological phenotypes.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



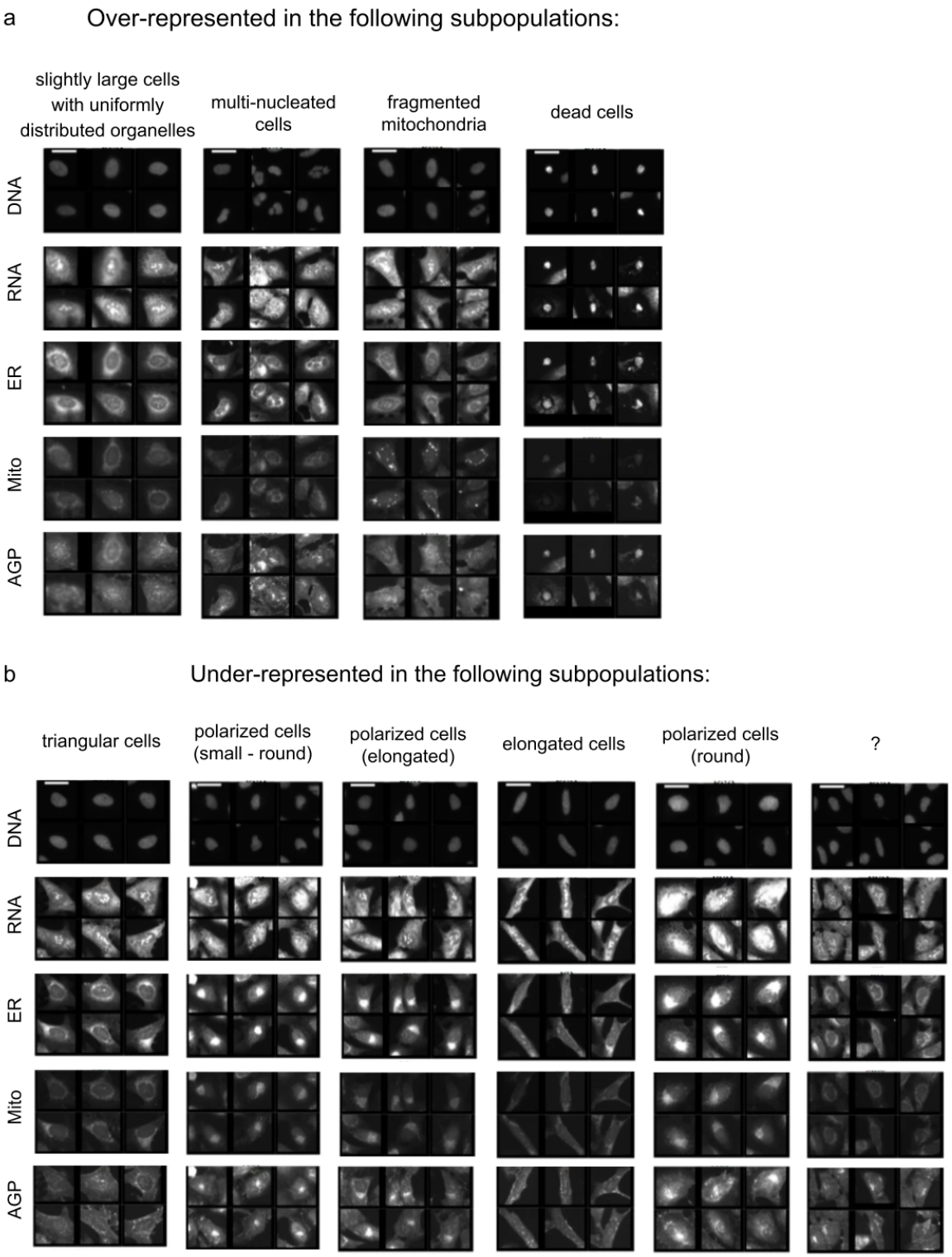
Extended Data Figure 2: Compounds yielding a low cell count may be toxic or proliferation-impeding but they display many distinguishable phenotypes. *Low-cell-proliferation or potentially toxic compounds (with the z-score for the sum of DNA content less than -3) are clustered, and show many different types of toxic phenotype. Various tight clusters mean the assay is specific and has sufficient resolution to distinguish types of toxicity.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



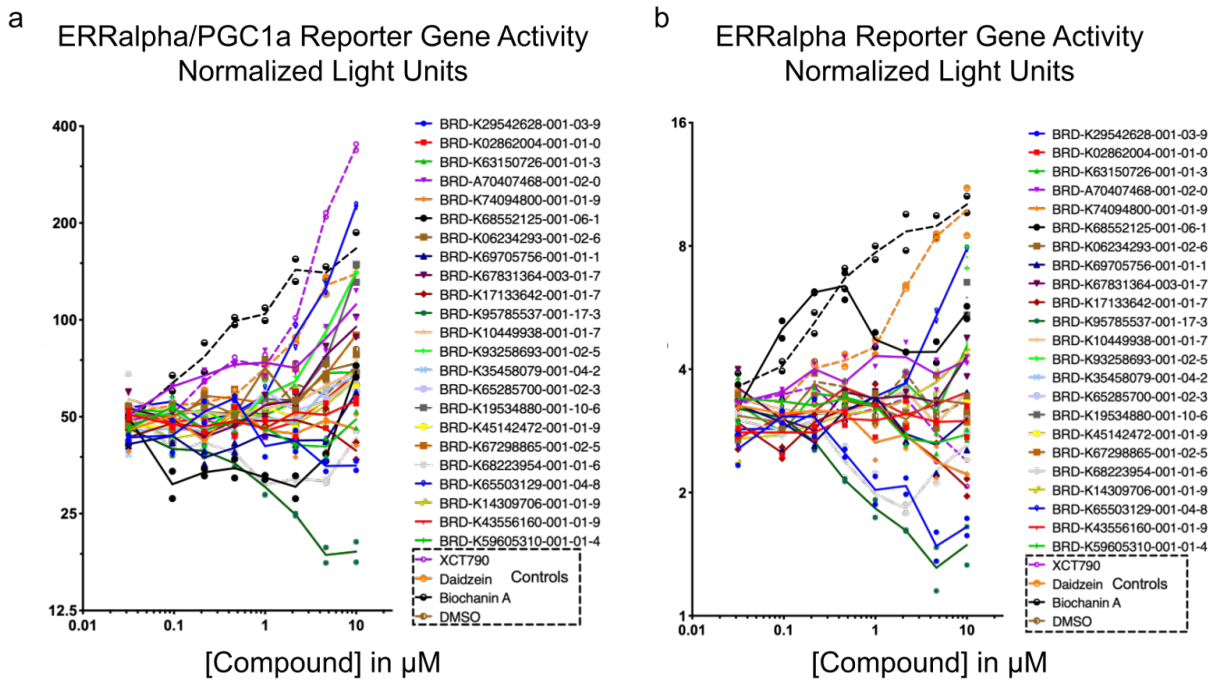
Extended Data Figure 3: Predicted compounds impact p38 activity in a single-cell reporter assay. *a) The same experiment as shown in Figure 2 is shown here, except using a Kolmogorov-Smirnoff (KS) analysis to detect differences in distribution instead of shifts in the mean. This raises an additional hit, K523. b-i) Single cell distribution plots show the shifts induced, at both 1μM and 10μM, by a known inhibitor of p38, SB202190 (b-c), by the two hits from the t-test in Figure 2 (d-g) and by the hit from the KS test (h-i).*

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



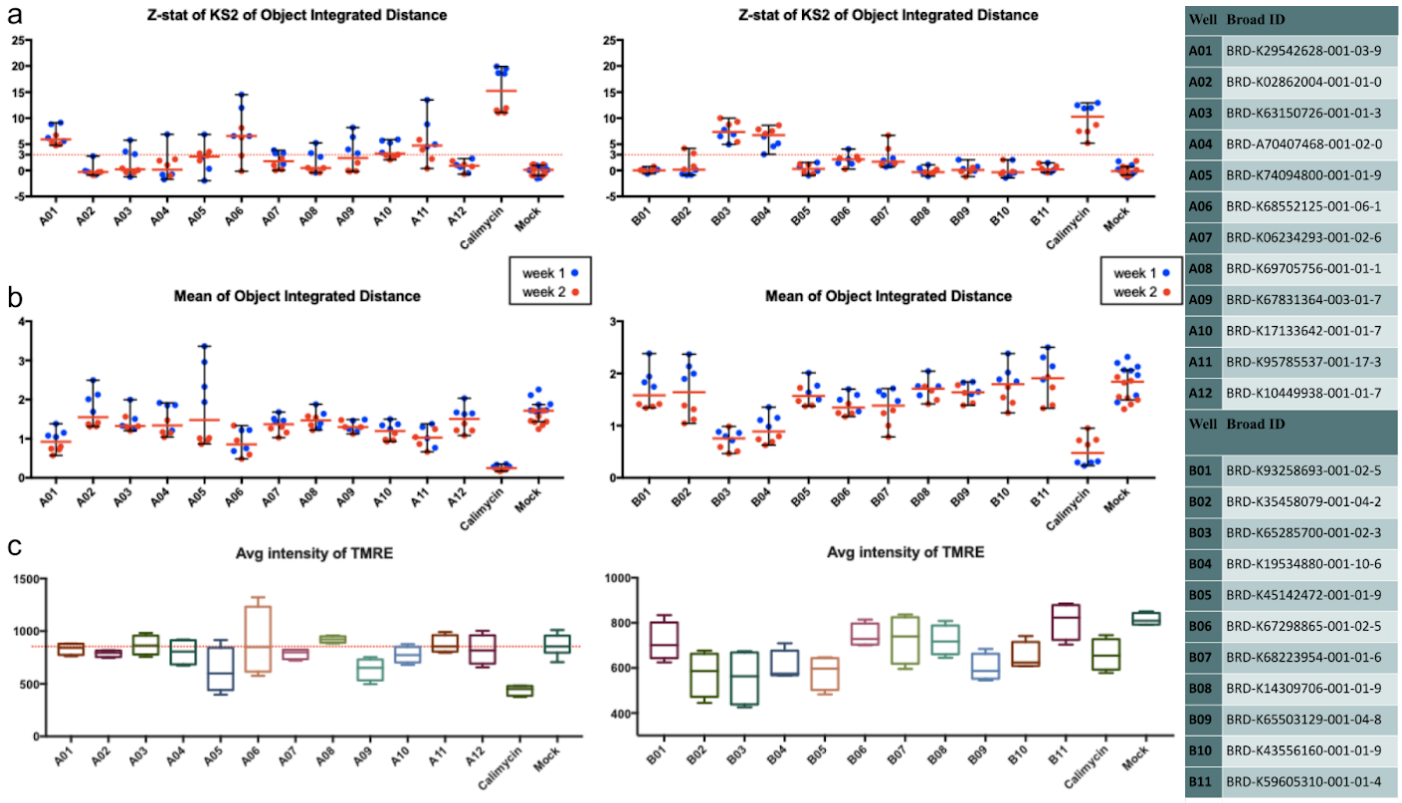
Extended Data Figure 4: Certain subpopulations of cells are over- or under-represented when PPARGC1A is overexpressed. *Following the procedure described previously[16] we clustered cells based on their morphological profiles, then identified which subpopulations were (a) over- or (b) under-represented when PPARGC1A is overexpressed. Scale bars = 39.36 μm.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Extended Data Figure 5: Compounds predicted to influence pathways containing PGC1α impact an ERRa reporter assay in 293T cells. *In this reporter system, a mammalian one-hybrid fusion protein containing the Gal4 DNA binding domain and the ERR alpha ligand binding domain is co-expressed with the Firefly luciferase gene under control of the Gal4 Upstream Activating Sequence. Renilla luciferase was included for normalization. The assay was performed in the presence (a) or absence (b) of ectopically expressed PGC1α; their behavior being similar in these two conditions suggests, but does not prove, that the compounds do not directly target PGC1α but instead modulate other targets in the relevant pathway, consistent with having been discovered by the morphological matching approach which assesses impact on the cell system rather than a particular desired target.*

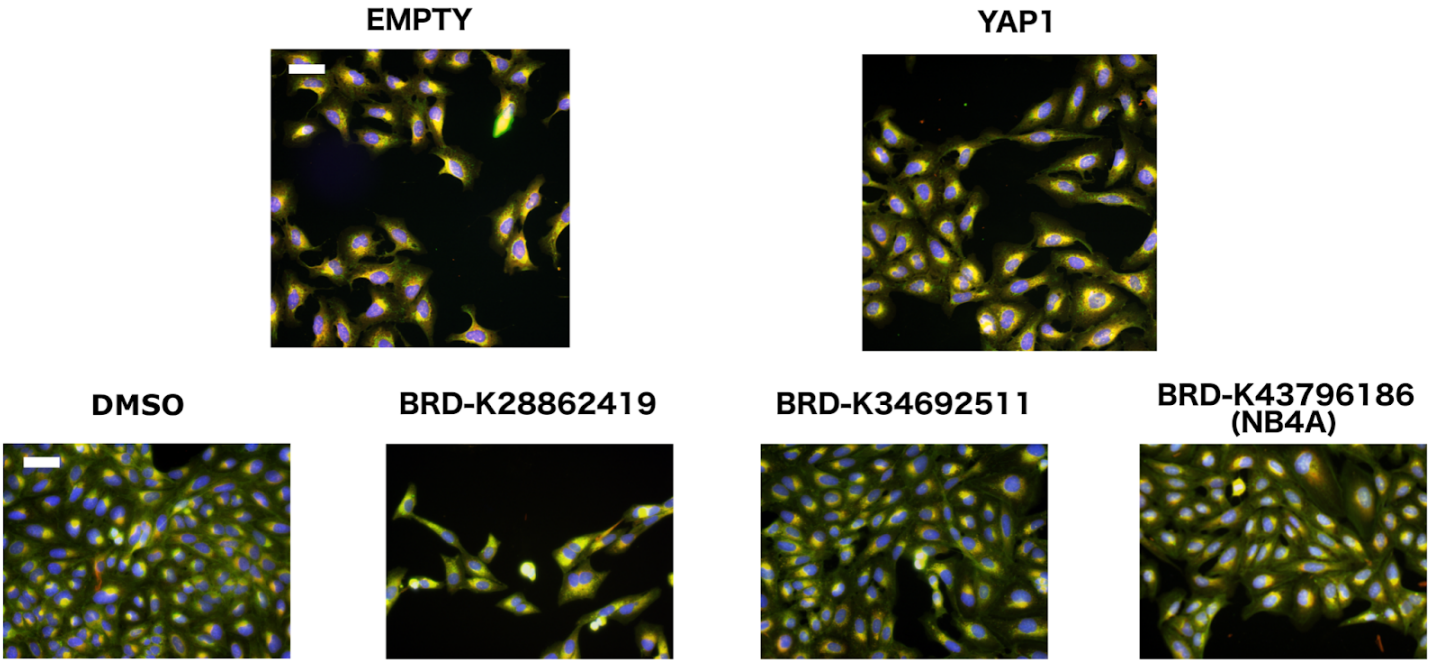
bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Well	Broad ID
A01	BRD-K29542628-001-03-9
A02	BRD-K02862004-001-01-0
A03	BRD-K63150726-001-01-3
A04	BRD-A70407468-001-02-0
A05	BRD-K74094800-001-01-9
A06	BRD-K68552125-001-06-1
A07	BRD-K06234293-001-02-6
A08	BRD-K69705756-001-01-1
A09	BRD-K67831364-003-01-7
A10	BRD-K17133642-001-01-7
A11	BRD-K95785537-001-17-3
A12	BRD-K10449938-001-01-7
Well	Broad ID
B01	BRD-K93258693-001-02-5
B02	BRD-K35458079-001-04-2
B03	BRD-K65285700-001-02-3
B04	BRD-K19534880-001-10-6
B05	BRD-K45142472-001-01-9
B06	BRD-K67298865-001-02-5
B07	BRD-K68223954-001-01-6
B08	BRD-K14309706-001-01-9
B09	BRD-K65503129-001-04-8
B10	BRD-K43556160-001-01-9
B11	BRD-K59605310-001-01-4

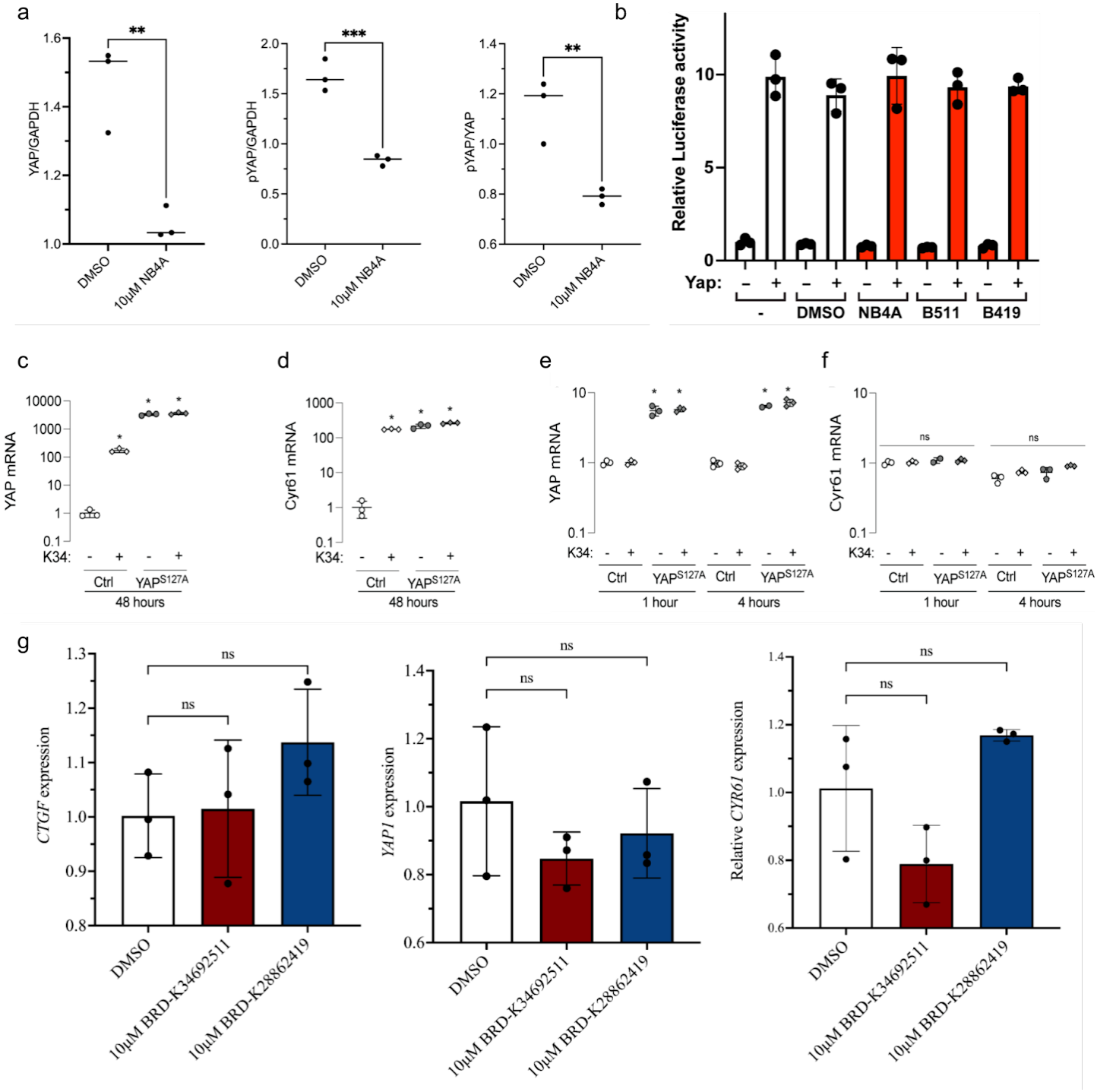
Extended Data Figure 6: Predicted compounds impact a mitochondrial motility assay in rat cortical neurons. *(a) For most compounds, the integrated distance traveled for each motile mitochondrion (the length of travel, or the sum of all movements, including changes in direction) is comparable to the negative control (Mock), but a few (A01, A06, A10, A11, B03, and B04) consistently have a z-score >3, as does the positive control, Calcimycin, a calcium ionophore that arrests mitochondria[59]. Two separate experiments are plotted (week 1 in blue and week 2 in red), and the values are the Z-prime factor of the Kolmogorov-Smirnov (KS) statistic calculated for each compound. The red line indicates the median +- 95% confidence interval. (b) Mean values of the mitochondrial distance; these are the values that underlie the statistical analysis in (a). The red line indicates the median +- 95% confidence interval. (c) The average intensity of TMRE reflects the mitochondrial membrane potential, a measure of mitochondrial function. Boxplots show the median and 25th/75th percentiles, with whiskers showing the most extreme observation less than or equal to the upper hinge + 1.5 * inter-quartile range. Interestingly, A01, A06 and A11 all show normal levels of TMRE staining, suggesting a specific effect on mitochondrial motility rather than a more general decrease in neuronal or mitochondrial health. This cannot be said for B03 and B04 (and A10 to a lesser extent), which apparently reduce membrane potential, although additional validation with TMRE is needed to conclude that they are in fact detrimental to cell health. Of note, four of these compounds were also active in the PPARG reporter assay (Figure 3c): A01 and A11 are structurally related molecules of the pyrazolo-pyrimidine family, 1-Naphthyl-PP1 and PP2, which are Src family kinase inhibitors with additional targets including TGFbeta receptors and others. A06 is Phorbol myristate acetate (aka TPA, PMA). B09 is annotated as an HSP-90 inhibitor CCT-018159. 23 compounds were tested because one of the original 24 tested in Figure 3c became unavailable.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Extended Data Figure 7: Cell Painting images related to the YAP1 pathway in U2OS cells. *Top: Cell Painting images for YAP1 overexpression compared to negative control (EMPTY, same image as in Figure 1c). Overexpressing YAP1 produces elongated cells with more cell protrusions, lower RNA staining, and disjoint, bright mitochondria patterns. Bottom: Cell Painting images for the negative control (DMSO, same image as in Figure 1c) and three compounds that correlated negatively or positively to the YAP1 overexpression profile. NB4A (BRD-K43796186) was positively correlated and the other two negatively correlated. Scale bars = 60 μm.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



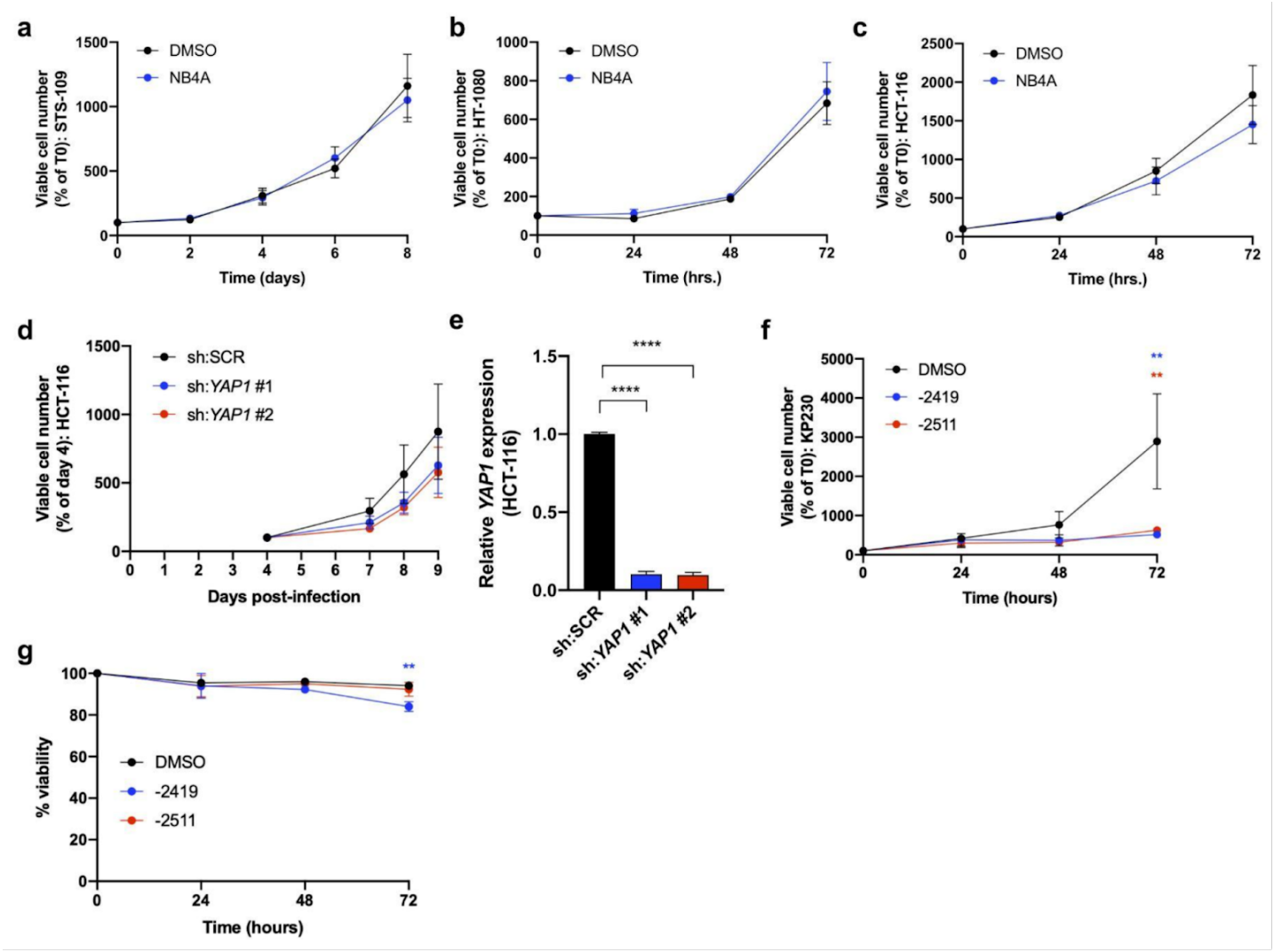
830

831 ***Extended Data Figure 8: Analysis of selected compounds in various YAP-related contexts.***
832 *a) Quantification of relative levels of total YAP1 and phospho-YAP1 in H9 hPSCs after treatment with DMSO or*
833 *NB4A for 24 hours. **P<0.01; ***P<0.001 (Two-tailed student's t-test). Mean ± SD. n = 3. A representative*
834 *example western blot is shown in Figure 4c. b) A TEAD luciferase reporter was co-transfected with or without a*
835 *Yap expression construct into HEK293T cells followed by treatment for 48 hours with DMSO or the indicated*
836 *compounds, which appear to have no effect. The data shown are the average of three samples within a*
837 *representative experiment ± SEM. c-f) BRD-K34692511 upregulates YAP1 and target-gene mRNA levels in*
838 *murine periosteal cells: c, d) YAP1 and Cyr61 mRNA levels in murine periosteal cells after 48 hours of*

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

*treatment with BRD-K34692511 (K34) in the presence or absence of doxycycline-induced YAP^{S127A}. e, f) YAP1 and Cyr61 mRNA levels after 1 and 4 hours of treatment. Gene expression was evaluated by one and two-way ANOVA with Tukey post hoc test n=3/group/time-point. * indicates p<0.05 compared to untreated controls. g) BRD-K28862419 and BRD-K34692511 did not dramatically impact mRNA levels of Hippo pathway members in hPSCs. Relative transcript levels of YAP1, CTGF, and CYR61 from H9 hPSCs treated with DMSO, BRD-K28862419, or BRD-K34692511 for 24 hrs. Error bars represent mean + SEM, from n=3 biological replicates (one-way ANOVA with Dunnett multiple comparison test).*

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.29.454377; this version posted July 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



850

851 ***Extended Data Figure 9: Predicted Hippo pathway-modulating compounds impact proliferation in a cell***
852 ***type-specific manner.*** *a, b) Growth curves of YAP1-dependent human sarcoma cells*[42,44] *treated with 10 μM*
853 *NB4A or DMSO control. c) Growth curve of HCT-116 colon cancer cells treated with 10 μM NB4A or DMSO*
854 *control. a-c are not significantly different at any time point (2-way ANOVA with Sidak's multiple comparisons*
855 *test). n = 3. Mean ± SEM. d) Growth curve of HCT-116 cells infected with YAP1-targeting shRNAs or*
856 *scrambled shRNA control (sh:SCR); no conditions were significantly different at any time point (vs. sh:SCR;*
857 *2-way ANOVA with Dunnett's multiple comparisons test). n = 3. Mean ± SEM. e) Relative YAP1 expression in*
858 *the cells depicted in panel d ****P<0.0001 vs. sh:SCR (1-way ANOVA with Dunnett's multiple comparisons*
859 *test). f) Growth curves of KP230 cells treated with 10 μM BRD-K28862419, BRD-K34692511, or DMSO*
860 *control. **P<0.01 vs. DMSO (72 hrs.; 2-way ANOVA with Dunnett's multiple comparisons test). n = 2 Mean ±*
861 *SEM. g) Percent viability of KP230 cells depicted in panel f **P<0.01 vs. DMSO (72 hrs.; 2-way ANOVA with*
862 *Dunnett's multiple comparisons test). n = 3. Mean ± SEM. For panels a, b, c, f, and g, cells were treated with*
863 *10 μM of the indicated inhibitor daily for 72 hours.*