Title: Virtual screening for small molecule pathway regulators by image profile matching

Authors: Mohammad H. Rohban[1], Ashley M. Fuller[2], Ceryl Tan[3,4], Jonathan T. Goldstein[5], Deepsing Syangtan[6], Amos Gutnick[7,8], Madhura P. Nijsure[9], Megan Rigby[10], Joshua R Sacher[11], Steven M. Corsello[5,12], Grace B. Peppler[13], Marta Bogaczynska[14], Andrew Boghossian[5], Gabrielle E. Ciotti[2], Ann DeVine[2], Minh Doan[1], Jennifer P. Gale[11], Rik Derynck[14], Thomas Turbyville[10], Joel D. Boerckel[9], Shantanu Singh[1], Laura L. Kiessling[6], Thomas L. Schwarz[7,8], Xaralabos Varelas[13], Florence F. Wagner[11], Ran Kafri[3,4], T.S. Karin Eisinger-Mathason[2], Anne E. Carpenter[1]*

Affiliations:

[1]Broad Institute of MIT and Harvard; Cambridge MA, USA; Department: Imaging Platform

[2]Abramson Family Cancer Research Institute, Department of Pathology & Laboratory Medicine, Penn Sarcoma Program, University of Pennsylvania Perelman School of Medicine; Philadelphia, PA, USA

[3]Department of Molecular Genetics, University of Toronto; Toronto, ON M5G 1A8, Canada

[4]Cell Biology, The Hospital for Sick Children; Toronto, ON M5G 0A4, Canada

[5]Broad Institute of MIT and Harvard; Cambridge MA, USA; Department: Cancer Program

[6]Massachusetts Institute of Technology, Cambridge; MA, USA; Department of Chemistry

[7]FM Kirby Neurobiology Center, Boston Children's Hospital; Boston, MA, USA

[8]Department of Neurobiology, Harvard Medical School; Boston, MA, USA

[9]Department of Orthopaedic Surgery & Bioengineering, University of Pennsylvania; Philadelphia, PA, USA

[10]Frederick National Laboratory for Cancer Research; Frederick, MD, USA

[11]Broad Institute of MIT and Harvard; Cambridge MA, USA; Department: Center for the Development of Therapeutics

[12]Department of Medical Oncology, Dana-Farber Cancer Institute; Boston, MA, USA

[13]Department of Biochemistry, Boston University School of Medicine; Boston, MA, USA

[14]Departments of Cell/Tissue Biology and Anatomy, University of California; San Francisco, CA, USA

*Corresponding author. Email: anne@broadinstitute.org

Abstract: Identifying chemical regulators of biological pathways is a time-consuming bottleneck in developing therapeutics and research compounds. 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 each disease. Here, our broad, virtual profile-based screening approach instead matches compounds to pathways based on phenotypic information in public cell image data, created using the Cell Painting assay. Our computational strategy efficiently uncovered the expected small molecule regulators for 19% of positive control genes. In discovery mode, we identified compounds related to three of seven tested genes, validated in subsequent gene-relevant assays. This image profile-based approach could replace many customized labor- and resource-intensive screens and accelerate the discovery of biologically and therapeutically useful compounds.

One-Sentence Summary: If a genetic perturbation impacts cell morphology, a computational query can reveal compounds whose morphology “matches”.

Main Text:

The pace of defining new diseases is rapidly accelerating (*1*), as is the cost and time required to develop novel therapeutics (*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 screening of small molecule modulators of pathways using the published literature or existing experimental data are beginning to emerge to accelerate drug discovery (*4*, *5*).

Here we develop a distinct computational strategy. We measure the complex morphological responses of cells to a genetic perturbation, then identify small molecules (i.e., chemical compounds) that produce the same (or opposite) response in the microscopy assay, Cell Painting (*6*). Conceptually similar to transcriptional profiling (*7*), Cell Painting is cheaper and useful in many applications (*8*, *9*), but its ability to identify compounds matching genes had not been explored.

Recent decades have given rise to an appealing, reductive ideal in the pharmaceutical industry: one drug that targets one protein to target one disease (*10*). However, diseases often involve many interacting proteins and successful drugs often impact multiple targets (*11–13*). There is therefore a renewed appreciation for identifying small molecules that can modulate pathways or networks in living cell systems to yield a desired phenotypic effect (*10*). Because genes in a pathway often show similar morphology (*14*) and compounds often show similar morphology based on their mechanism of action (*9*), we examined image profile matching as a promising but untested route to capturing perturbations at the pathway level and accelerating the screening step prior to identifying useful therapeutics and research tool compounds.

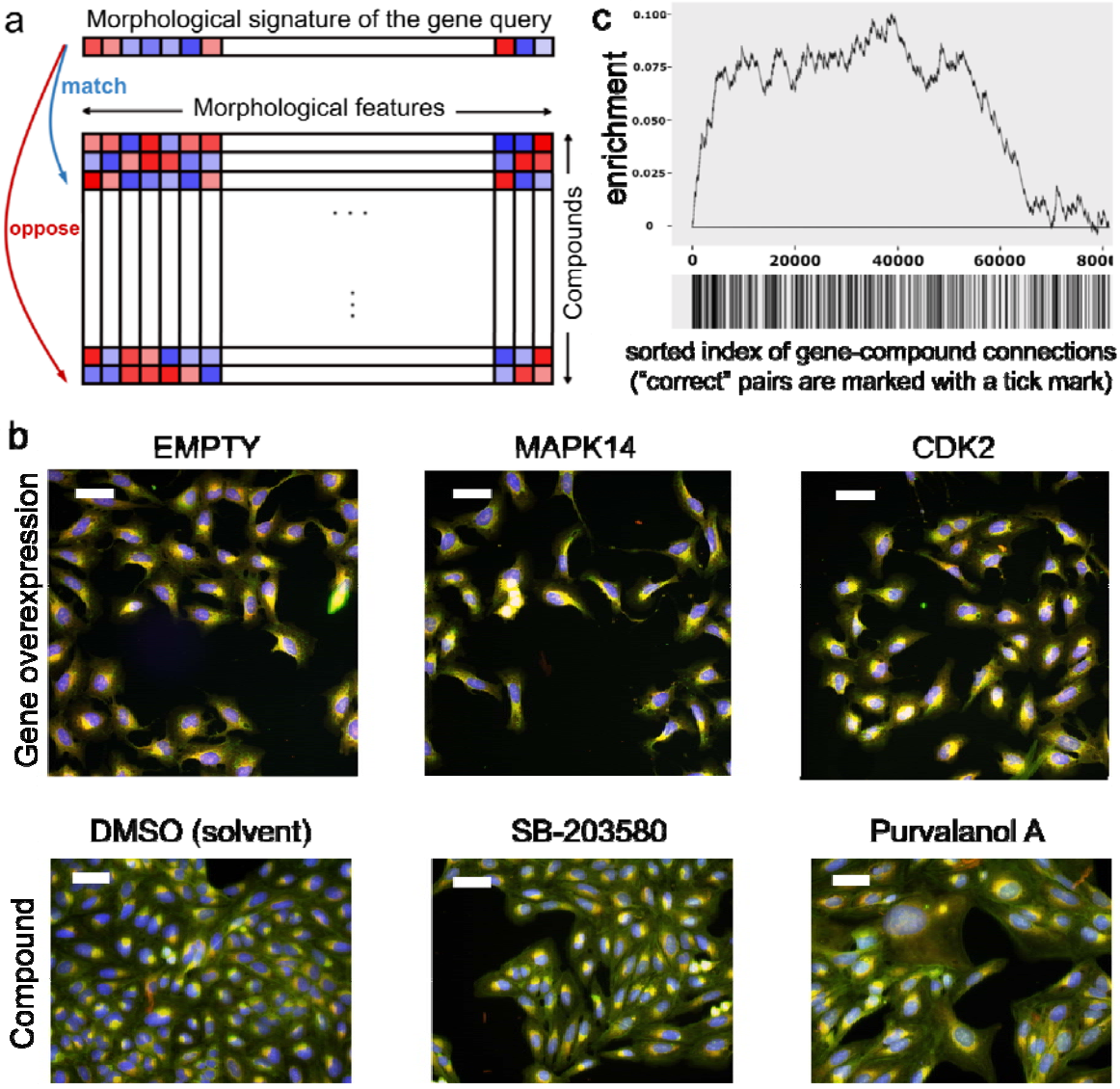


Figure 1: Image profile-based drug discovery offers efficient, virtual screening for pathway modulators. *a) If an overexpressed gene changes the morphology of cells, its image-based profile can be used as a query in a database of small molecule profiles, looking for those that match (positively correlate) or oppose (negatively correlate). b) Cell Painting images for two positive control gene-compound matches that yield observable morphological phenotypes (not all are expected to). EMPTY and DMSO are the negative controls in the gene overexpression and compound experiments, respectively; they differ in confluency and image acquisition conditions. The phenotype of p38□ (MAPK14) overexpression matches (correlates to) that of SB-203580, a known p38 inhibitor; in both, elongated 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 m. c) Enrichment plot of all gene-compound pairs sorted based on their absolute profile correlation. Starting from the left, the curve rises a unit if the gene-compound pair is a known/annotated connection, and goes down a unit otherwise. The units are normalized to the number of possible known connections, so the maximum height is one and ends in zero. The steep increase of the*

curve indicates enrichment of correct connections towards the top of the rank-ordered list of pairs.

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 (*14*). We matched their image-based profiles to our public Cell Painting profiles of 30,616 small molecules (*15*) (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 (Supplementary Figures S1 and S2).

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 (*9*), 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 S1).

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 (*11, 12*). 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 1b), 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, hundreds of virtual screens and cherrypicked validations could be done for less than the cost of a single traditional screening campaign.

In a rank-ordered list of gene-compound correlations, we found enrichment in the "correct" connections, using the 63 of the 69 genes that were annotated as targeted by a compound in the set as positive control pairs (Figure 1c). Although most correct gene-compound pairs did not correlate highly, as expected for the reasons described above, nevertheless we care most about the highest, left-most connections, where correct gene-compound pairings were 2.5-fold overrepresented among the strongest gene-compound pairings (correlation ≥ 0.35) (p-value = 0.007; Supplementary Table S2). For some matches, we visually confirmed that gene overexpression phenocopies or pheno-opposes the matching/opposing compound (Figure 1b),

although we emphasize that computationally-discovered phenotypes are not always visible to the human eye, particularly given cell heterogeneity.

We next 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 S3), 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). For comparison, an alternative supervised machine learning approach provided strong predictive ability in 8% of assays tested (*16*). Framing our approach as a virtual pre-screen, success even in 19% of cases would eliminate the need to carry out many dozens of large-scale customized screens in the pharmaceutical setting (representing hundreds of millions of dollars), instead advancing a few hundred compounds immediately to disease-relevant assays.

Image-based gene-compound matching: discovery

We next searched virtually for novel small molecule regulators of pathways (defined loosely here as networks involving a given gene). Throughout this study, we looked for compounds that both match (positively correlate) and oppose (negatively correlate) each overexpressed gene profile for several reasons. First, inhibitors and activators of a given pathway may both be of interest. Second, it is known that negative correlations among profiles can be biologically meaningful (*14*). Third, 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. Fourth, the impact of a gene or compound perturbation could be cell-type specific. Supporting these theoretical arguments, we found that, empirically, among the top 12 known gene-compound matches in our validation set, six showed correlation of the opposite directionality than expected (where expected is that an inhibitor's profile would have the opposite correlation to its overexpressed target gene).

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 (https://github.com/carpenterlab/2021_Rohban_submitted/blob/master/corr_mat.csv.zip). Because there is no systematic experiment to validate compounds impacting pathways, we took a customized expert-guided approach to ensure the results are biologically meaningful rather than just statistically significant. We found seven experts studying pathways with strong hits who were willing to conduct experiments; they chose the most relevant biological systems and readouts, rather than simply attempting to validate the original image-based finding.

Two cases yielded no confirmation (data not shown): RAS and SMAD3. 236 compounds with positive or negative correlations to the wild-type RAS or oncogenic HRAS G12V differential profile (see Methods) 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 (*17*). Nine compounds matching or opposing the SMAD3 overexpression profile failed to yield activity in a transcription reporter assay in A549 lung carcinoma cells 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 major 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.

A third case affirmed the approach but the novel compound identified was not very potent. We tested 17 compounds that negatively correlated with CSNK1E overexpression in a biochemical assay for the closely related kinase CSNK1A1. 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 (*18–20*). A fourth compound, BRD-K65952656, failed to bind any native kinases in a full KINOMEscan panel, suggesting it mimics CSNK1A1 inhibition via another molecular target. We chose not to pursue the expensive step of target deconvolution given its weak inhibition of CSNK1A1 (IC_{50} 12 uM).

A fourth case affirmed the approach but the novel compound failed to replicate following compound resynthesis, suggesting the desired activity, although validated, was not due to the expected structure, perhaps due to breakdown. We tested 16 compounds that positively correlated and 17 compounds that negatively correlated to GSK3B 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 $\leq$ 10 μM; the two most potent failed to show activity following resynthesis and hit expansion (testing of similarly-structured compounds) (Supplementary Table S4).

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

Discovery of hits modulating the p38□ (MAPK14) pathway

p38□ (MAPK14) inhibitors are sought for a variety of disorders, including cancers, dementia, asthma, and COVID-19 (*21*, *22*). We chose 20 compounds whose Cell Painting profile matched (n=9) or opposed (n=11) that of p38α overexpression in U2OS cells. In a single-cell p38 activity reporter assay in retinal pigment epithelial (RPE1) cells (*23*, *24*), we identified many activating compounds; these are less interesting given that the p38 pathway is activated by many stressors but rarely inhibited. We also found several inhibiting compounds and confirmed their activity (Figure 2, Supplementary Figure S3), including a known p38 MAPK inhibitor. Although the novel compounds are relatively weak, they nevertheless prove the principle that p38 pathway modulators can be found by image-profile matching, without a specific assay for the gene's function.

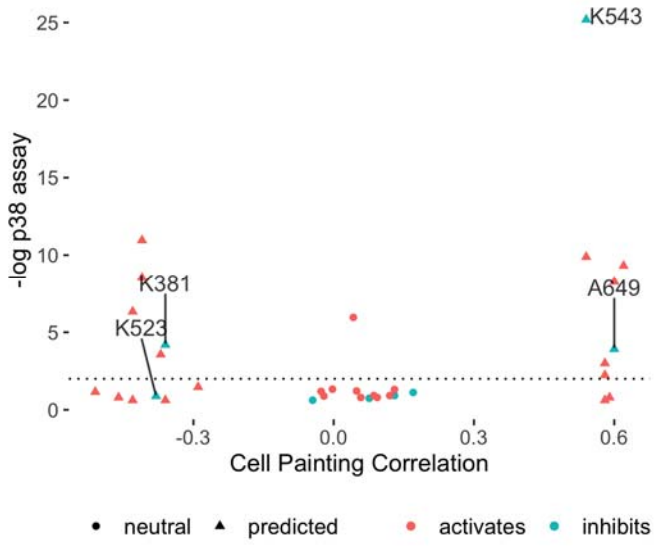


Figure 2: Cell Painting profiles identify compounds impacting the p38 pathway. *Compounds predicted to perturb p38 activity (triangles) and a set of 14 neutral compounds (Cell Painting profile correlations to p38α between -0.2 to 0.2; circles) were tested for their influence on p38 activity at 1 μM using a two-sided t-test on the single cell distributions of a p38 activity reporter (25) (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 (Supplementary Figure S3a, h-i). K543 (BRD-K54330070) denotes SB-202190, a known p38 inhibitor found as a match (other known inhibitors such as SB-203580 from Figure 1 were strong matches but excluded from this experiment because they were already known).*

Discovery of hits 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) (*26*), 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 (Figure 3a,b) without major changes to cell proliferation, size, or overall protein content. 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 (Supplementary Figure S4). More symmetric organelle morphology has been associated with reduced motility and PGC-1α overexpression (*27*). The role of PGC-1α in mitochondrial biogenesis is well-appreciated (*28*). The phenotype uncovered here using image profile matching is consistent with other recently discovered mitochondrial phenotypes associated with this gene (*29*).

We chose 24 compounds whose Cell Painting profiles correlated positively or negatively with PGC-1α overexpression in U2OS cells (Supplementary Table S5); 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 (*30*). We therefore tested compounds in a reporter assay representing FABP4, a prototypical target gene of the nuclear receptor, PPARG (*31*), 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 PGC1□ (*32*), as well as a novel analog thereof. Inhibitors uncovered were CCT018159 (BRD-K65503129) and Phorbol 12-myristate 13-acetate (BRD-K68552125). Many of the same compounds also showed activity in a ERRalpha reporter assay in 293T cells, albeit with differing effects (Supplementary Figure S5).

Encouraged by these results, we tested the impact of the compounds on mitochondrial motility, given the mitochondrial phenotype we observed and the role of PGC1□ in mitochondrial phenotypes and neurodegenerative disorders (*33*). In an automated imaging assay of rat cortical neurons (*34*), we found several compounds decreased mitochondrial motility; none increased motility (Supplementary Figure S6). 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%)(*34*).

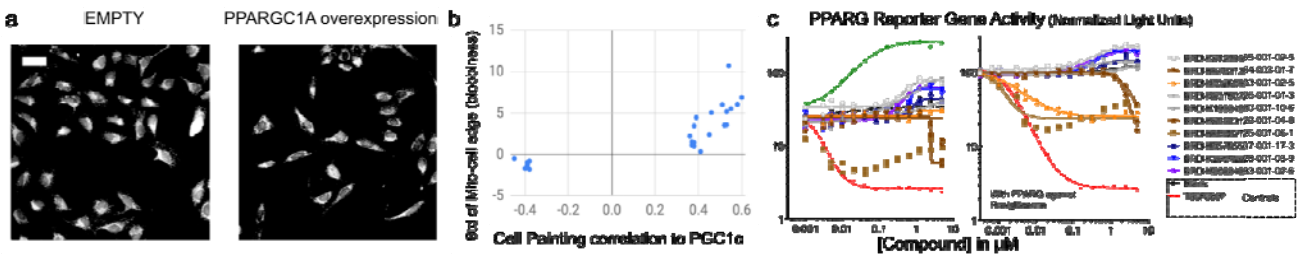


Figure 3: Cell Painting profiles identify compounds impacting PPARGC1A (PGC-1α) overexpression phenotypes. *a) Cell Painting images for PPARGC1A (PGC-1α) overexpression compared to negative control (EMPTY, same field as in Figure 1a). Scale bar = 60 μm. b) Compounds with high or low correlations of their 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). Correlation 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 (displayed 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.*

Discovery of small molecules impacting YAP1-related phenotypes

The Hippo pathway affects development, organ size regulation, and tissue regeneration. Small molecule regulators are highly sought for research and as potential therapeutics for cancer and other diseases; the pathway has been deemed relatively undruggable (*35*, *36*). We tested 30 compounds (Supplementary Table S6) 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 (*14*) (Supplementary Table S7, Supplementary Figure S7). 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 (*37*) (99.9 connectivity score in the Connectivity Map(*7*)). 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) (*38*, *39*), 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 YAP1 phenotypes. Accordingly, NB4A increased YAP1 nuclear localization (Figure 4b). While decreasing total YAP1 levels, NB4A also reduced YAP1 S127 phosphorylation (Figure 4c and Supplementary Figure S8a), which promotes YAP1 cytoplasmic sequestration (*40*).

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) (*7*). NB4A had no effect in a YAP1-responsive reporter assay following 48h of YAP overexpression in HEK-293 cells (Supplementary Figure S8b).

Compounds influencing the Hippo pathway might be therapeutic for undifferentiated pleomorphic sarcoma (UPS), an aggressive mesenchymal tumor that lacks targeted treatments (*41*). In UPS, YAP1 promotes tumorigenesis and is inversely correlated with patient survival (*41*). In KP230 cells, derived from a mouse model of UPS (*41*), Yap1 protein levels were reduced after 72 hours of NB4A 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. Interestingly, NB4A did not directly alter 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 S8). This indicates a potentially useful mechanism distinct from transcriptional regulation of *YAP1*.

Genetic and pharmacologic inhibition of Yap1 is known to suppress UPS cell proliferation *in vitro* and tumor initiation and progression *in vivo (41)*. 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 (*42*) (Figure 4i). NB4A did not affect the proliferation of two other YAP1-dependent lines, STS-109 human UPS cells (Supplementary Figure S9a) and HT-1080 fibrosarcoma cells (Supplementary Figure S9b) (*41*, *43*), nor YAP1-independent HCT-116 colon cancer cells (Supplementary Figure S9c-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) (Supplementary Figure S7). These compounds are not commercially available, limiting our experiments and past literature.

We assessed the compounds' impact on 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 (Supplementary Figure S8c-f). By contrast, the compounds had no effect on *YAP1* or its target genes in H9 hPSCs (Supplementary Figure S8g), nor in a 48 h YAP-responsive reporter assay following YAP overexpression in HEK-293 cells (Supplementary Figure S8b).

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(*44*). BRD-K34692511 had none of these impacts.

Both compounds had the desired effect of inhibiting KP230 cell proliferation (Supplementary Figure S9f). Also noteworthy, BRD-K28862419 modestly yet significantly reduced KP230 cell viability (Supplementary Figure S9g), 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 YAP1-related phenotypes, in particular an unusual ability to reduce growth of certain aggressive sarcoma lines. 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.

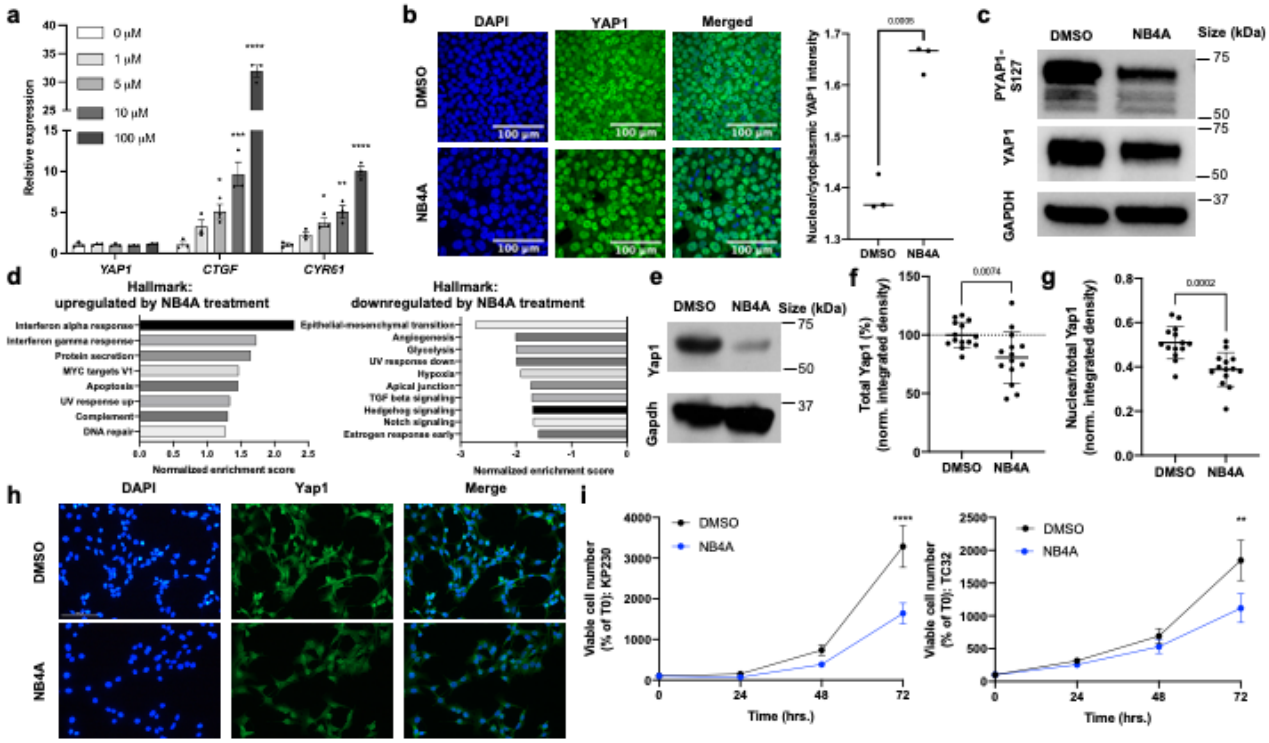


Figure 4: Cell Painting profiles identify compounds impacting YAP1 phenotypes. *a) Relative transcript levels of YAP1, CTGF, and CYR61 in H9 human pluripotent stem cells treated with NB4A or DMSO control for 24 hrs. *P<0.05; **P<0.01; ***P<0.001; ****P<0.0001 (one-way ANOVA with Dunnett's multiple comparisons test). Mean ± SEM. n = 3 biologically independent experiments . 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 biologically independent experiments; an average of mean intensities from 3 fields of each biological replicate is calculated. c) Representative blot of n = 3 biologically independent experiments for 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 Supplementary Figure S8a). 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.*

Discussion

We found that hit-stage 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 screening. We do not claim the particular compounds we uncovered are sufficiently potent, specific, and non-toxic for human therapeutics. As with all screening approaches, significant further work is necessary to develop hits into useful therapeutics; this includes confirming activity and directionality of hits in a relevant cell type or model system, improving potency and specificity, and identifying the molecular target(s) (including so-called off-target effects that may or may not be useful to achieve the desired phenotypic effect). Further, like all drug discovery, the eventual clinical success relies on the therapeutic hypothesis for the gene, pathway, and/or phenotype being correct, which is never guaranteed.

Even so, virtualizing a first-pass large-scale screen by computationally matching the phenotypic effect of compounds to that of gene manipulation will in many cases enable rapid and inexpensive identification of compounds with desired phenotypic impacts. Gene-compound matching might also be useful to identify which genes/pathways are targeted by novel small molecules of unknown mechanism of action, another significant bottleneck in the drug discovery process (*45*).

We expect future iterations of this strategy to be more successful. First, we would expect better-quality chemical matter from larger libraries; only 30,000 were screened in this work whereas a pharmaceutical screening campaign can test millions (*46*). Large-scale data production efforts are underway that will increase the potential for matching profiles against public data: the JUMP-Cell Painting Consortium is producing a public dataset of 140,000 chemical and genetic perturbations. 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, as could assessing whether gene knockdown profiles yield better results in practice than gene overexpression, or whether pathways where overexpression and knockdown give opposite profiles are even better starting points for virtual screening. More advanced computational methods are also on the horizon, from feature extraction (*47*) to machine learning on new benchmark datasets of gene-compound pairs (*48*); we would expect supervised machine learning to work better than our unsupervised correlation-based approach (*9*). We anticipate that image profile-based virtual screening provides a new, broad, and unbiased accelerant toward meeting the pressing need for novel therapeutics.

References and Notes

1. H. I. Roessler, N. V. A. M. Knoers, M. M. van Haelst, G. van Haaften, Drug Repurposing for Rare Diseases. *Trends Pharmacol. Sci.* **42**, 255–267 (2021).

2. O. J. Wouters, M. McKee, J. Luyten, Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA.* **323**, 844–853 (2020).

3. J. G. Moffat, F. Vincent, J. A. Lee, J. Eder, M. Prunotto, Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nat. Rev. Drug Discov.* **16**, 531–543 (2017).

4. P. Schneider, W. P. Walters, A. T. Plowright, N. Sieroka, J. Listgarten, R. A. Goodnow Jr, J. Fisher, J. M. Jansen, J. S. Duca, T. S. Rush, M. Zentgraf, J. E. Hill, E. Krutoholow, M. Kohler, J. Blaney, K. Funatsu, C. Luebkemann, G. Schneider, Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug Discov.* **19**, 353–364 (2020).

5. J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, B. Li, A. Madabhushi, P. Shah, M. Spitzer, S. Zhao, Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* **18**, 463–477 (2019).

6. M.-A. Bray, S. Singh, H. Han, C. T. Davis, B. Borgeson, C. Hartland, M. Kost-Alimova, S. M. Gustafsdottir, C. C. Gibson, A. E. Carpenter, Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* **11**, 1757–1774 (2016).

7. A. Subramanian, R. Narayan, S. M. Corsello, D. D. Peck, T. E. Natoli, X. Lu, J. Gould, J. F. Davis, A. A. Tubelli, J. K. Asiedu, D. L. Lahr, J. E. Hirschman, Z. Liu, M. Donahue, B. Julian, M. Khan, D. Wadden, I. C. Smith, D. Lam, A. Liberzon, C. Toder, M. Bagul, M. Orzechowski, O. M. Enache, F. Piccioni, S. A. Johnson, N. J. Lyons, A. H. Berger, A. F. Shamji, A. N. Brooks, A. Vrcic, C. Flynn, J. Rosains, D. Y. Takeda, R. Hu, D. Davison, J. Lamb, K. Ardlie, L. Hogstrom, P. Greenside, N. S. Gray, P. A. Clemons, S. Silver, X. Wu, W.-N. Zhao, W. Read-Button, X. Wu, S. J. Haggarty, L. V. Ronco, J. S. Boehm, S. L. Schreiber, J. G. Doench, J. A. Bittker, D. E. Root, B. Wong, T. R. Golub, A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*. **171**, 1437–1452.e17 (2017).

8. M. Lapins, O. Spjuth, Evaluation of Gene Expression and Phenotypic Profiling Data as Quantitative Descriptors for Predicting Drug Targets and Mechanisms of Action. *Cold Spring Harbor Laboratory* (2019), p. 580654.

9. S. N. Chandrasekaran, H. Ceulemans, J. D. Boyd, A. E. Carpenter, Image-based profiling for drug discovery: due for a machine-learning upgrade? *Nat. Rev. Drug Discov.*, 1–15 (2020).

10. R. E. Hughes, R. J. R. Elliott, J. C. Dawson, N. O. Carragher, High-content phenotypic and pathway profiling to advance drug discovery in diseases of unmet need. *Cell Chem Biol*. **28**, 338–355 (2021).

11. E. Proschak, H. Stark, D. Merk, Polypharmacology by Design: A Medicinal Chemist's Perspective on Multitargeting Compounds. *J. Med. Chem.* **62**, 420–444 (2019).

12. A. Lin, C. J. Giuliano, A. Palladino, K. M. John, C. Abramowicz, M. L. Yuan, E. L. Sausville, D. A. Lukow, L. Liu, A. R. Chait, Z. C. Galluzzo, C. Tucker, J. M. Sheltzer, Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Sci. Transl. Med.* **11** (2019), doi:10.1126/scitranslmed.aaw8412.

13. M. E. Bunnage, E. L. P. Chekler, L. H. Jones, Target validation using chemical probes. *Nat. Chem. Biol.* **9**, 195–199 (2013).

14. M. H. Rohban, S. Singh, X. Wu, J. B. Berthet, M.-A. Bray, Y. Shrestha, X. Varelas, J. S. Boehm, A. E. Carpenter, Systematic morphological profiling of human gene and allele function via Cell Painting. *Elife*. **6** (2017), doi:10.7554/eLife.24060.

15. M.-A. Bray, S. M. Gustafsdottir, M. H. Rohban, S. Singh, V. Ljosa, K. L. Sokolnicki, J. A. Bittker, N. E. Bodycombe, V. Dancík, T. P. Hasaka, C. S. Hon, M. M. Kemp, K. Li, D. Walpita, M. J. Wawer, T. R. Golub, S. L. Schreiber, P. A. Clemons, A. F. Shamji, A. E. Carpenter, A dataset of images and morphological profiles of 30 000 small-molecule treatments using the Cell Painting assay. *Gigascience*. **6**, 1–5 (2017).

16. J. Simm, G. Klambauer, A. Arany, M. Steijaert, J. K. Wegner, E. Gustin, V. Chupakhin, Y. T. Chong, J. Vialard, P. Buijnsters, I. Velter, A. Vapirev, S. Singh, A. E. Carpenter, R. Wuyts, S. Hochreiter, Y. Moreau, H. Ceulemans, Repurposing High-Throughput Image Assays Enables Biological Activity Prediction for Drug Discovery. *Cell Chem Biol*. **25**, 611–618.e3 (2018).

17. M. Drosten, A. Dhawahir, E. Y. M. Sum, J. Urosevic, C. G. Lechuga, L. M. Esteban, E. Castellano, C. Guerra, E. Santos, M. Barbacid, Genetic analysis of Ras signalling pathways in cell proliferation, migration and survival. *EMBO J.* **29**, 1091–1104 (2010).

18. M. I. Davis, J. P. Hunt, S. Herrgard, P. Ciceri, L. M. Wodicka, G. Pallares, M. Hocker, D. K. Treiber, P. P. Zarrinkar, Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **29**, 1046–1051 (2011).

19. S. Klaeger, S. Heinzlmeir, M. Wilhelm, H. Polzer, B. Vick, P.-A. Koenig, M. Reinecke, B. Ruprecht, S. Petzoldt, C. Meng, J. Zecha, K. Reiter, H. Qiao, D. Helm, H. Koch, M. Schoof, G. Canevari, E. Casale, S. R. Depaolini, A. Feuchtinger, Z. Wu, T. Schmidt, L. Rueckert, W. Becker, J. Huenges, A.-K. Garz, B.-O. Gohlke, D. P. Zolg, G. Kayser, T. Vooder, R. Preissner, H. Hahne, N. Tõnisson, K. Kramer, K. Götze, F. Bassermann, J. Schlegl, H.-C. Ehrlich, S. Aiche, A. Walch, P. A. Greif, S. Schneider, E. R. Felder, J. Ruland, G. Médard, I. Jeremias, K. Spiekermann, B. Kuster, The target landscape of clinical kinase drugs. *Science*. **358** (2017), doi:10.1126/science.aan4368.

20. N. P. Shanware, L. M. Williams, M. J. Bowler, R. S. Tibbetts, Non-specific in vivo inhibition of CK1 by the pyridinyl imidazole p38 inhibitors SB 203580 and SB 202190. *BMB Rep.* **42** (2009), doi:10.5483/bmbrep.2009.42.3.142.

21. MAPK14 - ClinicalTrials.Gov, (available at https://clinicaltrials.gov/ct2/results?term=MAPK14&Search=Apply&age_v=&gndr=&type=

&rslt=).

22. A. Martínez-Limón, M. Joaquin, M. Caballero, F. Posas, E. de Nadal, The p38 Pathway: From Biology to Cancer Therapy. *Int. J. Mol. Sci.* **21** (2020), doi:10.3390/ijms21061913.

23. S. Regot, J. J. Hughey, B. T. Bajar, S. Carrasco, M. W. Covert, High-sensitivity measurements of multiple kinase activities in live single cells. *Cell*. **157**, 1724–1734 (2014).

24. S. Liu, M. B. Ginzberg, N. Patel, M. Hild, B. Leung, Z. Li, Y.-C. Chen, N. Chang, Y. Wang, C. Tan, S. Diena, W. Trimble, L. Wasserman, J. L. Jenkins, M. W. Kirschner, R. Kafri, Size uniformity of animal cells is actively maintained by a p38 MAPK-dependent regulation of G1-length. *Elife*. **7** (2018), doi:10.7554/eLife.26947.

25. T. Kaufman, E. Nitzan, N. Firestein, M. Ginzberg, S. Iyengar, N. Patel, R. Ben-Hamo, Z. Porat, A. Hilfinger, R. Kafri, R. Straussman, Visual barcodes for multiplexing live microscopy-based assays (2020), , doi:10.21203/rs.3.rs-67883/v1.

26. National Center for Biotechnology Information, PubChem Bioassay Record for AID 651723, Source: Broad Institute, (available at https://pubchem.ncbi.nlm.nih.gov/bioassay/651723).

27. H.-J. Lee, Y. Su, P.-H. Yin, H.-C. Lee, C.-W. Chi, PPAR(gamma)/PGC-1(alpha) pathway in E-cadherin expression and motility of HepG2 cells. *Anticancer Res.* **29**, 5057–5063 (2009).

28. C. Luo, H. R. Widlund, P. Puigserver, PGC-1 Coactivators: Shepherding the Mitochondrial Biogenesis of Tumors. *Trends Cancer Res.* **2**, 619–631 (2016).

29. J. F. Halling, H. Pilegaard, PGC-1α-mediated regulation of mitochondrial function and physiological implications. *Appl. Physiol. Nutr. Metab.* **45**, 927–936 (2020).

30. C. Handschin, B. M. Spiegelman, Peroxisome proliferator-activated receptor gamma coactivator 1 coactivators, energy homeostasis, and metabolism. *Endocr. Rev.* **27**, 728–735 (2006).

31. J. T. Goldstein, A. C. Berger, J. Shih, F. F. Duke, L. Furst, D. J. Kwiatkowski, A. D. Cherniack, M. Meyerson, C. A. Strathdee, Genomic Activation of PPARG Reveals a Candidate Therapeutic Axis in Bladder Cancer. *Cancer Res.* **77**, 6987–6998 (2017).

32. M. Usui, M. Uno, E. Nishida, Src family kinases suppress differentiation of brown adipocytes and browning of white adipocytes. *Genes Cells*. **21**, 302–310 (2016).

33. A. A. Nierenberg, S. A. Ghaznavi, I. Sande Mathias, K. K. Ellard, J. A. Janos, L. G. Sylvia, 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).

34. E. Shlevkov, H. Basu, M.-A. Bray, Z. Sun, W. Wei, K. Apaydin, K. Karhohs, P.-F. Chen, J. L. M. Smith, O. Wiskow, K. Roet, X. Huang, K. Eggan, A. E. Carpenter, R. J. Kleiman, T. L. Schwarz, A High-Content Screen Identifies TPP1 and Aurora B as Regulators of Axonal Mitochondrial Transport. *Cell Rep.* **28**, 3224–3237.e5 (2019).

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

36. T. T. Tang, A. W. Konradi, Y. Feng, X. Peng, M. Ma, J. Li, F.-X. Yu, K.-L. Guan, L. Post, Small Molecule Inhibitors of TEAD Auto-palmitoylation Selectively Inhibit Proliferation and Tumor Growth of NF2-deficient Mesothelioma. *Mol. Cancer Ther.* **20**, 986–998 (2021).

37. M. Nishio, K. Sugimachi, H. Goto, J. Wang, T. Morikawa, Y. Miyachi, Y. Takano, H. Hikasa, T. Itoh, S. O. Suzuki, H. Kurihara, S. Aishima, A. Leask, T. Sasaki, T. Nakano, H. Nishina, Y. Nishikawa, Y. Sekido, K. Nakao, K. Shin-Ya, K. Mimori, A. Suzuki, 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).

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

39. S. Musah, P. J. Wrighton, Y. Zaltsman, X. Zhong, S. Zorn, M. B. Parlato, C. Hsiao, S. P. Palecek, Q. Chang, W. L. Murphy, L. L. Kiessling, 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).

40. B. Zhao, X. Wei, W. Li, R. S. Udan, Q. Yang, J. Kim, J. Xie, T. Ikenoue, J. Yu, L. Li, P. Zheng, K. Ye, A. Chinnaiyan, G. Halder, Z.-C. Lai, K.-L. Guan, Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. *Genes Dev.* **21**, 2747–2761 (2007).

41. S. Ye, M. A. Lawlor, A. Rivera-Reyes, S. Egolf, S. Chor, K. Pak, G. E. Ciotti, A. C. Lee, G. E. Marino, J. Shah, D. Niedzwicki, K. Weber, P. M. C. Park, M. Z. Alam, A. Grazioli, M. Haldar, M. Xu, J. A. Perry, J. Qi, T. S. K. Eisinger-Mathason, YAP1-Mediated Suppression of USP31 Enhances NFκB Activity to Promote Sarcomagenesis. *Cancer Res.* **78**, 2705–2720 (2018).

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

43. T. S. K. Eisinger-Mathason, V. Mucaj, K. M. Biju, M. S. Nakazawa, M. Gohil, T. P. Cash, S. S. Yoon, N. Skuli, K. M. Park, S. Gerecht, M. C. Simon, 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).

44. K. M. Elias, M. M. Emori, T. Westerling, H. Long, A. Budina-Kolomets, F. Li, E. MacDuffie, M. R. Davis, A. Holman, B. Lawney, M. L. Freedman, J. Quackenbush, M. Brown, R. Drapkin, Epigenetic remodeling regulates transcriptional changes between ovarian cancer and benign precursors. *JCI Insight.* **1** (2016), doi:10.1172/jci.insight.87988.

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

46. A. Mullard, Machine learning brings cell imaging promises into focus. *Nat. Rev. Drug*

Discov. **18**, 653–655 (2019).

47. A. Pratapa, M. Doron, J. C. Caicedo, Image-based cell phenotyping with deep learning. *Curr. Opin. Chem. Biol.* **65**, 9–17 (2021).

48. S. N. Chandrasekaran, B. A. Cimini, A. Goodale, L. Miller, Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations (2021) (available at https://openreview.net/forum?id=rCRyg1-Yovi).

49. REPURPOSING, (available at https://clue.io/repurposing-app).

50. DrugBank Online, (available at https://www.drugbank.ca).

51. M. T. Lab, BioGRID Interaction Database, (available at https://thebiogrid.org).

52. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

53. X. H. Feng, Y. Zhang, R. Y. Wu, R. Derynck, 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).

54. F. F. Wagner, L. Benajiba, A. J. Campbell, M. Weïwer, J. R. Sacher, J. P. Gale, L. Ross, A. Puissant, G. Alexe, A. Conway, M. Back, Y. Pikman, I. Galinsky, D. J. DeAngelo, R. M. Stone, T. Kaya, X. Shi, M. B. Robers, T. Machleidt, J. Wilkinson, O. Hermine, A. Kung, A. J. Stein, D. Lakshminarasimhan, M. T. Hemann, E. Scolnick, Y.-L. Zhang, J. Q. Pan, K. Stegmaier, E. B. Holson, 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), doi:10.1126/scitranslmed.aam8460.

55. A. Zetser, E. Gredinger, E. Bengal, p38 mitogen-activated protein kinase pathway promotes skeletal muscle differentiation. Participation of the Mef2c transcription factor. *J. Biol. Chem.* **274**, 5193–5200 (1999).

56. J. T. Rodgers, C. Lerin, W. Haas, S. P. Gygi, B. M. Spiegelman, P. Puigserver, Nutrient control of glucose homeostasis through a complex of PGC-1alpha and SIRT1. *Nature*. **434**, 113–118 (2005).

57. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*. **14**, 417–419 (2017).

58. L. A. Marcaurelle, E. Comer, S. Dandapani, J. R. Duvall, B. Gerard, S. Kesavan, M. D. Lee 4th, H. Liu, J. T. Lowe, J.-C. Marie, C. A. Mulrooney, B. A. Pandya, A. Rowley, T. D. Ryba, B.-C. Suh, J. Wei, D. W. Young, L. B. Akella, N. T. Ross, Y.-L. Zhang, D. M. Fass, S. A. Reis, W.-N. Zhao, S. J. Haggarty, M. Palmer, M. A. Foley, An aldol-based build/couple/pair strategy for the synthesis of medium- and large-sized rings: discovery of macrocyclic histone deacetylase inhibitors. *J. Am. Chem. Soc.* **132**, 16962–16976 (2010).

59. M. E. Fitzgerald, C. A. Mulrooney, J. R. Duvall, J. Wei, B.-C. Suh, L. B. Akella, A. Vrcic, L. A. Marcaurelle, Build/couple/pair strategy for the synthesis of stereochemically diverse macrolactams via head-to-tail cyclization. *ACS Comb. Sci.* **14**, 89–96 (2012).

60. X. Wang, T. L. Schwarz, in *Methods in Enzymology* (Academic Press, 2009), vol. 457, pp. 319–333.

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

Funding: 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.

Author contributions:

Conceptualization: AEC

Project Management: MHR, AEC

Experimental Design: MD, TSKE-M, LLK, DS, XV, JDB, JTG, AG, RD, JRS, FFW, SMC

Experiments: AMF, GEC, AD, GBP, MPN, RK, CT, JTG, AG, MR, MB, JRS, JPG

Data Analysis: MHR, SS, AMF, GEC, GBP, MPN, RK, CT, JRS, JPG, FFW, SMC, AB

Data Interpretation: MHR, AEC, SS, MD, TSKE-M, AMF, LLK, DS, XV, JDB, RK, CT, JTG, AG, TLS, TT, RD, MB, JRS, JPG, FFW, SMC, AB

Writing: MHR, AEC, MD, TSKE-M, AMF, LLK, DS, XV, JDB, RK, CT, JTG, AG, JRS

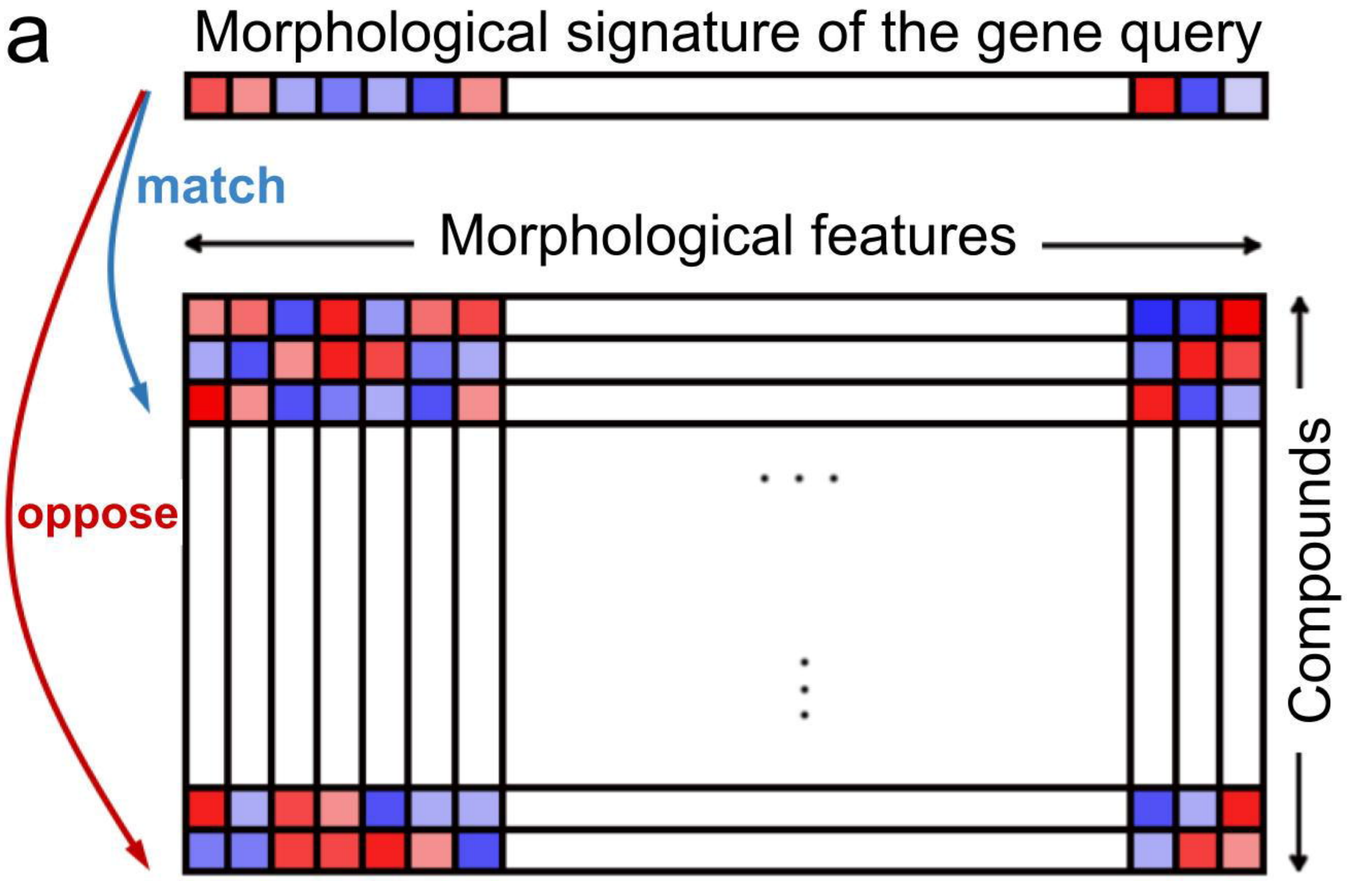
Funding: AEC

Supervision: TLS, TT

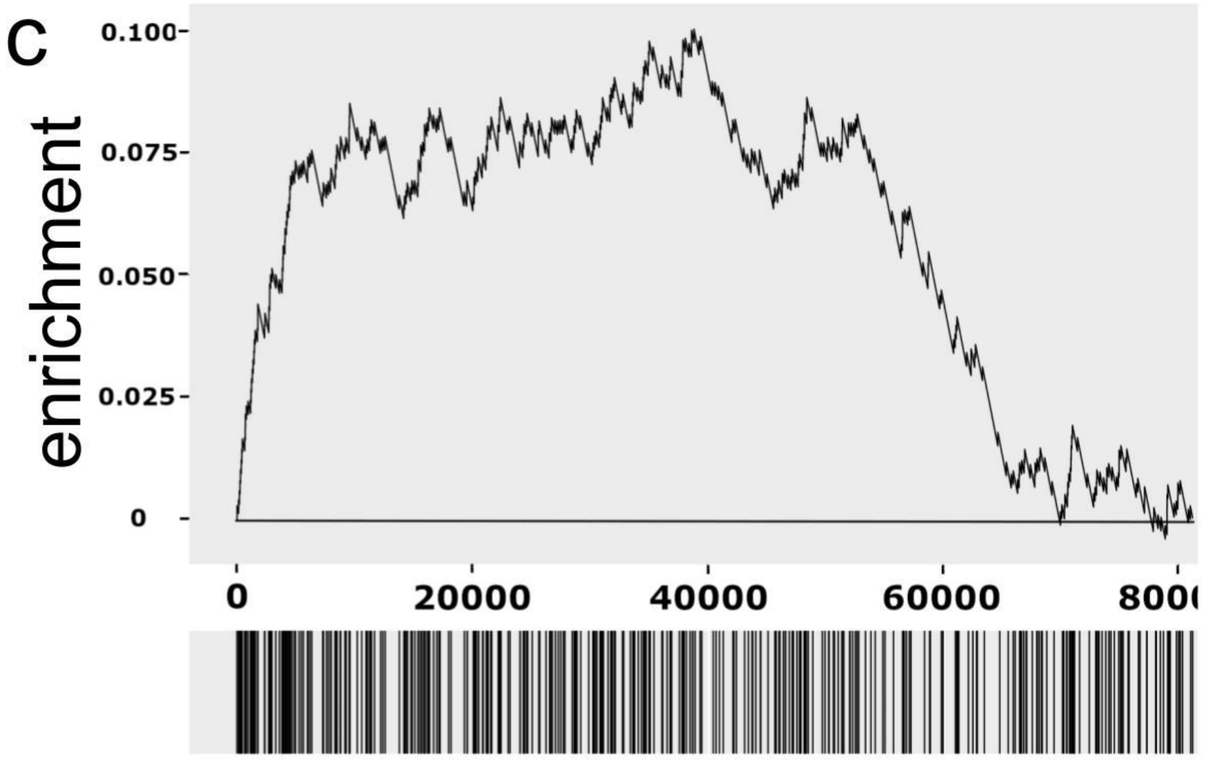
Competing interests: The Broad Institute and the University of Pennsylvania have filed a patent on the described compounds related to YAP1 overexpression. AEC has ownership interest in Recursion, a publicly-traded biotech company using images for drug discovery. JTG reports receiving a commercial research grant from Bayer AG. SMC reports receiving research funding from Bayer and Calico Life Sciences.

Data and materials availability: All data, code, and materials used in the analysis are available to any researcher for purposes of reproducing or extending the analysis.

a

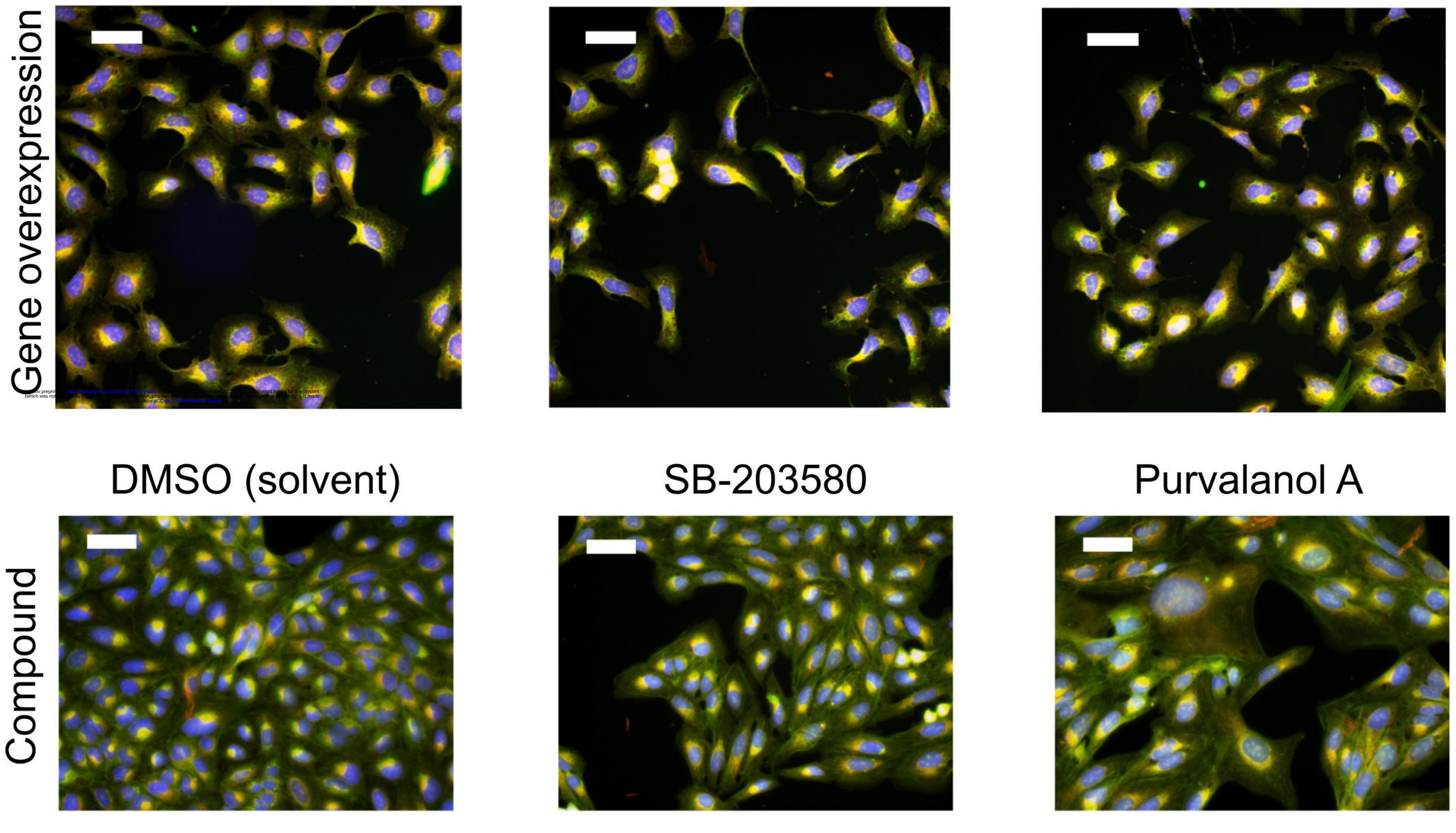


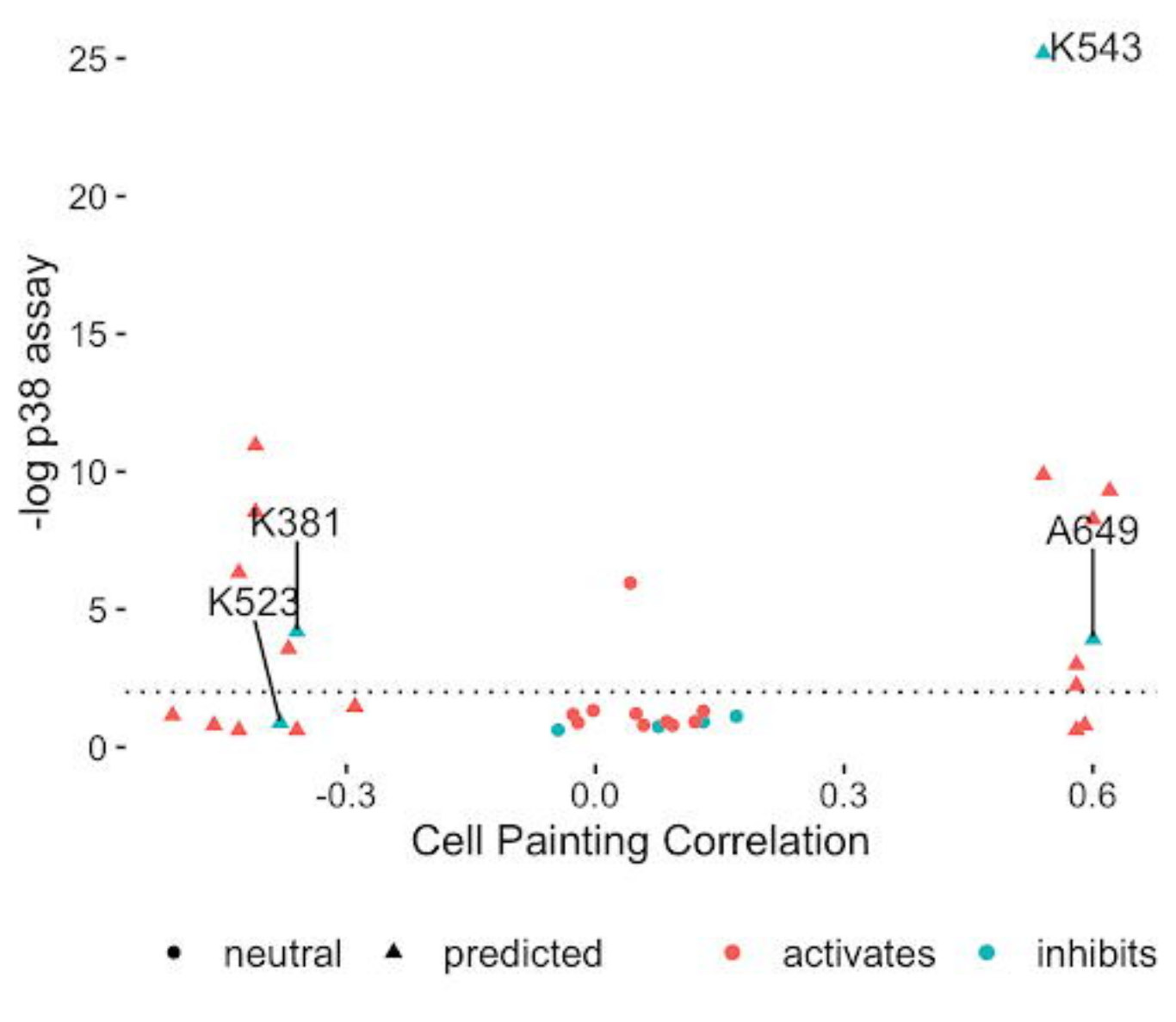
c

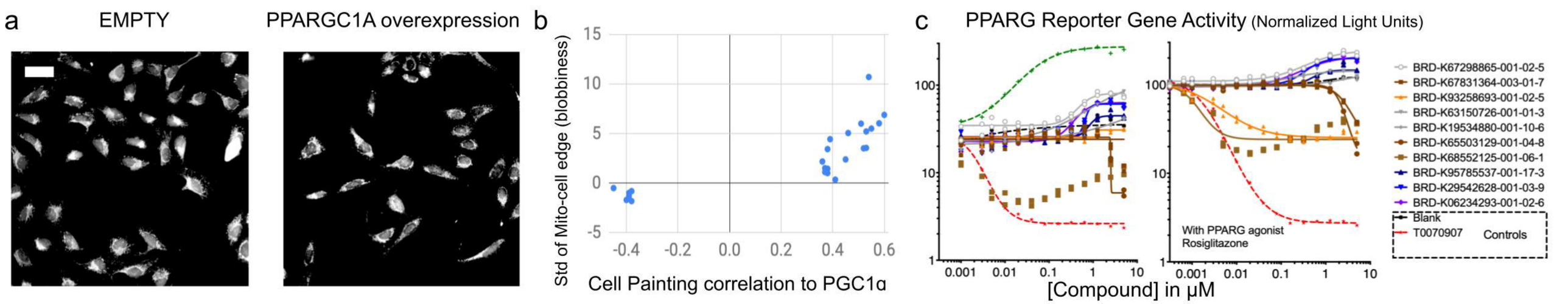


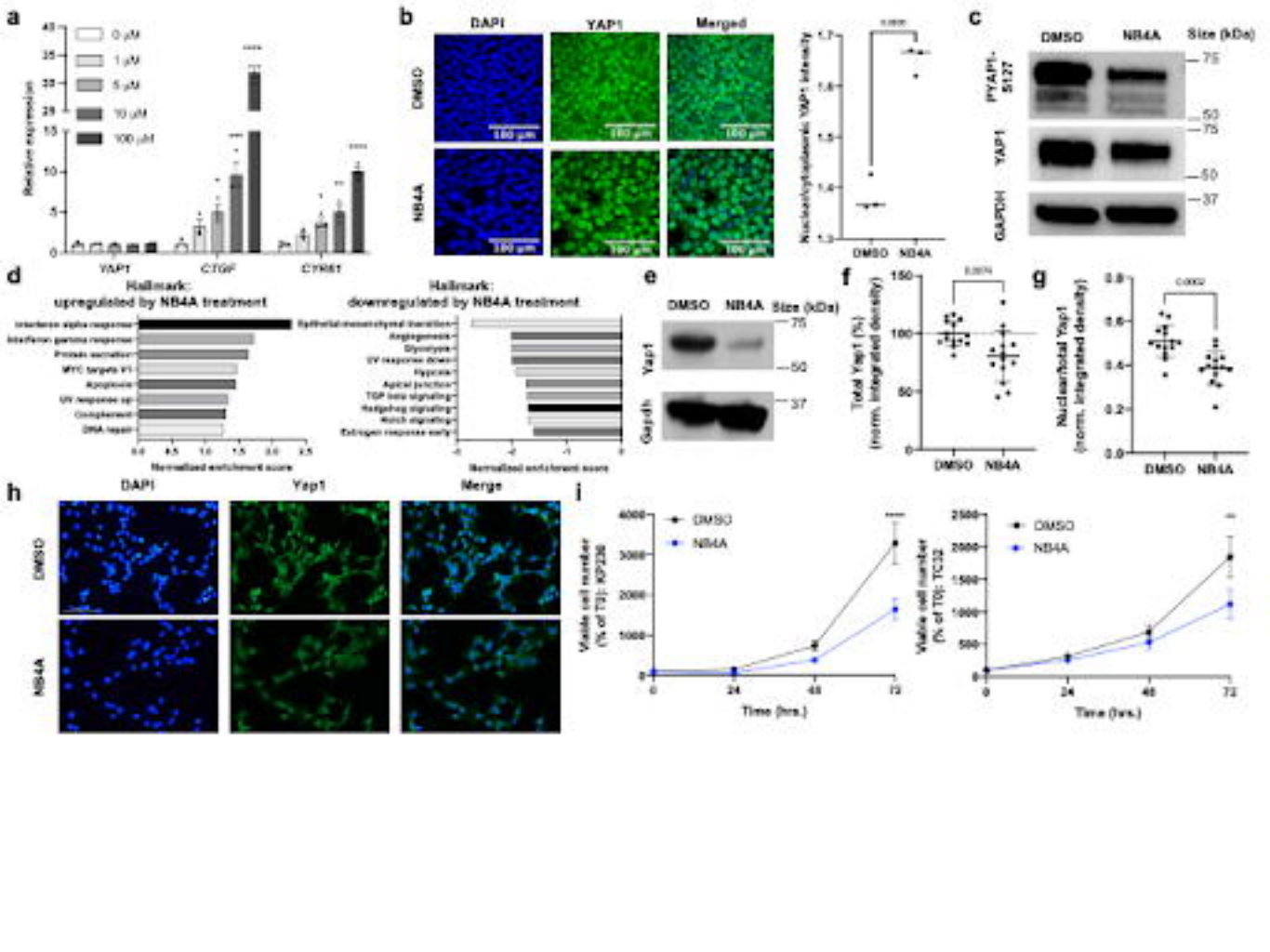
sorted index of gene-compound connections
("correct" pairs are marked with a tick mark)

b




K543
25
20
15
10
5
0
-log p38 assay
K381
K523
A649
-0.3
0.0
0.3
0.6
Cell Painting Correlation
neutral
predicted
activates
inhibits


a
EMPTY
PPARGC1A overexpression
b
Std of Mito-cell edge (blobbiness)
Cell Painting correlation to PGC1α
c
PPARG Reporter Gene Activity (Normalized Light Units)
With PPARG agonist Rosiglitazone
[Compound] in μM
BRD-K67298865-001-02-5
BRD-K67831364-003-01-7
BRD-K93258693-001-02-5
BRD-K63150726-001-01-3
BRD-K19534880-001-10-6
BRD-K65503129-001-04-8
BRD-K68552125-001-06-1
BRD-K95785537-001-17-3
BRD-K29542628-001-03-9
BRD-K06234293-001-02-6
Blank
T0070907
Controls


a
0 μM
1 μM
5 μM
10 μM
100 μM
Relative expression
YAP1
CTGF
CYR61
b
DAPI
YAP1
Merged
DMSO
NB4A
100 μm
Nuclear/cytoplasmic YAP1 intensity
DMSO
NB4A
c
DMSO
NB4A
Size (kDa)
PYAP1-S127
YAP1
GAPDH
75
50
37
d
Hallmark:
upregulated by NB4A treatment
Interferon alpha response
Interferon gamma response
Protein secretion
MYC targets V1
Apoptosis
UV response up
Complement
DNA repair
Normalized enrichment score
Hallmark:
downregulated by NB4A treatment
Epithelial mesenchymal transition
Angiogenesis
Glycolysis
UV response down
Hypoxia
Apical junction
TGF beta signaling
Hedgehog signaling
Notch signaling
Estrogen response early
Normalized enrichment score
e
DMSO
NB4A
Size (kDa)
Yap1
Gapdh
75
50
37
f
Total Yap1 (%)
(norm. integrated density)
DMSO
NB4A
g
Nuclear/total Yap1
(norm. integrated density)
DMSO
NB4A
h
DAPI
Yap1
Merge
DMSO
NB4A
i
DMSO
NB4A
Viable cell number
(% of T0): KP326
Time (hrs.)
DMSO
NB4A
Viable cell number
(% of T0): TC32
Time (hrs.)