Systematic morphological profiling of human gene and allele function via Cell Painting

MH Rohban[1], S Singh[1], X Wu[1], JB Berthet[2], M-A Bray[1,3], Y Shrestha[1], X Varelas[2], JS Boehm[1], AE Carpenter[1*]

Affiliations:

1) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
2) Department of Biochemistry, Boston University School of Medicine, Boston MA USA
3) Present address: Novartis Institutes for BioMedical Research, Cambridge MA USA

* Corresponding author: AE Carpenter: anne@broadinstitute.org

Abstract

We hypothesized that human genes and disease-associated alleles might be systematically functionally annotated using morphological profiling of cDNA constructs, via a microscopy-based Cell Painting assay. Indeed, 50% of the 220 tested genes yielded detectable morphological profiles, which grouped into biologically meaningful gene clusters consistent with known functional annotation (e.g., the RAS-RAF-MEK-ERK cascade). We used novel subpopulation-based visualization methods to interpret the morphological changes for specific clusters. This unbiased morphologic map of gene function revealed TRAF2/c-REL negative regulation of YAP1/WWTR1-responsive pathways. We confirmed this discovery of functional connectivity between the NF-κB pathway and Hippo pathway effectors at the transcriptional level, thereby expanding knowledge of these two signaling pathways that critically regulate tumor initiation and progression. We make the images and raw data publicly available, providing an initial morphological map of major biological pathways for future study.

Introduction

The dramatic increase in human genome sequence data has created a significant bottleneck. The number of genes and variants known to be associated with most human diseases has increased dramatically (Amberger et al. 2015). Unfortunately, the next step - understanding the function of each gene and the mechanism of each allele in the disease - typically remains non-systematic and labor-intensive. Most commonly, researchers painstakingly design, develop, and apply a disease-specific or biological process-specific assay.

Over 30% of genes in the human genome are of unknown function (Leonetti et al. 2016) and even annotated genes have additional functions yet to be uncovered. Furthermore, even when a gene's normal functions are known, methods are lacking to predict the functional impact of the millions of

genetic variants found in patients. These gaps must be filled in order to convert the promise of human genome sequence data into clinical treatments.

Therefore, there is a widespread need for systematic approaches to functionally annotate genes and variants therein, regardless of the biological process or disease of interest. One general approach depends on guilt-by-association, linking unannotated genes to annotated ones based on properties such as protein-protein interaction data, sequence similarity, or, most convincingly, functional similarity (Shehu, Barbará, and Molloy 2016). In the latter category are profiling techniques, where dozens to hundreds of measurements are made for each gene perturbation and the resulting profile is compared against profiles for annotated genes. Various data sources can be used for profiling; gene expression is one that can be performed in relatively high-throughput and it has been proven useful in predicting gene function (Lamb et al. 2006). In fact, high-throughput mRNA profiles were recently used to cluster alleles found in lung adenocarcinoma based on their functional impact, a precursor to therapeutic strategy for variants of previously unknown significance (Berger et al. 2016).

Images are a less mature data source for profiling but show tremendous promise. Morphological profiling data is complementary to transcriptional profiling data (Wawer et al. 2014) and is less expensive. Morphological profiling has succeeded across several applications, including grouping small-molecule perturbations based on their mechanism of action (Caicedo, Singh, and Carpenter 2016; Bougen-Zhukov et al. 2016), and grouping genes based on morphological profiles derived from cells perturbed by RNA interference (RNAi) (Mukherji et al. 2006; Boutros and Ahringer 2008; Fuchs et al. 2010; Pau et al. 2013). One limitation of RNAi for morphological profiling is that the number of measurements must be limited or else the resulting profiles are dominated by off-target effects, especially seed effects (Singh et al. 2015). Some computational solutions have shown some promise in overcoming this problem for gene expression profiling (Schmich et al. 2015), but their utility is unproven for image-based profiling, and regardless RNAi does not permit analysis of gene variants, only knockdown. Modification of genes via CRISPR will require new libraries of reagents and is as yet untested in morphological profiling.

In the proof-of-concept work presented here, we tested morphological profiling using overexpression in human cells as a general approach to annotate gene and allele function. We profiled a reference series of well-known genes, and a small number of variants thereof, by Cell Painting. In particular, we wondered whether the information content of this strategy would outweigh potential limitations (e.g., due to cellular context or expression level). We found that the approach successfully clustered genes and alleles based on functional similarity, revealed specific morphological changes even when present in only a subpopulation of heterogeneous cells, and uncovered novel functional connections between important biological pathways.

Results

Morphological profiles from Cell Painting of expression constructs are sensitive and reproducible

To profile each exogenously expressed gene (or allele therein), we used our previously developed image-based profiling assay, called Cell Painting (Gustafsdottir et al. 2013; Bray et al. 2016). This microscopy-based assay consists of six stains imaged in five channels and revealing eight cellular components: DNA, mitochondria, endoplasmic reticulum, Golgi, cytoplasmic RNA, nucleoli, actin,

and plasma membrane (Fig. 1A). In five replicates in 384-well plate format, we infected U-2 OS cells (human bone osteosarcoma cells), chosen for their flat morphology and previous validation in the assay, with an arrayed “reference” expression library of 323 open reading frame (ORF) constructs of partially characterized functions (Supplementary file 1A), a subset of which have been previously described (E. Kim et al. 2016). Of these, we prioritized analysis of the 220 constructs that were most closely representative of the annotated full length transcripts (see Methods). Morphological profiles were extracted using CellProfiler for image processing, yielding 1,384 morphological features per cell, and Python/R scripts for data processing, including feature selection and dimensionality reduction (Fig. 1B, and see Methods). This computational pipeline yielded a 158-dimensional profile for each of 5 replicates for each gene or allele tested.

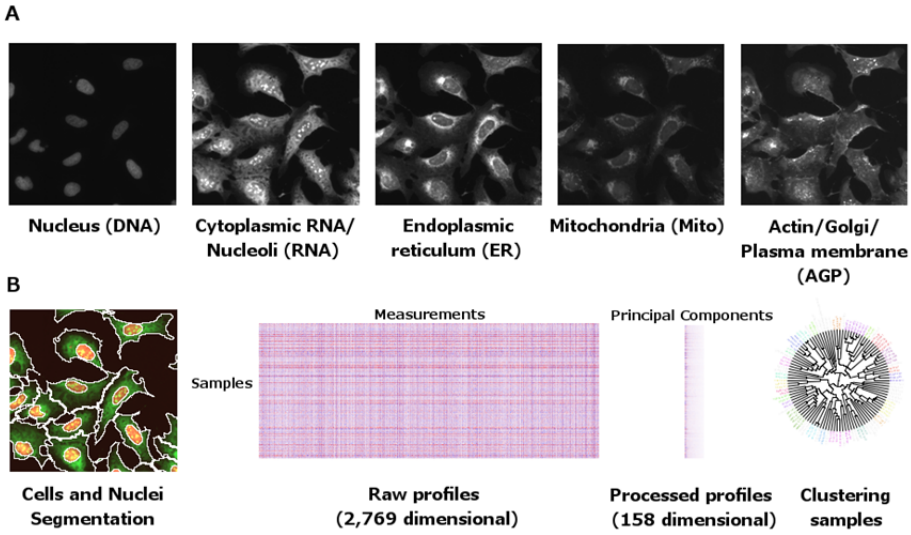


Figure 1: Morphological profiling by Cell Painting. (A) Example Cell Painting images from each of the five channels for a negative control sample (no gene introduced). (B) From left to right : Cell and nucleus outlines found by segmentation in CellProfiler; raw profiles (2,769 dimensional) containing median and median absolute deviation of each of 1,384 measurements over all the cells in a sample, plus cell count; processed profiles which are made less redundant by feature selection and Principal Component Analysis; dendrogram constructed based on the processed profiles (see Fig. 3). Replicates are merged to produce a profile for each gene which is then compared against others in the experiment to look for similarities and differences.

Not all genes are likely to impact cellular morphology given the limitation of our experiment--using a single cell line at a single time point under a single set of conditions and stained with six fluorescent labels. We therefore first asked what fraction of these ORFs impacted morphology. Surprisingly, we found that 50% (110/220) of these ORF constructs induced reproducible morphological profiles distinct from negative control profiles (Fig. 2A, and see Methods). Next, we ruled out the possibility that position artifacts may have artificially inflated this result by taking an alternative pessimistic null distribution which takes well position into account (Fig. 2-figure supplement 1). Therefore, we conclude that a single “generic” morphological profiling assay can detect signal from a substantial proportion of genes in our reference set. We next turned to testing whether those signals are biologically meaningful and can lead to novel, unbiased discoveries about gene function.

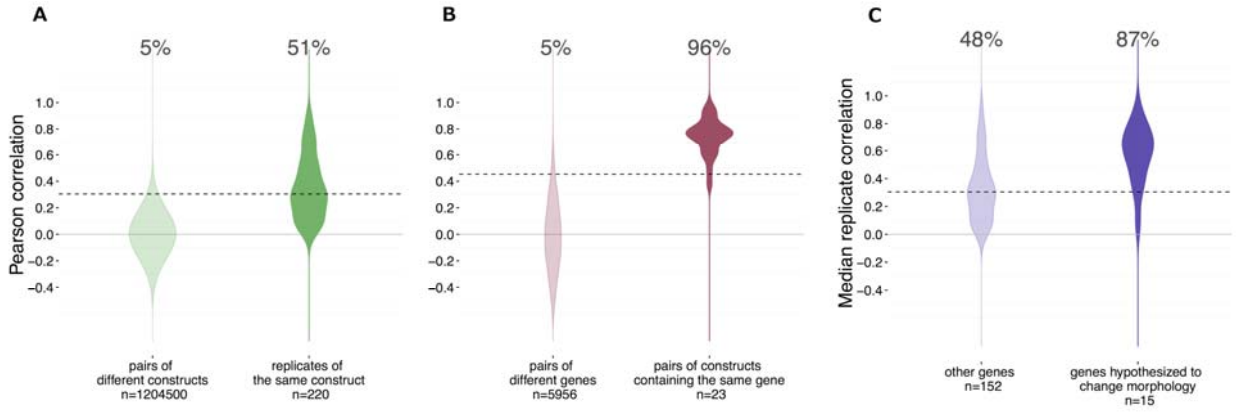


Figure 2: (A) 50% of the gene overexpression constructs produced a detectable phenotype by image-based profiling. Constructs yielding a reproducible phenotype ought to have a median correlation among replicates that is higher than the 95th percentile of correlations seen for pairs of different constructs; this is true for 51% (112 out of 220) of the constructs (as shown). Additionally, we removed two constructs that passed that filter but whose profiles were highly similar to negative control profiles (not shown), leaving 110 constructs (50%) for further analysis. **(B) Of wild-type ORF pairs that both yielded a distinguishable phenotype, 96% showed significant correlation to each other.** Correlations between the 23 pairs of constructs that are clones of the same gene (although with potential sequence variation or possibly different isoforms) were almost always much higher than correlations between pairs of constructs related to different genes. The threshold, shown as the dashed line, is set to 95th percentile of profile correlation for pairs of different genes. Profile correlation of these 23 pairs lie above the threshold. **(C) Genes in pathways thought to regulate morphology were more likely to yield detectable phenotypes** vs. the remainder of genes in the experiment. The same cutoff as in (A) is used to identify percentage of genes with a detectable phenotype. This percentage is 87% for the genes hypothesized to change morphology, while it is 48% for the other genes.

Morphological profiling is robust, showing expected relationships

Given that technical replicates produce similar morphological profiles, we next evaluated whether similarities between profiles induced by different constructs are meaningful. We began with the simplest case: for a subset of genes in the experiment, a “wild-type” sequence (see Methods for important definitions) was captured in more than one ORF construct (23 pairs). These pairs either correspond to different physical cloning events and preparations but with highly similar full-length sequence (as defined in Methods; category a: 9 pairs), or a substantive difference in their nucleotide sequence, for example, isoforms (category b: 14 pairs). We found that, as expected, the phenotypes of over-expressed wild-type ORFs of the same gene were more similar to each other, on average, than to randomly selected genes. Of the 23 pairs for which both wild-type ORFs yielded a phenotype distinguishable from negative controls, 22 (~96%) of the pairs’ profiles were correlated more than expected by chance (Fig. 2B, the one pair not meeting that threshold was in category b), confirming that different constructs with biological similarity indeed produce similar morphological profiles.

This result also confirms that the sequence differences seen in separately cloned wild-type constructs do not generally have a major functional impact, but we caution that any individual construct of interest may have an impactful mutation; thus the raw sequence data should be

examined and testing alternate constructs for a gene may be recommended. Note that if, for example, only 50% of wild-type pairs showed high profile correlation, it would remain ambiguous whether it was caused by the poor assay quality or the constructs' sequence mismatches. But in this particular case the mentioned near perfect consistency rules out either of the two possibilities. We also note that the 23 pairs analyzed here are located in different well locations on each plate; this result therefore also rules out widespread artifacts, such as plate position effects or metadata errors.

We suspected that the small number of engineered constitutively activating alleles for certain genes would, on average, yield a stronger phenotype than their wild-type counterparts. We indeed found that correlations between replicates of the constitutively activating allele were typically higher than correlations between replicates of the wild-type version of a gene (Supplementary file 1B; p-value = 0.012, one-sided paired t-test).

We hypothesized that genes in pathways known to affect cellular morphology (RAC1, KRAS, CDC42, RHOA, PAK1, and genes related to the Hippo pathway) would be more likely to yield a morphological phenotype distinguishable from negative controls than other genes in the analysis. Indeed, we found this to be true (Fisher's test p-value = 3.7×10^{-3}) (Fig. 2C). Reassured by this validation, we were curious which pathways would be most and least likely to yield detectable morphological phenotypes, recognizing that "pathways" are neither separate nor well-defined entities. We found genes manually annotated as being in the Hippo, Hedgehog, cytoskeletal reorganization, and Mitogen-activated protein kinases (MAPK) pathways were more likely to result in a phenotype, whereas genes annotated as belonging to the JAK/STAT, hypoxia, and BMP pathways were among the least likely to yield a phenotype under the conditions tested (Fig. 2-figure supplement 2 and Supplementary file 1C). Nevertheless, the majority of pathways could be interrogated by morphological profiling.

Morphological signature similarity captures known gene-gene relationships

Given the caveats and limitations of overexpressing genes (see Discussion), we next tested whether image-based profiling of expression constructs could capture relationships among genes known to be functionally related. Because a reliable and complete map of all gene-gene connections is not available, we evaluated the accuracy of our results via two approaches.

First, we compared our data to protein-protein interaction data from BioGRID (Stark et al. 2006). This is imperfect ground truth for judging our predictions because two proteins might physically interact without producing the same morphological phenotype when overexpressed, and genes in the same pathway might regulate the same phenotype without any physical interaction. Nevertheless, we expect that the corresponding proteins of gene pairs with highest profile similarity are more likely than average to physically interact. Indeed, looking at wild-type versions of genes showing a detectable phenotype (the 73 genes represented in the 110 constructs), the ratio of verified gene connections among the top 5% correlated gene pairs (9%, 13 verified out of 143 possible combinations) is significantly higher than that of other gene pairs (5%, 128 verified out of 2,485 possible; Fisher's test p-value = 0.04; Supplementary file 1D).

Second, we manually annotated each gene for the pathway with which it is associated. This approach is based on expert opinion and thus imperfect knowledge of all genes' function; furthermore many pathways interrelate, and genes in the same pathway are not expected to have

identical phenotypes given that their functions are rarely identical (most notably, overexpression of some may activate while others suppress a biological pathway or process). Nonetheless, we expect pairs of genes whose morphological profiles correlate highly to be more likely than average to be annotated in the same pathway vs. different pathways. Using the same 73 genes as in the previous analysis, the ratio of gene connections with the same-pathway annotation in the top 5% most-correlated gene pairs was 20% (29 pairs out of 143), significantly higher than the ratio for the remaining pairs (6%, 139 pairs out of 2,485; Fisher's test p-value = 7.53×10^{-9}; Supplementary file 1E).

An initial morphological map of gene function

Having quantitatively established that morphological profiling is sensitive, robust, and captures known gene-gene relationships, we explored these relationships in a correlation matrix (Fig. 3 bottom left and Fig. 3-figure supplement 1). The overall structure, with multiple groupings along the diagonal, is consistent with the fact that the 110 constructs (73 unique genes) that showed a phenotype had been annotated as representing 19 different pathways. That is, we did not see large, homogeneous clusters, as would be expected if morphological profiling was sensitive to perturbation but not highly specific. This rules out uniform toxicity induced by a large number of genes, for example. Neither did we see only signal along the diagonal, which would have indicated no strong similarity between any gene pairs.

We next created a dendrogram (Fig. 3) and defined 25 clusters (see Methods and Fig. 3-figure supplement 2) to explore the similarities among genes. Pairs of wild-type ORFs almost always clustered adjacently, consistent with our quantitative analysis described above (Fig. 2B). After retaining only one copy of replicate ORFs, we found that the majority of clusters (19 out of the 22 clusters containing more than one gene) were enriched for one or more Gene Ontology terms (Supplementary file 1F), indicating shared biological functions within each cluster.

Using this dendrogram, we began by interrogating three clusters that conformed well to prior biological knowledge. First, we analyzed Cluster 20, containing the two canonical Hippo pathway members YAP1 and WWTR1 (more detail in Supplementary file 2-20A and Supplementary file 2-20B PDFs, and in a later section of the text). Both are known to encode core transcriptional effectors of the Hippo pathway (Johnson and Halder 2014), and a negative regulator of these proteins, STK3 (also known as MST2), is the strongest anti-correlating gene for the cluster (Supplementary file 2-20A PDF, panel c1).

Second, we noted Cluster 21 is comprised of the two phosphatidylinositol 3-kinase signaling/Akt (PI3K) regulating genes, PIK3R1 and PTEN, both frequently mutated across 12 cancer types in The Cancer Genome Atlas (TCGA) (Kandoth et al. 2013). These results are consistent with previous observations that certain isoforms of PIK3R1 reduce levels of activated Akt, a dominant negative effect (Abell et al. 2005) AKT3 is in a cluster anti-correlated to the Cluster 21 (Supplementary file 2-21A PDF, panel b1).

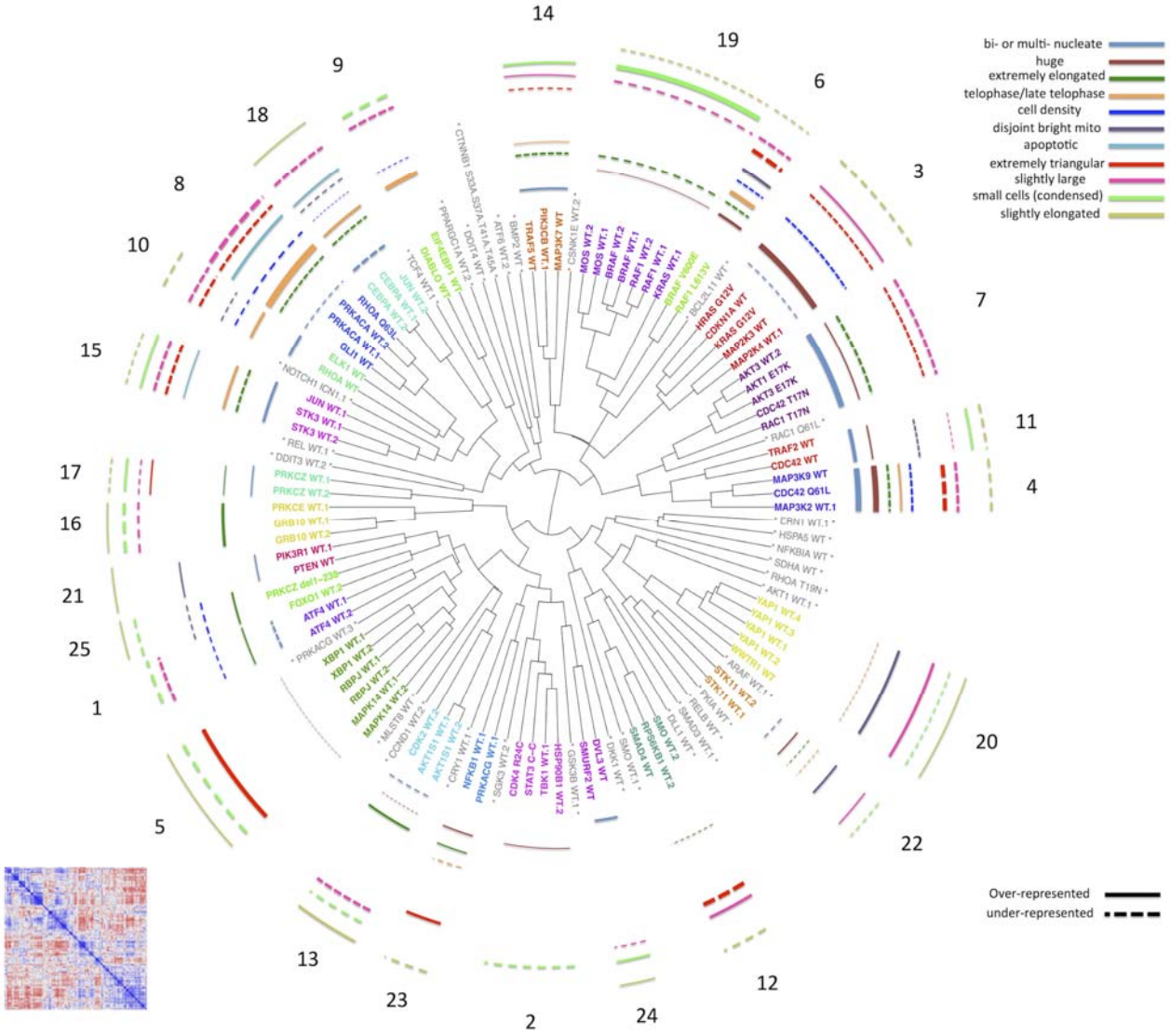


Figure 3: Morphological relationships among overexpressed genes/alleles, determined by Cell Painting. Correlations between pairs of genes/alleles were calculated and displayed in a correlation matrix (bottom left inset, full resolution is available as Fig. 3-figure supplement 1). Only the 110 genes/alleles with a detectable morphological phenotype were included. The rows and columns are ordered based on a hierarchical clustering algorithm such that each blue submatrix on the diagonal shows a cluster of genes resulting in similar phenotypes. The correlations were then used to create a dendrogram (main panel) where the radius of the subtree containing a cluster shows the strength of correlation. The 25 clusters containing at least two constructs are printed on the dendrogram in arbitrary colored fonts, while gene names colored gray and marked by asterisks are those that do not correlate as strongly with their nearest neighbors (i.e., they are singletons or fall below the threshold used to cut the dendrogram for clustering). Each colored arc corresponds to a cell subpopulation as noted in the legend. Line thickness indicates the strength of enrichment of the subpopulation in the cluster samples compared to the negative control. Solid vs. dashed lines indicate the over- vs. under-representation of the corresponding subpopulation in a cluster, respectively. Note that the number next to each cluster in the dendrogram is referenced in the main text and corresponds to the numbered supplemental data file for each cluster.

Third, we examined three Clusters (19, 6 and 3) that included many MAPK-related genes. Cluster 19 is the largest example of a tight cluster of genes already known to be associated; it includes four activators in the RAS-RAF-MEK-ERK cascade: KRAS, RAF1 (CRAF), BRAF, and MOS. Notably, two constitutively active alleles of these genes, $BRAF^{V600E}$ (H. Davies et al. 2002) and $RAF1^{L613V}$ (Wu et al. 2011), form a separate cluster (Cluster 6) adjacent to their wild-type counterparts. Furthermore, the constitutively active RAS alleles $HRAS^{G12V}$ and $KRAS^{G12V}$ (McCOY, Bargmann, and Weinberg 1984) are in the next-closest cluster (Cluster 3), which also contains MAP2K4 and MAP2K3 (known to be activated by Ras (Shin et al. 2005)), as well as CDKN1A (Jalili et al. 2012). By contrast, MAPKs that are known to be unrelated to the RAS-RAF-MEK-ERK cascade, such as MAPK14 in Cluster 5, are far away in the dendrogram.

Overall, these results support the notion that connections between genes can be efficiently discovered using our approach.

Visualization approaches to assist interpretation of morphological signatures

We hypothesized that the specific morphologic features that segregated each of the clusters would provide insight into gene function. Examining images (Supplementary file 2-19A PDF, panel 3) or rank-ordered lists of features that distinguish individual profiles or clusters (Supplementary file 1G) is tedious and lacks sensitivity for all but the most obvious of phenotypes, confirming that quantitative morphological profiling is more sensitive than the human visual system.

We therefore devised several strategies to enhance biological interpretability from these experiments and applied these in combination. First, we grouped features into meta-features based on their type of measurement, i.e., shape, texture, intensity, etc., and the cell constituents to which they are related, to create a Feature Grid (Fig. 4A). Second, we performed unsupervised grouping of features by mapping the top 20 most-distinguishing features for each cluster onto a plane, creating a Feature Map (Fig. 4B), in which highly correlated features are mapped nearby each other (see “Feature Interpretation” in Methods for an explanation of individual feature names). In certain cases, these visualizations revealed the nature of the morphological phenotype (e.g., nuclear shape abnormalities distinguishing Supplementary file 2-7A PDF), but for others these approaches did not suffice to yield an obvious phenotypic conclusion (e.g., for Cluster 19, Fig. 4A and 4B).

Third, we hypothesized that leveraging the single-cell resolution of image-based profiling might be highly sensitive in enhancing interpretation, particularly for cases where only a subset of cells is distinctive from negative controls. To test this, for each given cluster of genes together with negative controls we identified 20 subpopulations using k-means clustering on single cell data. We calculated the abundance of cells in each of the 20 subpopulations to determine which are over/under-represented relative to controls for the given cluster (corresponding images are shown; Supplementary file 2-1B, 2B, … 25B PDFs. For example, the MAPK pathway activators in Cluster 19 show increased prevalence of a subpopulation of cells with strongly asymmetric ER, mitochondria, and Golgi staining, indicating a cell polarization phenotype (Fig. 4C, and Supplementary file 2-19B PDF, Categories 1 and 2), for which there is evidence in the literature (Šamaj, Baluška, and Hirt 2004; Elsum, Martin, and Humbert 2013; Godde et al. 2014). This phenotype was not captured by manual inspection nor the first two approaches (e.g., Supplementary file 2-19A PDF, panels a2 and b2).

Encouraged by this, we supplemented the morphological map by compiling these and other visualizations into PDF files for each cluster, summarized in Fig. 5 and provided in full as Supplementary file 2. We also noticed that certain subpopulations were similar across several clusters (Fig. 3-figure supplement 3 shows sample cell images of each such subpopulation); we annotated their enrichment/de-enrichment on the dendrogram (Fig. 3).

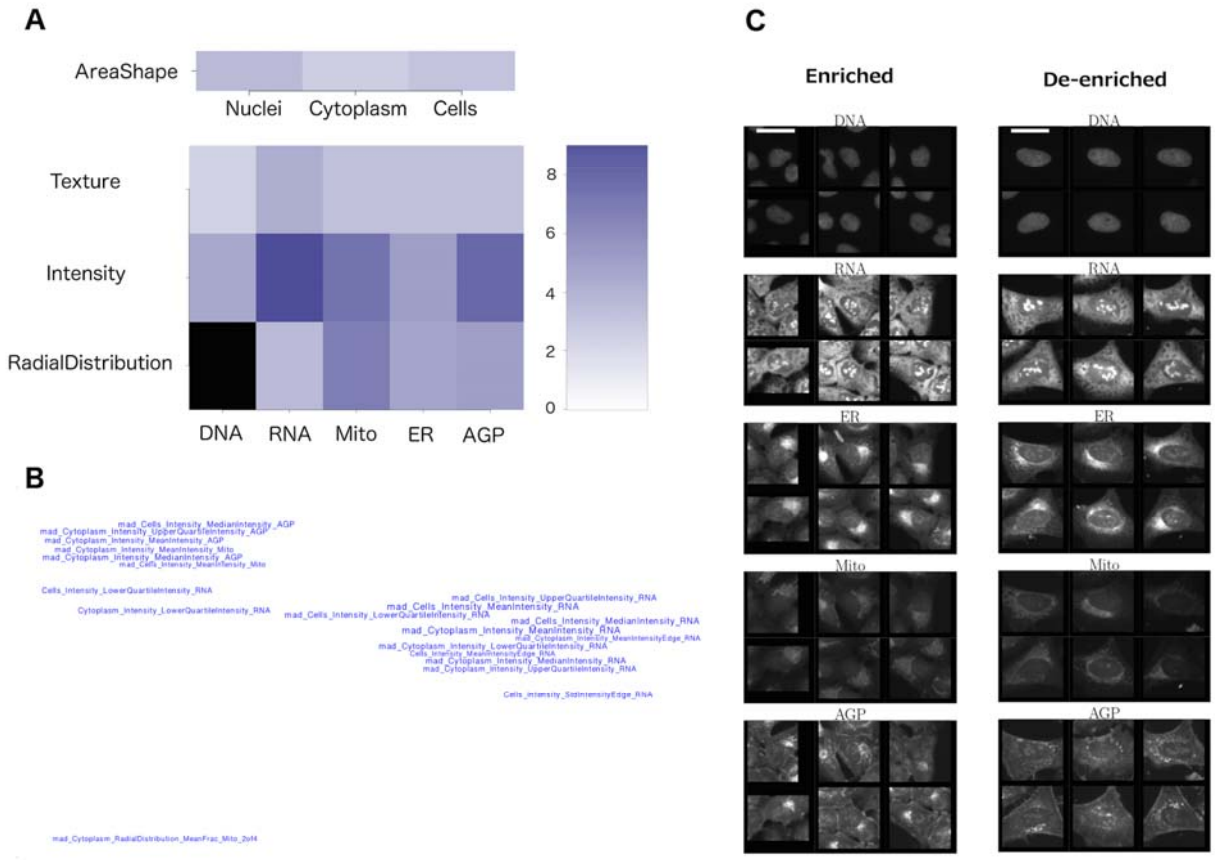


Figure 4: Visualizations used to interpret morphology of Cluster 19 (for other clusters, see Supplementary file 2-1A, …, 25A PDF files). (A) Feature Grid. RNA and AGP (actin, Golgi, plasma membrane) intensity contribute most to the genes in Cluster 19 (KRAS, RAF1, BRAF, and MOS). Dark blue colors indicate higher median z-score of the relevant measurements for genes in the cluster relative to negative controls. As "RadialDistribution" features do not exist for the DNA channel, it is colored in black. **(B) Feature Map.** The feature names showing the greatest difference between the cluster and negative controls are shown, based on largest absolute value of z-scores (full resolution version is available in Cluster 19A PDF). They are mapped in 2D space such that features that are highly correlated with each other across all genes' profiles are placed close together and thus can be interpreted together. Blue/red colored names indicate positive/negative sign of the z-score (i.e., blue indicates that the cluster shows higher values than controls). According to this map, the average intensity of AGP, RNA and Mito shows high variation for cells within samples in Cluster 19 (e.g., large mad_Cytoplasm_Intensity_MeanIntensity_AGP, where the prefix "mad" refers to median absolute deviation, a robust form of standard deviation). **(C) Sample images of a subpopulation of cells enriched and de-enriched for all genes in Cluster 19.** Cells with asymmetric organelle distribution are highly over-represented for genes in the cluster, and cells with more even distribution of organelles are less abundant. Note that the exemplar cells are shown at the center of the patches. This explains the duplications observed in some patches. Scale bars are 39.36 μm long. Pixel intensities are multiplied by 5 for display.

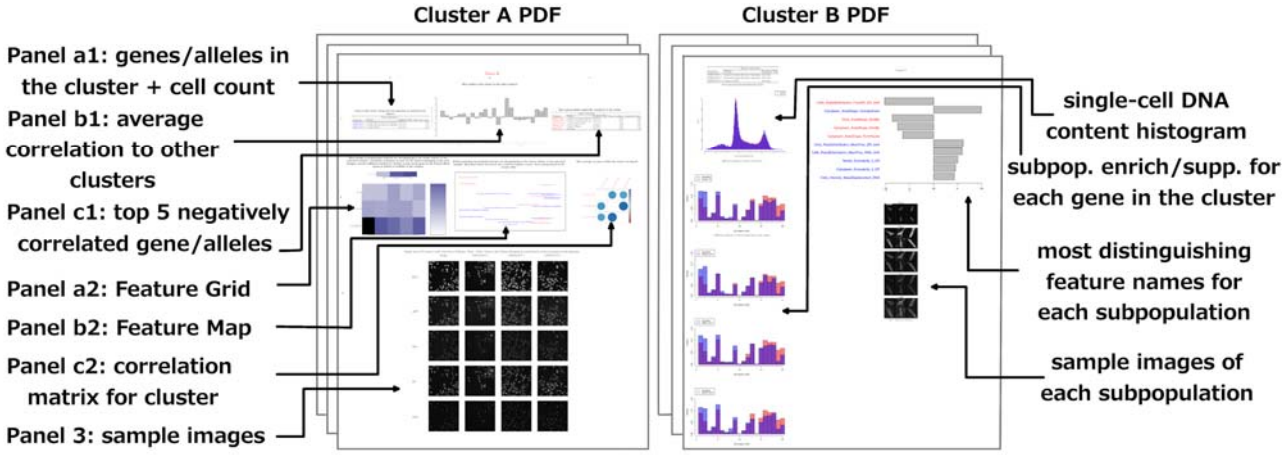


Figure 5: Data and visualizations supporting the morphological map for each cluster. For all 25 clusters, there are two corresponding Supplemental PDF files. **Left: Supplementary file 2-type A PDFs** (e.g., "1A.pdf") provide an overview of data about the cluster. **Panel a1** lists the genes/alleles in the cluster as well as expert annotations regarding related pathways and the cell count (as a z-score) for each gene/allele. **Panel b1** contains the average correlation of the cluster to other clusters, indicating uniqueness of the cluster's morphological phenotype. **Panel c1** lists the top five negatively correlated gene/alleles to the cluster. **Panel a2** shows the Feature Grid summarizing categories of morphological features distinguishing the cluster from the negative control. **Panel b2** shows the Feature Map displaying the names of the top 20 morphological features distinguishing the cluster from the negative control, positioned based on similarity. Explanations for feature names can be found in the Methods section. **Panel c2** shows a correlation matrix for just those genes/alleles in the cluster. **Panel 3** contains sample images of fields of view of cells expressing each gene/allele in the cluster, along with images of the control for comparison. **Right: Supplementary file 2-type B PDFs** contain multiple plots aiming to illustrate the phenotype based on single-cell data, including cell subpopulation enrichment/suppression in the cluster. First, a histogram of single-cell DNA content is shown for all cells from all genes/allele treatments in the cluster, indicating the overall cell cycle distribution. Next, bar plots show (for the cluster overall and for each gene in the cluster) which of 20 subpopulations of cells are enriched and suppressed relative to negative controls. Finally, each subsequent page of the PDF is devoted to the subpopulations whose representation differs from negative controls in a statistically significant way, whether enriched or suppressed (subpopulations which are very small in both the cluster and negative control samples are omitted). For each subpopulation, a bar plot shows the top 10 most-distinguishing feature names (versus negative control cells). Then, sample images are shown of individual representative cells from each subpopulation.

Using these visualizations, we began by interrogating three adjacent and correlating clusters (Clusters 4, 7, and 11) contain wild-type and mutant alleles of CDC42, a gene encoding a Rho family GTPase with diverse roles in cell polarity, morphology, and migration (Melendez, Grogg, and Zheng 2011; Martin 2015). Cluster 4 contains the constitutively active mutant CDC42 Q61L (Nobes and Hall 1999) as well as MAP3K2 and MAP3K9. The highly similar Cluster 7 contains the dominant negative alleles CDC42 T17N (Nobes and Hall 1999) and RAC1 T17N (S. Zhang et al. 1995), a related RAS superfamily member. That activating and inhibiting alleles would yield similar phenotypes when overexpressed is not surprising for CDC42 (Melendez, Grogg, and Zheng 2011). Cluster 7 also contains isoforms and alleles of AKT: specifically, AKT3 and the constitutively active E17K alleles of both AKT1 and AKT3 (M. S. Kim et al. 2008; M. A. Davies et al. 2008). Akt is known to be essential for certain Cdc42-regulated functions (Higuchi et al. 2001) and vice versa

(Stengel and Zheng 2012). Finally, the nearby Cluster 11 (which is discussed in more detail later) contains the wild-type form of CDC42 as well as TRAF2, a canonical NF-κB activator; these two are known to interact and share functions in actin remodeling (Marivin et al. 2014). We also note that anti-correlating genes to these clusters (generally in Clusters 13 and 21) are consistent with existing knowledge, including (a) AKT family member AKT1S1 (a Proline rich AKT substrate, PRAS40 (Kovacina et al. 2003; Wiza et al. 2014), Supplementary file 2-7A PDF panels b1 and c1) (b) CDK2 (a known target of Akt (Maddika et al. 2008)), (c) PIK3R1 and PTEN in Cluster 21, described previously, which have known interactions with AKT (Cheung and Mills 2016; Hemmings and Restuccia 2015). Thus, all of these connections have previously been identified.

Subpopulation visualization revealed that Clusters 4, 7, and 11 are enriched in cells that are huge and binucleate (Fig. 3, example images shown in Supplementary file 2-4B PDF). Genes in all three clusters also show irregularities in DNA content, namely, an enrichment in cells with sub-2N DNA content, a decrease in cells with 2N DNA content, and, for most genes, a decrease in cells with S phase and 4N DNA content, indicating a significant amount of DNA fragmentation and thus apoptosis (DNA histograms in Supplementary file 2-4B, 7B, and 11B PDFs). These phenotypes are consistent with these genes' known role in the cell cycle and cell polarity (Chircop 2014).

As a second test case, we examined Cluster 8, which contains PRKACA (the catalytic subunit α of protein kinase A, PKA) and two of its known substrates: GLI1 (a transcription factor mediating Hedgehog signaling)(Asaoka 2012), and $RHOA^{Q63L}$ (a Ras homolog gene family member)(Lang et al. 1996; Rolli-Derkinderen et al. 2005). The highly similar Cluster 10 contains the wild-type RHOA, as well as ELK1 which is also linked to the Rho GTPase family and PKA (Bachmann et al. 2013; Murai and Treisman 2002).

We investigated the morphological changes causing these genes to cluster. RhoA is a known regulator of cell morphology and cell rounding is a known related phenotype (Oishi et al. 2012). We found that indeed all members of Clusters 8 and 10 significantly induce cell rounding (Supplementary file 1H). Although cell count is lower for genes Clusters 8 and 10, the degree varies greatly (from z-score -0.67 to -3.02, Supplementary file 2-8A and 10A PDFs, panel a1), ruling out that simple sparseness of cells explains their high similarity in the assay. As well, the overall DNA content distribution of the cell populations appears relatively normal (Supplementary file 2-8B and 10B PDFs). Subpopulation extraction provides a satisfying biological explanation for these clusters' distinctive phenotype: the increased roundness and strong variation in intensity levels (per the Feature Grid) across the population stems from an increased proportion of telophase, anaphase, and apoptotic cells (Fig. 3 and Supplementary file 2-8B and 10B PDFs).

We therefore conclude that the morphological map can link related genes to each other and that the morphological data can provide insight into their functions, particularly with the help of subpopulation visualization.

An unexpected relationship between the Hippo pathway and regulators of NF-κB signaling (Clusters 11, 20, and 22)

We wondered whether novel relationships might emerge from our unbiased classification of gene and allele function based on morphologic profiling. We noticed that the known regulator of NF-κB signaling, TRAF2 (in Cluster 11, together with CDC42) (Grech et al. 2004; Tada et al. 2001), yields a signature strongly anti-correlated to YAP1/WWTR1 (Cluster 20), which encode the transcriptional

effectors of the Hippo pathway, YAP (Yes-associated protein) and TAZ (Transcriptional co-activator with a PDZ-domain). The Hippo pathway and NF-κB signaling are critical regulators of cell survival and differentiation, and dysregulation of these pathways is implicated in a number of cancers (Varelas 2014; Hoesel and Schmid 2013; Tornatore et al. 2012), but we found no evidence in the literature (in particular through BioGRID) of physical interaction between the proteins encoded by Cluster 11 genes and Cluster 20 genes. Confirming our approach, a functional connection between CDC42 (Cluster 11) and YAP1 (Cluster 20) has been identified: deletion of CDC42 phenocopies the loss of YAP1 in kidney-specific conditional knockouts in mice (Reginensi et al. 2013). Still, the NF-κB pathway (and in particular the Cluster 11 member TRAF2), has not been closely tied to YAP and TAZ in human cells (see Discussion).

We first wanted to characterize Clusters 11 and 20 to confirm that relationships within each cluster are supported in the literature. Indeed we found evidence for most of the within-cluster connections. CDC42 and TRAF2 (Cluster 11) physically interact and share functions in actin remodeling (Marivin et al. 2014). As described in a prior section YAP/TAZ (Cluster 20) are known to share functional similarities in the Hippo pathway, being regulated by, and also regulating, cytoskeletal dynamics. Consistent with these known functions, we found that core effector of the Hippo pathway which functions to restrict YAP/TAZ nuclear activity, STK3 (which encodes the Mst2 kinase) (Meng, Moroishi, and Guan 2016), has a morphological signature strongly anti-correlated to YAP1/WWTR1 (Supplementary file 2-20A PDF, panel c1). We note that although STK3 and TRAF2 are both moderately anti-correlated with YAP/TAZ (Cluster 20), STK3 and TRAF2 are not themselves highly correlated, indicating each has a different subset of phenotypes that anti-correlate to YAP/TAZ. We also note that two clones of another protein known to influence YAP activity, STK11, form Cluster 22 which falls nearby YAP1/WWTR1; a connection between STK11 and YAP has been identified (albeit with opposite directionality, identified via knockdown of STK11 (Mohseni et al. 2014)). Further, YAP1 is among the highest anti-correlating genes to REL (data not shown; REL is a singleton in the dendrogram and thus not in a cluster), whose protein product, c-Rel, has a known connection to TRAF2 (Jin et al. 2015). These results reaffirm that the Cell Painting-based morphological signatures are a useful reporter of biologically meaningful connections among genes in these pathways.

Given the striking inverse correlation between YAP1/WWTR1 and TRAF2, we sought to confirm a negative regulatory relationship between the Hippo and NF-κB pathways by multiple orthogonal methods.

First, we explored the observed inverse morphological impact using the Cell Painting data. The morphological impact of genes in Cluster 11 and 20 is quite strong (median replicate correlation is at the 74th and 81st percentile, and average within-group correlations are 0.66 and 0.73). Subpopulation analysis showed that Cluster 20 (YAP1, WWTR1) is enriched for cells that are slightly large, slightly elongated, and have disjoint, bright mitochondria patterns, whereas Cluster 11 (TRAF2, CDC42) is de-enriched for those subpopulations and instead enriched for binucleate cells, very large cells, and small cells with asymmetric organelles (Fig. 3 and 6A and 6B).

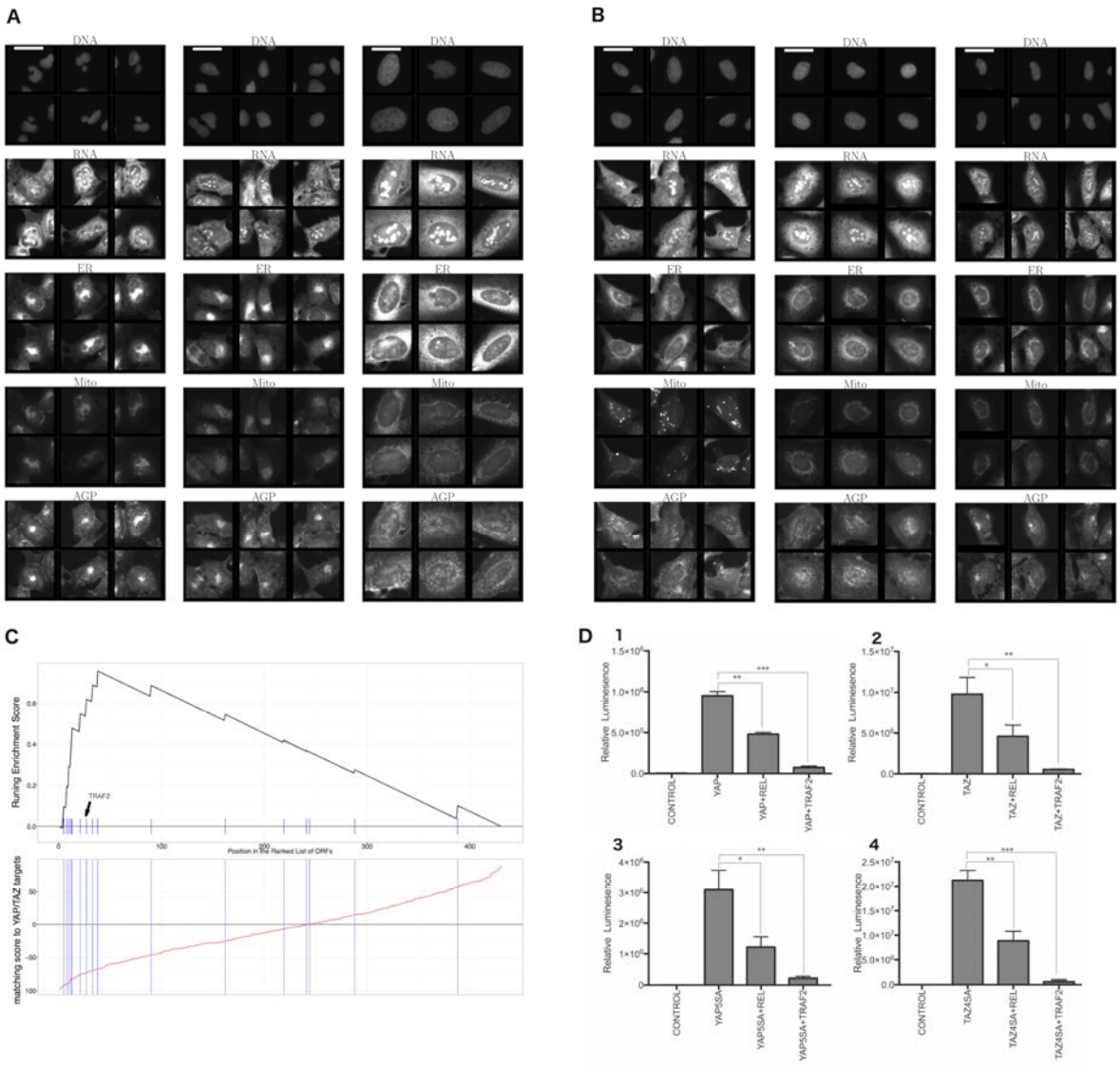


Figure 6: Morphological and transcriptional cross-talk between the Hippo pathway and regulators of NF-κB signaling. (A) The TRAF2/CDC42 cluster (Cluster 11) is enriched for bi-nucleate cells, small cells with asymmetric organelles, and huge cells. Note that exemplar images shown are not labeled with the actual gene they are associated with. Rather they are only supposed to provide a visual insight of the cell morphologies which are enriched in the gene cluster. **(B)** The YAP1/WWTR1 cluster (Cluster 20) is enriched for cells with bright disjoint mitochondria patterns, slightly large cells, and slightly elongated cells. Scale bars are 39.36 μm long. Pixel intensities are multiplied by 5 for display. **(C)** Gene Set Enrichment Analysis (GSEA) reveals that gene overexpression leading to down-regulation of YAP1 targets (CTGF, CYR61, and BIRC5) are enriched for regulators of the NF-κB pathway (Enrichment Score p-value = 8.19×10^{-5}). The horizontal axis gives the index of ORFs sorted based on the average amount of down-regulation of the YAP1 targets. Each blue hash mark on this axis indicates an NF-κB pathway member. The running enrichment score, which can range from -1 to 1, is plotted on the vertical axis and quantifies the accumulation of NF-κB pathways members on the sorted list of ORFs. **(D)** TRAF2 and REL suppress YAP and TAZ transcriptional activity. REL and TRAF2 suppress the ability of wild-type **(D1)** YAP and **(D2)** TAZ to drive the expression of a TEAD-regulated luciferase reporter. Activity of nuclear active mutants of **(D3)** YAP (5SA) and **(D4)** TAZ (4SA) are similarly suppressed. Luciferase reporter activity was measured in HEK293T cells co-transfected with expression constructs as indicated and a TEAD luciferase reporter was used to measure YAP-directed transcription. (* p-value < 0.05, ** p-value = 0.001, *** p-value < 0.0001)

Second, given that YAP/TAZ are transcriptional regulators, we analyzed gene expression data. Using the same constructs as in our Cell Painting experiment, we found an anti-correlated relationship at the mRNA level, consistent with the anti-correlation we had seen in morphological space. To do this, we used Gene Set Enrichment Analysis (Subramanian et al. 2005) and publicly available data, which includes data from four to nine different cell lines at one to four time points (http://lincscloud.org). Time point refers to the duration of treating the cells with over-expression constructs until the time gene expression readouts are made. This analysis revealed that the NF-κB pathway is the pathway most enriched among genes whose overexpression results in down-regulation of known YAP1 targets, CTGF, CYR61, and BIRC5 (Zhao et al. 2008) (Benjamini and Hochberg (BH) adjusted p-value = 2×10^{-8} in Supplementary file 1I, and Fig. 6C), with TRAF2 being among the genes contributing to this enrichment (Supplementary file 1I). We also saw enrichment of NF-κB pathway members when testing a data-driven set of targets of YAP1/TAZ (Fig. 6-figure supplement 1, see Methods). In the inverse analysis, genes that alter the levels of TRAF2/REL common targets are weakly enriched in Hippo pathway members (Fig. 6-figure supplement 2, see Methods). This is consistent with the hypothesis that NF-κB members can downregulate YAP/TAZ targets but not strongly vice versa.

Finally, we more directly confirmed negative crosstalk between NF-κB effectors and YAP/TAZ using a synthetic TEAD luciferase reporter that is YAP/TAZ responsive (Dupont et al. 2011). Importantly, these confirmatory experiments used different cellular contexts and perturbation constructs versus the original Cell Painting data. Co-expression of the NF-κB pathway effectors TRAF2 or C-REL with YAP or TAZ led to significantly lower reporter activity than expression of YAP or TAZ alone (Fig. 6D1 and 6D2). Intriguingly, mutants of YAP or TAZ that are insensitive to negative regulation by the Hippo pathway (YAP-5SA and TAZ-4SA; (Zhao et al. 2008)) remained sensitive to suppression of transcriptional activity by TRAF2 and C-REL, indicating that the negative relationship we identified may be independent of canonical upstream Hippo pathway signals (Fig. 6D3 and 6D4).

Discussion

We conclude that connections among genes can be profitably analyzed using morphological profiling of overexpressed genes via the Cell Painting assay. In a single inexpensive experiment, we were able to rediscover a remarkable number of known biological connections among the genes tested. Further, we found that morphological data from the Cell Painting assay, together with novel subpopulation visualization methods, can be used to flesh out the functionality of particular genes and/or clusters of interest.

By adopting a two-pronged approach, merging this Cell Painting morphological analysis with transcriptional data, we were able to identify an unexpected relationship in human cells between two major signaling pathways, Hippo and NF-κB, both under intense study recently for their involvement in cancer. Through validation of these clustered genes, we have identified that YAP/TAZ-directed transcription is negatively regulated by NF-κB pathway effectors and our data suggests a novel regulatory mechanism that is independent of upstream Hippo kinases.

To date, there has been little evidence of the intersection between these important signaling pathways. Recent work examining osteoclast-osteoblast differentiation has suggested that Hippo

pathway kinases, such as Mst2, may affect the NF-κB pathway through phosphorylation of IkB proteins, thereby promoting nuclear translocation of NF-κB transcription factors (Lee et al. 2015). TAZ was found to be a direct target of NF-κB transcription factors and its expression is regulated via NF-κB signaling (Cho et al. 2010). Our work, however, supports a possible additional mode of interaction, whereby regulators of NF-κB signaling directly regulate the function of Yap and Taz as transcriptional co-factors. Recent work has demonstrated, in *Drosophila*, that NF-κB activation via Toll receptor signaling negatively regulates the transcriptional activity of Yorkie, the homolog of YAP/TAZ, through activation of canonical hippo pathway kinases (B. Liu et al. 2016). The work described here identifies, for the first time in a mammalian system, that a negative regulatory relationship exists between NF-κB activation and YAP/TAZ transcriptional function. Furthermore, we have identified that this regulation of YAP/TAZ occurs in a manner that is independent of Hippo pathway-mediated phosphorylation events on YAP/TAZ, suggesting a more direct relationship between NF-κB and YAP/TAZ signaling.

In this work, we tested quantitatively and explored qualitatively the connections among genes revealed by morphological profiling. Our underlying hypothesis was that functionally similar genes would generally yield morphologically similar cells when overexpressed, and indeed we found this to be the case. Still, some discussion of this point is warranted. Most commonly, gene overexpression will result in activation of the corresponding pathway via amplification of the endogenous gene's function. However, it is important to note that the profiling strategy to discover functional relationships does not assume or require this. For example, overexpression could also disrupt a protein complex, producing a trans-dominant negative effect that results in precisely the opposite phenotypic effect (Veitia 2007). In still other cases, overexpression of a particular gene may not affect any of the normal functions of the gene (producing a false negative signal), or trigger a stress response (yielding a confounded profile), or produce a complicated response, due to feedback loops. Further, artifactual phenotypes could be seen, e.g., if overexpression yields a non-physiological interaction among proteins or toxic aggregates. Nevertheless, despite these caveats and complications, our results indicate that valuable information could be gleaned from the similarity and dissimilarity of the morphological perturbations induced by gene overexpression. Using overexpression avoids the complications of RNAi off-target effects (often due to seed effects), which were far more prevalent (impacting 90% of constructs in our recent study (Singh et al. 2015)).

In addition to functionally annotating genes, as demonstrated here, one particularly appealing application enables personalized medicine: it should be feasible to use morphological profiling to predict the functional impact of various disease alleles, particularly rare variants of unknown significance. This has recently been successful using mRNA profiles (Berger et al. 2016). Thus, an even more exciting prospect would be to combine mRNA profiles with morphological profiles to better predict groups of alleles of similar mechanism, and ultimately to predict effective therapeutics for each group of corresponding patients.

We make all raw images, extracted cellular features, calculated profiles, and interpretive visualizations publicly available, providing an initial morphological map for several major signaling pathways, including several unexplored connections among genes for further study (see Supplementary file 2). Expanding this map to full genome scale could prove an enormously fruitful resource.

Materials and Methods

cDNA constructs used for expression

The Reference Set of human cDNA clones utilized here has been previously described (E. Kim et al. 2016); ~90% of these constructs induce expression of the intended gene greater than 2 standard deviations above the control mean. Briefly, wild-type ORF constructs were obtained as Entry clones from the human ORFeome library version 8.1 (http://horfdb.dfci.harvard.edu) with additional templates generously provided by collaborating laboratories, and cloned into the pDONR223 Gateway Entry vector. In addition, here, to maximize coverage of cellular pathways, we included additional clones with minimal sequence deviations from the intended templates. Sanger sequencing of Entry clones verified the intended transcripts and, if applicable, the intended mutation. Entry constructs and associated sequencing data will be publicly available via www.addgene.org and may also be available via members of the ORFeome Collaboration (http://www.orfeomecollaboration.org/), including the Dana-Farber/Harvard Cancer Center (DF/HCC) DNA Resource Core DNA Repository (http://www.dfhcc.harvard.edu/core-facilities/dna-resource/) and the DNASU Plasmid Repository at ASU Biodesign Institute (http://dnasu.asu.edu/DNASU/Home.jsp). Clone requests must include the unique clone identifier numbers provided in the last column of Supplementary file 1A (e.g. ccsbBroadEn_12345 as an example for a specific entry clone and ccsbBroad304_12345 as an example for a specific expression clone). ORFs were transferred to the pLX304 lentiviral expression vector (Yang et al. 2011) by LR (attL x attR) recombination.

For simplicity, throughout this paper “wild-type” refers to ORFs found in the original collection without a particular known mutation intentionally engineered. Due to natural human variation, and occasional cloning artifacts, there are often non-identical matches of such constructs to reference sequence; these differences are fully documented for each construct and sequence data will be publicly available through AddGene, in addition to the sequencing data for the original Entry clones for the genome-scale library (Yang et al. 2011).

Cell lines

U-2 OS cells (human bone osteosarcoma cells), RRID:CVCL_0042, were obtained from ATCC and propagated in the William Hahn lab; they were not additionally authenticated prior to this experiment. The cell line tested negative for mycoplasma prior to this experiment. HEK293T cells, RRID:CVCL_0063, were obtained from ATCC. The cell line was validated by STR profiling (Genetica DNA Laboratories) and was negative for mycoplasma as measured by MycoAlert Mycoplasma Detection Kit (Lonza).

Lentiviral transduction for morphological profiling

We followed our previously described protocol (E. Kim et al. 2016; Berger et al. 2016) except for durations of some steps. Briefly, cells were plated in 384-well plates and transduced with lentiviral particles carrying ORF constructs the next day. Viral particles were removed 18–24 hr post-infection and cells cultured for 48 hr until staining and imaging (72 hr total post-transduction). The

experiment was conducted in five replicates, each in a different plate. The number of replicates being five was decided based on prior experiments (Bray et al. 2016).

Cell staining and imaging

The Cell Painting assay followed our previously published protocol (Bray et al. 2016). Briefly, eight different cell components and organelles were stained with fluorescent dyes: nucleus (Hoechst 33342), endoplasmic reticulum (concanavalin A/AlexaFluor488 conjugate), nucleoli and cytoplasmic RNA (SYTO14 green fluorescent nucleic acid stain), Golgi apparatus and plasma membrane (wheat germ agglutinin/AlexaFluor594 conjugate, WGA), F-actin (phalloidin/AlexaFluor594 conjugate) and mitochondria (MitoTracker Deep Red). WGA and MitoTracker were added to living cells, with the remaining stains carried out after cell fixation with 3.2% formaldehyde. Images from five fluorescent channels were captured at 20x magnification on an ImageXpress Micro epifluorescent microscope (Molecular Devices): DAPI (387/447 nm), GFP (472/520 nm), Cy3 (531/593 nm), Texas Red (562/624 nm), Cy5 (628/692 nm). Nine sites per well were acquired, with laser based autofocus using the DAPI channel at the first site of each well.

Image processing and feature extraction

The workflow for image processing and cellular feature extraction has been described elsewhere (Bray et al. 2016), but we describe it briefly here. CellProfiler (Carpenter et al. 2006) software version 2.1.0 was used to correct the image channels for uneven illumination, and identify, segment, and measure the cells. An image quality workflow (Bray et al. 2012) was applied to exclude saturated and/or out-of focus wells; six wells containing blurry images were excluded, retaining 1,914 plate/well combinations in the experiment. Cellular morphological, intensity, textural and adjacency statistics were then measured for the cell, nuclei and cytoplasmic sub-compartments. The 1,402 cellular features thus extracted were normalized as follows: For each feature, the median and median absolute deviation were calculated across all untreated cells within a plate; feature values for all the cells in the plate were then normalized by subtracting the median and dividing by the median absolute deviation (MAD) times 1.4826 (Chung et al. 2008). Features having MAD = 0 in any plate were excluded, retaining 1,384 features in all. The image data along with the extracted morphological features at the per-cell level were made publicly available in the Image Data Repository under DOI http://dx.doi.org/10.17867/10000105.

Profiling and data preprocessing

The code repository for the profiling and all the subsequent analysis will be made publicly available at https://github.com/carpenterlab/2016_rohban_submitted. We will next explain details of each analysis step implemented in the code. Single cell measurements in each well and plate position are summarized into the profiles by taking their median and median absolute deviation (abbreviated as “MAD” or “mad” in some tables) over all the cells. Although this method does not explicitly capture population heterogeneity, no alternate method has yet been proven more effective (Ljosa et al. 2013). We also include the cell count in a sample as an additional feature. This results in a vector of 2,769 elements describing the summarized morphology of cells in a sample. We then use the median polishing algorithm immediately after obtaining the summarized profiles, to remove and correct for any plate position artifacts. For each feature, the algorithm de-trends the rows, i.e. by subtracting the row median from the corresponding feature of each profile in that particular row. Next, it de-trends the columns in a similar way using column medians. The

row and column de-trending is repeated until convergence is reached in all the features. For the rest of the analysis we considered only the constructs which have more than 99% sequence identity to both the intended protein and gene transcript, to avoid testing uncharacterized mutations/truncations.

Not all of the morphological features contain useful reproducible information. We first filter out features for which their replicate correlation across all samples (except the negative controls) is less than 0.30, retaining 2,200 features. Subsequently, a feature selection method is used (Fischer et al. 2015). Briefly, starting with features (measurements) that we identify as essential, a new feature that contributes the most information with respect to those that have been chosen, is added to the set. The contribution of each feature to the already-selected features is measured by the replicate correlation of the residue when the feature is regressed on the already selected features. This is repeated until the incremental information added drops below a threshold. The original method proposed in (Fischer et al. 2015) overfits in its regression step when the original data is very high dimensional. As a remedy, in the regression step we only use features that have a Pearson correlation of more than 0.50 with the selected features thus far. This prevents overfitting of regression when the dimensionality of selected features grows. We stop feature selection when the maximum replicate correlation of residue is less than 0.30.

The feature selection method greatly removes redundancy, but because of the non-optimal "greedy" strategy, some redundancy remains. Principal component analysis is then applied to keep 99% of variance in data, resulting in 158 principal components being selected.

Feature interpretation

The features measured using CellProfiler follow a standard naming convention. Each feature name is made up of several tokens separated by underscores, in the following order:

- **Prefix** which could be either empty or "mad". This means that the feature is calculated either by taking median (no prefix) or median absolute deviation ("mad" prefix) of the relevant measurement over all the cells in a sample.
- **Cellular compartment** in which the measurement related to the feature is made, i.e., "Cells", "Cytoplasm", or "Nuclei". Note that features labeled "Nuclei" are based on segmentation of nuclei using Hoechst staining, "Cells" are based on segmentation of the cell edges using the RNA channel, and "Cytoplasm" is the subtraction of the aforementioned compartments.
- **Measurement type**, which can be either "Intensity", "Texture", "RadialDistribution", "AreaShape", "Correlation_Correlation", "Granularity", and "Neighbors". Note that "Correlation_Correlation" measures, within a cellular compartment, the correlation between gray level intensities of corresponding pixel pairs across two channels (specified in the next tokens in the feature name). Note also that the relative positioning of a cell is measured in the "Neighbors" category.
- **Name(s) of channels** in which the measurement is made, if appropriate (omitted for AreaShape and Neighbors).
- **Feature name.** The precise measurement name appears at the end. A description of each metric can be found in the CellProfiler manual (http://d1zymp9ayga15t.cloudfront.net/CPmanual/index.html)

Identifying ORF constructs that are distinguishable from negative controls

Our method to identify which genes produce a discernable profile involves first normalizing each profile to the negative controls, such that a treatment's median replicate correlation becomes a surrogate for phenotype strength. In the case that a treatment does not show a phenotype different from the negative control, its replicates would center around the origin in the feature space. This would consequently decrease the median replicate correlation. On the other hand, a phenotype which is consistently observed in the replicates and is significantly different from the controls results in the replicates to concentrate in a region far from the origin in the feature space, and hence a high median replicate correlation value.

The cutoff for "discernible" is set based on the top 5th percentile of a null distribution. The null distribution is defined based on the correlations between non-replicates (that is, different constructs) in the experiment. Treatments whose replicate correlations are greater than the 95th percentile of the null distribution are considered as "hits" that have a morphological phenotype that is highly reproducible (Fig. 2A).

At this point, for strong treatments, all profiles of the replicates are collapsed by taking the average of individual features. 110 out of the 112 selected ORFs were significantly different from the untreated profiles in the feature space. That is, their average Euclidean distances to the untreated profiles were higher than 95th percentile of untreated profile distances to themselves. This shows these two alternative notions of phenotype strength–replicate reproducibility and distance to negative control–are consistent. We restrict all the remaining analyses to the 110 ORFs.

Comparison of morphological connections between genes to protein-protein interaction data and pathway annotations

In this analysis, mutant alleles were removed and we considered only one wild-type allele for each gene with a detectable phenotype, retaining 73 genes. We calculated a threshold to identify significantly correlated gene pairs. We picked the threshold to minimize the probability of error in classifying wild-type clone pairs versus different-gene pairs. To do so, we found the value at which the probability density functions of the two groups intersect; this value (here, 0.43) can be proved to have the desired property (Duda, Hart, and Stork 2012). This approach results in about 5% of the gene pairs being categorized as highly correlated. We next formed a 2 by 2 contingency table, where the rows correspond to two groups of gene pairs, determined by whether they have high profile correlation or not. Similarly, the columns also correspond to two groups of gene pairs, determined by whether the corresponding proteins have been reported to interact in BioGRID (or alternatively have been annotated to be in the same pathway; Supplementary files 1C and 1D). This table was then used to perform a one-tailed Fisher's exact test.

Creation of a dendrogram relating genes to each other, and agglomerative clustering by cutting the dendrogram

A dendrogram was created based on the Pearson correlation distance and average linkage, using the hclust function in R (Fig. 3).

Gene clusters were formed by cutting the dendrogram at a fixed correlation level, 0.522, which was chosen using a stability-based measure. The measure is defined as follows: the local clustering

stability is measured for a range of candidate cutoffs, from 0.43 (used earlier to test consistency to protein interaction data) to 0.70. The point with highest stability was chosen (Fig. 3-figure supplement 2), and the stability measure was defined as the proportion of treatments whose clusters do not change if the cutoff is slightly changed by a small amount, $\epsilon = .002$.

Subpopulation extraction

In order to extract cell categories (subpopulations) and subpopulation enrichment laid over the dendrogram in Fig. 3, we applied k-means clustering on the normalized single cell data for each gene cluster and the control. Data normalization was carried out on a plate-wise basis by z-scoring each feature using the control samples as reference. In order to avoid curse of dimensionality, we restricted the dataset to the features obtained from the feature selection step mentioned earlier. We set k=20 to be the number of subpopulations. The algorithm was run for at most 5000 iterations. Each cell was assigned to the subpopulation for which it has the shortest Euclidean distance to its center. Then, the number of cells belonging to each cell subpopulation was counted and the proportion in each subpopulation for genes in the cluster was compared against that of the control. If the change in proportion of a cell category was consistent across the genes in the cluster, the cell category is shown in the Supplementary file 2-type B PDFs. To quantify this consistency, we used the inverse coefficient of variation of the change in a category proportion. If this quantity exceeded 1, we called the change consistent and included the corresponding cell category in the PDFs. Images of cells which have highest similarity to the category center in the feature space are then used to interpret and give name to each cell category (Fig. 3-figure supplement 3)

Identifying targets of a gene using a data-driven approach

For this purpose, we used a replicate of the original experiment but with L1000 gene-expression readouts, which is provided in the supplemental data; i.e. cell line, time point, and ORF constructs are the same. This data is different from the data used in creating GSEA plots, which entails multiple cell lines and time points. The mRNA levels are all normalized with respect to the negative control. For each replicate of the overexpression construct, we sort the expression levels of landmark genes and take the list of top and bottom 50 landmark genes. Then, to find targets of the gene related to the construct, we find the landmark genes among this list which has shown up at least in p% of replicates/clones of the gene. In particular, we set p to 33% for YAP1, 50% for WWTR1, TRAF2, and REL. Then, we simply take the intersection of predicted targets of YAP1 and WWTR1 (and similarly TRAF2 and REL, separately) to get their common targets. These targets are then used to produce Fig. 6-figure supplements 1-2.

Gene Set Enrichment Analysis

In order to produce Fig. 6C, we specified the three known targets of YAP/WWTR1 (CYR61, CTGF, and BIRC5) and queried for ORFs resulting in down-regulation of these genes. This scores each ORF (out of the 430 in the dataset) based on the observed change in mRNA level of the specified YAP/WWTR1 targets, across between four to nine different cell lines and between one to four time points. For each ORF, we then sought the summarized score which takes the mean of 4 largest scores across time point/cell line combinations. Finally, the ORFs were sorted based on the summarized score, and top 30 ORFs were tested for enrichment in different pathways

(Supplementary file 1l). We used the “clusterProfiler” package in R and the KEGG pathway enrichment analysis implemented in it for creating the GSEA plot (Yu et al. 2012).

Luciferase Reporter assay

Wild-type and mutant sequences of WWTR1 (TAZ) (4SA: S66A, S89A, S117A, and S311A) and YAP1 (5SA: S61A, S109A, S127A, S164A, and S397A) were previously generated and cloned into the pCMV5 backbone; these constructs are distinct from those used in the original Cell Painting data set. TRAF2 and REL were cloned from the original constructs (using Broad ID# ccsbBroadEn_01710 and ID# ccsbBroadEn_11094, respectively), into pCMV5 expression vectors. These were sequenced and confirmed to BLAST against the appropriate Broad clone ID. The empty pCMV5 backbone was used as the control condition. The Tead luciferase reporter construct, 8xGTIIC-luciferase was a gift from Stefano Piccolo (Addgene plasmid # 34615).

HEK293T cells, RRID:CVCL_0063, were transfected using Turbofect (ThermoFisher Scientific) according to manufacturer’s protocol. All cells were co-transfected with a β-galactosidase reporter plasmid (pCMV-LacZ from Clontech) as a transfection control. Cells were lysed 48 hours following transfection. Lysates were mixed with firefly luciferase (Promega) according to the manufacturer’s protocol and luminescence was measured using a luminometer (BioTek). Lysates were mixed with o-nitrophenyl-β-D-galactoside (ONPG) and β-galactosidase expression was determined spectrophotometrically by measurement of absorbance at 405nm following ONPG cleavage. All luciferase readings were normalized to β-galactosidase expression for the sample. Statistical analysis was conducted using a two tailed unpaired Student’s t test. The data shown in Fig. 6D are from triplicate samples within a single experiment and is representative of replicate experiments.

Acknowledgements

The authors gratefully acknowledge contributions from members of the Carpenter laboratory, especially Steven A. Moore. Funding for this work was provided by the National Science Foundation (NSF CAREER DBI 1148823 to AEC), a BroadNext10 grant from the Broad Institute, and the Slim Initiative for Genomic Medicine, a project funded by the Carlos Slim Foundation in Mexico.

Competing interests

The authors declare no competing financial or non-financial interests.

References

Abell, Kathrine, Antonio Bilancio, Richard W. E. Clarkson, Paul G. Tiffen, Anton I. Altaparmakov, Thomas G. Burdon, Tomoichiro Asano, Bart Vanhaesebroeck, and Christine J. Watson. 2005. “Stat3-Induced Apoptosis Requires a Molecular Switch in PI (3) K Subunit Composition.” *Nature Cell Biology* 7 (4). Nature Publishing Group: 392–98. http://www.nature.com/ncb/journal/v7/n4/abs/ncb1242.html.

Alexa, Adrian, and Jörg Rahnenführer. 2009. “Gene Set Enrichment Analysis with topGO.” Available. https://bioconductor.riken.jp/packages/3.2/bioc/vignettes/topGO/inst/doc/topGO.pdf.

Amberger, Joanna S., Carol A. Bocchini, François Schiettecatte, Alan F. Scott, and Ada Hamosh. 2015. “OMIM. Org: Online Mendelian Inheritance in Man (OMIM®), an Online Catalog of Human Genes and Genetic Disorders.” *Nucleic Acids Research* 43 (D1). Oxford Univ Press: D789–98. http://nar.oxfordjournals.org/content/43/D1/D789.short.

Asaoka, Yoshinari. 2012. “Phosphorylation of Gli by cAMP-Dependent Protein Kinase.” *Vitamins and*

773 *Hormones* 88: 293–307. doi:10.1016/B978-0-12-394622-5.00013-4.
774 Bachmann, Verena A., Anna Riml, Roland G. Huber, George S. Baillie, Klaus R. Liedl, Taras Valovka, and
775 Eduard Stefan. 2013. “Reciprocal Regulation of PKA and Rac Signaling.” *Proceedings of the National*
776 *Academy of Sciences of the United States of America* 110 (21): 8531–36.
777 doi:10.1073/pnas.1215902110.
778 Berger, Alice H., Angela N. Brooks, Xiaoyun Wu, Yashaswi Shrestha, Candace Chouinard, Federica
779 Piccioni, Mukta Bagul, et al. 2016. “High-Throughput Phenotyping of Lung Cancer Somatic Mutations.”
780 *Cancer Cell* 30 (2): 214–28. doi:10.1016/j.ccell.2016.06.022.
781 Bougen-Zhukov, Nicola, Sheng Yang Loh, Hwee Kuan Lee, and Lit-Hsin Loo. 2016. “Large-Scale Image-
782 Based Screening and Profiling of Cellular Phenotypes.” *Cytometry. Part A: The Journal of the*
783 *International Society for Analytical Cytology*, July. doi:10.1002/cyto.a.22909.
784 Boutros, Michael, and Julie Ahringer. 2008. “The Art and Design of Genetic Screens: RNA Interference.”
785 *Nature Reviews. Genetics* 9 (7): 554–66. doi:10.1038/nrg2364.
786 Bray, Mark-Anthony, Adam N. Fraser, Thomas P. Hasaka, and Anne E. Carpenter. 2012. “Workflow and
787 Metrics for Image Quality Control in Large-Scale High-Content Screens.” *Journal of Biomolecular*
788 *Screening* 17 (2): 266–74. doi:10.1177/1087057111420292.
789 Bray, Mark-Anthony, Shantanu Singh, Han Han, Chadwick T. Davis, Blake Borgeson, Cathy Hartland, Maria
790 Kost-Alimova, Sigrun M. Gustafsdottir, Christopher C. Gibson, and Anne E. Carpenter. 2016. “Cell
791 Painting, a High-Content Image-Based Assay for Morphological Profiling Using Multiplexed Fluorescent
792 Dyes.” *Nature Protocols* 11 (9): 1757–74. doi:10.1038/nprot.2016.105.
793 Caicedo, Juan C., Shantanu Singh, and Anne E. Carpenter. 2016. “Applications in Image-Based Profiling of
794 Perturbations.” *Current Opinion in Biotechnology* 39 (June): 134–42. doi:10.1016/j.copbio.2016.04.003.
795 Cantwell-Dorris, Emma R., John J. O’Leary, and Orla M. Sheils. 2011. “BRAFV600E: Implications for
796 Carcinogenesis and Molecular Therapy.” *Molecular Cancer Therapeutics* 10 (3): 385–94.
797 doi:10.1158/1535-7163.MCT-10-0799.
798 Carpenter, Anne E., Thouis R. Jones, Michael R. Lamprecht, Colin Clarke, In Han Kang, Ola Friman, David
799 A. Guertin, et al. 2006. “CellProfiler: Image Analysis Software for Identifying and Quantifying Cell
800 Phenotypes.” *Genome Biology* 7 (10): R100. doi:10.1186/gb-2006-7-10-r100.
801 Cheung, Lydia Wt, and Gordon B. Mills. 2016. “Targeting Therapeutic Liabilities Engendered by PIK3R1
802 Mutations for Cancer Treatment.” *Pharmacogenomics* 17 (3): 297–307. doi:10.2217/pgs.15.174.
803 Chircop, Megan. 2014. “Rho GTPases as Regulators of Mitosis and Cytokinesis in Mammalian Cells.” *Small*
804 *GTPases* 5 (July). doi:10.4161/sgtp.29770.
805 Cho, Hyun Hwa, Keun Koo Shin, Yeon Jeong Kim, Ji Sun Song, Jong Myung Kim, Yong Chan Bae, Chi Dae
806 Kim, and Jin Sup Jung. 2010. “NF-κB Activation Stimulates Osteogenic Differentiation of Mesenchymal
807 Stem Cells Derived from Human Adipose Tissue by Increasing TAZ Expression.” *Journal of Cellular*
808 *Physiology* 223 (1). Wiley Online Library: 168–77.
809 http://onlinelibrary.wiley.com/doi/10.1002/jcp.22024/full.
810 Chung, Namjin, Xiaohua Douglas Zhang, Anthony Kreamer, Louis Locco, Pei-Fen Kuan, Steven Bartz, Peter
811 S. Linsley, Marc Ferrer, and Berta Strulovici. 2008. “Median Absolute Deviation to Improve Hit Selection
812 for Genome-Scale RNAi Screens.” *Journal of Biomolecular Screening* 13 (2): 149–58.
813 doi:10.1177/1087057107312035.
814 Davies, Helen, Graham R. Bignell, Charles Cox, Philip Stephens, Sarah Edkins, Sheila Clegg, Jon Teague,
815 et al. 2002. “Mutations of the BRAF Gene in Human Cancer.” *Nature* 417 (6892): 949–54.
816 doi:10.1038/nature00766.
817 Davies, M. A., K. Stemke-Hale, C. Tellez, T. L. Calderone, W. Deng, V. G. Prieto, A. J. F. Lazar, J. E.
818 Gershenwald, and G. B. Mills. 2008. “A Novel AKT3 Mutation in Melanoma Tumours and Cell Lines.”
819 *British Journal of Cancer* 99 (8): 1265–68. doi:10.1038/sj.bjc.6604637.
820 Duda, R. O., P. E. Hart, and D. G. Stork. 2012. *Pattern Classification*. Wiley.
821 https://books.google.com/books?id=Br33IRC3PkQC.
822 Dupont, Sirio, Leonardo Morsut, Mariaceleste Aragona, Elena Enzo, Stefano Giulitti, Michelangelo
823 Cordenonsi, Francesca Zanconato, et al. 2011. “Role of YAP/TAZ in Mechanotransduction.” *Nature* 474
824 (7350): 179–83. doi:10.1038/nature10137.
825 Elsum, Imogen A., Claire Martin, and Patrick O. Humbert. 2013. “Scribble Regulates an EMT Polarity
826 Pathway through Modulation of MAPK-ERK Signaling to Mediate Junction Formation.” *Journal of Cell*
827 *Science* 126 (Pt 17): 3990–99. doi:10.1242/jcs.129387.
828 Eser, S., A. Schnieke, G. Schneider, and D. Saur. 2014. “Oncogenic KRAS Signalling in Pancreatic Cancer.”
829 *British Journal of Cancer* 111 (5): 817–22. doi:10.1038/bjc.2014.215.
830 Fischer, Bernd, Thomas Sandmann, Thomas Horn, Maximilian Billmann, Varun Chaudhary, Wolfgang

Huber, and Michael Boutros. 2015. “A Map of Directional Genetic Interactions in a Metazoan Cell.” *eLife* 4 (March). doi:10.7554/eLife.05464.

Fuchs, Florian, Gregoire Pau, Dominique Kranz, Oleg Sklyar, Christoph Budjan, Sandra Steinbrink, Thomas Horn, Angelika Pedal, Wolfgang Huber, and Michael Boutros. 2010. “Clustering Phenotype Populations by Genome-Wide RNAi and Multiparametric Imaging.” *Molecular Systems Biology* 6 (June): 370. doi:10.1038/msb.2010.25.

Godde, Nathan J., Julie M. Sheridan, Lorey K. Smith, Helen B. Pearson, Kara L. Britt, Ryan C. Galea, Laura L. Yates, Jane E. Visvader, and Patrick O. Humbert. 2014. “Scribble Modulates the MAPK/Fra1 Pathway to Disrupt Luminal and Ductal Integrity and Suppress Tumour Formation in the Mammary Gland.” *PLoS Genetics* 10 (5): e1004323. doi:10.1371/journal.pgen.1004323.

Grech, Adrian P., Michelle Amesbury, Tyani Chan, Sandra Gardam, Antony Basten, and Robert Brink. 2004. “TRAF2 Differentially Regulates the Canonical and Noncanonical Pathways of NF-kappaB Activation in Mature B Cells.” *Immunity* 21 (5): 629–42. doi:10.1016/j.immuni.2004.09.011.

Gustafsdottir, S. M., V. Ljosa, K. L. Sokolnicki, J. Anthony Wilson, D. Walpita, M. M. Kemp, K. Petri Seiler, et al. 2013. “Multiplex Cytological Profiling Assay to Measure Diverse Cellular States.” *PloS One* 8: e80999. doi:10.1371/journal.pone.0080999.

Hemmings, Brian A., and David F. Restuccia. 2015. “The PI3K-PKB/Akt Pathway.” *Cold Spring Harbor Perspectives in Biology* 7 (4). doi:10.1101/cshperspect.a026609.

Higuchi, M., N. Masuyama, Y. Fukui, A. Suzuki, and Y. Gotoh. 2001. “Akt Mediates Rac/Cdc42-Regulated Cell Motility in Growth Factor-Stimulated Cells and in Invasive PTEN Knockout Cells.” *Current Biology: CB* 11 (24): 1958–62. https://www.ncbi.nlm.nih.gov/pubmed/11747822.

Hoesel, Bastian, and Johannes A. Schmid. 2013. “The Complexity of NF-κB Signaling in Inflammation and Cancer.” *Molecular Cancer* 12 (August): 86. doi:10.1186/1476-4598-12-86.

Jalili, Ahmad, Christine Wagner, Mikhail Pashenkov, Gaurav Pathria, Kirsten D. Mertz, Hans R. Widlund, Mathieu Lupien, et al. 2012. “Dual Suppression of the Cyclin-Dependent Kinase Inhibitors CDKN2C and CDKN1A in Human Melanoma.” *Journal of the National Cancer Institute* 104 (21): 1673–79. doi:10.1093/jnci/djs373.

Jin, Jin, Yichuan Xiao, Hongbo Hu, Qiang Zou, Yanchuan Li, Yanpan Gao, Wei Ge, Xuhong Cheng, and Shao-Cong Sun. 2015. “Proinflammatory TLR Signalling Is Regulated by a TRAF2-Dependent Proteolysis Mechanism in Macrophages.” *Nature Communications* 6 (January): 5930. doi:10.1038/ncomms6930.

Johnson, Randy, and Georg Halder. 2014. “The Two Faces of Hippo: Targeting the Hippo Pathway for Regenerative Medicine and Cancer Treatment.” *Nature Reviews. Drug Discovery* 13 (1): 63–79. doi:10.1038/nrd4161.

Kandoth, Cyriac, Michael D. McLellan, Fabio Vandin, Kai Ye, Beifang Niu, Charles Lu, Mingchao Xie, et al. 2013. “Mutational Landscape and Significance across 12 Major Cancer Types.” *Nature* 502 (7471): 333–39. doi:10.1038/nature12634.

Kim, Eejung, Nina Ilic, Yashaswi Shrestha, Lihua Zou, Atanas Kamburov, Cong Zhu, Xiaoping Yang, et al. 2016. “Systematic Functional Interrogation of Rare Cancer Variants Identifies Oncogenic Alleles.” *Cancer Discovery* 6 (7): 714–26. doi:10.1158/2159-8290.CD-16-0160.

Kim, M. S., E. G. Jeong, N. J. Yoo, and S. H. Lee. 2008. “Mutational Analysis of Oncogenic AKT E17K Mutation in Common Solid Cancers and Acute Leukaemias.” *British Journal of Cancer* 98 (9): 1533–35. doi:10.1038/sj.bjc.6604212.

Kovacina, Kristina S., Grace Y. Park, Sun Sik Bae, Andrew W. Guzzetta, Erik Schaefer, Morris J. Birnbaum, and Richard A. Roth. 2003. “Identification of a Proline-Rich Akt Substrate as a 14-3-3 Binding Partner.” *The Journal of Biological Chemistry* 278 (12): 10189–94. doi:10.1074/jbc.M210837200.

Lamb, Justin, Emily D. Crawford, David Peck, Joshua W. Modell, Irene C. Blat, Matthew J. Wrobel, Jim Lerner, et al. 2006. “The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease.” *Science* 313 (5795): 1929–35. doi:10.1126/science.1132939.

Lang, P., F. Gesbert, M. Delespine-Carmagnat, R. Stancou, M. Pouchelet, and J. Bertoglio. 1996. “Protein Kinase A Phosphorylation of RhoA Mediates the Morphological and Functional Effects of Cyclic AMP in Cytotoxic Lymphocytes.” *The EMBO Journal* 15 (3): 510–19. http://www.ncbi.nlm.nih.gov/pubmed/8599934.

Lee, Jongwon, Bang Ung Youn, Kabsun Kim, Jung Ha Kim, Da-Hye Lee, Semun Seong, Inyoung Kim, et al. 2015. “Mst2 Controls Bone Homeostasis by Regulating Osteoclast and Osteoblast Differentiation.” *Journal of Bone and Mineral Research: The Official Journal of the American Society for Bone and Mineral Research* 30 (9): 1597–1607. doi:10.1002/jbmr.2503.

Leonetti, Manuel D., Sayaka Sekine, Daichi Kamiyama, Jonathan S. Weissman, and Bo Huang. 2016. “A

889 Scalable Strategy for High-Throughput GFP Tagging of Endogenous Human Proteins." *Proceedings of*
890 *the National Academy of Sciences of the United States of America* 113 (25): E3501–8.
891 doi:10.1073/pnas.1606731113.
892 Liu, Bo, Yonggang Zheng, Feng Yin, Jianzhong Yu, Neal Silverman, and Duojia Pan. 2016. "Toll Receptor-
893 Mediated Hippo Signaling Controls Innate Immunity in Drosophila." *Cell* 164 (3): 406–19.
894 doi:10.1016/j.cell.2015.12.029.
895 Liu, Chen-Ying, Xianbo Lv, Tingting Li, Yanping Xu, Xin Zhou, Shimin Zhao, Yue Xiong, Qun-Ying Lei, and
896 Kun-Liang Guan. 2011. "PP1 Cooperates with ASPP2 to Dephosphorylate and Activate TAZ." *The*
897 *Journal of Biological Chemistry* 286 (7): 5558–66. doi:10.1074/jbc.M110.194019.
898 Ljosa, Vebjorn, Peter D. Caie, Rob Ter Horst, Katherine L. Sokolnicki, Emma L. Jenkins, Sandeep Daya,
899 Mark E. Roberts, et al. 2013. "Comparison of Methods for Image-Based Profiling of Cellular
900 Morphological Responses to Small-Molecule Treatment." *Journal of Biomolecular Screening* 18 (10):
901 1321–29. doi:10.1177/1087057113503553.
902 Maddika, Subbareddy, Sudharsana Rao Ande, Emilia Wiechec, Lise Lotte Hansen, Sebastian Wesselborg,
903 and Marek Los. 2008. "Akt-Mediated Phosphorylation of CDK2 Regulates Its Dual Role in Cell Cycle
904 Progression and Apoptosis." *Journal of Cell Science* 121 (Pt 7): 979–88. doi:10.1242/jcs.009530.
905 Marivin, A., J. Berthelet, J. Cartier, C. Paul, S. Gemble, A. Morizot, W. Boireau, et al. 2014. "cIAP1 Regulates
906 TNF-Mediated cdc42 Activation and Filopodia Formation." *Oncogene* 33 (48): 5534–45.
907 doi:10.1038/onc.2013.499.
908 Martin, Sophie G. 2015. "Spontaneous Cell Polarization: Feedback Control of Cdc42 GTPase Breaks
909 Cellular Symmetry." *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*
910 37 (11): 1193–1201. doi:10.1002/bies.201500077.
911 McCOY, Melissa S., Cornelia I. Bargmann, and Robert A. Weinberg. 1984. "Human Colon Carcinoma Ki-
912 ras2 Oncogene and Its Corresponding Proto-Oncogene." *Molecular and Cellular Biology* 4 (8). Am Soc
913 Microbiol: 1577–82. http://mcb.asm.org/content/4/8/1577.short.
914 Melendez, Jaime, Matthew Grogg, and Yi Zheng. 2011. "Signaling Role of Cdc42 in Regulating Mammalian
915 Physiology." *The Journal of Biological Chemistry* 286 (4): 2375–81. doi:10.1074/jbc.R110.200329.
916 Meng, Zhipeng, Toshiro Moroishi, and Kun-Liang Guan. 2016. "Mechanisms of Hippo Pathway Regulation."
917 *Genes & Development* 30 (1): 1–17. doi:10.1101/gad.274027.115.
918 Mohseni, Morvarid, Jianlong Sun, Allison Lau, Stephen Curtis, Jeffrey Goldsmith, Victor L. Fox, Chongjuan
919 Wei, et al. 2014. "A Genetic Screen Identifies an LKB1-MARK Signalling Axis Controlling the Hippo-YAP
920 Pathway." *Nature Cell Biology* 16 (1): 108–17. doi:10.1038/ncb2884.
921 Mukherji, Mridul, Russell Bell, Lubica Supekova, Yan Wang, Anthony P. Orth, Serge Batalov, Loren Miraglia,
922 et al. 2006. "Genome-Wide Functional Analysis of Human Cell-Cycle Regulators." *Proceedings of the*
923 *National Academy of Sciences of the United States of America* 103 (40): 14819–24.
924 doi:10.1073/pnas.0604320103.
925 Murai, Kasumi, and Richard Treisman. 2002. "Interaction of Serum Response Factor (SRF) with the Elk-1 B
926 Box Inhibits RhoA-Actin Signaling to SRF and Potentiates Transcriptional Activation by Elk-1." *Molecular*
927 *and Cellular Biology* 22 (20): 7083–92. https://www.ncbi.nlm.nih.gov/pubmed/12242287.
928 Nobes, C. D., and A. Hall. 1999. "Rho GTPases Control Polarity, Protrusion, and Adhesion during Cell
929 Movement." *The Journal of Cell Biology* 144 (6): 1235–44.
930 https://www.ncbi.nlm.nih.gov/pubmed/10087266.
931 Oishi, Atsuro, Noriko Makita, Junichiro Sato, and Taroh Iiri. 2012. "Regulation of RhoA Signaling by the
932 cAMP-Dependent Phosphorylation of RhoGDIα." *The Journal of Biological Chemistry* 287 (46): 38705–
933 15. doi:10.1074/jbc.M112.401547.
934 Pau, Gregoire, Thomas Walter, Beate Neumann, Jean-Karim Hériché, Jan Ellenberg, and Wolfgang Huber.
935 2013. "Dynamical Modelling of Phenotypes in a Genome-Wide RNAi Live-Cell Imaging Assay." *BMC*
936 *Bioinformatics* 14 (October): 308. doi:10.1186/1471-2105-14-308.
937 Reginensi, Antoine, Rizaldy P. Scott, Alex Gregorieff, Mazdak Bagherie-Lachidan, Chaeuk Chung, Dae-Sik
938 Lim, Tony Pawson, Jeff Wrana, and Helen McNeill. 2013. "Yap- and Cdc42-Dependent Nephrogenesis
939 and Morphogenesis during Mouse Kidney Development." *PLoS Genetics* 9 (3): e1003380.
940 doi:10.1371/journal.pgen.1003380.
941 Rolli-Derkinderen, Malvyne, Vincent Sauzeau, Laurent Boyer, Emmanuel Lemichez, Céline Baron, Daniel
942 Henrion, Gervaise Loirand, and Pierre Pacaud. 2005. "Phosphorylation of Serine 188 Protects RhoA
943 from Ubiquitin/proteasome-Mediated Degradation in Vascular Smooth Muscle Cells." *Circulation*
944 *Research* 96 (11): 1152–60. doi:10.1161/01.RES.0000170084.88780.ea.
945 Šamaj, Jozef, František Baluška, and Heribert Hirt. 2004. "From Signal to Cell Polarity: Mitogen□activated
946 Protein Kinases as Sensors and Effectors of Cytoskeleton Dynamicity." *Journal of Experimental Botany*

947 55 (395): 189–98. doi:10.1093/jxb/erh012.
948 Schmich, Fabian, Ewa Szczurek, Saskia Kreibich, Sabrina Dilling, Daniel Andritschke, Alain Casanova,
949 Shyan Huey Low, et al. 2015. “gespeR: A Statistical Model for Deconvoluting off-Target-Confounded
950 RNA Interference Screens.” *Genome Biology* 16 (October): 220. doi:10.1186/s13059-015-0783-1.
951 Shehu, Amarda, Daniel Barbará, and Kevin Molloy. 2016. “A Survey of Computational Methods for Protein
952 Function Prediction.” In *Big Data Analytics in Genomics*, edited by Ka-Chun Wong, 225–98. Springer
953 International Publishing. doi:10.1007/978-3-319-41279-5_7.
954 Shin, Ilchung, Seonhoe Kim, Hyun Song, Hyeong-Reh Choi Kim, and Aree Moon. 2005. “H-Ras-Specific
955 Activation of Rac-MKK3/6-p38 Pathway: Its Critical Role in Invasion and Migration of Breast Epithelial
956 Cells.” *The Journal of Biological Chemistry* 280 (15): 14675–83. doi:10.1074/jbc.M411625200.
957 Singh, Shantanu, Xiaoyun Wu, Vebjorn Ljosa, Mark-Anthony Bray, Federica Piccioni, David E. Root, John G.
958 Doench, Jesse S. Boehm, and Anne E. Carpenter. 2015. “Morphological Profiles of RNAi-Induced Gene
959 Knockdown Are Highly Reproducible but Dominated by Seed Effects.” *PloS One* 10 (7): e0131370.
960 doi:10.1371/journal.pone.0131370.
961 Stark, Chris, Bobby-Joe Breitkreutz, Teresa Reguly, Lorrie Boucher, Ashton Breitkreutz, and Mike Tyers.
962 2006. “BioGRID: A General Repository for Interaction Datasets.” *Nucleic Acids Research* 34 (Database
963 issue): D535–39. doi:10.1093/nar/gkj109.
964 Stengel, Kristy R., and Yi Zheng. 2012. “Essential Role of Cdc42 in Ras-Induced Transformation Revealed
965 by Gene Targeting.” *PloS One* 7 (6): e37317. doi:10.1371/journal.pone.0037317.
966 Subramanian, Aravind, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A.
967 Gillette, Amanda Paulovich, et al. 2005. “Gene Set Enrichment Analysis: A Knowledge-Based Approach
968 for Interpreting Genome-Wide Expression Profiles.” *Proceedings of the National Academy of Sciences*
969 *of the United States of America* 102 (43): 15545–50. doi:10.1073/pnas.0506580102.
970 Tada, K., T. Okazaki, S. Sakon, T. Kobarai, K. Kurosawa, S. Yamaoka, H. Hashimoto, et al. 2001. “Critical
971 Roles of TRAF2 and TRAF5 in Tumor Necrosis Factor-Induced NF-Kappa B Activation and Protection
972 from Cell Death.” *The Journal of Biological Chemistry* 276 (39): 36530–34.
973 doi:10.1074/jbc.M104837200.
974 Tornatore, Laura, Anil K. Thotakura, Jason Bennett, Marta Moretti, and Guido Franzoso. 2012. “The Nuclear
975 Factor Kappa B Signaling Pathway: Integrating Metabolism with Inflammation.” *Trends in Cell Biology* 22
976 (11): 557–66. doi:10.1016/j.tcb.2012.08.001.
977 Varelas, Xaralabos. 2014. “The Hippo Pathway Effectors TAZ and YAP in Development, Homeostasis and
978 Disease.” *Development* 141 (8): 1614–26. doi:10.1242/dev.102376.
979 Veitia, Reiner A. 2007. “Exploring the Molecular Etiology of Dominant-Negative Mutations.” *The Plant Cell* 19
980 (12): 3843–51. doi:10.1105/tpc.107.055053.
981 Wawer, Mathias J., Kejie Li, Sigrun M. Gustafsdottir, Vebjorn Ljosa, Nicole E. Bodycombe, Melissa A.
982 Marton, Katherine L. Sokolnicki, et al. 2014. “Toward Performance-Diverse Small-Molecule Libraries for
983 Cell-Based Phenotypic Screening Using Multiplexed High-Dimensional Profiling.” *Proceedings of the*
984 *National Academy of Sciences of the United States of America* 111 (30): 10911–16.
985 doi:10.1073/pnas.1410933111.
986 Wiza, Claudia, Alexandra Chadt, Marcel Blumensatt, Timo Kanzleiter, Daniella Herzfeld De Wiza, Angelika
987 Horrighs, Heidi Mueller, et al. 2014. “Over-Expression of PRAS40 Enhances Insulin Sensitivity in
988 Skeletal Muscle.” *Archives of Physiology and Biochemistry* 120 (2): 64–72.
989 doi:10.3109/13813455.2014.894076.
990 Wu, Xue, Jeremy Simpson, Jenny H. Hong, Kyoung-Han Kim, Nirusha K. Thavarajah, Peter H. Backx,
991 Benjamin G. Neel, and Toshiyuki Araki. 2011. “MEK-ERK Pathway Modulation Ameliorates Disease
992 Phenotypes in a Mouse Model of Noonan Syndrome Associated with the Raf1L613V Mutation.”
993 doi:10.1172/JCI44929.
994 Yang, Xiaoping, Jesse S. Boehm, Xinping Yang, Kourosh Salehi-Ashtiani, Tong Hao, Yun Shen, Rakela
995 Lubonja, et al. 2011. “A Public Genome-Scale Lentiviral Expression Library of Human ORFs.” *Nature*
996 *Methods* 8 (8): 659–61. doi:10.1038/nmeth.1638.
997 Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. “clusterProfiler: An R Package for
998 Comparing Biological Themes among Gene Clusters.” *Omics: A Journal of Integrative Biology* 16 (5):
999 284–87. doi:10.1089/omi.2011.0118.
1000 Zhang, S., J. Han, M. A. Sells, J. Chernoff, U. G. Knaus, R. J. Ulevitch, and G. M. Bokoch. 1995. “Rho
1001 Family GTPases Regulate p38 Mitogen-Activated Protein Kinase through the Downstream Mediator
1002 Pak1.” *The Journal of Biological Chemistry* 270 (41): 23934–36.
1003 https://www.ncbi.nlm.nih.gov/pubmed/7592586.
1004 Zhang, Z-G, C. A. Lambert, S. Servotte, G. Chometon, B. Eckes, T. Krieg, C. M. Lapière, B. V. Nusgens, and

M. Aumailley. 2006. “Effects of Constitutively Active GTPases on Fibroblast Behavior.” *Cellular and Molecular Life Sciences: CMLS* 63 (1): 82–91. doi:10.1007/s00018-005-5416-5.

Zhao, Bin, Xin Ye, Jindan Yu, Li Li, Weiquan Li, Siming Li, Jianjun Yu, et al. 2008. “TEAD Mediates YAP-Dependent Gene Induction and Growth Control.” *Genes & Development* 22 (14): 1962–71. doi:10.1101/gad.1664408.

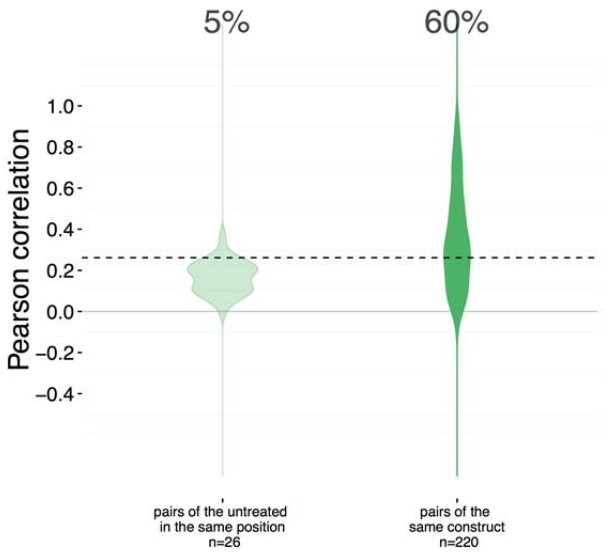


Figure 2-figure supplement 1: Position artifacts do not contribute to the hit rate seen in the experiment. We were concerned that position artifacts may result in overestimating the replicate correlations because replicates of the same treatment are assigned to the same well location across different plates in the experiment (that is, it was infeasible to scramble well locations). We ruled out this possibility by taking an alternative pessimistic null distribution which takes well position into account. In contrast to Figure 2A, which shows a 51% hit rate, a more pessimistic alternative null distribution is shown here (left), calculated based on the replicate correlation of pairs of negative controls in the same position only. We consider this less reliable because the number of such pairs is small (26) and we excluded edge wells; nevertheless the hit rate increases slightly, to 60%.

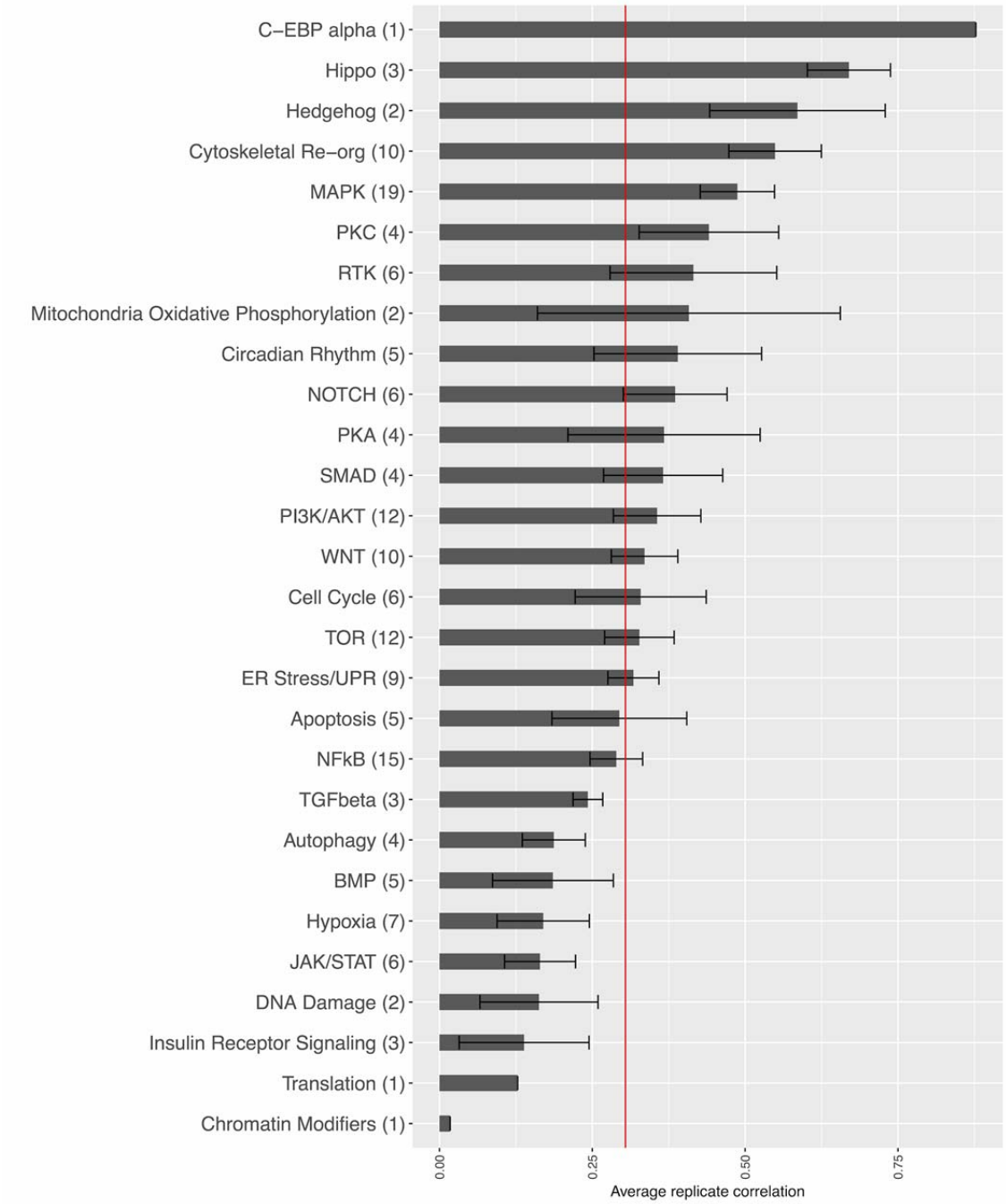


Figure 2-figure supplement 2: Strength of morphological phenotypes, according to annotated pathway. Morphological phenotype strength is calculated as the average replicate correlation for genes that experts manually annotated genes as belonging to each pathway. The number of genes tested in each category is shown in parentheses after the pathway name (two wild-type clones of the same gene are only counted once, but a mutant allele is counted separately). The red line shows the threshold beyond which an individual gene's profile would be considered to yield a distinguishable phenotype. Error bars indicate the deviation in the replicate correlation among the genes associated with the pathway.

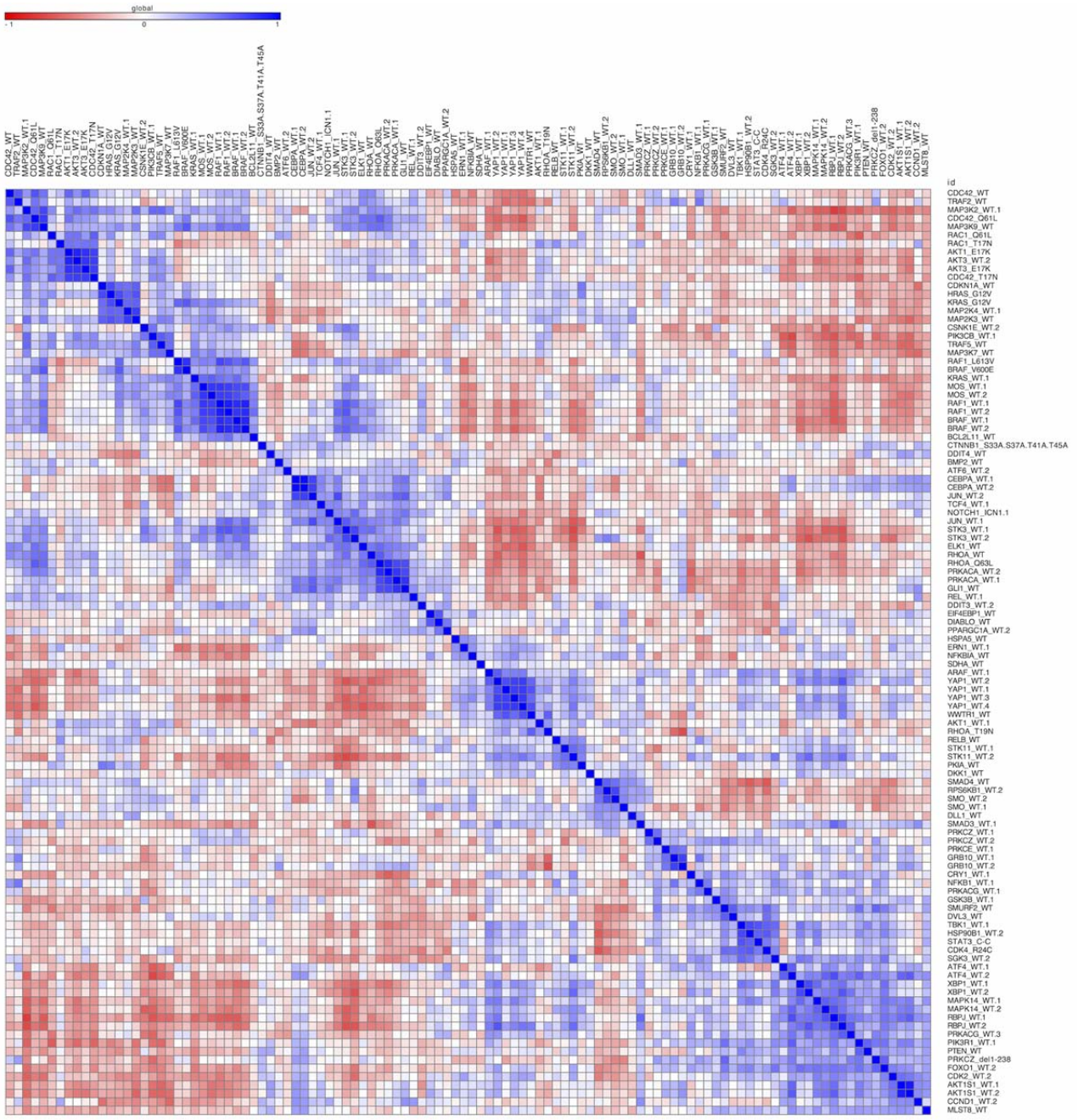


Figure 3-figure supplement 1: Correlation among the 110 genes/alleles with a detectable morphological phenotype. The rows and columns are ordered based on a hierarchical clustering algorithm such that each blue submatrix on the diagonal shows a cluster of genes resulting in similar phenotypes. The scale bar depicts Pearson correlation.

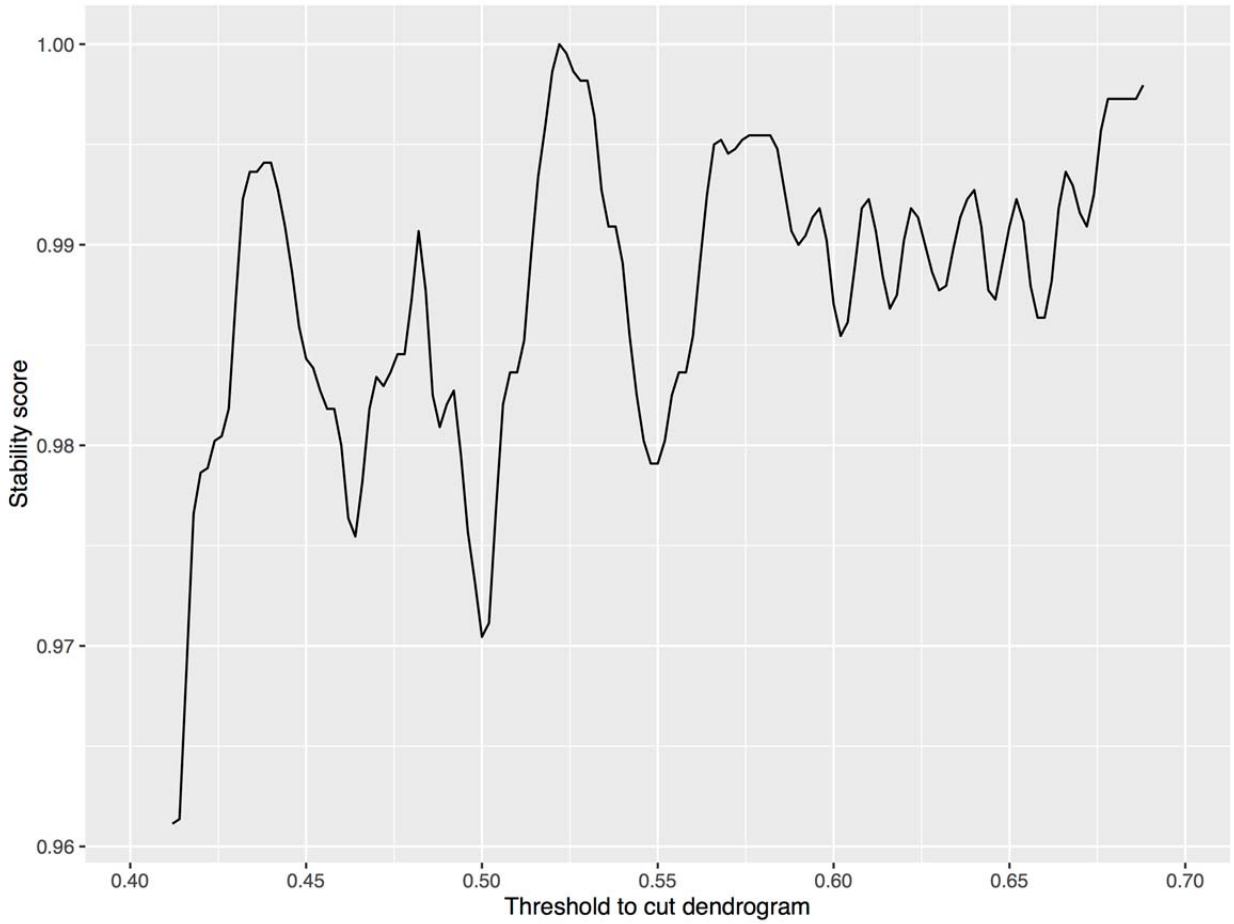


Figure 3-figure supplement 2: Smoothed stability score across different cutoffs, in order to choose a threshold for cutting the dendrogram to form clusters. The maximum occurs at threshold = 0.522. Smoothing is done by taking the moving average of order 0.02. The stability score is defined as the proportion of treatments whose clusters are not affected if the cutoff is increased or decreased by a small amount ($\epsilon = .002$).

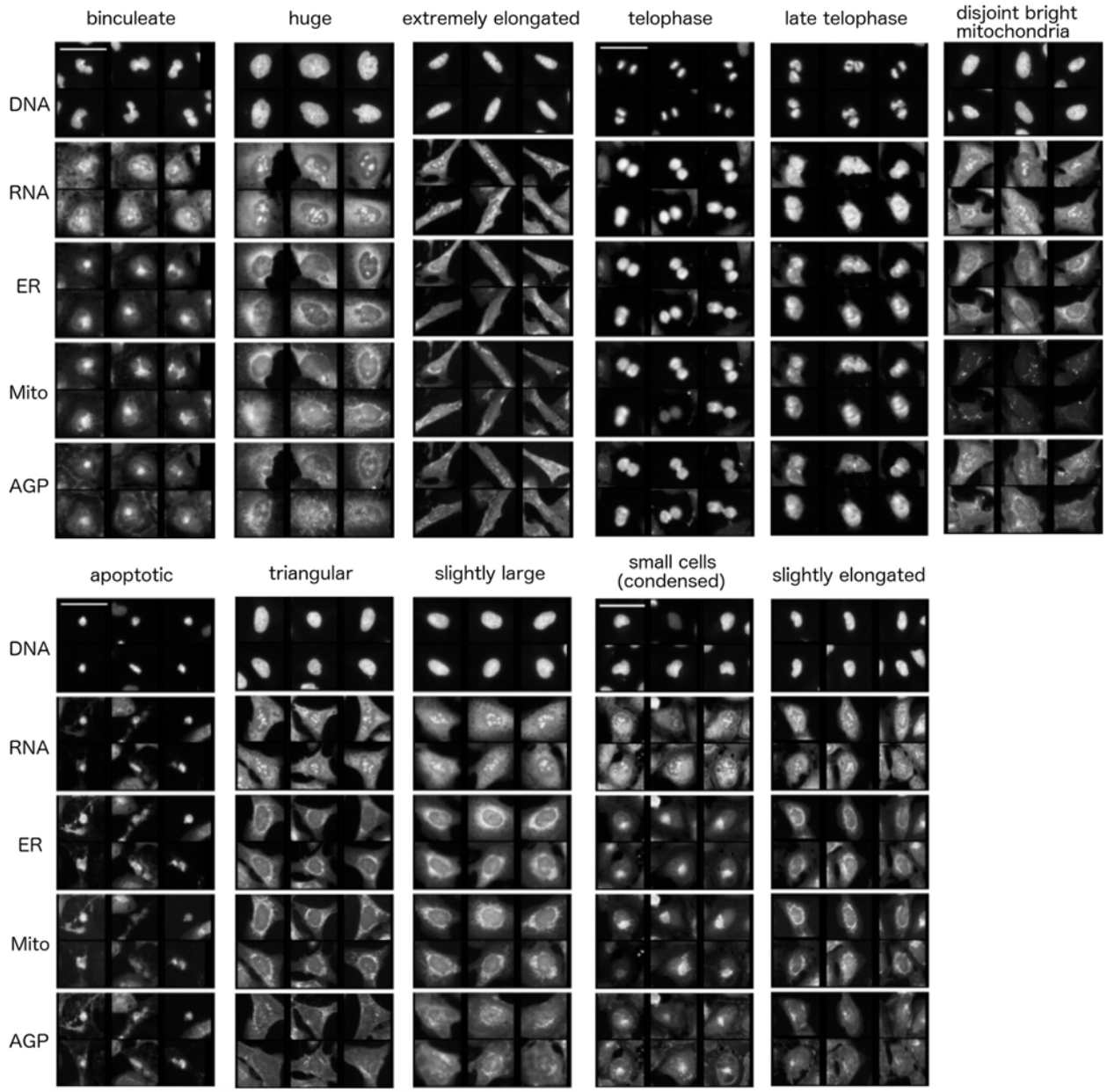


Figure 3-figure supplement 3: Common cell subpopulations seen across more than one cluster. These names are used to annotate clusters of genes in Fig. 3. Example images shown are taken from individual clusters. Scale bar is 63 μm and image intensities are log normalized. References to size and shape in the subpopulation legends refer to both the nucleus and cell borders, unless otherwise noted.

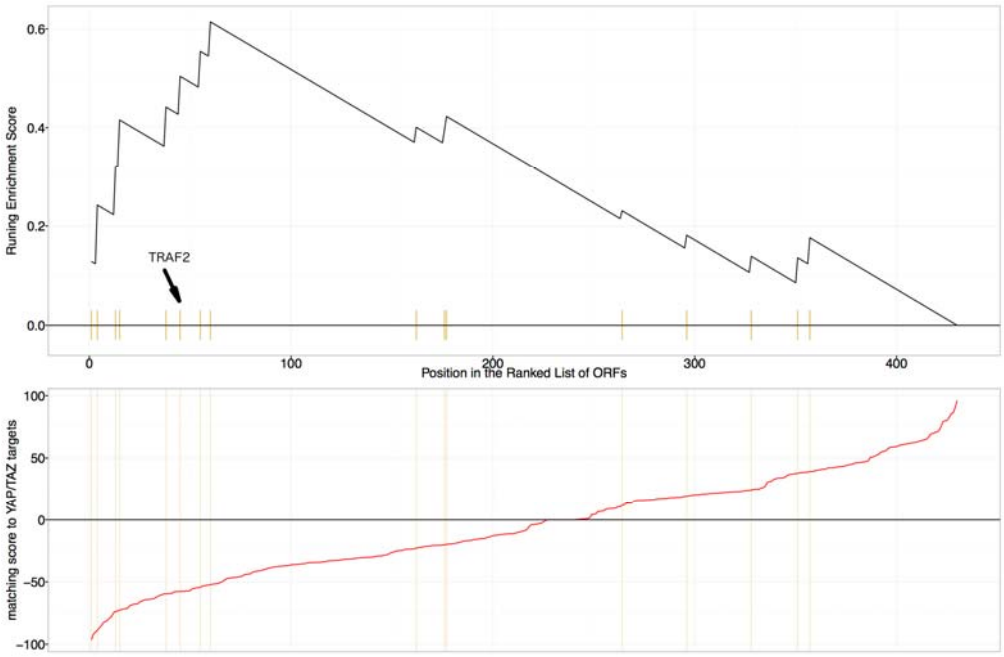


Figure 6-figure supplement 1: Gene Set Enrichment Analysis (GSEA) reveals that overexpression constructs sorted based on their similarity to YAP1/WWTR1 overexpression (in terms of impact on particular mRNA targets), are enriched for regulators of the NF-κB pathway (Enrichment Score p-value = 0.0019). mRNA targets common to both YAP1 and WWTR overexpression include INPP4B, MAP7, LAMA3, STMN1, and TRAM2, which are positively regulated, and SPP1, IER3, RAB31, and GPR56, which are negatively regulated.

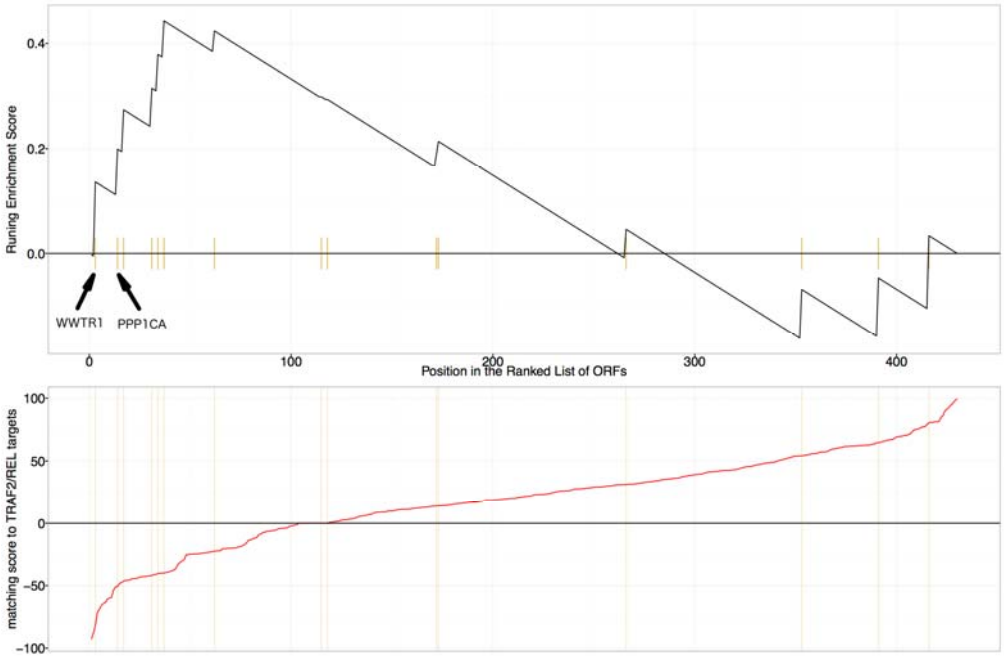


Figure 6-figure supplement 2: Gene Set Enrichment Analysis (GSEA) reveals that overexpression constructs sorted based on their similarity to TRAF2/REL overexpression (in terms of impact on particular mRNA targets), are weakly enriched for regulators of the Hippo pathway (Enrichment Score p-value = 0.024). mRNA targets common to both TRAF2 and REL overexpression include NFKBIA, IKBKE, AKAP8, and BIRC2, which are positively regulated and RPA3 which is negatively regulated. As compared to Fig. 6C and Fig. 6-figure supplement 1, this is a weaker/lower-confidence enrichment - note the lower maximum height (~ 0.44 compared to > 0.6) and higher p-value (0.024 compared to < 0.002). Still, we note that WWTR1 and PPP1CA are the top two matches among those annotated as related to the Hippo pathway in KEGG; PPP1CA (also known as PP-1A) activates TAZ (C.-Y. Liu et al. 2011).

Supplementary Files

Supplementary File 1

We have released the data tables related to graphs and analyses presented in the paper in the ZIP format. This file includes several PDFs listed below:

- 1A: List of all the 323 constructs used in the experiment along with the target transcript and their public clone ID.
- 1B: Replicate correlation is higher in the constitutively active mutant allele compared to the wild-type allele, except for AKT3_E17K. Constitutively active mutant annotations were obtained by literature search for all the mutants in the experiment showing a detectable phenotype. Genes shown here are only those where either the wild-type gene or its constitutively activating allele yielded a phenotype distinct from controls.
- 1C: Pathways sorted based on proportion of their associated gene showing a detectable phenotype.
- 1D: Highly correlated proteins (according to morphology in the Cell Painting assay) that have also been reported to interact physically.
- 1E: Highly correlated genes (according to morphology in the Cell Painting assay) that have also been annotated to be related to the same pathway.
- 1F: Gene Ontology terms associated with each gene cluster.
- 1G: Rank ordered list of distinctive features based on their z-scores for Cluster 19.
- 1H: All genes/alleles in Cluster 8 and 10 induce cell rounding.
- 1I: The NF-□B signaling pathway is the most enriched when searching for gene overexpressions that downregulate known YAP/TAZ targets (CYR61, CTGF, and BIRC5).

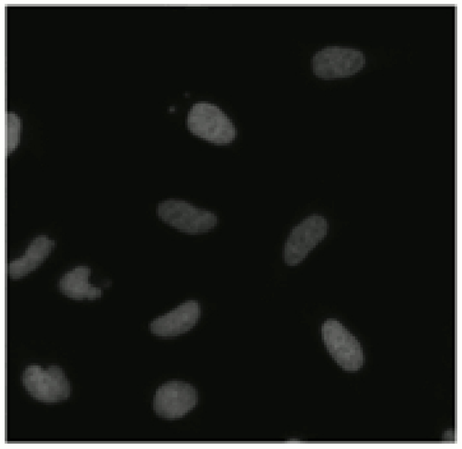
Supplementary File 2

Type A and B PDFs are collected in a ZIP file in supplementary file 2. The details of the contents have been described in Fig. 5.

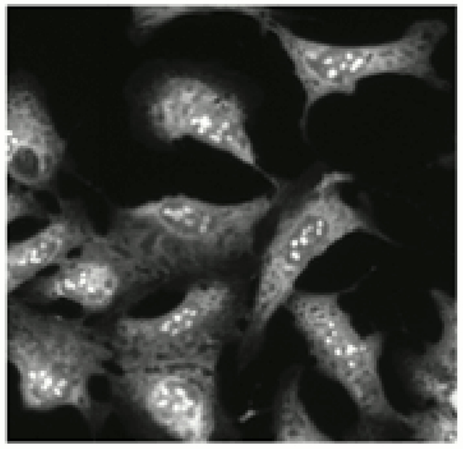
Supplementary File 3

The CellProfiler pipeline used to process the images is released as the supplementary file 3.

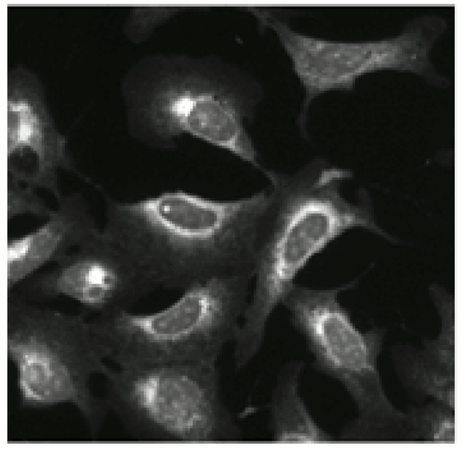
A



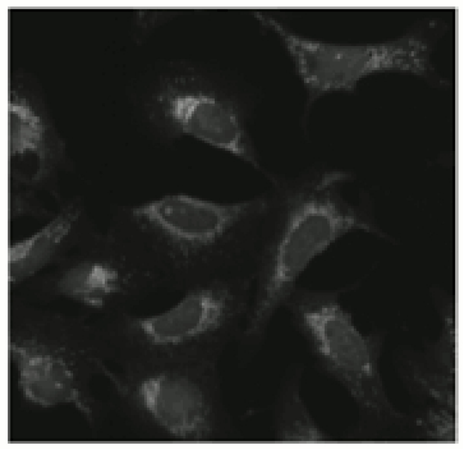
Nucleus (DNA)



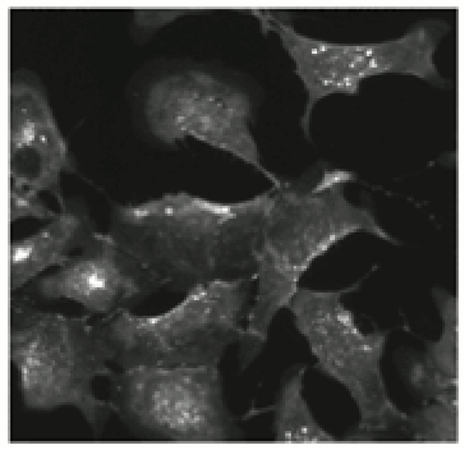
Cytoplasmic RNA/
Nucleoli (RNA)



Endoplasmic
reticulum (ER)

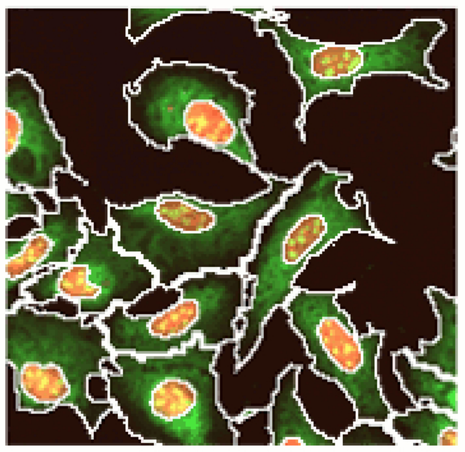


Mitochondria (Mito)



Actin/Golgi/
Plasma membrane
(AGP)

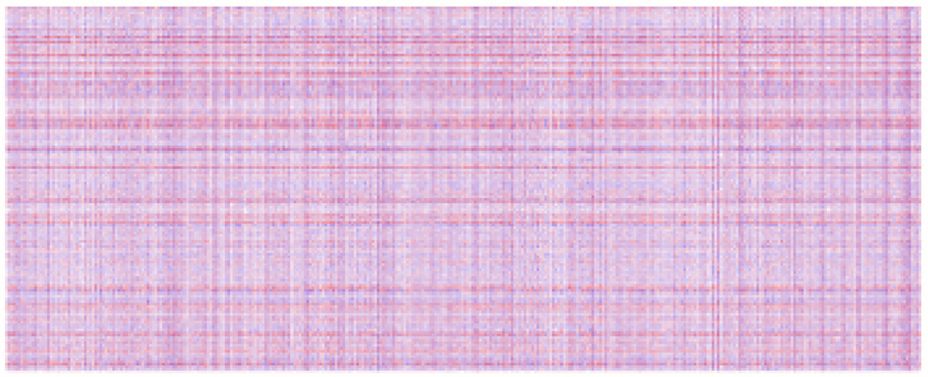
B



Cells and Nuclei
Segmentation

Samples

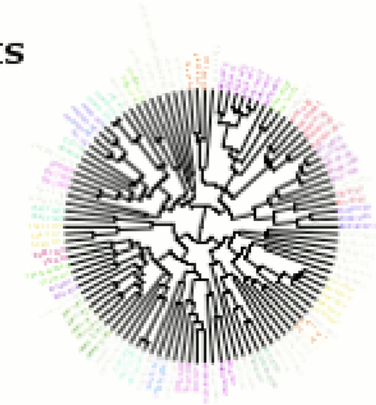
Measurements



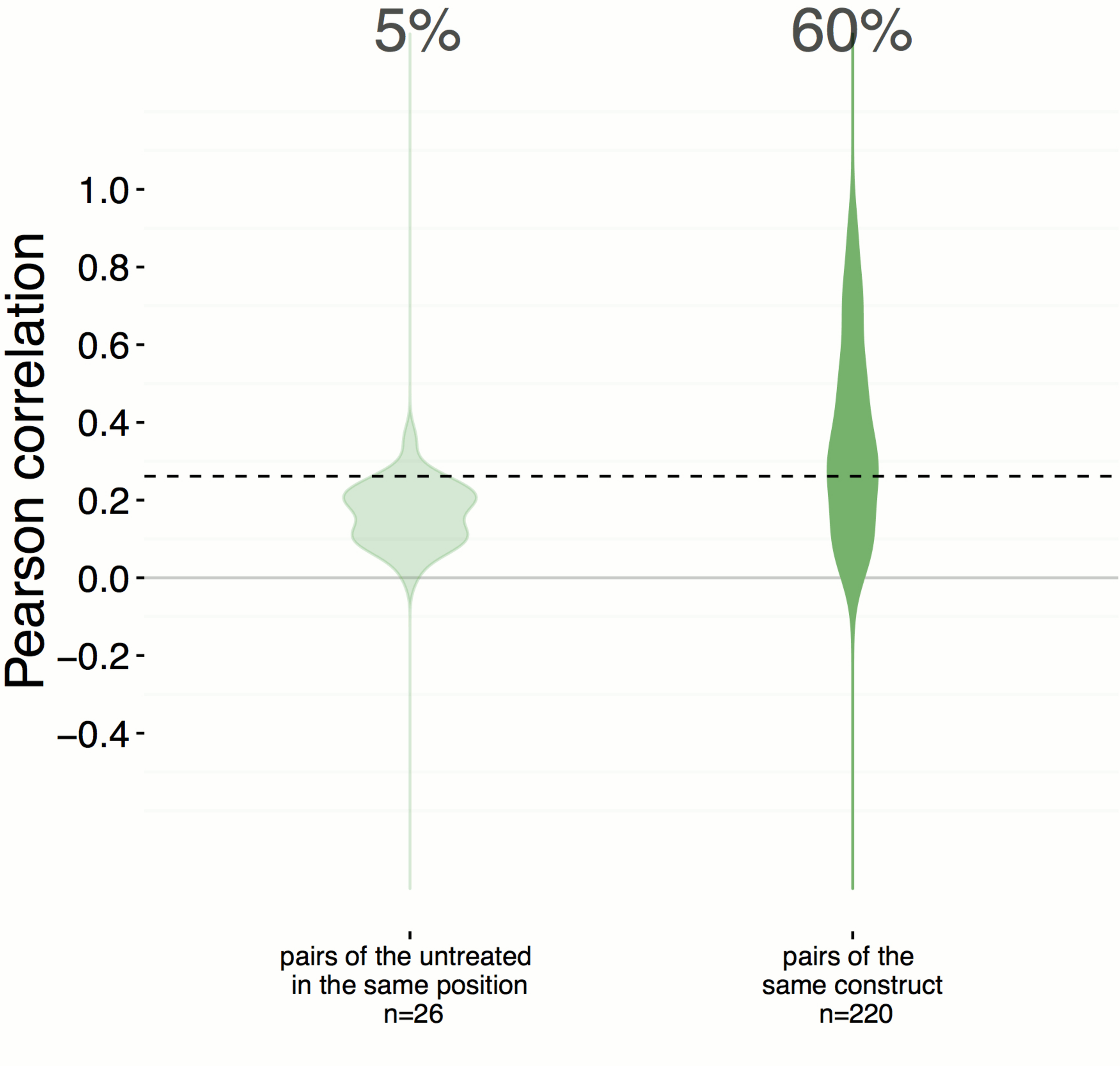
Raw profiles
(2,769 dimensional)

Principal Components

Processed profiles
(158 dimensional)



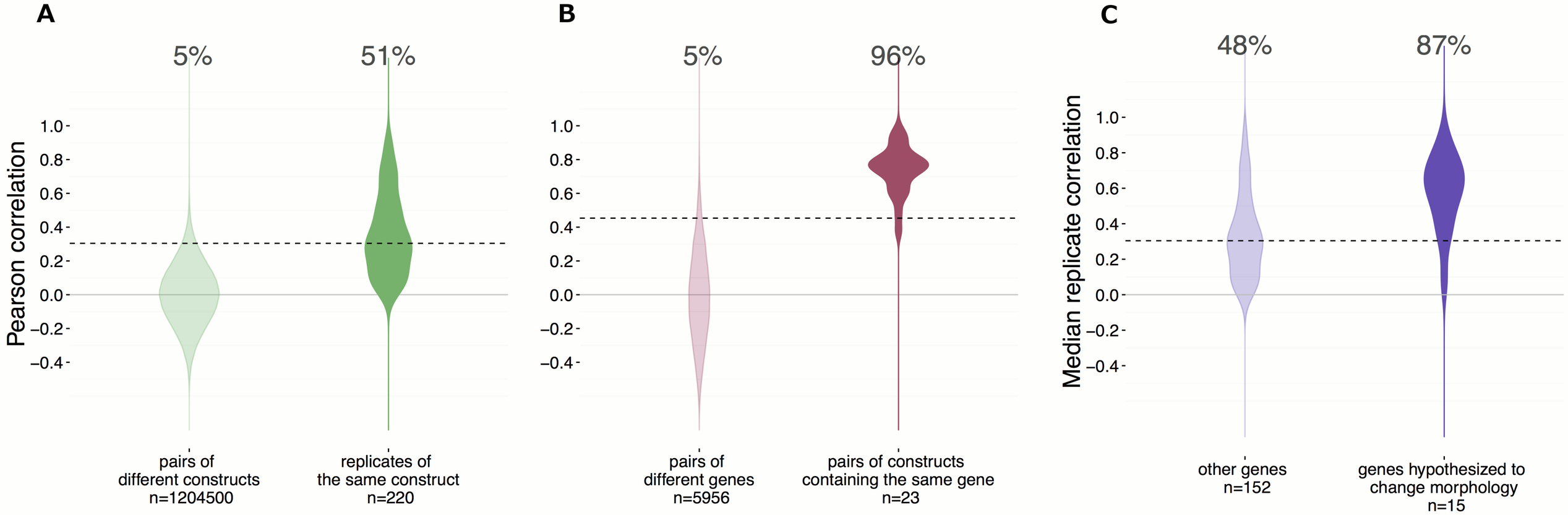
Clustering
samples

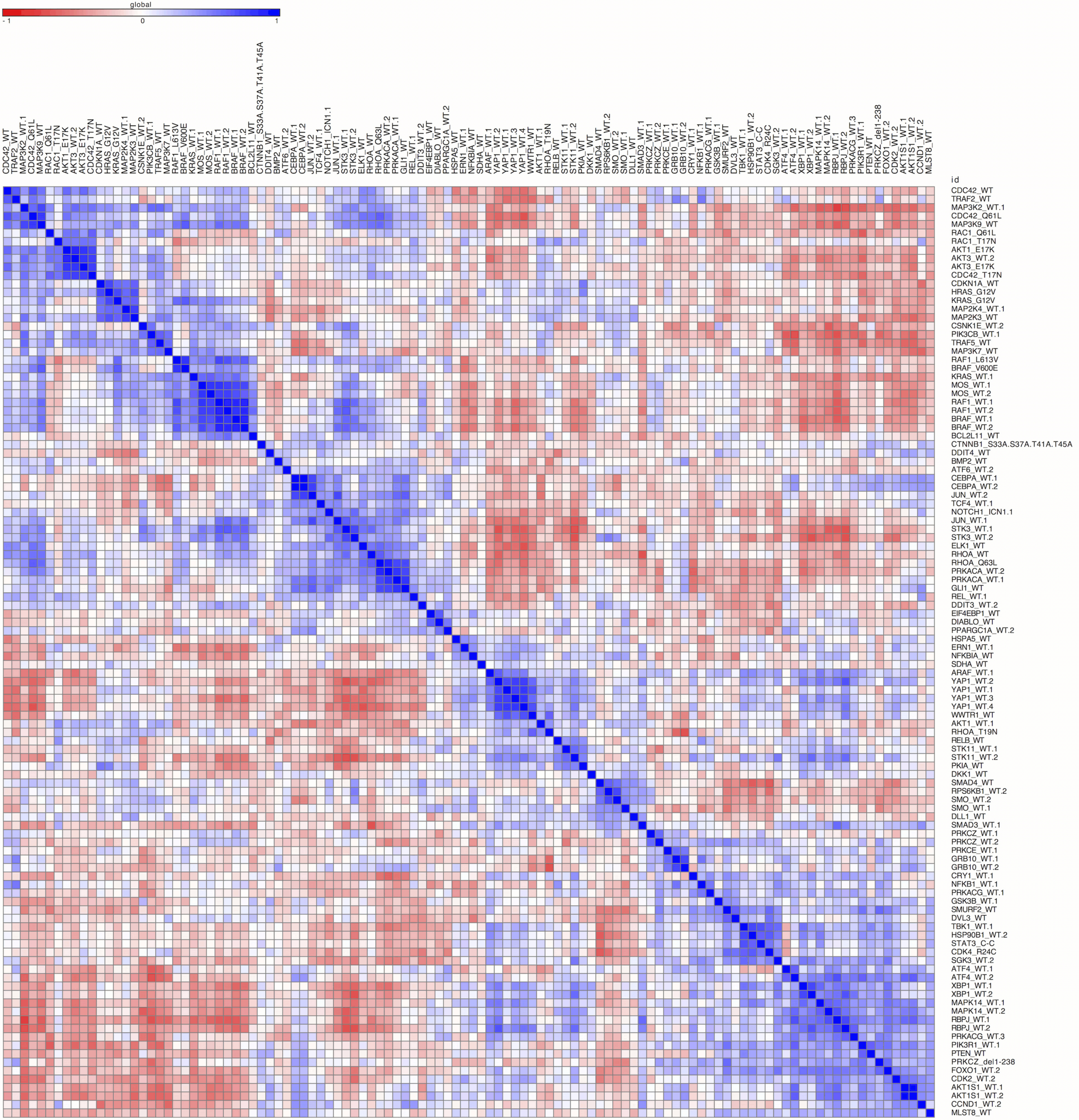

5%
60%
Pearson correlation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
pairs of the untreated
in the same position
n=26
pairs of the
same construct
n=220

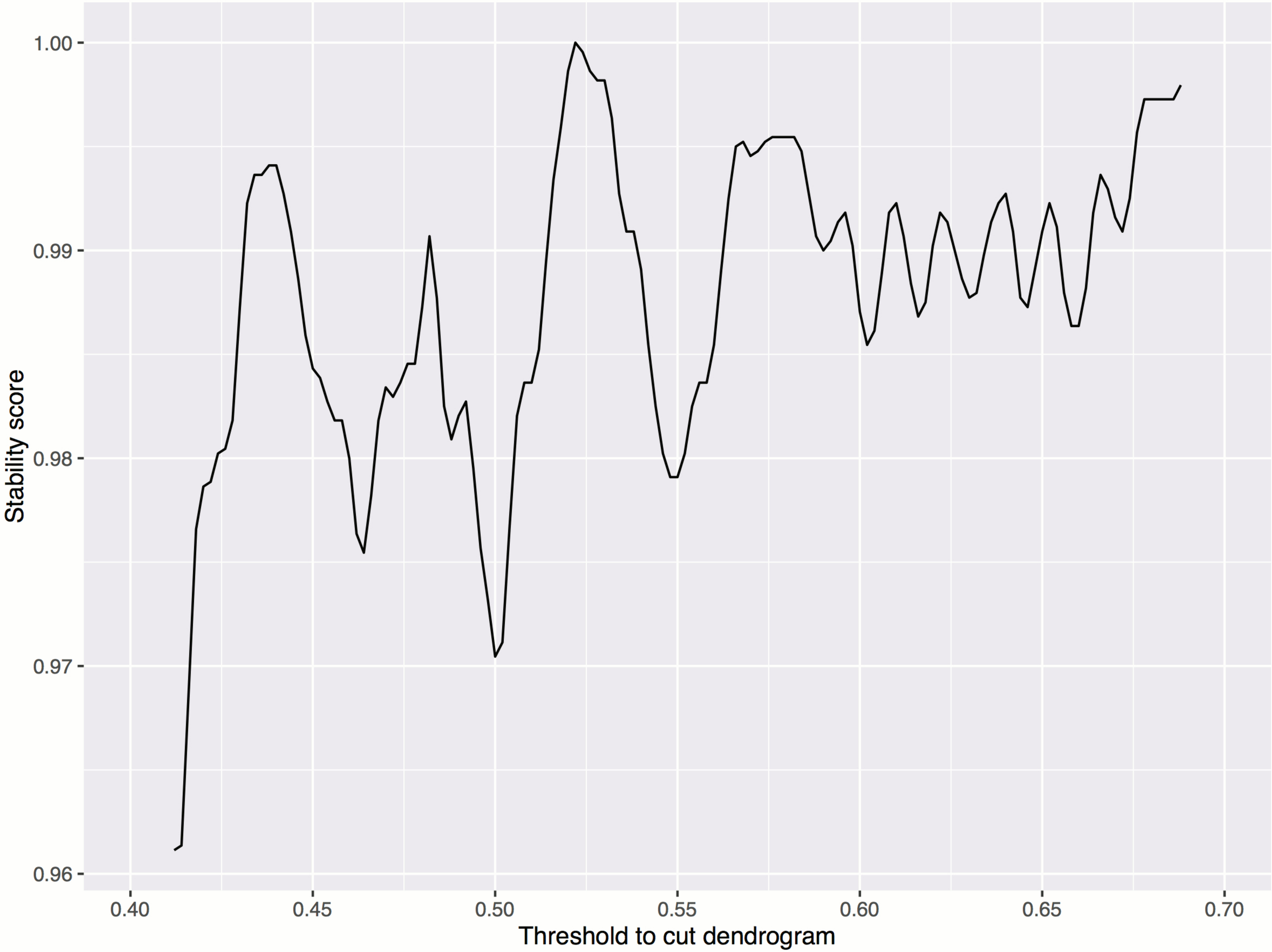
C–EBP alpha (1)
Hippo (3)
Hedgehog (2)
Cytoskeletal Re–org (10)
MAPK (19)
PKC (4)
RTK (6)
Mitochondria Oxidative Phosphorylation (2)
Circadian Rhythm (5)
NOTCH (6)
PKA (4)
SMAD (4)
PI3K/AKT (12)
WNT (10)
Cell Cycle (6)
TOR (12)
ER Stress/UPR (9)
Apoptosis (5)
NFkB (15)
TGFbeta (3)
Autophagy (4)
BMP (5)
Hypoxia (7)
JAK/STAT (6)
DNA Damage (2)
Insulin Receptor Signaling (3)
Translation (1)
Chromatin Modifiers (1)

0.00
0.25
0.50
0.75

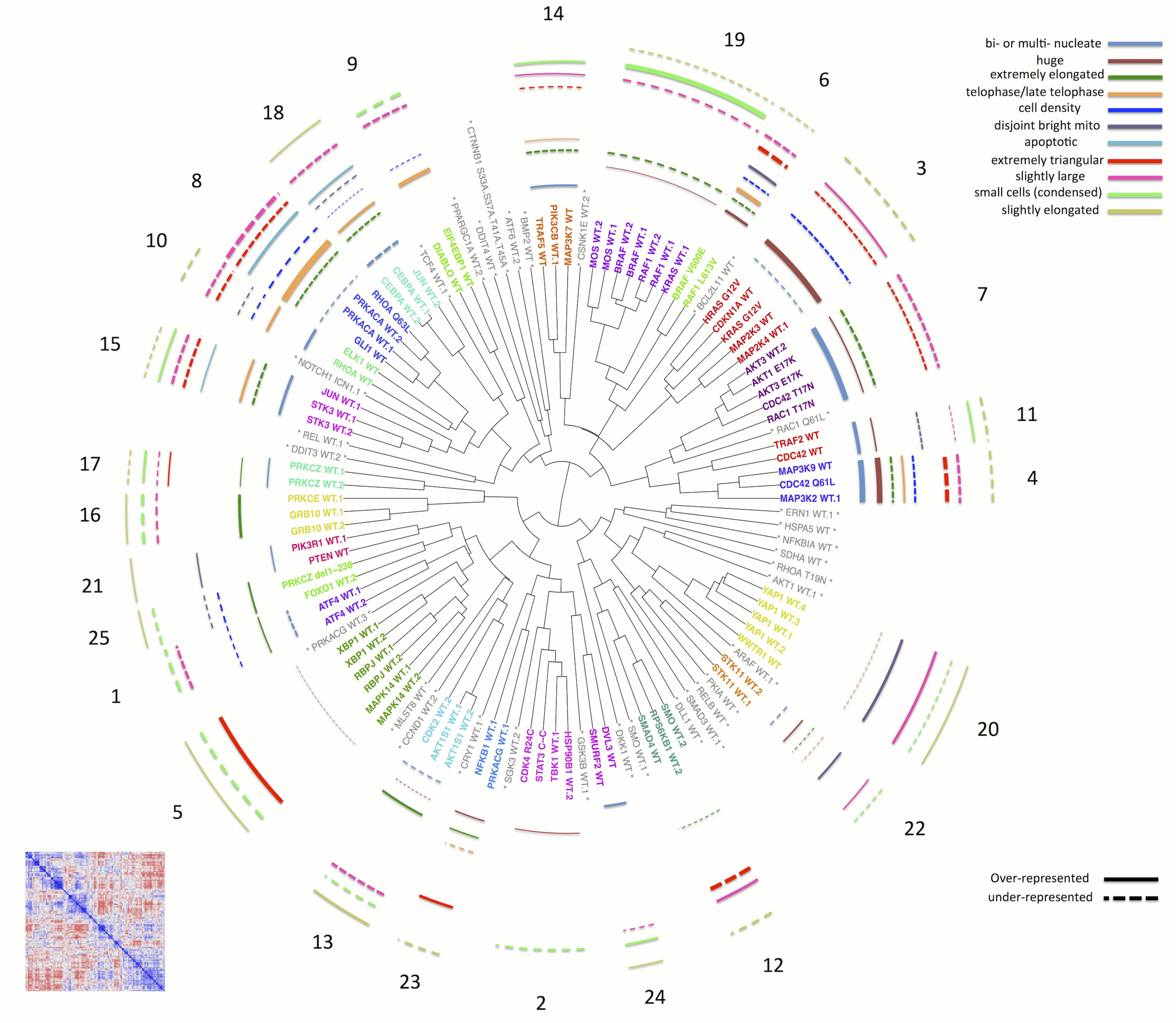
Average replicate correlation


A
5%
51%
Pearson correlation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
pairs of
different constructs
n=1204500
replicates of
the same construct
n=220
B
5%
96%
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
pairs of
different genes
n=5956
pairs of constructs
containing the same gene
n=23
C
48%
87%
Median replicate correlation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
other genes
n=152
genes hypothesized to
change morphology
n=15

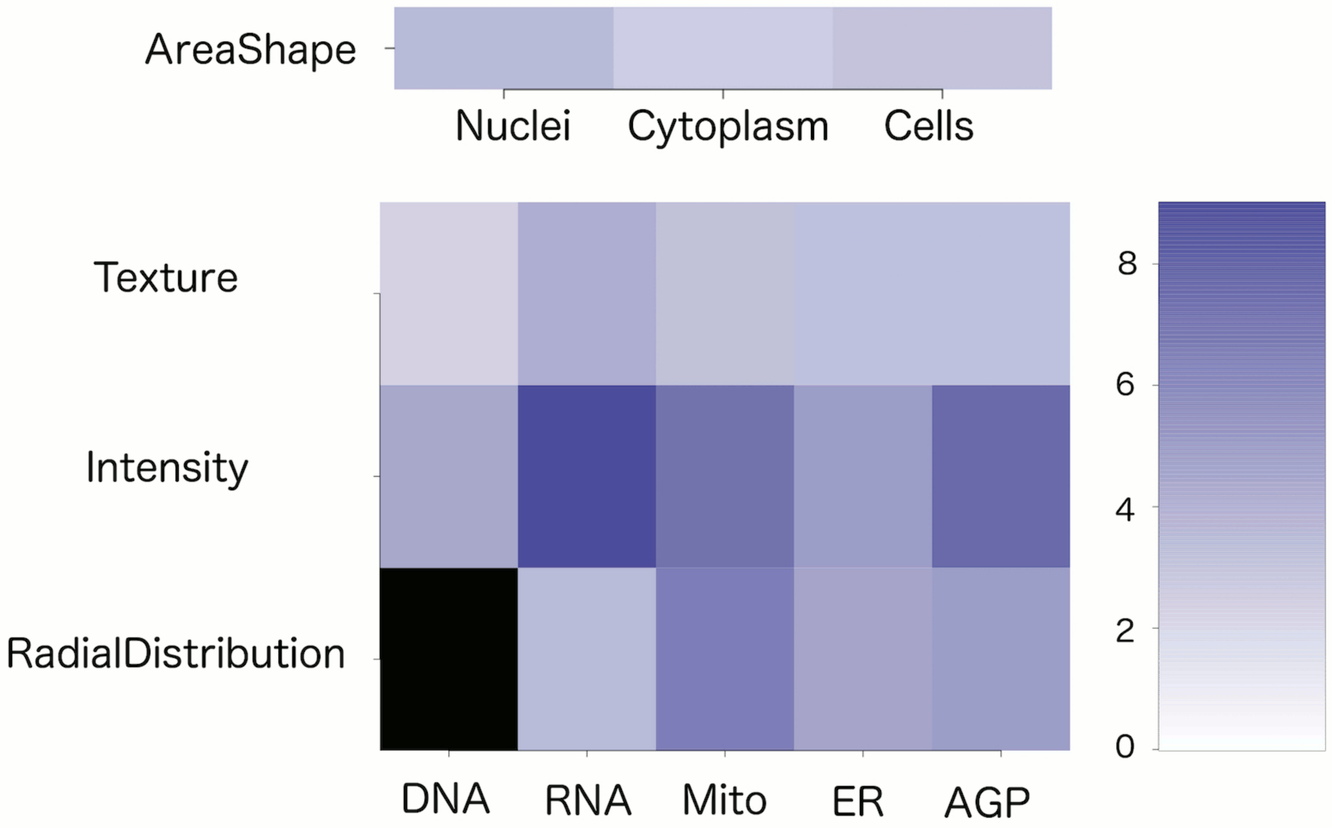

global
- 1
0
1
CDC42_WT
TRAF2_WT
MAP3K2_WT.1
CDC42_Q61L
MAP3K9_WT
RAC1_Q61L
RAC1_T17N
AKT1_E17K
AKT3_WT.2
AKT3_E17K
CDC42_T17N
CDKN1A_WT
HRAS_G12V
KRAS_G12V
MAP2K4_WT.1
MAP2K3_WT
CSNK1E_WT.2
PIK3CB_WT.1
TRAF5_WT
MAP3K7_WT
RAF1_L613V
BRAF_V600E
KRAS_WT.1
MOS_WT.1
MOS_WT.2
RAF1_WT.1
RAF1_WT.2
BRAF_WT.1
BRAF_WT.2
BCL2L11_WT
CTNNB1_S33A.S37A.T41A.T45A
DDIT4_WT
BMP2_WT
ATF6_WT.2
CEBPA_WT.1
CEBPA_WT.2
JUN_WT.2
TCF4_WT.1
NOTCH1_ICN1.1
JUN_WT.1
STK3_WT.1
STK3_WT.2
ELK1_WT
RHOA_WT
RHOA_Q63L
PRKACA_WT.2
PRKACA_WT.1
GLI1_WT
REL_WT.1
DDIT3_WT.2
EIF4EBP1_WT
DIABLO_WT
PPARGC1A_WT.2
HSPA5_WT
ERN1_WT.1
NFKBIA_WT
SDHA_WT
ARAF_WT.1
YAP1_WT.2
YAP1_WT.1
YAP1_WT.3
YAP1_WT.4
WWTR1_WT
AKT1_WT.1
RHOA_T19N
RELB_WT
STK11_WT.1
STK11_WT.2
PKIA_WT
DKK1_WT
SMAD4_WT
RPS6KB1_WT.2
SMO_WT.2
SMO_WT.1
DLL1_WT
SMAD3_WT.1
PRKCZ_WT.1
PRKCZ_WT.2
PRKCE_WT.1
GRB10_WT.1
GRB10_WT.2
CRY1_WT.1
NFKB1_WT.1
PRKACG_WT.1
GSK3B_WT.1
SMURF2_WT
DVL3_WT
TBK1_WT.1
HSP90B1_WT.2
STAT3_C-C
CDK4_R24C
SGK3_WT.2
ATF4_WT.1
ATF4_WT.2
XBP1_WT.1
XBP1_WT.2
MAPK14_WT.1
MAPK14_WT.2
RBPJ_WT.1
RBPJ_WT.2
PRKACG_WT.3
PIK3R1_WT.1
PTEN_WT
PRKCZ_del1-238
FOXO1_WT.2
CDK2_WT.2
AKT1S1_WT.1
AKT1S1_WT.2
CCND1_WT.2
MLST8_WT
id
CDC42_WT
TRAF2_WT
MAP3K2_WT.1
CDC42_Q61L
MAP3K9_WT
RAC1_Q61L
RAC1_T17N
AKT1_E17K
AKT3_WT.2
AKT3_E17K
CDC42_T17N
CDKN1A_WT
HRAS_G12V
KRAS_G12V
MAP2K4_WT.1
MAP2K3_WT
CSNK1E_WT.2
PIK3CB_WT.1
TRAF5_WT
MAP3K7_WT
RAF1_L613V
BRAF_V600E
KRAS_WT.1
MOS_WT.1
MOS_WT.2
RAF1_WT.1
RAF1_WT.2
BRAF_WT.1
BRAF_WT.2
BCL2L11_WT
CTNNB1_S33A.S37A.T41A.T45A
DDIT4_WT
BMP2_WT
ATF6_WT.2
CEBPA_WT.1
CEBPA_WT.2
JUN_WT.2
TCF4_WT.1
NOTCH1_ICN1.1
JUN_WT.1
STK3_WT.1
STK3_WT.2
ELK1_WT
RHOA_WT
RHOA_Q63L
PRKACA_WT.2
PRKACA_WT.1
GLI1_WT
REL_WT.1
DDIT3_WT.2
EIF4EBP1_WT
DIABLO_WT
PPARGC1A_WT.2
HSPA5_WT
ERN1_WT.1
NFKBIA_WT
SDHA_WT
ARAF_WT.1
YAP1_WT.2
YAP1_WT.1
YAP1_WT.3
YAP1_WT.4
WWTR1_WT
AKT1_WT.1
RHOA_T19N
RELB_WT
STK11_WT.1
STK11_WT.2
PKIA_WT
DKK1_WT
SMAD4_WT
RPS6KB1_WT.2
SMO_WT.2
SMO_WT.1
DLL1_WT
SMAD3_WT.1
PRKCZ_WT.1
PRKCZ_WT.2
PRKCE_WT.1
GRB10_WT.1
GRB10_WT.2
CRY1_WT.1
NFKB1_WT.1
PRKACG_WT.1
GSK3B_WT.1
SMURF2_WT
DVL3_WT
TBK1_WT.1
HSP90B1_WT.2
STAT3_C-C
CDK4_R24C
SGK3_WT.2
ATF4_WT.1
ATF4_WT.2
XBP1_WT.1
XBP1_WT.2
MAPK14_WT.1
MAPK14_WT.2
RBPJ_WT.1
RBPJ_WT.2
PRKACG_WT.3
PIK3R1_WT.1
PTEN_WT
PRKCZ_del1-238
FOXO1_WT.2
CDK2_WT.2
AKT1S1_WT.1
AKT1S1_WT.2
CCND1_WT.2
MLST8_WT


1.00
0.99
0.98
0.97
0.96
Stability score
0.40
0.45
0.50
0.55
0.60
0.65
0.70
Threshold to cut dendrogram

binculeate
huge
extremely elongated
telophase
late telophase
disjoint bright mitochondria
DNA
RNA
ER
Mito
AGP
apoptotic
triangular
slightly large
small cells (condensed)
slightly elongated
DNA
RNA
ER
Mito
AGP


bi- or multi- nucleate
huge
extremely elongated
telophase/late telophase
cell density
disjoint bright mito
apoptotic
extremely triangular
slightly large
small cells (condensed)
slightly elongated
Over-represented
under-represented
14
19
9
6
18
3
8
10
7
15
11
17
4
16
21
25
1
20
5
22
13
12
23
24
2
* CTNNB1 S33A.S37A.T41A.T45A
* PPARGC1A WT.2
* DDIT4 WT
* ATF6 WT.2
* BMP2 WT
PIK3CB WT.1
TRAF5 WT
MAP3K7 WT
* CSNK1E WT.2
MOS WT.2
MOS WT.1
BRAF WT.2
BRAF WT.1
RAF1 WT.2
RAF1 WT.1
KRAS WT.1
BRAF V600E
RAF1 L613V
* BCL2L11 WT
HRAS G12V
CDKN1A WT
KRAS G12V
MAP2K3 WT
MAP2K4 WT.1
AKT3 WT.2
AKT1 E17K
AKT3 E17K
CDC42 T17N
RAC1 T17N
* RAC1 Q61L
TRAF2 WT
CDC42 WT
MAP3K9 WT
CDC42 Q61L
MAP3K2 WT.1
* ERN1 WT.1
* HSPA5 WT
* NFKBIA WT
* SDHA WT
* RHOA T19N
* AKT1 WT.1
YAP1 WT.4
YAP1 WT.3
YAP1 WT.1
YAP1 WT.2
WWTR1 WT
* ARAF WT.1
STK11 WT.2
STK11 WT.1
* PKIA WT
* RELB WT
* SMAD3 WT.1
SMO WT.2
RPS6KB1 WT.2
SMAD4 WT
* SMO WT.1
* DKK1 WT
DVL3 WT
SMURF2 WT
* GSK3B WT.1
HSP90B1 WT.2
TBK1 WT.1
STAT3 C–C
CDK4 R24C
* SGK3 WT.2
PRKACG WT.1
NFKB1 WT.1
* CRY1 WT.1
AKT1S1 WT.2
AKT1S1 WT.1
CDK2 WT.2
* CCND1 WT.2
* MLST8 WT
MAPK14 WT.2
MAPK14 WT.1
RBPJ WT.2
RBPJ WT.1
XBP1 WT.2
XBP1 WT.1
* PRKACG WT.3
ATF4 WT.2
ATF4 WT.1
FOXO1 WT.2
PRKCZ del1–238
PTEN WT
PIK3R1 WT.1
GRB10 WT.2
GRB10 WT.1
PRKCE WT.1
PRKCZ WT.2
PRKCZ WT.1
* DDIT3 WT.2
* REL WT.1
STK3 WT.2
STK3 WT.1
JUN WT.1
* NOTCH1 ICN1.1
RHOA WT
ELK1 WT
GLI1 WT
PRKACA WT.1
PRKACA WT.2
RHOA Q63L
CEBPA WT.2
CEBPA WT.1
JUN WT.2
* TCF4 WT.1
DIABLO WT
EIF4EBP1 WT

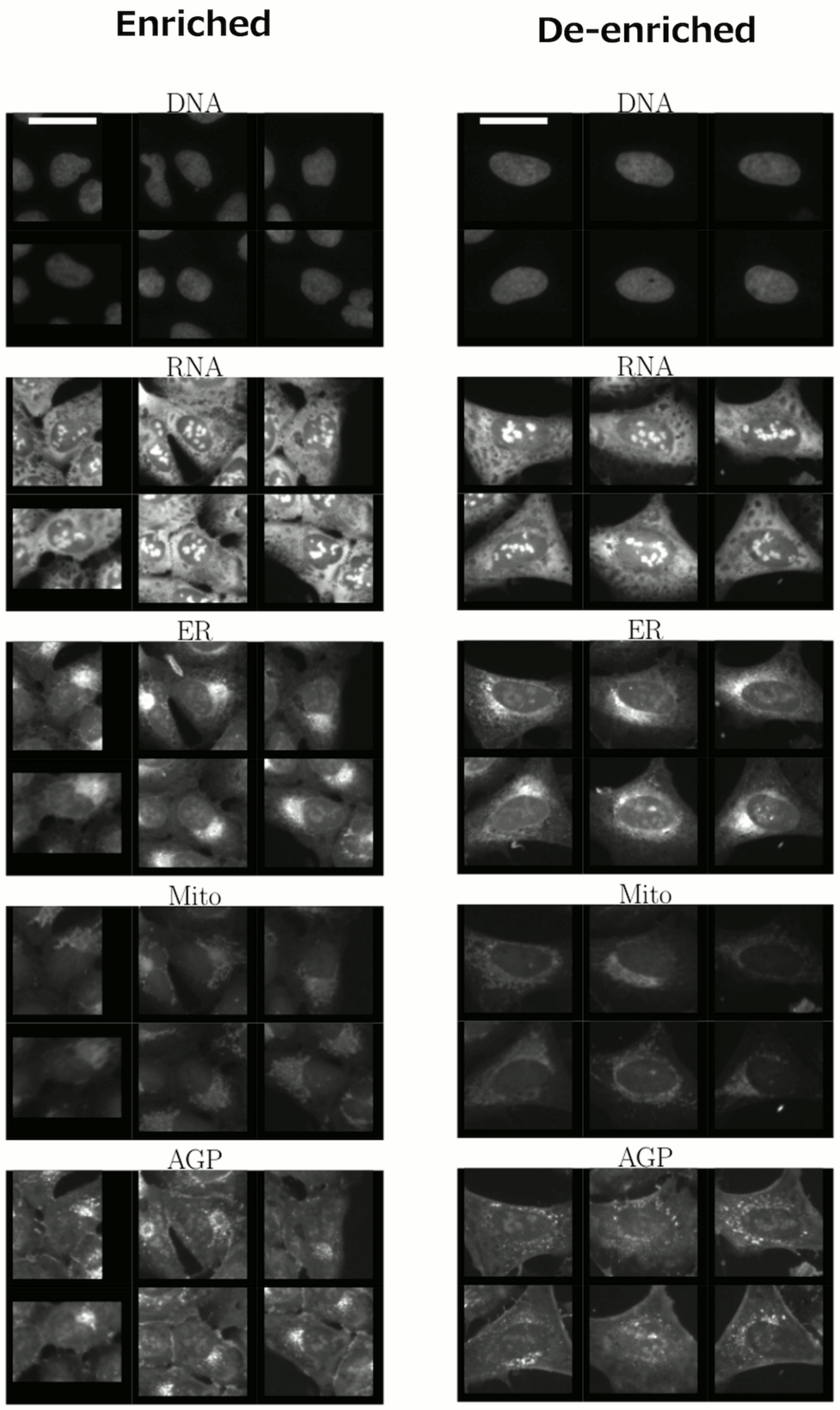
A

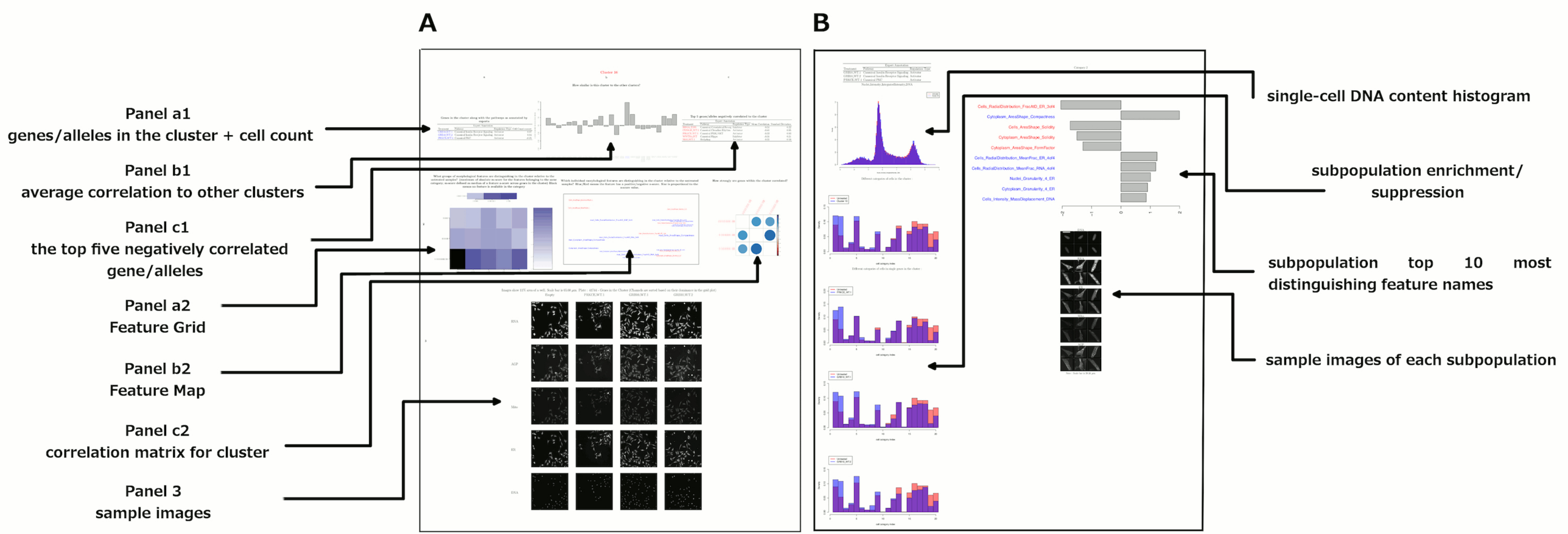

AreaShape
Nuclei
Cytoplasm
Cells
Texture
Intensity
RadialDistribution
DNA
RNA
Mito
ER
AGP
8
6
4
2
0

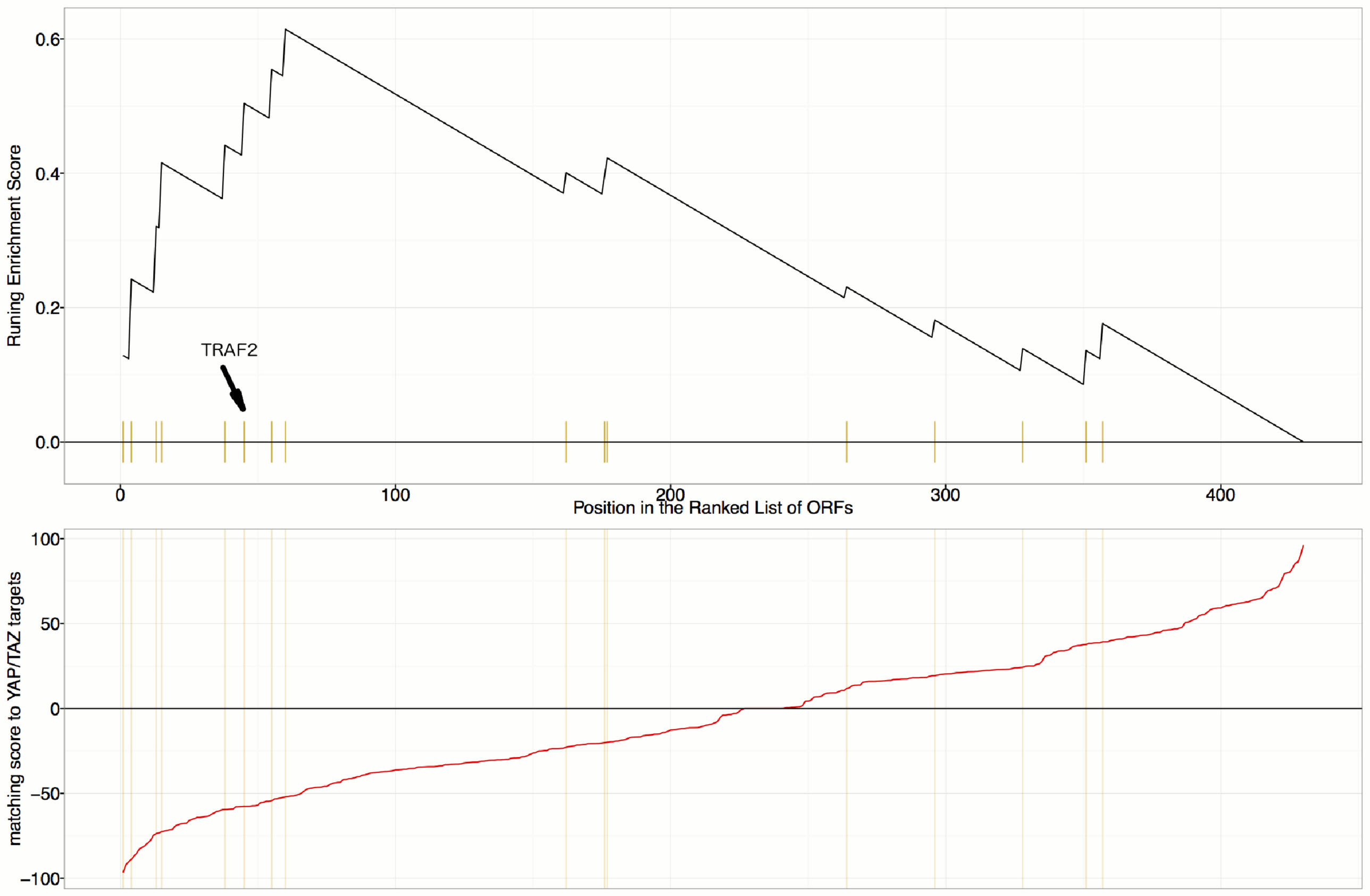
B

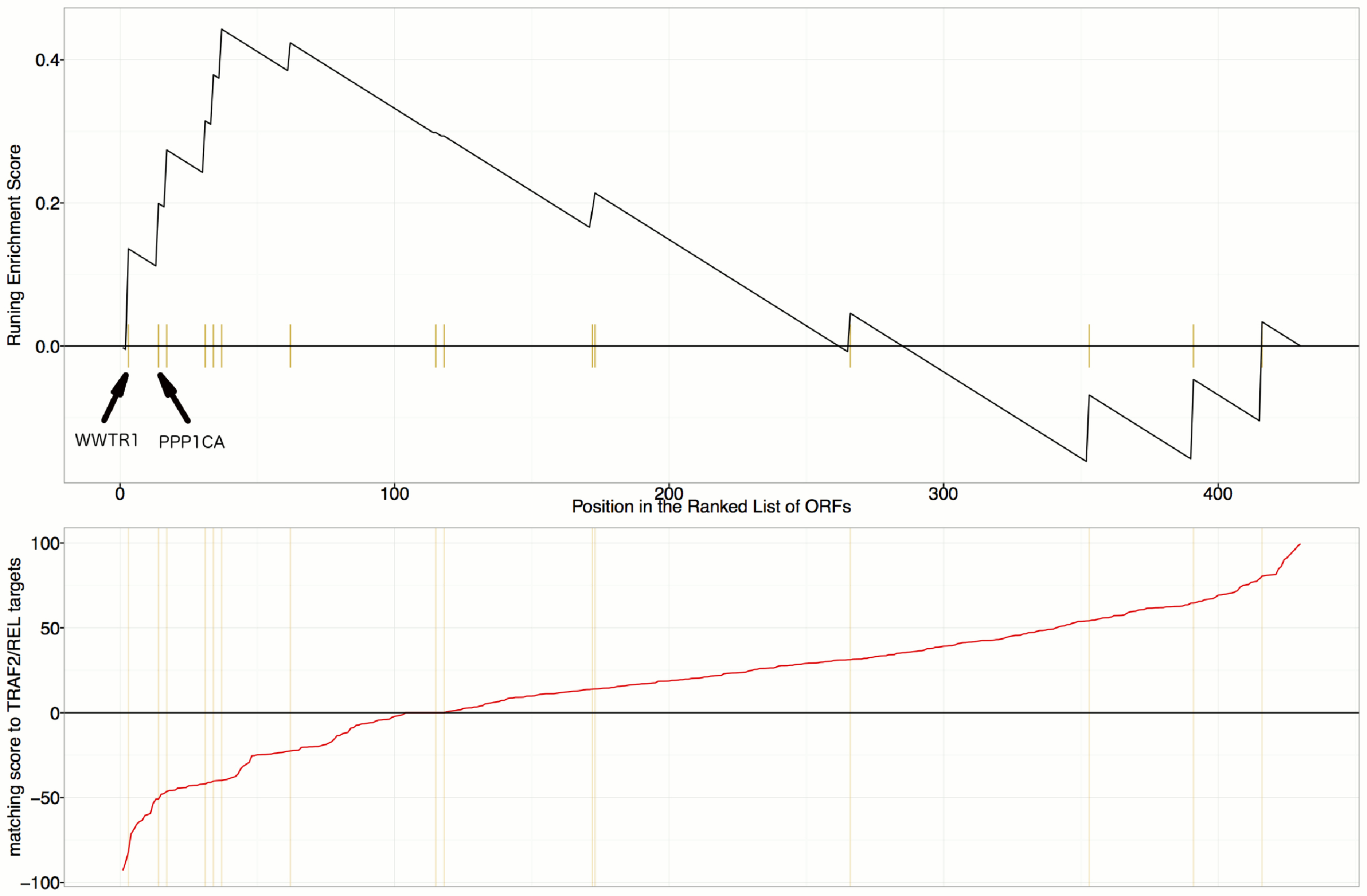

mad_Cells_Intensity_MedianIntensity_AGP
mad_Cytoplasm_Intensity_UpperQuartileIntensity_AGP
mad_Cytoplasm_Intensity_MeanIntensity_AGP
mad_Cytoplasm_Intensity_MeanIntensity_Mito
mad_Cytoplasm_Intensity_MedianIntensity_AGP
mad_Cells_Intensity_MeanIntensity_Mito
Cells_Intensity_LowerQuartileIntensity_RNA
Cytoplasm_Intensity_LowerQuartileIntensity_RNA
mad_Cells_Intensity_UpperQuartileIntensity_RNA
mad_Cells_Intensity_MeanIntensity_RNA
mad_Cells_Intensity_LowerQuartileIntensity_RNA
mad_Cells_Intensity_MedianIntensity_RNA
mad_Cytoplasm_Intensity_MeanIntensity_RNA
mad_Cytoplasm_Intensity_MeanIntensityEdge_RNA
mad_Cytoplasm_Intensity_LowerQuartileIntensity_RNA
Cells_Intensity_MeanIntensityEdge_RNA
mad_Cytoplasm_Intensity_MedianIntensity_RNA
mad_Cytoplasm_Intensity_UpperQuartileIntensity_RNA
Cells_Intensity_StdIntensityEdge_RNA
mad_Cytoplasm_RadialDistribution_MeanFrac_Mito_2of4

C

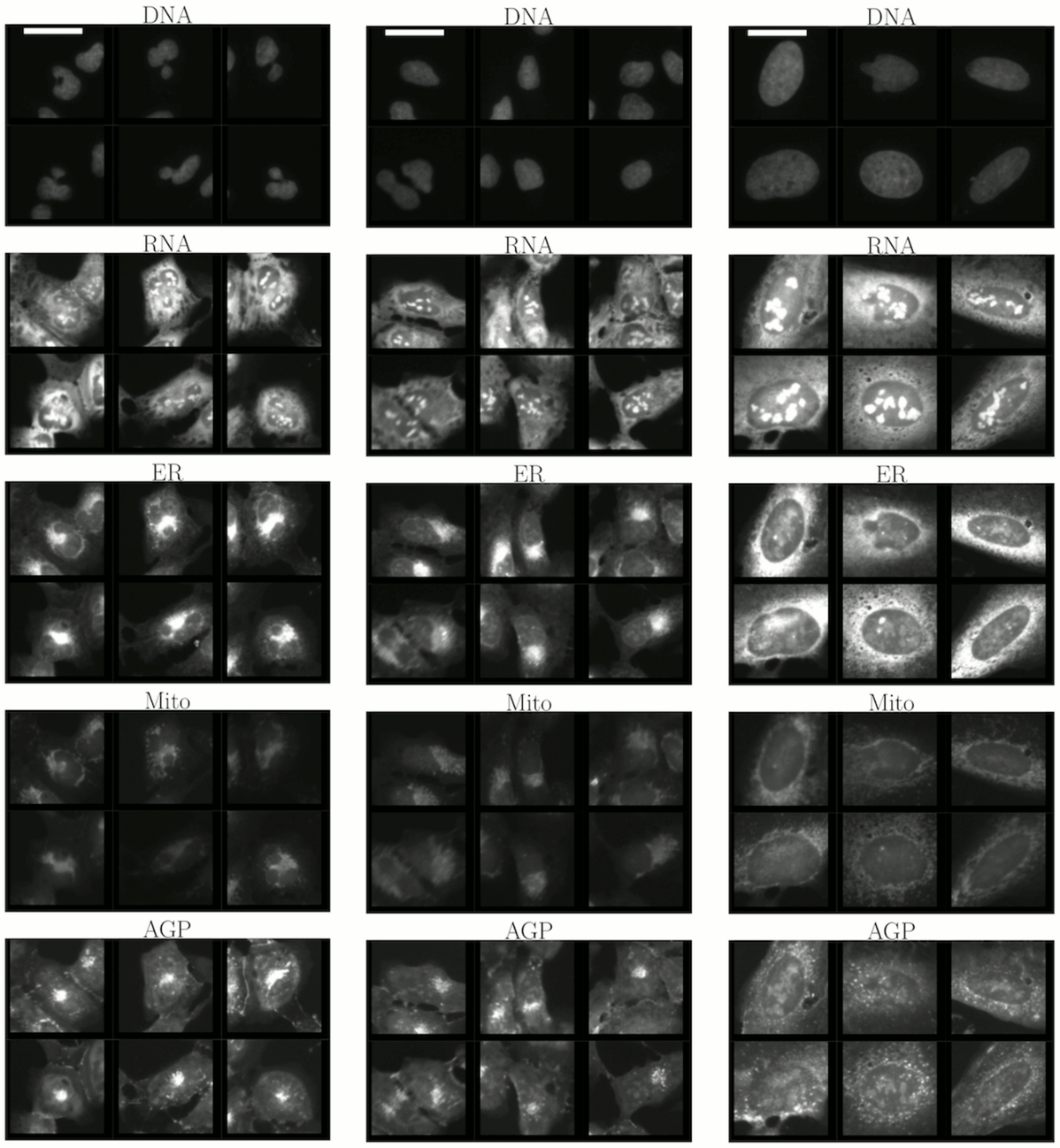

Enriched
De-enriched
DNA
RNA
ER
Mito
AGP


A
B
Panel a1
genes/alleles in the cluster + cell count
Panel b1
average correlation to other clusters
Panel c1
the top five negatively correlated gene/alleles
Panel a2
Feature Grid
Panel b2
Feature Map
Panel c2
correlation matrix for cluster
Panel 3
sample images
single-cell DNA content histogram
subpopulation enrichment/ suppression
subpopulation top 10 most distinguishing feature names
sample images of each subpopulation

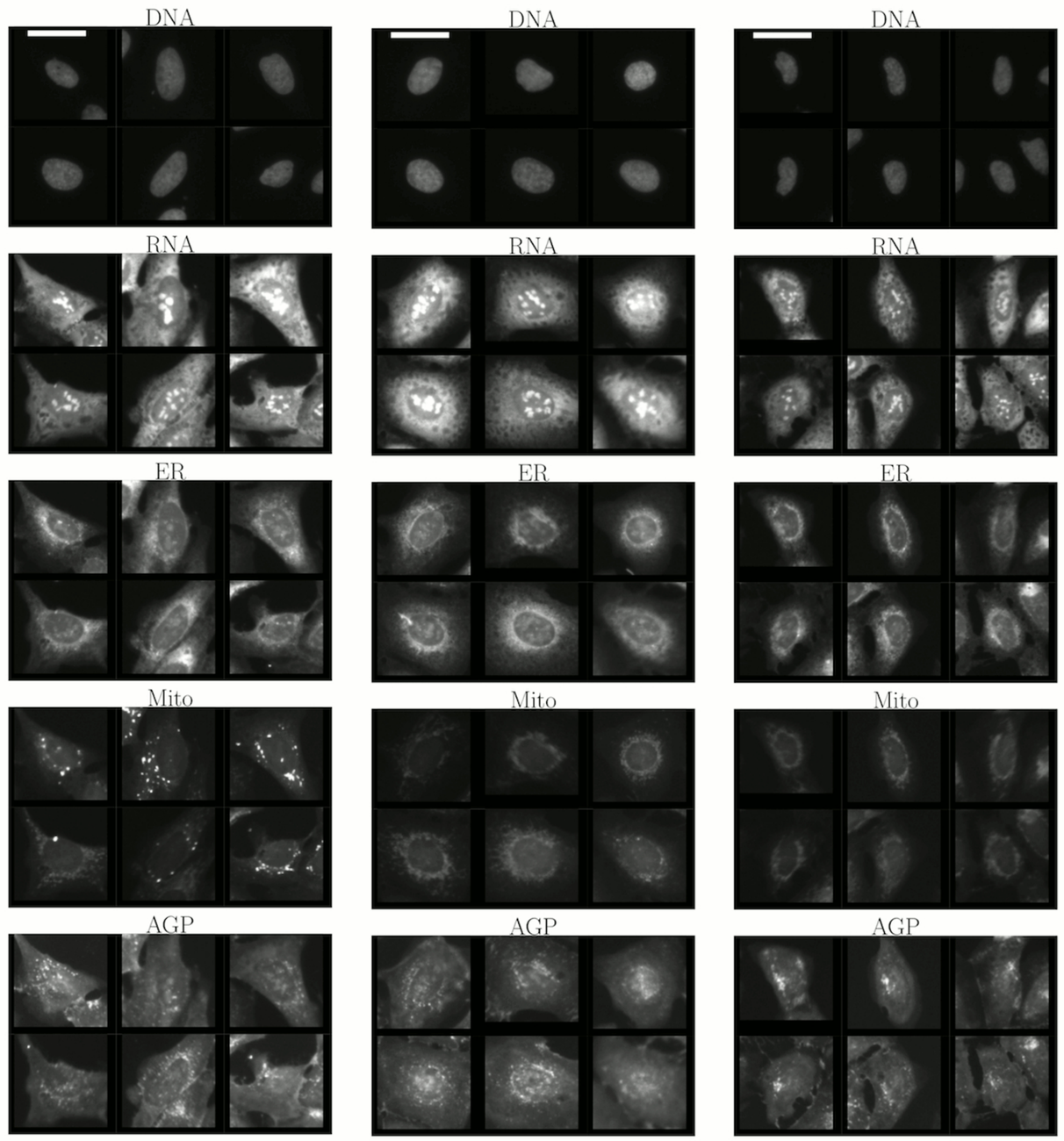

Runing Enrichment Score
0.6
0.4
0.2
0.0
TRAF2
0
100
200
300
400
Position in the Ranked List of ORFs
matching score to YAP/TAZ targets
100
50
0
−50
−100


Runing Enrichment Score
0.4
0.2
0.0
WWTR1
PPP1CA
0
100
200
300
400
Position in the Ranked List of ORFs
matching score to TRAF2/REL targets
100
50
0
−50
−100

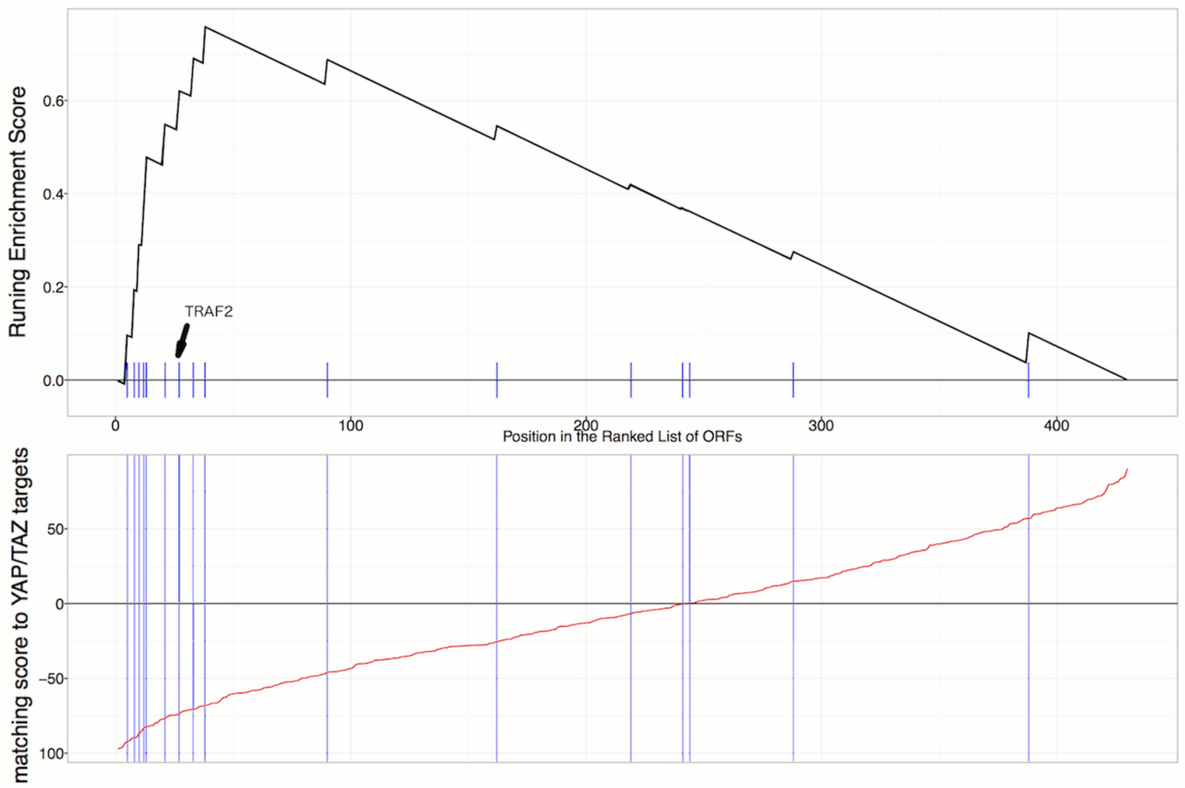
A


DNA
RNA
ER
Mito
AGP

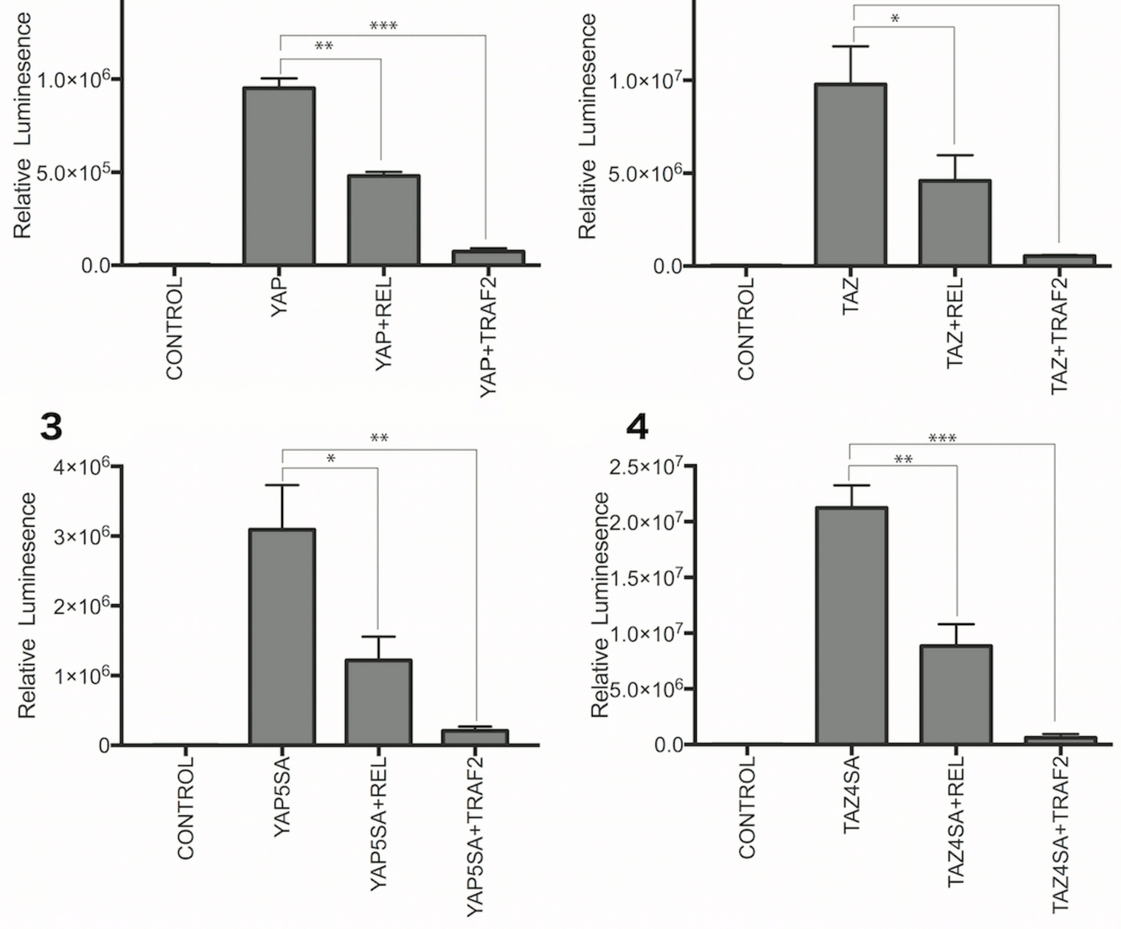
B


DNA
RNA
ER
Mito
AGP

C


Runing Enrichment Score
TRAF2
Position in the Ranked List of ORFs
matching score to YAP/TAZ targets

D


1
Relative Luminesence
CONTROL
YAP
YAP+REL
YAP+TRAF2
2
TAZ
TAZ+REL
TAZ+TRAF2
3
YAP5SA
YAP5SA+REL
YAP5SA+TRAF2
4
TAZ4SA
TAZ4SA+REL
TAZ4SA+TRAF2