Article:
Pahl, A., Schölermann, B., Lampe, P. et al. (7 more authors) (2023) Morphological subprofile analysis for bioactivity annotation of small molecules. Cell Chemical Biology, 30 (7). P839-P853.E7. ISSN 2451-9456

https://doi.org/10.1016/j.chembiol.2023.06.003

Morphological Subprofile Analysis for Bioactivity Annotation of Small Molecules

Axel Pahl[1], Beate Schölermann[1], Philipp Lampe[1], Marion Rusch[1], Mark Dow[2], Christian Hedberg[1†], Adam Nelson[2], Sonja Sievers[1], Herbert Waldmann[1,3] and Slava Ziegler[1]

[1] Max-Planck Institute of Molecular Physiology, Department of Chemical Biology, Otto-Hahn-Strasse 11, Dortmund 44227, Germany

[2] School of Chemistry and Astbury Centre for Structural Molecular Biology, University of Leeds, LS2 9JT, UK

[3] Technical University Dortmund, Faculty of Chemistry and Chemical Biology, Otto-Hahn-Strasse 6, Dortmund 44227, Germany

† Current address: Umeå University, Faculty of Science and Technology, Department of Chemistry

Correspondence: axel.pahl@mpi-dortmund.mpg.de, slava.ziegler@mpi-dortmund.mpg.de

Lead contact: slava.ziegler@mpi-dortmund.mpg.de

Summary

Fast prediction of mode of action for bioactive compounds would immensely foster bioactivity annotation in compound collections and may early on reveal off-targets in chemical biology research and drug discovery. Morphological profiling, e.g., using the Cell Painting assay, offers a fast, unbiased assessment of compound activity on various targets in one experiment. However, due to incomplete bioactivity annotation and unknown activities of reference compounds, prediction of bioactivity is not straightforward. Here we introduce the concept of subprofile analysis to map the mode of action (MoA) for both, reference and unexplored compounds. We defined MoA clusters and extracted cluster subprofiles that contain only a subset of morphological features. Subprofile analysis allows for assignment of compounds to, currently, twelve targets or MoA. This approach enables rapid bioactivity annotation of compounds and will be extended to further clusters in the future. The data is accessible via https://github.com/mpimp-comas/2022_pahl_ziegler_subprofiles and the web app https://cpcse.pythonanywhere.com/.

Keywords: Morphological profiling, Cell Painting assay, Mode-of-action prediction, Bioactivity, Clusters, Subprofiles, Small molecules

Introduction

The design, synthesis and biological investigation of compound collections is at the heart of chemical biology research and drug discovery. Target- and cell-based assays are frequently employed to detect modulators of disease-relevant targets or processes. However, these assays are usually biased towards the kind of the sought bioactivity. Profiling approaches, which monitor hundreds of parameters like gene or protein expression or morphological features, provide a broader view on biological states in cells and organisms.[1] Profiling of small molecules in cells provides an unbiased snapshot of perturbed processes and may uncover novel and off-targets. Morphological profiling, e.g., using the Cell Painting assay (CPA), monitors the change in morphological features, has higher throughput than gene and protein expression profiling and has been used to detect bioactivity in compound collections.[1] The CPA employs six different dyes for detection of cell organelles and components (DNA, RNA, mitochondria, Golgi, plasma membrane, endoplasmic reticulum, actin cytoskeleton).[2,3] Generation of CPA profiles for reference compounds is crucial and a prerequisite for the analysis of bioactivity of uncharacterized small molecules. Ideally, reference compounds sharing the same target shall give rise to similar CPA profiles and this profile biosimilarity can then be used to predict a target for a new compound. However, reference compounds often lack complete annotation, may display polypharmacology and address unknown targets which significantly hampers the generation of target hypotheses based solely on biosimilarity.[4-6]

Herein we introduce the concept of subprofile analysis to define cluster subprofiles for bioactivity clusters that are detectable using the Cell Painting assay. These cluster subprofiles consist only of features that are common to reference compounds within one bioactivity cluster. Biosimilarity to the cluster subprofiles is then sufficient to reliably predict a mode of action (MoA) related to biological targets such as AKT/PI3K/MTOR, Aurora kinases, BET, HDAC, HSP90, Na^+/K^+ ATPase and tubulin or processes like DNA synthesis,

lysosomotropism/cholesterol homeostasis regulation, protein synthesis, pyrimidine biosynthesis and uncoupling of the mitochondrial proton gradient. Non-cluster features, i.e., features from the full profile that do not belong to the cluster subprofile, can in addition be employed to map activity that is not related to the cluster features or to group compounds within one MoA cluster. The subprofile approach allows for easy and rapid annotation of bioactivity of unexplored compounds using the Cell Painting assay without prior knowledge of the top biosimilar reference compounds and can detect polypharmacology.

Results

For hypothesis generation in the biological analysis of our in-house compound collection, we explored a set of 4,256 different reference compounds from multiple suppliers (mainly Prestwick, LOPAC, Selleckchem, MedChemExpress, GSK and Structural Genomics Consortium (SGC)) using the CPA in a wide range of concentrations. Thus far, we have obtained 10,565 data points or 'measurements' (a measurement describes the profile of a given compound of a given batch and at a given concentration, thus, the number of measurements is higher than to the total number of profiled compounds). Therefore, U2OS cells were treated with the compounds for 20 h prior to staining of cell compartments and components such as DNA, RNA, actin, plasma membrane, Golgi, mitochondria and endoplasmic reticulum.[2,3] Morphological profiles consist of Z scores of 579 features, which represent the difference for each feature relative to the DMSO control.[7] The percentage of altered features defined as the induction (in %) is used as a measure of bioactivity and compounds are considered active for induction ≥ 5%. Profile similarity (termed biosimilarity, in %, see the Methods section) is used for profile comparison and profiles are similar to each other if biosimilarity is higher than 75 %. Of the 4,256 tested references, 4,167 were tested in a concentration range between 2 and 10 µM. 1,417 (34%) of them were non-toxic (cell count ≥ 50 % compared to the DMSO control)

and displayed activity with an induction between 5 and 85%. The final data set used for the present analysis, which contains also measurements outside of the 2-10 μM concentration range, is composed of 1,888 different reference compounds and three internal compounds (in total 3,572 measurements at different concentrations) that were non-toxic and showed an induction between 5 and 85 %. Analysis of this data set revealed only in few cases biosimilarity for compounds with similar target annotation, e.g., for inhibitors of Aurora kinase, BET, HDAC, HSP90, Na^+/K^+ ATPase and protein synthesis. Instead, for a given reference compound, we detected diverse annotated activities for the most highly biosimilar compounds. Whereas biosimilarity of MTOR inhibitors to PI3K and AKT is most likely attributed to modulating the same pathway (AKT/PI3K/MTOR), we recently detected biosimilarity of the iron chelator Deferoxamine to nucleoside analogues, antifolates and CDK inhibitors.[5] The biosimilarity of these compounds is not attributed to impairment of the same target but rather the same process, namely DNA synthesis and, thus, the cell cycle. We recently reported that CPA can detect modulators of *de novo* pyrimidine biosynthesis and the identified cluster includes compounds that interfere with the activity of different enzymes or complex III of the mitochondrial electron transport chain.[8] Furthermore, we noticed several reference compounds that share similar CPA profiles with Nocodazole, a well-studied tubulin inhibitor, although for most of these compounds the bioactivity annotation was different from tubulin targeting.[9] A literature search revealed impairment of microtubules for several compounds and we experimentally validated tubulin targeting for reference compounds that had not been previously linked to tubulin. In addition, 12 % of the reference set (27 % of the active references) belong to a large cluster with deregulation of cholesterol biosynthesis as a common denominator.[6] This cluster contains compounds with very diverse targets and the activity of approximately 75 % of the compounds is most likely due to their physicochemical properties that lead to accumulation of compounds in lysosomes due to protonation, thereby raising lysosomal pH and affecting the activity of lysosomes, lysosomal enzymes and lysosome-

dependent processes such as autophagy. We also detected a cluster of biosimilar compounds related to the profile of the uncoupling reagent FCCP (see Figure S1). FCCP is biosimilar to Tyrphostin AG 879 and Tyrphostin A9 that are known to be kinase inhibitors. However, tyrphostins also disturb the mitochondrial proton gradient.[10,11] Moreover, the antimicrobial agent Triclosan and the IKK kinase inhibitor IMD0354 were biosimilar to FCCP and uncoupling activity for these compounds has been reported.[12,13] although this information rarely is included in the compound annotation. Thus, Cell Painting detects uncoupling of the mitochondrial proton gradient as the common mode of action for these small molecules, as observed also by different morphological profiling approaches.[1] Consequently, for the CPA profile of an uncharacterized small molecule and even a reference compound, a simple inspection of the top biosimilar reference compounds often would not lead to generation of target hypothesis as frequently compounds are biosimilar to several apparently not related reference compounds. One possibility is to map the location of the profile in a lower dimension plot (e.g., PCA, tSNE or UMAP), however, bioactivity clusters have to be defined beforehand and, more importantly, polypharmacological compounds may fail to localize in a separate cluster as recently shown for the BET inhibitor CF53.[9]

To enhance the correct assignment to bioactivity (MoA) clusters, we set out to determine the characteristic profile of each cluster that potentially could be employed to assess biosimilarity to a predefined cluster rather than to single references. As compounds change only a subset of the 579 morphological features, we envisioned that we can extract these altered features for a given bioactivity cluster to yield a cluster 'subprofile'. If such subprofiles represented the MoA of the clusters, they could successfully be used to simultaneously map biosimilarity of compounds to all clusters defined thus far. To generate the subprofiles, first the dominating features for a set of biosimilar compounds are extracted and then a representative consensus subprofile is defined, which then in turn could be used to determine the biological similarity of

new compounds to a respective cluster (Figure 1A). For subprofile identification, the following procedure was applied:

In Pseudo-code:

```
 01: function cluster_features(list_of_cluster_defining_compounds):
02:     list_of_cluster_features = empty_list()
03:     number_of_defining_compounds =
04:         length(list_of_cluster_defining_compounds)
05:     for each feature in all_features:
06:         count_plus = 0
07:         count_minus = 0
08:         for each compound in list_of_cluster_defining_compounds:
09:             if feature_value_of(compound) >= 0:
10:                 count_plus += 1
11:             else:
12:                 count_minus += 1
13:
14:         fraction = max(count_plus, count_minus) /
15:             number_of_defining_compounds
16:         if fraction >= 0.85:
17:             list_of_cluster_features.append(feature)
18:     return list_of_cluster_features
```

The pseudo-code of the function that extracts the dominating features of a given biological cluster listed above is explained as follows (line numbers in curly brackets):

The function takes a list of Cell Painting measurements for the compounds that define the cluster (concrete: a Pandas dataframe) {1}. An empty list is defined to hold the features that will be collected for the cluster {2}. The number of measurements that define this cluster is counted. This is equal to number of rows in the dataframe that was passed to the function {3-4}.

A loop is then run over each of the 579 measured features {5}. Within this loop, the variables for counting the positive values and for counting the negative values are set to zero {6-7}. Now, for each of the measured compounds that were passed to the function, the measured value of the current feature in the loop is assessed whether it is positive or negative and the respective

counter is increased accordingly {8-12}. After iteration over all defining compounds, the maximum of the two counters is determined and divided by the total number of defining measurements {14-15}. If this fraction is ≥ 0.85 then the given feature is added to the collected list of features for this cluster {16-17}, which is finally returned as the result when the loop over all 579 features has finished {18}.

(The Python implementation can be found in the function jupy_tools.cpa.cluster_features of the Github repository https://github.com/mpimp-comas/2022_pahl_ziegler_subprofiles)

From the list of cluster features subsequently a representative median subprofile for the cluster can be calculated by taking the median values over all compounds from the defining set for every given feature and combining them into a new reduced profile (Figure 1A). This median (*consensus*) subprofile can then be used for the calculation of biosimilarity of test compounds to the cluster (Figure 1A). The dominating features for a cluster are then defined as all features of the full Cell Painting profile, where the values of 85% of the compounds from the cluster-defining set have the same sign (are all either positive or negative).

Initially, we extracted the subprofile features for the tubulin cluster based on the three highly biosimilar compounds Vinblastine, Vincristine and Nocodazole.[9] The resulting median subprofile consisted of 531 features instead of 579 features of the full profiles (Figure 1B). We then searched for reference compounds that are biosimilar to the tubulin cluster subprofile by calculating the similarity only between the extracted 531 features for all compounds. Several known tubulin-targeting small molecules displayed subprofile biosimilarity of more than 85%, e.g., Rotenone, Tyk2-IN-2, 2-Methoxy-estradiol, and Albendazole (Figure 1C). However, the subprofiles for Fluvastatin (6 µM), CAY10603 (0.2 µM) and Tetracaine (30 µM), which are reported to impair microtubules[14-16], showed biosimilarity lower than 75% to the tubulin cluster

subprofile (see Figure 1C). As the profiles for Vinblastine, Vincristine and Nocodazole are highly biosimilar (≥ 95%), the tubulin cluster subprofile contains 92% of all features. Considering that many references that target tubulin have a nominal target different from tubulin (the nominal target is the target that is commonly associated with the compound[17]), their profiles may also contain the morphological signature that is caused by the nominal target, which may partly overlay the tubulin subprofile. To account for polypharmacology responses in the Cell Painting analysis, we selected a diverse set of 34 profiles from 26 confirmed tubulin-targeting agents, many of which are known to target also different proteins (see cluster 'Tubulin' in Table S1), to generate the tubulin signature.[9] The resulting tubulin cluster subprofile includes 424 features, i.e., 107 features less than the initially created tubulin cluster profile based only on Vinblastine, Vincristine and Nocodazole (Figure 1D). Using this tubulin cluster subprofile, we detected biosimilarities to the tubulin cluster of higher than 85% for Fluvastatin, CAY10603 and Tetracaine, which would confirm the compounds as tubulin-targeting agents (Figure 1E).

We applied the same strategy to extract subprofiles of clusters that are based on modulation of AKT/PI3K/MTOR, Aurora kinase, BET, DNA synthesis, HDAC, HSP90, lysosomotropism/cholesterol homeostasis, Na^+/K^+ ATPase, protein synthesis, pyrimidine biosynthesis, tubulin or uncoupling of the mitochondrial proton gradient. For this purpose, for each of the clusters, a set of compounds with confirmed activity was defined (see Table S1 for the respective set of reference compounds that were used to define the clusters for subprofile extraction).

We then extracted the cluster features and generated the cluster subprofiles of the twelve identified biological clusters, which were of different length (see Figure 2A): whereas the DNA synthesis or Aurora kinase subprofile are characterized by 288 and 358 features, respectively, the lysosomotropism/cholesterol homeostasis (L/CH) cluster subprofile contains 504 features.

A biosimilarity search using the obtained subprofiles successfully identified the references of the respective compound set for each cluster (Table S1). These twelve clusters were mapped in a lower dimension plot, using Uniform Manifold Approximation and Projection (UMAP[18], see Figure 2B). Several clusters, although separated from each other, were located close together as also demonstrated by cross-correlation analysis (Figure 2C and Figure S2), which may complicate or even prevent MoA assignment for compounds based only on the location in the plot. The high cross-correlation of the subprofiles for cholesterol homeostasis and HDAC and BET clusters most likely is a result of the similar transcriptional alterations caused by BET and HDAC inhibitors, which both also impair the transcription of cholesterol biosynthesis genes and, thus, cholesterol homeostasis.[19,20] However, when only the overlapping clusters are explored in the lower dimension space, they can clearly be differentiated (Figure S2B-S2C). Moreover, inspection of the biosimilar references may also provide useful hints towards one of these specific clusters.

Shared subprofile biosimilarity is expected to uncover the mode of action for unexplored compounds or to find unanticipated activity for reference compounds. For instance, a biosimilarity search using the tubulin cluster subprofile identified the PDGFRβ and B-Raf inhibitor KG 5 and the JAK inhibitor AZ 960 as potential microtubule-interfering compounds (Figure 3A). The cluster biosimilarity heatmap revealed high similarity to the tubulin cluster (Figure 3B). Whereas the full profile of KG 5 shares 94% biosimilarity to the profile of Nocodazole, the full profile biosimilarity for AZ 960 is 75 %, and, thus, at the lower limit (Figure 3C and 3D). However, the subprofile analysis clearly indicates similarity of AZ 960 to the tubulin cluster (cluster biosimilarity of 90%). Indeed, upon treatment with the compounds, microtubule organization was disturbed (AZ 960) or microtubules were depolymerized (KG 5) (Figure 3E and Figure S3). These results confirm impairment of the tubulin cytoskeleton by

KG 5 and AZ 960 at low micromolar concentrations, which should be considered when these compounds are used to modulate their nominal targets.

Subprofile analysis predicted HDAC-related activity for two yet unexplored compounds (compounds **1** and **2**, Figure 4A and 4B) both bearing a trifluoromethyl ketone. This moiety is present in some HDAC inhibitors and is involved in zinc (II) coordination.[21] Compounds **1** and **2** differ by only one methylene group and share high biosimilarity with the HDAC inhibitor trichostatin A (TSA, see Figure 4C) and to the HDAC cluster subprofile (Figure 4B and 4D). Therefore, the influence of **1** and **2** on HDAC activity was explored in nuclear extracts from HeLa cells. Both compounds dose-dependently inhibited *in vitro* deacetylation (Figure 4E). When U2OS cells were incubated with the compounds for 2 h, HDAC activity was reduced down to ca. 20 % (Figure 4F), thus confirming inhibition of HDAC activity by **1** and **2**.

Furthermore, the subprofile analysis identified the ALK2- and ALK3 inhibitor LDN193189 and macrocycle **3** as potential protein synthesis inhibitors (Figure 5A and 5B). At a concentration of 1 µM, LDN193189 shares 91 % subprofile biosimilarity to the protein synthesis cluster (Figure 5B-5D). The subprofile of compound **3**[22] was 87 % biosimilar to the subprofile of protein synthesis inhibitors (Figure 5B-5D). Indeed, both compounds inhibited *in vitro* protein translation (Figure S4A). Compound **3** and to a lesser extent LDN193189 suppressed protein synthesis also in cells (Figure 5E and 5F, higher concentrations than 1 µM of LDN193189 could not be used due to substantially reduced cell count). These results confirm inhibition of protein translation by LDN193189 and compound **3**. LDN193189 is widely used in stem cell culture to induce differentiation, albeit at lower concentrations than 1 µM[23,24]. Of note, no similarity to the protein synthesis subprofile was detected for 0.3 and 0.5 µM LDN193189 (Figure S4B and S4C). To the best of our knowledge, no impairment of protein translation by LDN193189 has been reported thus far. However, gene expression analysis using L1000, a platform for exploring the expression of 1000 landmark genes that represent the transcriptome[25], suggests

connectivity between LDN193189 and protein synthesis in HT29, MCF7 and PC3 cells (Figure S4D) that further supports our findings. Correlation between morphological- and gene expression profiles has previously been reported[26-28], and we detected similarity in L1000 signatures for several reference compounds that define the different clusters (Table S2).

Compounds may induce distinct morphological changes as a result of target modulation. Graded phenotypes would correspond to CPA profiles that are similar at different compound concentrations. Increasing the dose is expected to enhance the changes that are detected at lower concentrations, as observed for the iron chelator deferoxamine (DFO, Figure 6A and 6B).[5] The profiles of DFO at 2 to 20 µM are biosimilar to each other, particularly for 3 µM to 30 µM (Figure 6B and 6C). At 2 µM DFO, we observed low induction of 9 % and biosimilarity to 3 µM DFO. The subprofile analysis indicates similarity to the DNA synthesis cluster for 3 µM to 30 µM DFO that is line with our previous results (Figure 6D)[5]. At 2 µM, the similarity to the DNA synthesis cluster is 78 %, thus indicating that the phenotype starts to evolve already at this concentration (Figure 6D). Interestingly, we noticed that for some compounds the profiles at different concentration are dissimilar. The profiles of the PI3K and MTOR inhibitor PI-103 at 2 µM and 10 µM are not biosimilar and are assigned to two different clusters, i.e., AKT/PI3K/MTOR and L/CH respectively (Figure 6E-6G). The profile of 3 µM PI-103 resembles features of the profiles at 2 and 10 µM and, thus, constitutes a mixed profile. A similar observation was made for the pan-AKT inhibitor Afuresertib and the Aurora kinase inhibitor BI-847325 (Figure S5): cluster biosimilarity assigned the profiles at low doses for both compounds to the cluster of their nominal targets, whereas at higher concentrations the phenotypes changed and biosimilarity to the L/CH cluster becomes evident. Both, Afuresertib and BI-847325, share physicochemical properties of lysosomotropic compounds (i.e., clogP >2 and bpKa > 6.5)[29] that may explain the phenotype at higher concentration. However, PI-103 has a clogP of 3.53 and bpKa of 2.4 and does not classify as a lysosomotropic compound. The

L/CH phenotype at 10 µM PI-103 most likely results from the modulation of its nominal target as MTOR activity is directly linked to regulation of lysosomes[30].

As a subprofile constitutes only part of the full profile, we used the non-cluster features to gain further bioactivity information. We previously assigned a large number of reference compounds which have very diverse annotated nominal targets to the lysosomotropism/cholesterol homeostasis cluster.[6] Potentially, the nominal activity of these compounds may be detectable using CPA but masked by the dominant profile of disturbing cholesterol homeostasis. Therefore, we searched for biosimilar reference compounds using only the non-cluster features. For example, the NUAK1 inhibitor HTH-01-015 shares high biosimilarity to the lysosomotropism/cholesterol homeostasis cluster at 10 µM but not at 3 µM (Figure 7A and 7B). However, bioactivity analysis for HTH-01-015 at 10 µM using the subprofile of only the 75 non-cluster features revealed biosimilarity to the respective subprofiles of the Aurora kinase inhibitors SNS-314, ZM-447439 and Barasertib that are highly similar to the Aurora cluster, whereas the full profiles were dissimilar (Figure 7C and Figure S6A).

Targeting of Aurora kinases by HTH-01-015 has not been demonstrated thus far.[32] Analysis of Aurora kinase A-, B- and C activity using an *in vitro* kinase assay revealed dose-dependent inhibition of Aurora kinases with IC_{50} values of 5.6 µM (AURKA), 0.95 µM (AURKB) and 0.97 µM (AURKC) respectively. Of note, the cluster biosimilarity heatmap for HTH-01-015 at 3 µM suggests a potential inhibition of Aurora kinases (Figure 7B). The full profiles for HTH-01-015 at 3 and 10 µM are not biosimilar (Figure 7C), whereas biosimilarity is detected when only the non-L/CH features are compared but not when L/CH or Aurora cluster subprofiles were used (Figure S6B-4F). HTH-01-015 has a basic pKa of 10.08 and clogP of 2.28 and, thus, shares physicochemical properties with lysosomotropic compounds[29]. Lysosomotropic properties are often detected at higher concentrations, e.g., 10 µM, which explains the different

morphological profiles for HTH-01-015 at 3 and 10 µM. These results demonstrate that non-cluster features can be employed for the detection of further bioactivities.

We envisioned that the non-profile features can further be used to differentiate between different mechanisms of action within a cluster that is based on a common mode of action, for instance the cluster of DNA synthesis/cell cycle inhibitors[5]. This cluster consists of compounds with different targets such as iron chelators, CDK and topoisomerase inhibitors as well as nucleosides and antifolates. Hierarchical clustering using the non-cluster profiles for compounds in this cluster revealed a clear separation of profiles of compounds sharing the same mechanism of action (see Figure 7F) and additionally suggests an unanticipated mechanism for compound **4** and BMS-265246, both reported to be CDK inhibitors.

Discussion

The identification of small-molecule modulators of disease-relevant targets is a focus of chemical biology research and drug discovery. There is a high demand for bioactivity annotations in compound screening collections, which additionally may facilitate target deconvolution in phenotypic screening. Conducting numerous target- and cell-based assays is time- and labor-consuming and still cannot cover the sheer variety of bioactivity that can be exerted by small molecules in cells. In contrast, profiling approaches, such as gene and protein expression profiling or morphological profiling, provide a more holistic view on the bioactivity of small molecules in cells by detecting hundreds of features. Proteomics and transcriptomics have not been suitable for high-throughput screening, although approaches enabling transcriptomics for screening have recently emerged[33]. Morphological profiling can be performed in medium to high throughput and, thus, is particularly suitable for bioactivity detection. In contrast to transcriptomics and proteomics, linking altered morphological features

to upstream regulation is challenging or even impossible without additional information and data. Therefore, morphological profiling relies on the profiles of reference compounds for the generation of target- or mode-of-action hypothesis. However, we and others noticed that profiles of reference compounds with similar target annotation can differ as the activity may be governed not by the nominal target but rather an off-target.[6,9,34-36] Moreover, reference compounds with diverse annotation may form a morphological cluster that is based on the same mode of action rather than on the same target.[5] Assignment of bioactivity by simply referring to the annotated MoA of the top most biosimilar compounds may be in many cases misleading. Thus, bioactivity clusters have to be defined *a priori* to allow bioactivity prediction for uncharacterized compounds, which may also reveal unanticipated activity for reference compounds. Mapping the location of a compound profile in a lower dimension plot may provide insight into the MoA if the profile is located in a predefined cluster. However, due to polypharmacology, profiles may locate in the space between the clusters[9], thus hampering MoA assignment. To facilitate profile analysis, we introduce the concept of subprofiles that contain only features describing the bioactivity of a predefined cluster. Subprofiles are extracted from the full profiles of a set of reference compounds and are specific for each cluster. Thus far, we defined twelve morphological clusters based on targeting AKT/PI3K/MTOR, Aurora kinases, BET, DNA synthesis, HDAC, HSP90, lysosomotropism/cholesterol homeostasis, Na^+/K^+ ATPase, protein synthesis, pyrimidine biosynthesis, tubulin and uncoupling of the mitochondrial proton gradient. These cluster subprofiles are of different length and compile the features that are characteristic for each individual cluster. The generated cluster subprofiles can be employed to simultaneously assess the biosimilarity of small-molecule profiles to thus far twelve clusters and, thereby to rapidly and reliably predict bioactivity related to one or more clusters. Whereas for full profiles we consider compounds as biosimilar if their profiles share > 75 % biosimilarity, for subprofiles this threshold lies at 80-85% as subprofiles contain less features. We demonstrate that the assignment of bioactivity based on the cluster subprofile

analysis can subsequently be experimentally confirmed as exemplified for three uncharacterized compounds targeting HDAC or protein synthesis, which were synthesized in a different context than the detected bioactivity.

The subprofile approach substantially simplifies the analysis of the multidimensional profiles from the Cell Painting assay and allows scientists with different backgrounds (e.g. chemists, biologists and computational scientists) to easily obtain mode-of-action information without the need to inspect all biosimilar profiles and references. Moreover, the subprofile analysis bypasses the issue of incomplete bioactivity annotation and polypharmacology that in many cases can be misleading and generate wrong conclusions. In addition, subprofile analysis may be superior to cluster analysis using dimensionality reduction as polypharmacology may be reflected in the profiles and such profiles would not be assigned to one given cluster in a lower dimension plot as recently reported for CF53[9].

Subprofile analysis allows detection of dose-dependent phenotypes and different concentrations of a given compound may cause graded phenotypes, i.e., for all tested doses the CP profiles are biosimilar and assigned to the same cluster. Furthermore, the subprofile approach can reveal dose-dependent changes in the phenotype, e.g., due to polypharmacology that is detectable at higher concentrations or if the extent of target inhibition causes different morphological changes as demonstrated for the PI3K- and MTOR inhibitor PI-103. Profiling compounds at different concentration is, therefore, crucial for correct MoA or target hypothesis generation using reference compounds.

Besides exploring various MoA that induce morphological changes, it is useful to disclose bioactivities that cannot be detected using CPA in U2OS cells. For instance, we recently reported on the identification of potent modulators of the aryl hydrocarbon receptor (AhR) or the indoleamine 2,3-dioxygenase 1 (IDO1). The AhR agonist Picoberin[37], although exhibiting

picomolar activity with regard to AhR, is not active in CPA below 10 µM. The structurally different AhR agonist FICZ shows some activity in CPA only at 50 µM (Figure S7A-S7C) and the profile is not biosimilar to Picoberin. This matches the observation that AhR is hardly expressed in U2OS cells.[38,39] Similar observations were made for very potent IDO1 inhibitors like Epacadostat, Linrodostat and our in-house compound Apoxidole[40]: whereas Linrodostat is active at 2 µM in CPA, Epacadostat and Apoxidole are inactive up to 10 µM and 50 µM, respectively (Figure S7D-S7F). IDO1 is not expressed in U2OS cells[41] which explains the lack of an IDO1-specific profile in this cell line for structurally different IDO1 inhibitors.

Not only the cluster features can be used for bioactivity analysis but also the non-cluster features may provide further information for bioactivity annotation as compounds may have more than one target. Impairment of more than one target may lead to complex, mixed profiles. Deconvolution of the underlying MoA of mixed profiles is challenging and the subprofile approach promises to simplify their analysis. A subprofile biosimilarity search with only the non-cluster features may reveal a second target or MoA beyond the one that determines the respective cluster bioactivity as demonstrated for the NUAK1 inhibitor HTH-01-015, which CPA assigned to the lysosomotropism/cholesterol homeostasis cluster. Analysis of the non-cluster features suggested inhibition of Aurora kinases which was subsequently confirmed. Moreover, non-cluster features can be employed to explore the mechanisms of action of a cluster that unites compounds with different targets such as the DNA synthesis cluster. This strategy allows for even more precise target prediction that may significantly shorten the target validation process.

There is a widespread interest in academia and pharmaceutical industry for thorough bioactivity annotation of compound collections. Detailed knowledge of the bioactivity of compounds may guide the hit triage process for selecting the most promising hits in screens.[42] In this regard, cluster subprofile analysis will early on point towards potential targets of a hit compound, which

may spur or prevent its priorization for further studies. Cluster subprofile analysis will complement structure-activity relationship (SAR) studies on a given target or process and may provide insights into unexpected behavior of derivatives in cellular assays. Moreover, this approach may indicate an off-target that may account for the activity observed in cell-based assays and will assist correct data interpretation, which is essential for proper hit selection. To support the chemical biology community, we disclose the cluster biosimilarity for our reference compound set that is accessible via https://github.com/mpimp-comas/2022_pahl_ziegler_subprofiles and the web app tool https://cpcse.pythonanywhere.com/.

In summary, we employed morphological subprofiles to define consensus cluster features and cluster subprofiles. Cluster subprofile analysis allows for easy and fast annotation of thus far twelve different modes of action for compound collections without the need for exploring the profiles of the most biosimilar reference compounds. The remaining non-cluster features may guide the mapping of further bioactivities and enable differentiation between mechanisms of action within a cluster that is based on a common mode of action.

Limitation of the Study

The reported subprofile analysis uses twelve defined bioactivity clusters that can be mapped using Cell painting. Looking at cluster biosimilarities to some extent takes away the unbiased nature of the CPA. However, comparison of full profiles can still be performed in parallel. For uncharacterized compounds, the lack of similarity to any of these clusters does not mean inactivity in the morphological profiling. In such cases, the most similar compounds based on profile comparison can be used for generation of target or MoA hypotheses. Additional bioactivity clusters have to be defined to cover wider target/MoA space. Therefore, the

reference compound set should be continuously expanded to facilitate the identification of further clusters. The similarity of many reference compounds with different target annotation to the lysosomotropism/cholesterol homeostasis cluster precludes the exploration of their nominal targets in CPA. Furthermore, targets or MoA should be defined by the community, whose modulation do not lead to morphological changes and, thus, to CPA profiles.

Significance

A detailed understanding of all processes modulated by small molecules in cells would immensely contribute to chemical biology and medicinal chemistry research. Bioactive compounds are usually identified using target-based or phenotypic assays. However, their precise impact on cells can hardly be predicted, and, therefore, there is a high demand to explore compound-perturbed states in an unbiased manner and at scale. Profiling approaches, like transcriptomics, proteomics and morphological profiling, provide a broader view on the bioactivities of small molecules in cells. Morphological profiling is amenable to high throughput, can identify biologically active compounds and predict their targets or mode of action (MoA). Generation of target or MoA hypotheses relies on the profiling of reference compounds with known activities. Often, morphological profiles are caused by off-targets or unanticipated targets rather than the nominal target. Therefore, for precise prediction of target or MoA, bioactivity clusters have to be defined a priori. Here we introduce the concept of subprofile analysis for the definition of cluster subprofiles using the Cell Painting assay. We determined a consensus subprofile for a group of biosimilar reference compounds, which is used to assess the biosimilarity of compound subprofiles to this cluster. This approach allows for easy and rapid assignment of bioactivity to twelve different clusters related to AKT/PI3K/MTOR, Aurora kinases, BET, DNA synthesis, HDAC, HSP90, lysosomotropism/cholesterol homeostasis, Na^+/K^+ ATPase, protein synthesis, pyrimidine

biosynthesis, tubulin and uncoupling of the mitochondrial protein gradient. The generated target or MoA hypotheses for three uncharacterized compounds was validated experimentally. Moreover, the non-cluster features may uncover bioactivity beyond the target or MoA and may hint towards a specific mechanism of action within a cluster that is based on the same MoA.

Acknowledgments

Research at the Max Planck Institute of Molecular Physiology was supported by the Max Planck Society. This work was co-funded by the European Union (Drug Discovery Hub Dortmund (DDHD), EFRE-0200481) and Innovative Medicines Initiative (grant agreement number 115489) resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in-kind contribution, and by EPSRC (grants EP/N025652/1 and EP/F043503/1). This work was funded from the programme " Netzwerke 2021", an initiative of the Ministry of Culture and Science of the State of Northrhine Westphalia. Colin Ziegler is acknowledged for the design of the web app tool. Soheila Rezaei Adariani is acknowledged for experimental assistance. Michael Grigalunas is acknowledged for the analysis of the chemistry data. The compound management and screening center (COMAS) in Dortmund is acknowledged for performing the high-throughput screening.

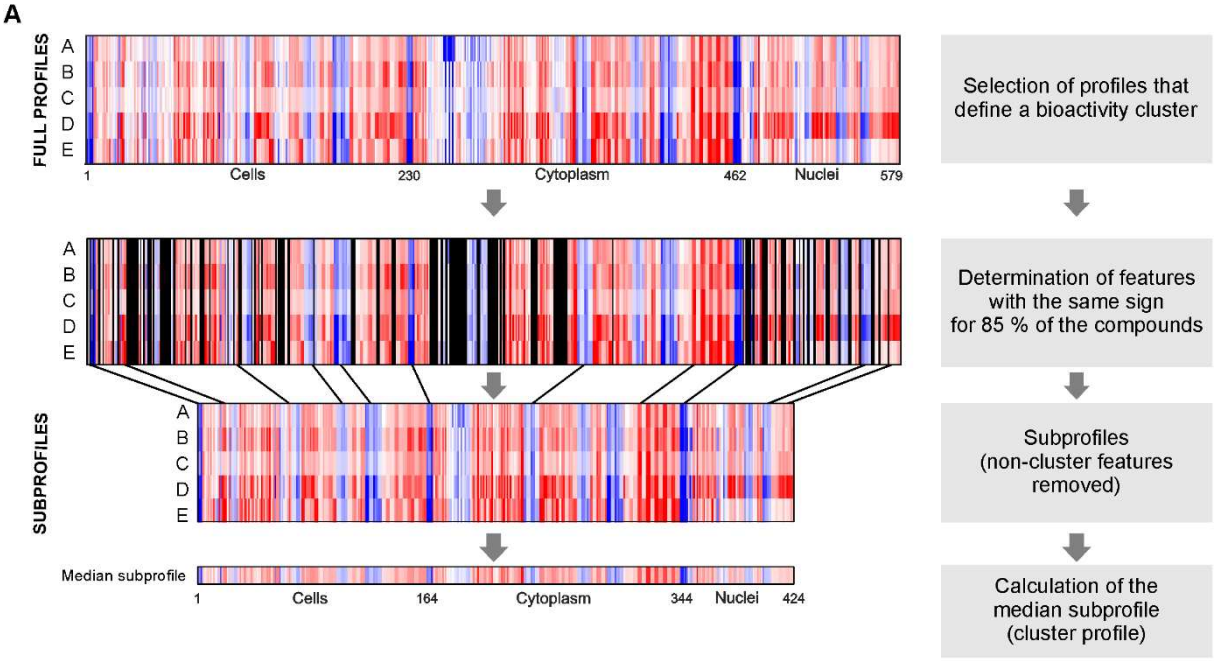
Author Contributions

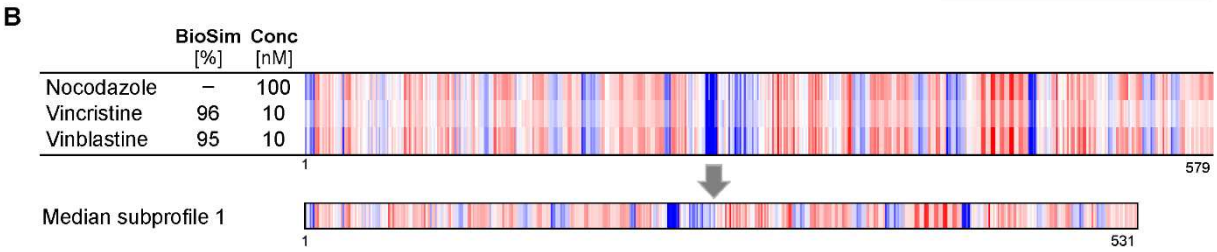
A.P. and S.Z. designed the research. B.S. and P.L. performed the biological experiments. S.S. and A.P. performed and processed CPA. M.R. and M.D. synthesized the compounds. A.N. and C.H. supervised the compound syntheses.. H.W. initiated and supervised the Cell Painting study. A.P and S.Z. analyzed the CPA data. A.P. and S.Z. wrote the manuscript. All authors discussed the results and commented on the manuscript.

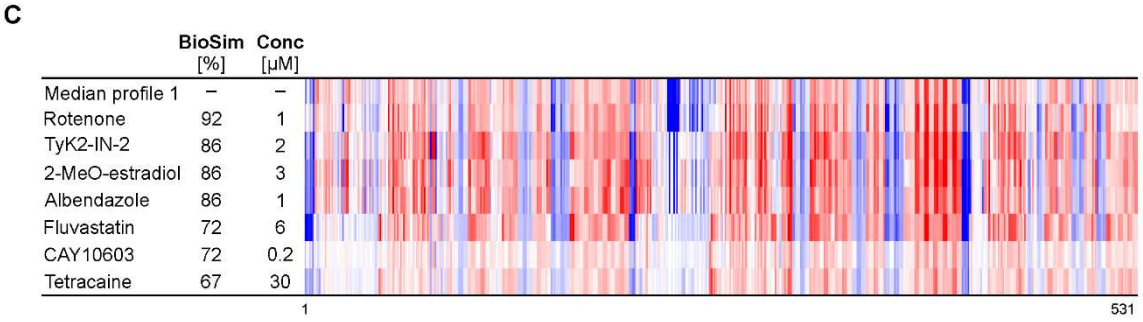
Declaration of Interests

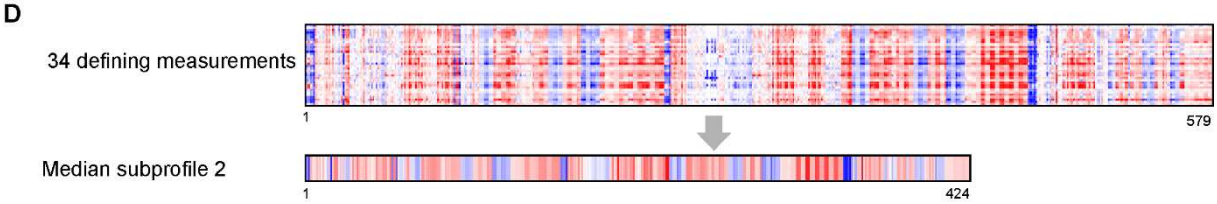
A.N. is partially seconded (20%) to Rosalind Franklin Institute, UK. M.D. is currently an employee of AstraZeneca, UK and his contribution to this study is not related to AstraZeneca, but solely to the School of Chemistry and Astbury Centre for Structural Molecular Biology, University of Leeds. M.R. is currently an employee of Roche, Switzerland, and her contribution to this study is not related to Roche, but solely to the Max Planck Institute of molecular physiology, Dortmund.

Figures


A
FULL PROFILES
A
B
C
D
E
1
Cells
230
Cytoplasm
462
Nuclei
579
SUBPROFILES
Median subprofile
1
Cells
164
Cytoplasm
344
Nuclei
424
Selection of profiles that define a bioactivity cluster
Determination of features with the same sign for 85 % of the compounds
Subprofiles (non-cluster features removed)
Calculation of the median subprofile (cluster profile)


B
BioSim [%]
Conc [nM]
Nocodazole – 100
Vincristine 96 10
Vinblastine 95 10
1
579
Median subprofile 1
1
531


C
BioSim [%]
Conc [µM]
Median profile 1 – –
Rotenone 92 1
TyK2-IN-2 86 2
2-MeO-estradiol 86 3
Albendazole 86 1
Fluvastatin 72 6
CAY10603 72 0.2
Tetracaine 67 30
1
531


D
34 defining measurements
1
579
Median subprofile 2
1
424

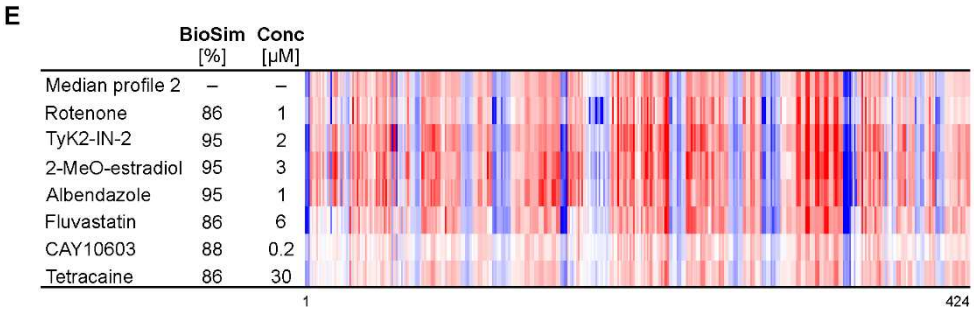

E
BioSim [%]
Conc [µM]
Median profile 2 – –
Rotenone 86 1
TyK2-IN-2 95 2
2-MeO-estradiol 95 3
Albendazole 95 1
Fluvastatin 86 6
CAY10603 88 0.2
Tetracaine 86 30
1
424

Figure 1. Extraction of cluster subprofiles. (A) Workflow for extracting relevant features from a set of defining features and calculation of median subprofile. From the full morphological profiles for biosimilar reference compounds (compounds A-E), features whose values do not have the same sign for 85% of the defining compounds will be removed (larger areas of these features that will be removed are marked in black). This results in reduced (shorter) profiles, i.e. subprofiles, containing only the homogeneous features (here 424). (B-E) Extraction of tubulin cluster subprofile. (B) Narrow definition of the cluster using only Nocodazole, Vincristine and Vinblastine. The high biosimilarity of these three compounds results in a long median cluster subprofile with 531 features (92 % of the full profile). (C) Biosimilarity to the cluster of known tubulin-targeting compounds shows a low value for some compounds. (D) Broader definition of the tubulin cluster using 34 profiles from 26 confirmed tubulin-targeting agents results in a shorter median cluster subprofile with 424 features. (E) Biosimilarity of the same compounds as in (C) now shows values ≥ 85 % for all examples. Blue color: decreased feature, red color: increased feature.

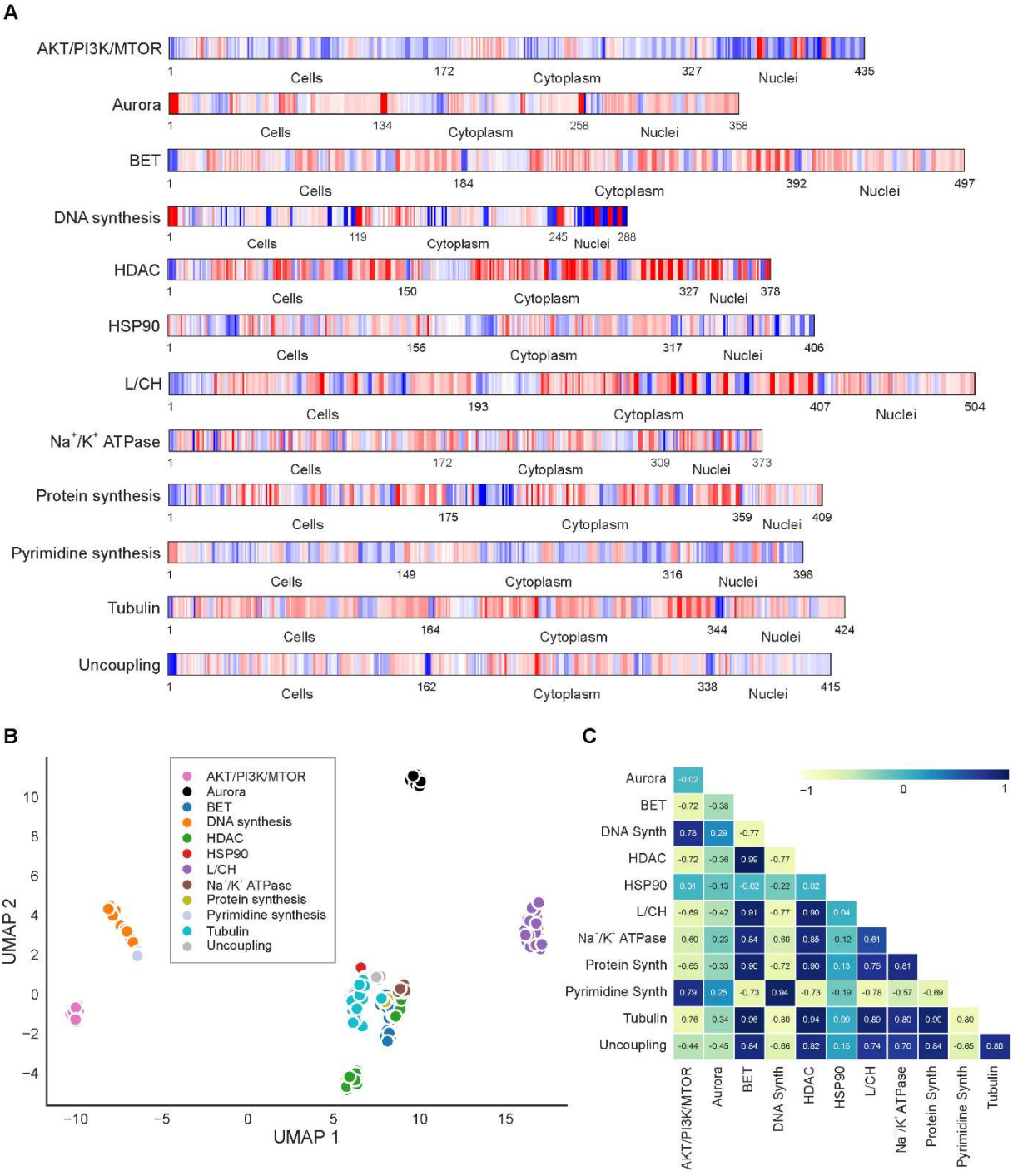


Figure 2. Subprofiles of the twelve defined bioactivity clusters. (A) Median cluster subprofiles of the defined clusters. Blue color: decreased feature, red color: increased feature. (B) UMAP plot using the full profiles of the reference compounds that were used to define the bioactivity clusters. Not normalized data, 15 neighbors. (C) Cluster subprofile cross-correlation using Pearson correlation. L/CH: lysosomotropism/cholesterol homeostasis cluster; Synth: synthesis. See also Figure S1 and S2 and Table S1 and S2.

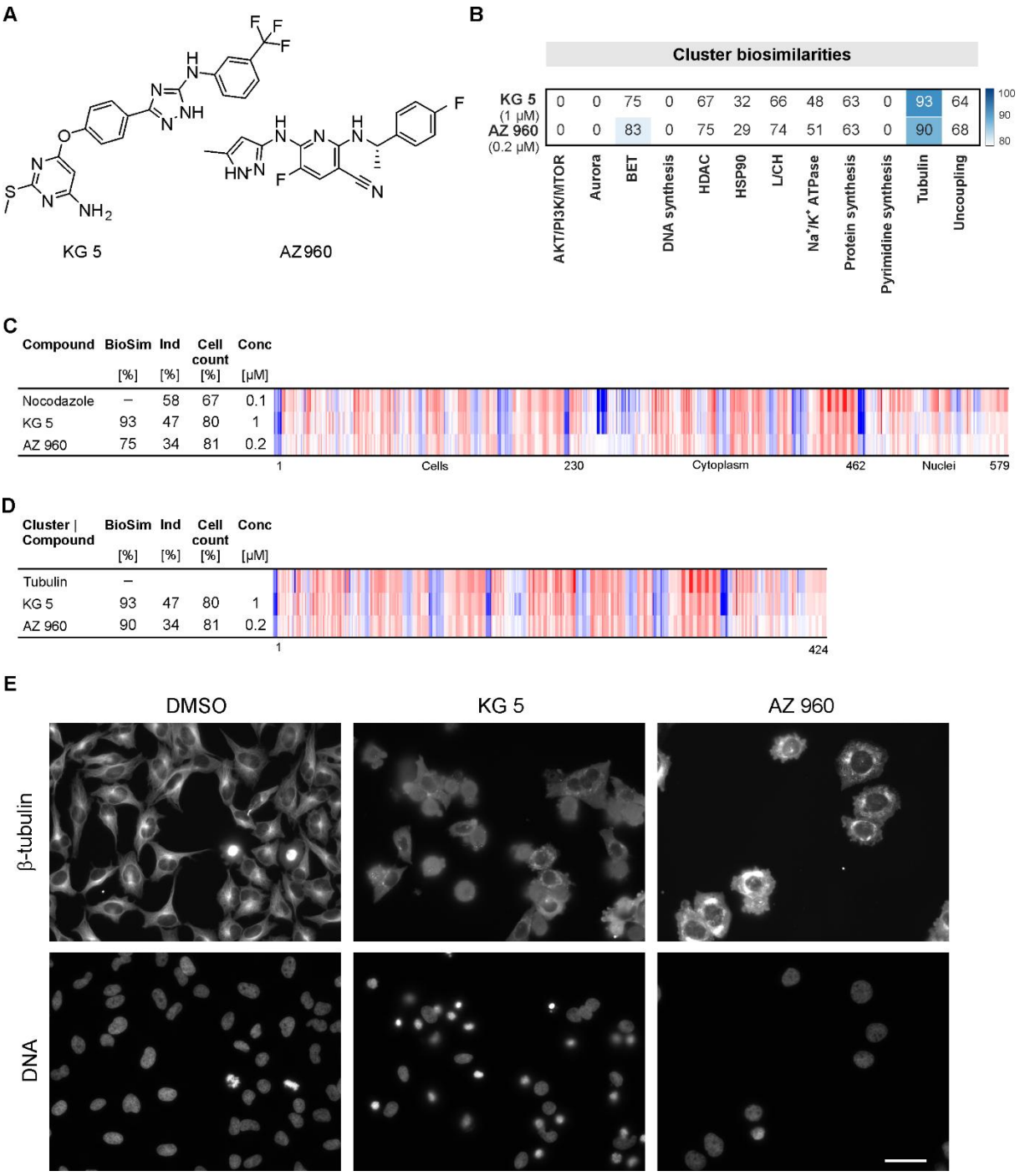


Figure 3. Identification of microtubule inhibitors. (A) Structures of KG 5 and AZ 960. (B) Cluster biosimilarity heatmap for KG 5 and AZ 960. (C) Biosimilarity of KG 5 and AZ 960 to Nocodazole. The top line profile is set as a reference profile (100 % biological similarity, BioSim) to which the following profiles are compared. Blue color: decreased feature, red color: increased feature. BioSim: biosimilarity, Ind: induction, Conc: concentration. (D) Biosimilarity of KG-5 and AZ 960 to the tubulin cluster subprofile. (E) Influence of KG 5 (10 µM) and AZ

960 (6 μM) on microtubules in U2OS cells. Cells were treated with the compounds for 24 h prior to staining for DNA and tubulin. Scale bar: 50 μm. See also Figure S3.

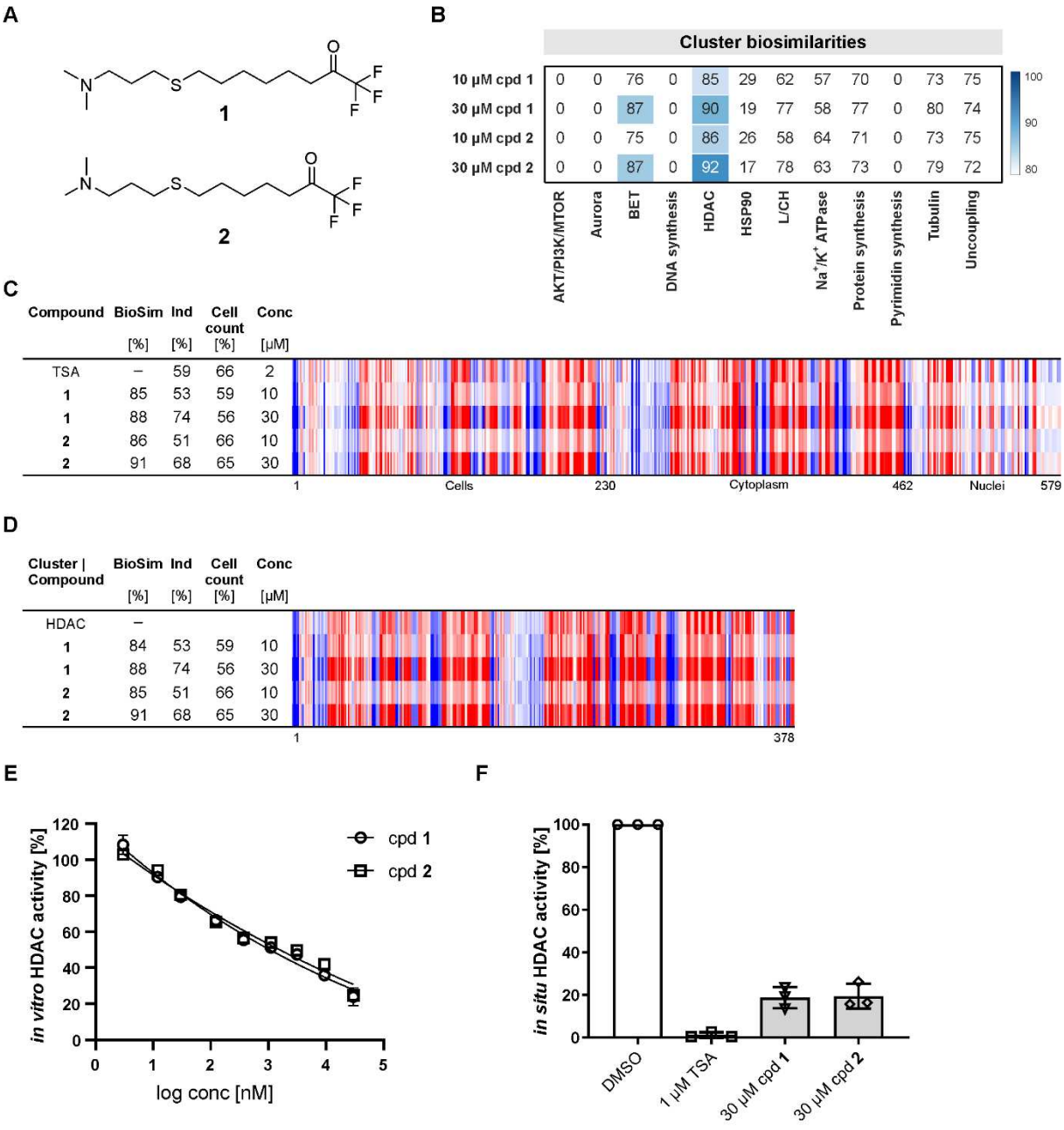


Figure 4: Inhibition of HDAC activity by compounds 1 and 2. (A) Structures of compounds **1** and **2**. (B) Cluster biosimilarity heatmap for compound **1** and **2**. (C) Biosimilarity of **1** and **2** to trichostatin A (TSA). The top line profile is set as a reference profile (100 % biological similarity, BioSim) to which the following profiles are compared. Blue color: decreased feature, red color: increased feature. BioSim: biosimilarity, Ind: induction, Conc: concentration. (D) Biosimilarity of **1** and **2** to HDAC cluster subprofile. (E) Influence on *in vitro* HDAC activity.

Data are mean values (n = 3) ± SD. (F) Influence on HDAC activity in U2OS cells upon treatment with the compounds for 2 h. Data are mean values (n = 3) ± SD.

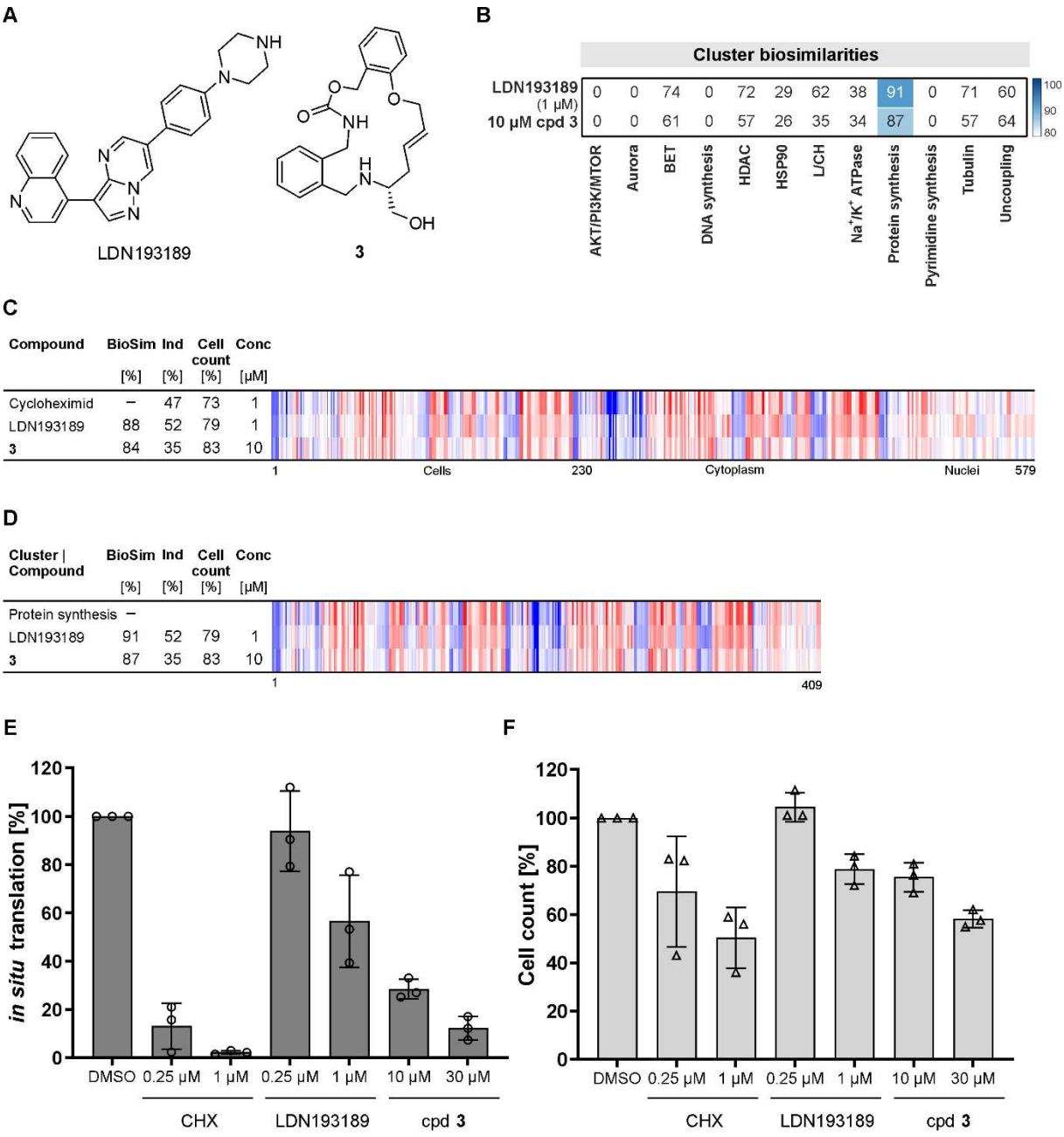


Figure 5. Identification of protein synthesis inhibitors. (A) Structures of the ALK2 and ALK3 inhibitor LDN193189 and compound **3**. (B) Cluster biosimilarity heatmap for LDN193189 and compound **3**. (C) Biosimilarity of LDN193189 and compound **3** to cycloheximide. The top line profile is set as a reference profile (100 % biological similarity, BioSim) to which the following profiles are compared. Values were normalized to the DMSO control. Blue color: decreased feature, red color: increased feature. The set of 579 features is divided in features related to the cell (1–229), cytoplasm (230–461) and nuclei (462–579). BioSim: biosimilarity, Ind: induction, Conc: concentration. (D) Biosimilarity of LDN193189

and compound **3** the protein synthesis cluster subprofile. (E and F) Influence on protein synthesis in HeLa cells. Cells were treated for 24 h with the compounds or cycloheximide (CHX) and DMSO as controls prior to detection of protein synthesis (D) or cell count (E). Data are mean values (n = 3) ± SD. See also Figure S4 and Table S2.

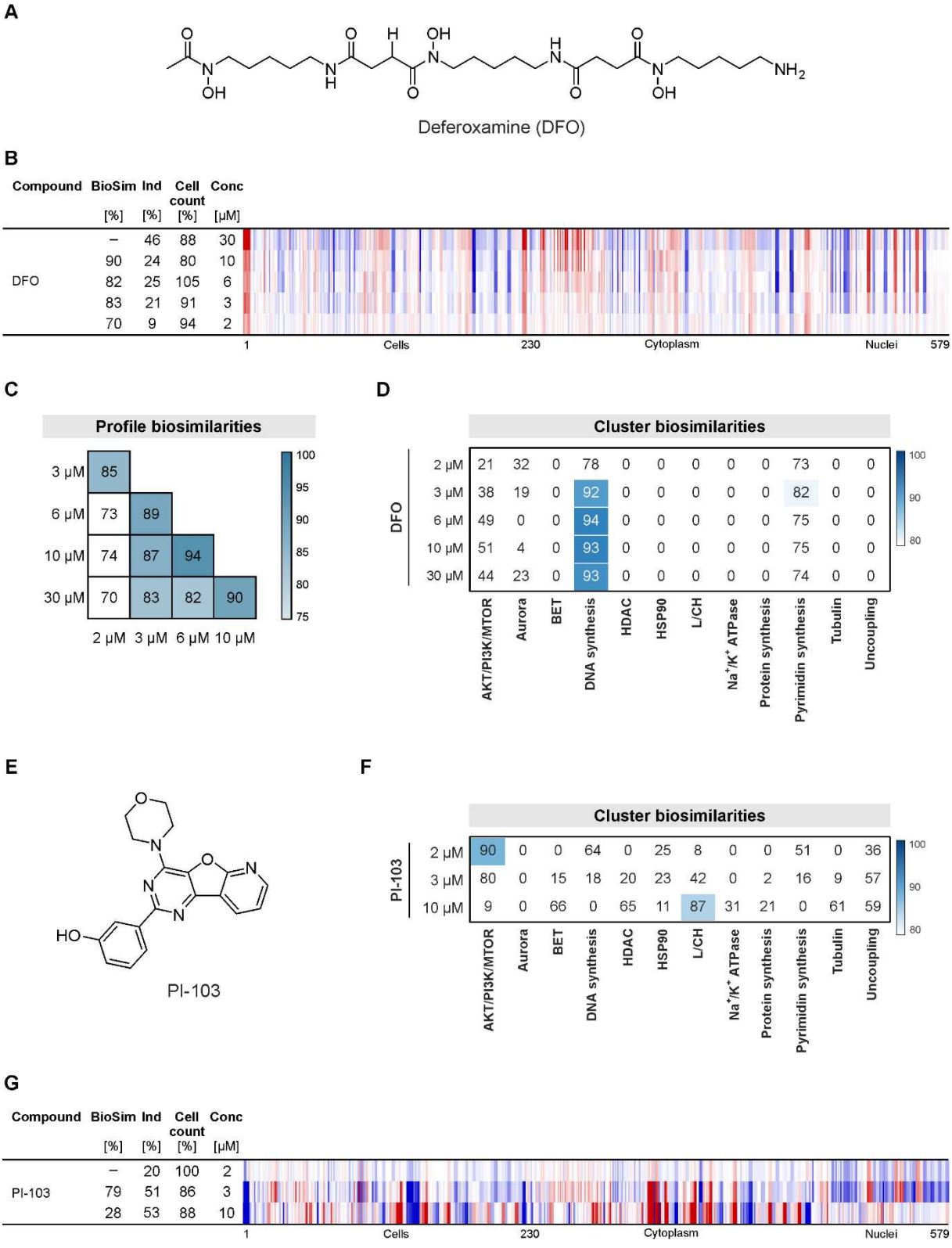
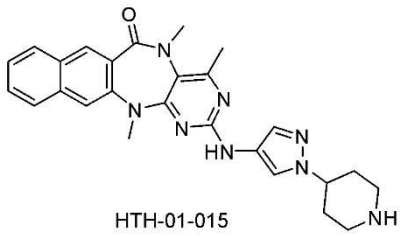


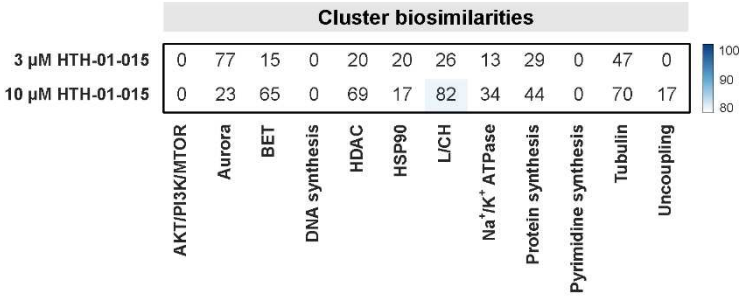
Figure 6. Dose-response phenotypes and cluster shifts. (A-D) Biosimilar profiles for Deferoxamine (DFO) at different concentrations. (A) Structure of DFO. (B) Biosimilarity of DFO profiles at different concentrations. (C) Profile biosimilarity cross correlation for DFO. (D) Cluster biosimilarity heatmap for DFO. (E-G) Dissimilar profiles for PI-103. (E) Structure

of PI-103. (F) Cluster biosimilarity heatmap for PI-103. (G) Biosimilarity of PI-103 profiles at different concentrations. For B and G: The top line profile is set as a reference profile (100 % biological similarity, BioSim) to which the following profiles are compared. Values were normalized to the DMSO control. Blue color: decreased feature, red color: increased feature. BioSim: biosimilarity, Ind: induction, Conc: concentration. See also Figure S5.

A

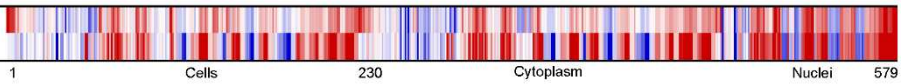


B



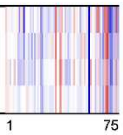
C

Compound	BioSim [%]	Ind [%]	Cell count [%]	Conc [µM]
HTH-01-015	–	62	68	3
HTH-01-015	70	74	70	10

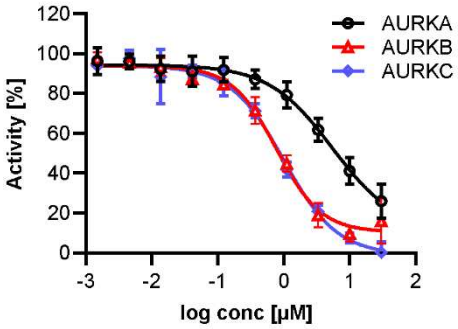


D

Compound	BioSim [%]	BioSim AURK cluster [%]	Ind [%]	Cell count [%]	Conc [µM]
HTH-01-015	–	23	74	70	10
SNS-314	86	96	41	69	0.2
ZM-447439	84	89	39	69	2
Barasertib	82	93	27	77	0.2



E



F

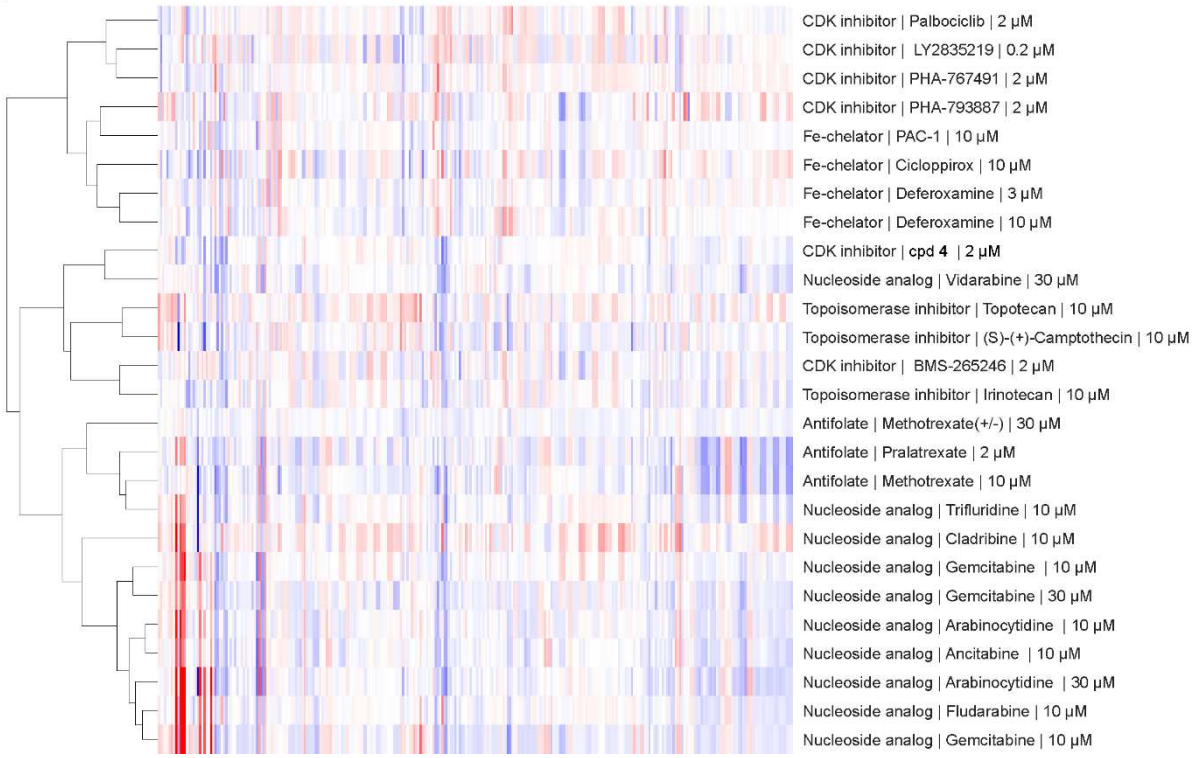


Figure 7. Non-cluster features for bioactivity assessment. (A) Structure of HTH-01-015. (B) Cluster biosimilarity heatmap for HTH-01-015. (C) Biosimilarity of HTH-01-015 profiles at 3 and 10 μM. (D) Biosimilarity of HTH-01-015 to Aurora inhibitors using only non-L/CH features. The top line profile is set as a reference profile (100 % biological similarity, BioSim) to which the following profile is compared. BioSim to Aurora cluster was calculated using the aurora cluster features. Blue color: decreased feature, red color: increased feature. BioSim: biosimilarity, Ind: induction, Conc: concentration. (E) Dose-dependent inhibition of Aurora kinases by HTH-01-015. Data are mean values (n = 2) ± SD. See also Figure S6. (F) Hierarchical clustering for subprofiles of compounds of the DNA synthesis cluster. Only non-cluster features were used. See Figure S6 for the structure of compound **4**[31].

STAR Methods

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Slava Ziegler (slava.ziegler@mpi-dortmund.mpg.de)

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The code to calculate subprofiles and profile biosimilarities, and to produce the plots in the figures can be found on GitHub. The repository also contains the full processed Cell painting data set used in this analysis (3,572 processed profiles at different concentrations). DOIs are listed in the key resources table. Cluster biosimilarity for reference compounds can be accessed via the web app tool https://cpcse.pythonanywhere.com/.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

U2OS cells (Cat#300364, Cell Line Service, Germany, RRID:CVCL_0042, female) and HeLa cells (Cat# ACC 57 DSMZ, Germany, RRID: CVCL_0030, female) were cultured in DMEM supplemented with 10% FBS, 2 mM L-glutamine, 1 mM sodium pyruvate and non-essential amino acids. Cells were maintained at 37 °C and 5% CO_2 in a humidified atmosphere. Cell lines were regularly assayed for mycoplasma and were confirmed to be mycoplasma-free.

MATERIALS

Dulbecco's Modified Eagle's medium (DMEM), L-glutamine, sodium pyruvate and non-essential amino acids were obtained from PAN Biotech, Germany. Roswell Park Memorial Institute (RPMI) 1640 medium without methionine and fetal bovine serum (FBS) was obtained from Gibco, Thermo Fisher Scientific Inc., USA. Anti-alpha-tubulin-FITC mouse mAb (#F2168, RRID:AB_476967) and 4',6-diamidine-20-phenylindole dihydrochloride (DAPI, #10236276001) were purchased from Sigma Aldrich, Germany. HDAC Activity Assay Kit (#566328) and *In Situ* HDAC Activity Fluorometric Assay Kit (#EPI003) were purchased from Sigma Aldrich, Germany. Click-iT™ HPG Alexa Fluor™ 488 Protein Synthesis Assay Kit (#C10428) and 1-Step Human Coupled IVT Kit – DNA (#88882) were obtained from Thermo Fisher Scientific Inc., USA. Microplates (384 well, white, low volume #3826, 96 well; black, clear bottom #3340) were obtained from Corning, Sigma Aldrich, Germany, and glass bottom plates (P96-1-N) from Cellvis, USA.

METHOD DETAILS

Cell Painting assay

The described assay follows closely the method described by Bray et al.[2]. Initially, 5 µl U2OS medium were added to each well of a 384-well plate (PerkinElmer CellCarrier-384 Ultra). Subsequently, U2OS cell were seeded with a density of 1600 cells per well in 20 µl medium. The plate was incubated for 10 min at the ambient temperature, followed by an additional 4 h incubation (37 °C, 5% CO_2). Compound treatment was performed with the Echo 520 acoustic dispenser (Labcyte). Different concentrations of DMSO were used as controls dependent on the used compound concentration, e.g., 0.1 % DMSO was used as a control for the profiling of compounds at 10 µM. Samples at a given compound concentration were compared to the DMSO sample of the same DMSO concentration. Incubation with compound was performed for 20 h (37 °C, 5% CO_2). Subsequently, mitochondria were stained with Mito Tracker Deep Red (Thermo Fisher Scientific, Cat. No. M22426). The Mito Tracker Deep Red stock solution (1 mM) was diluted to a final concentration of 100 nM in prewarmed medium. The medium was removed from the plate leaving 10 µl residual volume and 25 µl of the Mito Tracker solution were added to each well. The plate was incubated for 30 min in darkness (37 °C, 5% CO_2). To fix the cells 7 µl of 18.5 % formaldehyde in PBS were added, resulting in a final formaldehyde concentration of 3.7 %. Subsequently, the plate was incubated for another 20 min in darkness (RT) and washed three times with 70 µl of PBS. (Biotek Washer Elx405). Cells were permeabilized by addition of 25 µl 0.1% Triton X-100 to each well, followed by 15 min incubation (RT) in darkness. The cells were washed three times with PBS leaving a final volume of 10 µl. To each well 25 µl of a staining solution were added, which contains 1% BSA, 5 µl/ml Phalloidin (Alexa594 conjugate, Thermo Fisher Scientific, A12381), 25 µg/ml Concanavalin A (Alexa488 conjugate, Thermo Fisher Scientific, Cat. No. C11252), 5 µg/ml Hoechst 33342 (Sigma, Cat. No. B2261-25mg), 1.5 µg/ml WGA-Alexa594 conjugate (Thermo Fisher

Scientific, Cat. No. W11262) and 1.5 µM SYTO 14 solution (Thermo Fisher Scientific, Cat. No. S7576). The plate is incubated for 30 min (RT) in darkness and washed three times with 70 µl PBS. After the final washing step, the PBS was not aspirated. The plates were sealed and centrifuged for 1 min at 500 rpm.

The plates were prepared in triplicates with shifted layouts to reduce plate effects and imaged using a Micro XL High-Content Screening System (Molecular Devices) in 5 channels (DAPI: Ex350-400/ Em410-480; FITC: Ex470-500/ Em510-540; Spectrum Gold: Ex520-545/ Em560-585; TxRed: Ex535-585/ Em600-650; Cy5: Ex605-650/ Em670-715) with 9 sites per well and 20x magnification (binning 2).

The generated images were processed with the *CellProfiler* package (https://cellprofiler.org/, version 3.0.0)[43] on a computing cluster of the Max Planck Society to extract 1716 cell features per microscope site. The data was then further aggregated as medians per well (9 sites -> 1 well), then over the three replicates.

Further analysis was performed with custom *Python* (https://www.python.org/) scripts using the *Pandas* (https://pandas.pydata.org/) and *Dask* (https://dask.org/) data processing libraries as well as the *Scientific Python* (https://scipy.org/) package.

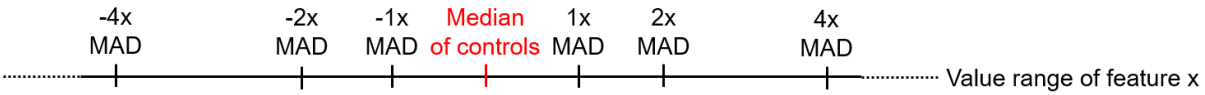
From the total set of 1716 features, a subset of highly reproducible and robust features was determined using the procedure described by [35] in the following way:

Two biological repeats of one plate containing reference compounds were analysed. For every feature, its full profile over each whole plate was calculated. If the profiles from the two repeats showed a similarity >= 0.8 (see below), the feature was added to the set.

This procedure was only performed once and resulted in a set of 579 robust features out of the total of 1716 that was used for all further analyses.

The phenotypic profiles were compiled from the Z-scores of all individual cellular features, where the Z-score is a measure of how far away a data point is from a median value.

Specifically, Z-scores of test compounds were calculated relative to the Median of DMSO controls. Thus, the Z-score of a test compound defines how many MADs (Median Absolute Deviations) the measured value is away from the Median of the controls as illustrated by the following formula:



$$z - score = \frac{value_{meas.} - Median_{Controls}}{MAD_{Controls}}$$

The phenotypic compound profile is then determined as the list of Z-scores of all features for one compound.

In addition to the phenotypic profile, an induction value was determined for each compound as the fraction of significantly changed features, in percent:

$$Induction\ [\%] = \frac{number\ of\ features\ with\ abs.\ \ values\ >\ 3}{total\ number\ of\ features}$$

Similarities of phenotypic profiles (termed *Biosimilarity*) were calculated from the correlation distances (CD) between two profiles (https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.correlation.html) [44]:

$$CD = 1 - \frac{(u - \bar{u}) \cdot (v - \bar{v})}{\|(u - \bar{u})\|_2 \|(v - \bar{v})\|_2}$$

where $\bar{x}$ is the mean of the elements of x, $x \cdot y$ is the dot product of x and y, and $\|x\|_2$ is the Euclidean norm of x:

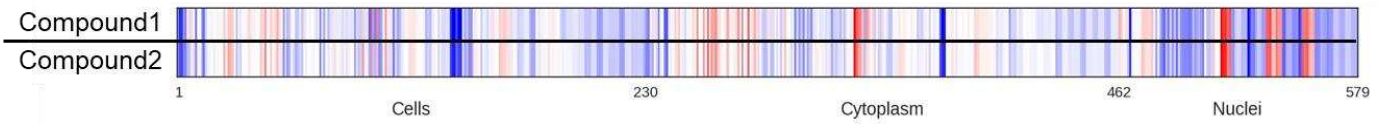
$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

The Biosimilarity is then defined as:

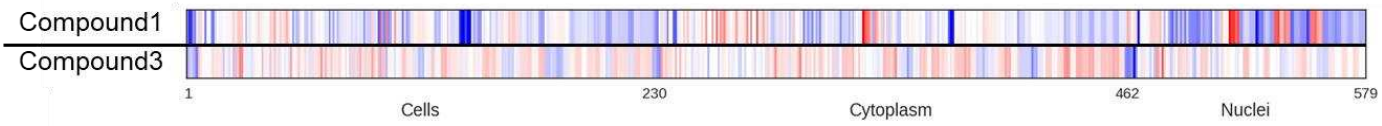
$$Biosimilarity = 1 - CD$$

Biosimilarity values smaller than 0 are set to 0 and the Biosimilarity is expressed in percent (0-100).

An example for two compounds with highly similar profiles (96% Biosimilarity):



An example for two compounds with low similarity profiles (0% Biosimilarity):



Each colored band represents one Z-score of a feature.

Immunocytochemistry

5,000 U2OS cells were seeded per well in a black 96-well plate with clear bottom and incubated overnight. Cells were treated with compounds or DMSO as a control for 24 hours. Cells were then fixed using 3.7 % paraformaldehyde in phosphate-buffered saline (PBS) and permeabilized with 0.1% Triton X-100 (in PBS) prior to blocking nonspecific binding with 2% bovine serum umin (BSA in PBS). Staining with DAPI to visualize DNA and anti-tubulin-FITC antibody was performed in blocking buffer overnight at 4 C°. Images were acquired using Observer Z1 (Carl Zeiss, Germany) using 40x objective (LD Plan-Neofluar).

***In vitro* HDAC Activity Assay**

HDAC activity was assayed *in vitro* using the Histone deacetylase activity assay kit (#566328, Sigma Aldrich) according to the manufacturer's protocol. HeLa Cell Nuclear Extract (Kit

component No. KP31841 in 0.1 M KCl, 20 mM HEPES/NaOH pH 7.9, 20% (v/v) glycerol, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF) was diluted 1:1000 in assay buffer (50 mM Tris/HCl, pH 8.0, 137 mM NaCl, 2.7 mM KCl, 1 mM $MgCl_2$ and 0.01% NP40S) and acts as a source for HDAC enzyme activity. A white 384- well plate (Corning #4513) with 4 µl of HeLa Cell Nuclear Extract per well was incubated with different concnetrations of the compounds for 30 min at room temperature. An equal volume of substrate (Kit Component No. KP31842, 5 µM in assay buffer), which contains an acetylated lysine side chain, was added to the mixture to start the reaction. After 30 min incubation at room temperature, 8 µl of developing solution (1/200) was added to the samples to convert the deacetylated substrate to a fluorophore. Fluorescence intensity was measured at ex/em 360/460 nm using a Paradigm plate reader (Molecular Devices).

***In Situ* Histone Deacetylase (HDAC) Activity Fluorometric Assay**

The assay was performed using the *In Situ* HDAC Activity Fluorometric Assay Kit (#EPI003, Sigma Aldrich) according to the manufacturer's instructions. 20,000 U2OS cells per well were seeded into a black 96-well plate (clear bottom) and incubated overnight at 37°C and 5% CO_2. The next day, cells were incubated for 2 h with the compounds or DMSO as a control together with a cell-permeable HDAC substrate, which contains an acetylated lysine side chain and a fluorophore that is quenched when bound to the substrate. Developer solution is added to lyse the cells and cleave the deacetylated HDAC substrate to release the fluorophore. After incubation for 30 min, fluorescence intensity was measured at ex/em 368/442 nm using the Tecan Spark plate reader.

Click-it HPG Alexa Fluor Protein Synthesis Assay Kit

The assay was performed using the Click-iT™ HPG Alexa Fluor™ 488 Protein Synthesis Assay Kit (# C10428, Sigma Aldrich) according to the manufacturer's instructions. 5,000 HeLa cells per well were seeded into a black 96-well plate (clear bottom) and incubated overnight at 37°C and 5% CO_2.The next day, cells were treated with compound or DMSO as a control for 24 h. For degrading already synthesized protein cells were washed three times with phosphate-buffered saline (PBS) and incubated in methionine-free medium (RPMI 1640) supplemented with 10 % FBS for 30 min in the presence of compound or DMSO as a control. L-homopropargylglycine incorporation was started by adding Click-iT® HPG reagent (50 µM) in methionine-free medium (RPMI 1640) supplemented with 10 % FBS for 45 min. Fixation, permeabilization and Click-iT® HPG detection were performed according to the manufacturer's protocol. HCS NuclearMask™ BlueStain was used to visualized DNA. Axiovert 200M microscope (Carl Zeiss, Germany) equipped with 10x objective was used to quantify Alexa Fluor® 488 using the software MetaMorph 7. Protein synthesis was assessed by determining signal intensity considering the total cell count as determined using the DNA.

1-Step Human Coupled IVT Kit

The assay was performed using the 1-Step Human Coupled IVT Kit (# 88882, Thermo Fisher Scientific Inc.) according to the manufacturer's instructions using 'turbo-type' green fluorescent protein (tGFP) mRNA (# 88880 Thermo Fisher Scientific Inc.) as a template for translation. 5 µl HeLa lysate was preincubated with 1 µl of accessory protein for 10 min at room temperature in PCR plates (HSP9601, Bio-Rad). Reaction was started by adding 2 µl reaction mix, 1.2 µl tGFP mRNA (0.75 mg/µl) and 0.8 µl compound or DMSO as a control. Expression of tGFP was monitored for 5 h at 30°C using the CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories, Inc.).

In vitro Aurora kinase assay

Aurora kinase assays were performed using the SelectScreen services (Thermo) by means of a Z'-Lyte assay. Briefly, upon phosphorylation of a synthetic FRET peptide by the kinases, a development reagent A is added that contains a protease, which specifically cleaves non-phosphorylated peptides. The FRET peptide is labelled with coumarine and fluorescein, which make up a FRET pair. The cleavage disrupts FRET between the donor and the acceptor (https://www.thermofisher.com/de/de/home/industrial/pharma-biopharma/drug-discovery-development/target-and-lead-identification-and-validation/kinasebiology/kinase-activity-assays/z-lyte.html).

For Aurora A, a 10 µL kinase reaction consisted of 0.75 - 3 ng Aurora A and 2 µM Ser/Thr 01 in 50 mM HEPES pH 7.5, 0.01% BRIJ-35, 10 mM $MgCl_2$, 1 mM EGTA. ATP concentration of Km ATP = 10 µM was used. After the 1 h of incubation, 5 µL of a 1:4096 dilution of Development Reagent A was added to the reaction.

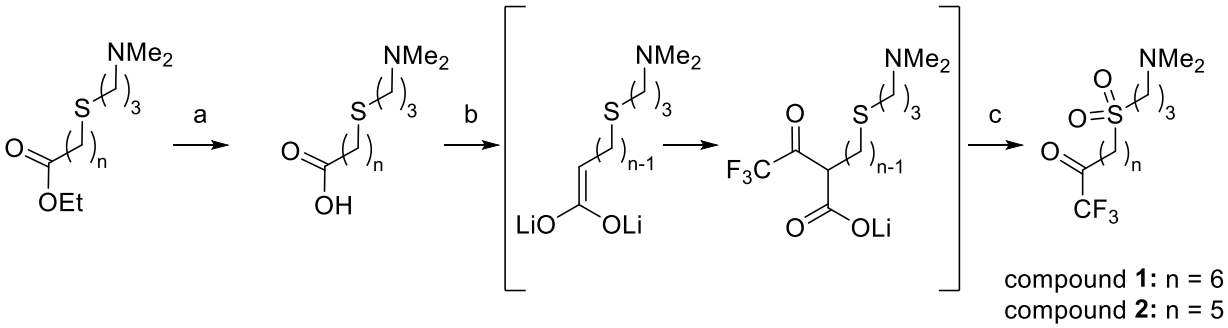
For Aurora B, a 10 µL kinase reaction consisted of 3.5 - 18 ng Aurora B and 2 µM Ser/Thr 01 in 50 mM HEPES pH 7.5, 0.01% BRIJ-35, 10 mM $MgCl_2$, 1 mM EGTA. ATP concentration of Km ATP = 81 µM was used. After the 1 h incubation, 5 µL of a 1:4096 dilution of Development Reagent A was added to the reaction.

For Aurora C, a 10 µL kinase reaction consisted of 2 - 20 ng Aurora C and 2 µM Ser/Thr 19 in 50 mM HEPES pH 7.5, 0.01% BRIJ-35, 10 mM $MgCl_2$, 1 mM EGTA. ATP concentration of Km ATP = 26 µM was used. After the 1 h incubation, 5 µL of a 1:256 dilution of Development Reagent A was added to the reaction.

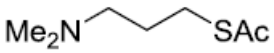
Quantification and statistical analysis

Data were either representative of three independent experiments or expressed as mean ± SD. All statistical details of the conducted experiments can be found in the respective figure caption. N: number of technical replicates, n: number of biological replicates.

Compound synthesis



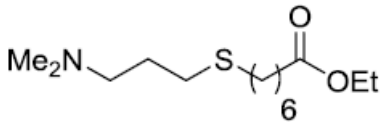
Synthesis of compounds **1** and **2**. Reagents and conditions: a) NaOH (1.0 eqv.), H_2O:EtOH, rt, 2h. b) Diisopropylamine (3.5 eqv.), n-Buli (3.4 eqv.), -78°C, carboxylic acid (1.0 eqv.), THF, rt (4h), CF_3COOEt (3.0 eqv.,), -78°C (15 min), 6N HCl, 45-53 % yield (over two steps). c) Oxone® (3.0 eqv.), MeOH/H_2O (3/2), overnight, rt, 24h, 35% yield.



***N,N*-(Dimethylamino)propyl thioacetate (S1)**

Compound **S1** was synthesized according to Hedberg et al. (Hedberg et al., 2011)

^{1}H NMR (400 MHz, $CDCl_3$): δ 2.90 (t, *J* = 7.2, 2H, -C*H*$_2$S), 2.31 (t, *J* = 7.2, 2H, -C*H*$_2$N), 2.32 (s, 3H, COC*H*$_3$), 2.21 (s, 6H, N(C*H*$_3$)2), 1.74 (p, *J* = 7.2, 2H). **^{13}C NMR** (125 MHz, $CDCl_3$): δ 195.7, 58.5, 45.5, 30.7, 27.8, 27.1. HRMS (ESI) calc. for $C_7H_{15}NOS$ $[M+H]^+$ 162.0947, found 162.0946. GCMS found 161 for $[M^+.]$.

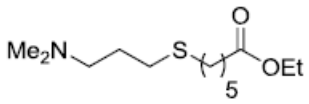


Ethyl 7-(3-(dimethylamino)propylthio)heptanoate (**S2**).

To a two-neck round bottom flask containing a stir bar and a solution of thioacetate **S1** (52 mmol, 1.0 eqv.) in anhydrous ethanol (80 ml) was added Cs_2CO_3 (58 mmol, 1.1 eqv.). The flask was equipped with a reflux condenser, and the reaction was flushed with Ar and kept under a positive pressure of Ar for the remainder of the reaction. The reaction was stirred at reflux until TLC analysis indicated that all of thioacetate **S1** had been consumed with the formation of the more polar thiol intermediate (EtOAc/CyHex 1/2). After this time, the reaction was cooled to 0 °C and a solution of ethyl 7-iodoheptanoate (58 mmol, 1.1 eqv.) in anhydrous ethanol (20 ml) was added dropwise. The reaction was then stirred at 22 °C for 1 h and then heated to 40 °C until the thiol intermediate was consumed (about 12 hours, TLC conditions: EtOAc/CyHex 1/2). After this time, the reaction was concentrated and EtOAc/H_2O 1/2 (400ml) was added to the crude residue and stirred until a solution was formed. The phases were separated and the aq. layer was washed twice with EtOAc (100 ml). The organic layers were combined, washed with water (150 ml), brine (270 ml), dried over $MgSO_4$, and concentrated to afford compound **S2** without any further purification.

Yield = 75% (slight yellow oil). R*f* = 0.56 (AcOH/EtOAc/MeOH/H_2O 3/3/3/2). IR: 1734 cm-1 (ester, C=O stretch). **^{1}H NMR** (400 MHz, $CDCl_3$): δ 4.10 (q, *J* = 7.2, 2H, -COC*H*$_2$CH$_3$), 2.52 (t, *J* = 7.2, 2H, -C*H*$_2$S),2.49 (t, *J* = 7.4, 2H, -C*H*$_2$S), 2.39 (t, *J* = 7.6, 2H, -C*H*$_2$COOEt), 2.27 (t, *J* = 7.6, 2H, -C*H*$_2$NMe$_2$), 2.25 (s,6H, N(C*H*$_3$)$_2$), 1.76 – 1.66 (m, 2H), 1.65 – 1.50 (m, 4H), 1.41–1.20 (m, 4H), 1.22 (t, *J* = 7.2, 3H, -COCH$_2$C*H*$_3$). **^{13}C NMR** (125 MHz, $CDCl_3$): δ 173.6, 60.1,

58.7, 45.4, 34.2, 32.1, 30.0, 29.4, 28.7, 28.5, 27.7, 24.8, 14.2. HRMS (ESI) calc. for $C_{14}H_{29}NO_2S$ $[M+H]^+$ 276.1992, found 276.1992. GCMS found 275 for $[M^+.]$.

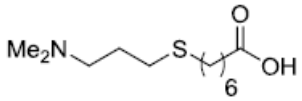


Ethyl 6-(3-(dimethylamino)propylthio)hexanoate (**S3**).

To a two-neck round bottom flask containing a stir bar and a solution of thioacetate **S1** (72 mmol, 1.0 eqv.) in anhydrous ethanol (108 ml) was added Cs_2CO_3 (78 mmol, 1.1 eqv.). The flask was equipped with a reflux condenser, and the reaction was flushed with Ar and kept under a positive pressure of Ar for the remainder of the reaction. The reaction was stirred at reflux until TLC analysis indicated that all of thioacetate **S1** had been consumed with the formation of the more polar thiol intermediate (EtOAc/CyHex 1/2). After this time, the reaction was cooled to 0 °C and a solution of ethyl 7-iodohexanoate (78 mmol, 1.1 eqv.) in anhydrous ethanol (27 ml) was added dropwise. The reaction was then stirred at 22 °C for 1 h and then heated to 40 °C until the thiol intermediate was consumed (about 12 hours, TLC conditions: EtOAc/CyHex 1/2). After this time, the reaction was concentrated and EtOAc/H_2O 1/2 (540 ml) was added to the crude residue and stirred until a solution was formed. The phases were separated and the aq. layer was washed twice with EtOAc (150 ml). The organic layers were combined, washed with water (150 ml), brine (270 ml), dried over $MgSO_4$, and concentrated to afford compound **S3** without any further purification.

Yield = 77% (slight yellow oil). R*f* = 0.56 (AcOH/EtOAc/MeOH/H_2O 3/3/3/2). IR: 1733 cm 1 (ester, C=O stretch). **^{1}H NMR** (400 MHz, $CDCl_3$): δ 4.01 (q, *J* = 7.2, 2H, -COC*H*$_2$CH$_3$), 2.40

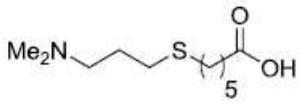
(t, *J* = 7.4, 2H, -C*H*$_2$S), 2.39 (t, *J* = 7.4, 2H, -C*H*$_2$S), 2.21 (t, *J* = 7.2, 2H, -C*H*$_2$COOEt), 2.17 (t, *J* = 7.4, 2H, -C*H*$_2$NMe$_2$), 2.09 (s,6H, N(C*H*$_3$)$_2$), 1.65 – 1.57 (m, 2H), 1.55 – 1.43 (m, 4H), 1.34 – 1.24 (m, 2H), 1.13 (t, *J* = 7.2, 3H, -COCH$_2$C*H*$_3$). **^{13}C NMR** (125 MHz, $CDCl_3$): δ 173.2, 59.9, 58.4, 45.2, 33.9, 31.7, 29.7, 29.0, 28.1, 27.5, 24.3, 14.0. HRMS (ESI) calc. for $C_{13}H_{27}NO_2S$ $[M+H]^+$ 262.1835, found 262.1837. GCMS found 261 for $[M^+.]$.



7-(3-(Dimethylamino)propylthio)heptanoic acid (**S4**).

To a 100 ml round bottomed flask equipped with a stir bar and a solution of ethyl ester **S2** (5.6 mml, 1 eqv. in 7 ml EtOH) was added 2 M NaOH (5.6 mmol, 1 eqv.). The reaction was stirred at 22 °C for two hours. Upon completion as determined by TLC (MeOH/DCM 2/8), the reaction was extracted with Et_2O and the organic layer was discarded. The aq. layer was then acidified using conc. HCl to approximately pH 3. The aq. phase was then lyophilized, resulting in compound **S4** without the need for further purification. The product contains some NaCl salts.

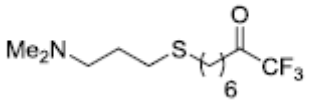
Yellow solid. R*f* = 0.71 (AcOH/EtOAc/MeOH/H_2O 3/3/3/2). IR: 1722 cm-1 (C=O). **1H NMR** (400 MHz, $CDCl_3$): δ 3.22 – 3.16 (m, 2H, -C*H*$_2$COOH), 2.88 (s, 6H, N(C*H*$_3$)2), 2.62 (t, *J* = 7.0, 2H, -C*H*$_2$S), 2.56 (t, *J* = 7.4, 2H, -C*H*$_2$S), 2.37 (t, *J* = 6.8, 2H, -C*H*$_2$NMe$_2$), 2.25 – 2.23 (m, 2H), 1.71 - 1.59 (m, 4H), 1.51 -1.36 (m, 4H). **^{13}C NMR** (125 MHz, $CDCl_3$): δ 176.9, 57.2, 43.4 (2C), 33.7, 30.8, 28.6, 28.2, 27.5, 27.2, 24.1, 23.9. HRMS (ESI) calc. for $C_{12}H_{25}NO_2S$ $[M+H]^+$ 248.1678, found 248.1679. LC-MS (ESI) found 248.04 for $[M+H]^+$.



6-(3-(Dimethylamino)propylthio)hexanoic acid (**S5**).

To a 100 ml round bottomed flask equipped with a stir bar and a solution of ethyl ester **S3** (5.6 mml, 1 eqv. in 7 ml EtOH) was added 2 M NaOH (5.6 mmol, 1 eqv.). The reaction was stirred at 22 °C for two hours. Upon completion as determined by TLC (MeOH/DCM 2/8), the reaction was extracted with Et_2O and the organic layer was discarded. The aq. layer was then acidified using conc. HCl to approximately pH 3. The aq. phase was then lyophilized, resulting in compound **S5** without the need for further purification. The product contains some NaCl salts.

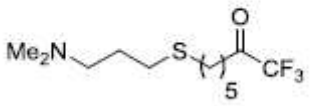
Yellow solid. R*f* =0.61 (AcOH/EtOAc/MeOH/H_2O 3/3/3/2). IR: 1729 cm-1 (C=O). **^{1}H NMR** (400 MHz, $CDCl_3$): δ 3.19 –3.13 (m, 2H, -CH_2COOH), 2.84 (s, 6H, N(CH_3)$_2$), 2.61 (t, *J* = 6.8, 2H, -CH_2S), 2.55 (t, *J* = 7.0, 2H, -CH_2S), 2.36 (t, *J* = 7.0, 2H, -CH_2NMe$_2$), 2.19 – 2.10 (m, 2H), 1.73 – 1.57 (m, 4H), 1.52 – 1.44 (m, 2H). **^{13}C NMR** (125 MHz, $CDCl_3$): δ 176.9, 57.1, 43.2 (2C), 33.7, 31.8, 28.7, 28.7, 27.7, 24.2, 24.0. HRMS (ESI) calc. for $C_{11}H_{23}NO_2S$ $[M+H]^+$ 234.1522, found 234.1522. LC-MS (ESI) found 234.05 for $[M+H]^+$.



8-(3-(Dimethylamino)propylthio)-1,1,1-trifluorooctan-2-one (**1**).

Compound **S4** (0.63 mmol, 1.0 eqv.) was suspended in anhydrous THF (1.5 ml) and cooled to -20 °C. In a separate reaction flask, a solution of LDA was prepared by adding n-BuLi (2.5 M in hexane, 2.2 mmol, 3.4 eqv.) to a solution of freshly distilled DIPA (2.26 mmol, 3.5 eqv. in THF (1.2 ml)) at -78 °C. The solution of LDA was added dropwise over a period of 10 min to the solution of **S4** and was warmed to 22 °C and reacted for 4 hours. In a third flask, the enediolate solution was added dropwise to a solution of CF_3CO_2Et (1.94 mmol, 3.0 eqv. in THF (0.7 ml)) at -78 °C. After stirring the reaction at -78 °C for 15 min, the reaction was quenched by adding 6M HCl (1.2 ml) dropwise. The reaction was diluted with EtOAc and the organic layer was isolated, dried ($MgSO_4$), and concentrated. The residue was purified by flash chromatography (1-8% MeOH in DCM) to afford compound **1**. Yield = 45% (over 2 steps, starting from ester **S2**, as yellow oil).

R_f = 0.73 (AcOH/EtOAc/MeOH/H_2O 3/3/3/2). IR: 1762 cm^{-1} (C=O). **^{1}H NMR** (400 MHz, $CDCl_3$): δ 2.70 (t, *J* = 7.2, 2H, -C*H*$_2$COCF$_3$), 2.53 (t, *J* = 7.4, 2H, -C*H*$_2$S), 2.51 (t, *J* = 7.4, 2H, -C*H*$_2$S), 2.35 (t, *J* = 7.4, 2H, - C*H*$_2$NMe$_2$), 2.23 (s, 6H, N(C*H*$_3$)$_2$), 1.79 – 1.53 (m, 6H), 1.46 - 1.26 (m, 4H). **^{13}C NMR** (125 MHz, $CDCl_3$): δ 191.4 (q, CO, J_2 (C-F) = 34.6Hz), 115.5 (q, CF_3, J_1 (C-F)= 290 Hz), 58.5, 45.2 (2C), 36.2, 31.9, 29.9, 29.2, 28.3, 28.2, 27.3, 22.1. **^{19}F NMR** (377 MHz, $CDCl_3$): δ - 79.7 (s, CF_3). HRMS (ESI) calc. for $C_{13}H_{24}F_3NOS$ $[M+H]_+$ 300.1603, found 300.1606. GCMS found 299 for $[M^+.]$.



7-(3-(Dimethylamino)propylthio)-1,1,1-trifluoroheptan-2-one (**2**).

Compound **S5** (0.63 mmol, 1.0 eqv.) was suspended in anhydrous THF (1.5 ml) and cooled to -20 °C. In a separate reaction flask, a solution of LDA was prepared by adding n-BuLi (2.5 M in hexane, 2.2 mmol, 3.4 eqv.) to a solution of freshly distilled DIPA (2.26 mmol, 3.5 eqv. in THF (1.2 ml)) at -78 °C. The solution of LDA was added dropwise over a period of 10 min to the solution of **S4** and was warmed to 22 °C and reacted for 4 hours. In a third flask, the enediolate solution was added dropwise to a solution of CF_3CO_2Et (1.94 mmol, 3.0 eqv. in THF (0.7 ml)) at -78 °C. After stirring the reaction at -78 °C for 15 min, the reaction was quenched by adding 6M HCl (1.2 ml) dropwise. The reaction was diluted with EtOAc and the organic layer was isolated, dried ($MgSO_4$), and concentrated. The residue was purified by flash chromatography (1-8% MeOH in DCM) to afford compound **2**. Yield = 52% (over 2 steps, starting from ester **S3**, as yellow oil).

R_f = 0.67 (AcOH/EtOAc/MeOH/H_2O 3/3/3/2). IR: 1762 cm^{-1} (C=O). **^{1}H NMR** (400 MHz, $CDCl_3$): δ 2.72 (t, *J* = 7.4, 2H, -C*H*$_2$COCF$_3$), 2.54 (t, *J* = 7.2, 2H, -C*H*$_2$S), 2.52 (t, *J* = 7.2, 2H, -C*H*$_2$S), 2.37 (t, *J* = 7.2, 2H, - C*H*$_2$NMe$_2$), 2.24 (s, 6H, N(C*H*$_3$)$_2$), 1.81 – 1.57 (m, 6H), 1.49 - 1.39 (m, 2H). **^{13}C NMR** (125 MHz, $CDCl_3$): δ 191.3 (q, CO, J_2 (C-F) = 34.7 Hz), 115.48 (q, CF_3, J_1 (C-F)= 290 Hz), 58.5, 45.2 (2C), 36.2, 31.7, 29.9, 29.1, 27.8, 27.4, 21.9. **^{19}F NMR** (377 MHz, $CDCl_3$): δ - 79.7 (s, CF_3). HRMS (ESI) calc. For $C_{12}H_{22}F_3NOS$ $[M+H]^+$ 286.1447, found 286.1449. GCMS found 285 for $[M^{+\cdot}]$.

Compound **3** was reported previously. All further compounds are reference compound and were obtained from commercial sources.

Supplemental Information

Figure S1-S7

Table S2

Compound synthesis and characterization

Table S1_Cluster defining compounds.xlsx: Cluster-defining compounds and similarity to the twelve clusters at the indicated concentrations. The table also includes the logP and pKa values as well as the cell count. Related to Figure 2.

References

1. Ziegler, S., Sievers, S., and Waldmann, H. (2021). Morphological profiling of small molecules. Cell Chem Biol *28*, 300-319. 10.1016/j.chembiol.2021.02.012.
2. Bray, M.A., Singh, S., Han, H., Davis, C.T., Borgeson, B., Hartland, C., Kost-Alimova, M., Gustafsdottir, S.M., Gibson, C.C., and Carpenter, A.E. (2016). Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc *11*, 1757-1774. 10.1038/nprot.2016.105.
3. Gustafsdottir, S.M., Ljosa, V., Sokolnicki, K.L., Wilson, J.A., Walpita, D., Kemp, M.M., Seiler, K.P., Carrel, H.A., Golub, T.R., Schreiber, S.L., et al. (2013). Multiplex Cytological Profiling Assay to Measure Diverse Cellular States. Plos One *8*, https://doi.org/10.1371/journal.pone.0080999. ARTN e80999
10.1371/journal.pone.0080999.
4. Akbarzadeh, M., Deipenwisch, I., Schoelermann, B., Pahl, A., Sievers, S., Ziegler, S., and Waldmann, H. (2022). Morphological profiling by means of the Cell Painting assay enables identification of tubulin-targeting compounds. Cell Chem Biol *29(6)*, 1053-1064. 10.1016/j.chembiol.2021.12.009.
5. Schneidewind, T., Brause, A., Pahl, A., Burhop, A., Mejuch, T., Sievers, S., Waldmann, H., and Ziegler, S. (2020). Morphological Profiling Identifies a Common Mode of Action for Small Molecules with Different Targets. Chembiochem *21*, 3197-3207. 10.1002/cbic.202000381.
6. Schneidewind, T., Brause, A., Scholermann, B., Sievers, S., Pahl, A., Sankar, M.G., Winzker, M., Janning, P., Kumar, K., Ziegler, S., and Waldmann, H. (2021). Combined morphological and proteome profiling reveals target-independent impairment of cholesterol homeostasis. Cell Chem Biol *28*, 1780-1794 e1785. 10.1016/j.chembiol.2021.06.003.
7. Christoforow, A., Wilke, J., Binici, A., Pahl, A., Ostermann, C., Sievers, S., and Waldmann, H. (2019). Design, Synthesis, and Phenotypic Profiling of Pyrano-Furo-Pyridone Pseudo Natural Products. Angew Chem Int Ed Engl *58*, 14715-14723. 10.1002/anie.201907853.
8. Schoelermann, B., Bonowski, J., Grigalunas, M., Burhop, A., Xie, Y., Hoock, J.G.F., Liu, J., Dow, M., Nelson, A., Nowak, C., et al. (2022). Identification of Dihydroorotate Dehydrogenase Inhibitors Using the Cell Painting Assay. Chembiochem *23*, e202200475. 10.1002/cbic.202200475.
9. Akbarzadeh, M., Deipenwisch, I., Schoelermann, B., Pahl, A., Sievers, S., Ziegler, S., and Waldmann, H. (2021). Morphological profiling by means of the Cell Painting assay enables identification of tubulin-targeting compounds. Cell Chem Biol. 10.1016/j.chembiol.2021.12.009.
10. Soltoff, S.P. (2004). Evidence that tyrphostins AG10 and AG18 are mitochondrial uncouplers that alter phosphorylation-dependent cell signaling. J Biol Chem *279*, 10910-10918. 10.1074/jbc.M305396200.
11. Childress, E.S., Alexopoulos, S.J., Hoehn, K.L., and Santos, W.L. (2018). Small Molecule Mitochondrial Uncouplers and Their Therapeutic Potential. J Med Chem *61*, 4641-4655. 10.1021/acs.jmedchem.7b01182.
12. Weatherly, L.M., Shim, J., Hashmi, H.N., Kennedy, R.H., Hess, S.T., and Gosse, J.A. (2016). Antimicrobial agent triclosan is a proton ionophore uncoupler of mitochondria in living rat and human mast cells and in primary human keratinocytes. J Appl Toxicol *36*, 777-789. 10.1002/jat.3209.

13. Lee, I.Y., Gruber, T.D., Samuels, A., Yun, M., Nam, B., Kang, M., Crowley, K., Winterroth, B., Boshoff, H.I., and Barry, C.E., 3rd (2013). Structure-activity relationships of antitubercular salicylanilides consistent with disruption of the proton gradient via proton shuttling. Bioorg Med Chem *21*, 114-126. 10.1016/j.bmc.2012.10.056.
14. Yoon, J.R., Whipple, R.A., Balzer, E.M., Cho, E.H., Matrone, M.A., Peckham, M., and Martin, S.S. (2011). Local anesthetics inhibit kinesin motility and microtentacle protrusions in human epithelial and breast tumor cells. Breast Cancer Res Treat *129*, 691-701. 10.1007/s10549-010-1239-7.
15. Ma, X.J., Xu, G., Li, Z.J., Chen, F., Wu, D., Miao, J.N., Zhan, Y., and Fan, Y. (2019). HDAC-selective Inhibitor Cay10603 Has Single Anti-tumour Effect in Burkitt's Lymphoma Cells by Impeding the Cell Cycle. Curr Med Sci *39*, 228-236. 10.1007/s11596-019-2024-4.
16. Ali, N., Allam, H., Bader, T., May, R., Basalingappa, K.M., Berry, W.L., Chandrakesan, P., Qu, D., Weygant, N., Bronze, M.S., et al. (2013). Fluvastatin interferes with hepatitis C virus replication via microtubule bundling and a doublecortin-like kinase-mediated mechanism. Plos One *8*, e80304. 10.1371/journal.pone.0080304.
17. Moret, N., Clark, N.A., Hafner, M., Wang, Y., Lounkine, E., Medvedovic, M., Wang, J.H., Gray, N., Jenkins, J., and Sorger, P.K. (2019). Cheminformatics Tools for Analyzing and Designing Optimized Small-Molecule Collections and Libraries. Cell Chemical Biology *26*, 765-777. 10.1016/j.chembiol.2019.02.018.
18. McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. doi.org/10.48550/arXiv.1802.03426.
19. Nunes, M.J., Moutinho, M., Gama, M.J., Rodrigues, C.M., and Rodrigues, E. (2013). Histone deacetylase inhibition decreases cholesterol levels in neuronal cells by modulating key genes in cholesterol synthesis, uptake and efflux. Plos One *8*, e53394. 10.1371/journal.pone.0053394.
20. Tonini, C., Colardo, M., Colella, B., Di Bartolomeo, S., Berardinelli, F., Caretti, G., Pallottini, V., and Segatto, M. (2020). Inhibition of Bromodomain and Extraterminal Domain (BET) Proteins by JQ1 Unravels a Novel Epigenetic Modulation to Control Lipid Homeostasis. Int J Mol Sci *21*. 10.3390/ijms21041297.
21. Frey, R.R., Wada, C.K., Garland, R.B., Curtin, M.L., Michaelides, M.R., Li, J.L., Pease, L.J., Glaser, K.B., Marcotte, P.A., Bouska, J.J., et al. (2002). Trifluoromethyl ketones as inhibitors of histone deacetylase. Bioorg Med Chem Lett *12*, 3443-3447. Pii S0960-894x(02)00754-0
Doi 10.1016/S0960-894x(02)00754-0.
22. Dow, M., Marchetti, F., Abrahams, K.A., Vaz, L., Besra, G.S., Warriner, S., and Nelson, A. (2017). Modular Synthesis of Diverse Natural Product-Like Macrocycles: Discovery of Hits with Antimycobacterial Activity. Chemistry *23*, 7207-7211. 10.1002/chem.201701150.
23. Konagaya, S., and Iwata, H. (2019). Chemically defined conditions for long-term maintenance of pancreatic progenitors derived from human induced pluripotent stem cells. Sci Rep *9*, 640. 10.1038/s41598-018-36606-7.
24. Horbelt, D., Boergermann, J.H., Chaikuad, A., Alfano, I., Williams, E., Lukonin, I., Timmel, T., Bullock, A.N., and Knaus, P. (2015). Small molecules dorsomorphin and LDN-193189 inhibit myostatin/GDF8 signaling and promote functional myoblast differentiation. J Biol Chem *290*, 3390-3404. 10.1074/jbc.M114.604397.
25. Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X.D., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017). A Next Generation

Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell *171*, 1437-1452. 10.1016/j.cell.2017.10.049.
26. Way, G.P., Natoli, T., Adeboye, A., Litichevskiy, L., Yang, A., Lu, X., Caicedo, J.C., Cimini, B.A., Karhohs, K., Logan, D.J., et al. (2022). Morphology and gene expression profiling provide complementary information for mapping cell state. Cell Syst *13*, 911-923 e919. 10.1016/j.cels.2022.10.001.
27. Wawer, M.J., Li, K., Gustafsdottir, S.M., Ljosa, V., Bodycombe, N.E., Marton, M.A., Sokolnicki, K.L., Bray, M.A., Kemp, M.M., Winchester, E., et al. (2014). Toward performance-diverse small-molecule libraries for cell-based phenotypic screening using multiplexed high-dimensional profiling. Proc Natl Acad Sci U S A *111*, 10911-10916. 10.1073/pnas.1410933111.
28. Seal, S., Carreras-Puigvert, J., Trapotsi, M.A., Yang, H.B., Spjuth, O., and Bender, A. (2022). Integrating cell morphology with gene expression and chemical structure to aid mitochondrial toxicity detection. Commun Biol *5*. ARTN 858
10.1038/s42003-022-03763-5.
29. Nadanaciva, S., Lu, S., Gebhard, D.F., Jessen, B.A., Pennie, W.D., and Will, Y. (2011). A high content screening assay for identifying lysosomotropic compounds. Toxicol In Vitro *25*, 715-723. 10.1016/j.tiv.2010.12.010.
30. Settembre, C., Zoncu, R., Medina, D.L., Vetrini, F., Erdin, S., Erdin, S., Huynh, T., Ferron, M., Karsenty, G., Vellard, M.C., et al. (2012). A lysosome-to-nucleus signalling mechanism senses and regulates the lysosome via mTOR and TFEB. EMBO J *31*, 1095-1108. 10.1038/emboj.2012.32.
31. Bramson, H.N., Corona, J., Davis, S.T., Dickerson, S.H., Edelstein, M., Frye, S.V., Gampe, R.T., Jr., Harris, P.A., Hassell, A., Holmes, W.D., et al. (2001). Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): design, synthesis, enzymatic activities, and X-ray crystallographic analysis. J Med Chem *44*, 4339-4358. 10.1021/jm010117d.
32. Banerjee, S., Buhrlage, S.J., Huang, H.T., Deng, X., Zhou, W., Wang, J., Traynor, R., Prescott, A.R., Alessi, D.R., and Gray, N.S. (2014). Characterization of WZ4003 and HTH-01-015 as selective inhibitors of the LKB1-tumour-suppressor-activated NUAK kinases. Biochem J *457*, 215-225. 10.1042/BJ20131152.
33. Srivatsan, S.R., McFaline-Figueroa, J.L., Ramani, V., Saunders, L., Cao, J.Y., Packer, J., Pliner, H.A., Jackson, D.L., Daza, R.M., Christiansen, L., et al. (2020). Massively multiplex chemical transcriptomics at single-cell resolution. Science *367*, 45-51. 10.1126/science.aax6234.
34. Cox, M.J., Jaensch, S., Vande Waeter, J., Cougnaud, L., Seynaeve, D., Benalla, S., Koo, S.J., VanDen Wyngaert, I., Neefs, J.M., Malkov, D., et al. (2020). Tales of 1,008 small molecules: phenomic profiling through live-cell imaging in a panel of reporter cell lines. Sci Rep-Uk *10*. 10.1038/s41598-020-69354-8.
35. Woehrmann, M.H., Bray, W.M., Durbin, J.K., Nisam, S.C., Michael, A.K., Glassey, E., Stuart, J.M., and Lokey, R.S. (2013). Large-scale cytological profiling for functional analysis of bioactive compounds. Mol Biosyst *9*, 2604-2617. 10.1039/c3mb70245f.
36. Breinig, M., Klein, F.A., Huber, W., and Boutros, M. (2015). A chemical-genetic interaction map of small molecules using high-throughput imaging in cancer cells. Mol Syst Biol *11*, 102-112. DOI 10.15252/msb.20156400.
37. Flegel, J., Shaaban, S., Jia, Z.J., Schulte, B., Lian, Y., Krzyzanowski, A., Metz, M., Schneidewind, T., Wesseler, F., Flegel, A., et al. (2022). The Highly Potent AhR Agonist Picoberin Modulates Hh-Dependent Osteoblast Differentiation. J Med Chem *65*, 16268-16289. 10.1021/acs.jmedchem.2c00956.
38. Vorontsova, J.E., Akishina, A.A., Cherezov, R.O., and Simonova, O.B. (2022). A new insight into the aryl hydrocarbon receptor/cytochrome 450 signaling pathway in MG63, HOS, SAOS2, and U2OS cell lines. Biochimie. 10.1016/j.biochi.2022.10.018.

39. Budin, C., Besselink, H., van Vugt-Lussenburg, B.M.A., Man, H.Y., van der Burg, B., and Brouwer, A. (2021). Induction of AhR transactivation by PBDD/Fs and PCDD/Fs using a novel human-relevant, high-throughput DR(human) CALUX reporter gene assay. Chemosphere *263*, 128086. 10.1016/j.chemosphere.2020.128086.
40. Davies, C., Dotsch, L., Ciulla, M.G., Hennes, E., Yoshida, K., Gasper, R., Scheel, R., Sievers, S., Strohmann, C., Kumar, K., et al. (2022). Identification of a Novel Pseudo-Natural Product Type IV IDO1 Inhibitor Chemotype. Angew Chem Int Ed Engl *61*, e202209374. 10.1002/anie.202209374.
41. www.proteinatlas.org/ Human Protein Atlas. https://www.proteinatlas.org/.
42. Vincent, F., Loria, P.M., Weston, A.D., Steppan, C.M., Doyonnas, R., Wang, Y.M., Rockwell, K.L., and Peakman, M.C. (2020). Hit Triage and Validation in Phenotypic Screening: Considerations and Strategies. Cell Chem Biol *27*, 1332-1346. 10.1016/j.chembiol.2020.08.009.
43. Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol *7*, R100. 10.1186/gb-2006-7-10-r100.
44. Grigalunas, M., Burhop, A., Zinken, S., Pahl, A., Gally, J.M., Wild, N., Mantel, Y., Sievers, S., Foley, D.J., Scheel, R., et al. (2021). Natural product fragment combination to performance-diverse pseudo-natural products. Nat Commun *12*, 1883. 10.1038/s41467-021-22174-4.