ACS chemical biology

pubs.acs.org/acschemicalbiology

Article

Enhancing the Small-Scale Screenable Biological Space beyond Known Chemogenomics Libraries with Gray Chemical Matter—Compounds with Novel Mechanisms from High-Throughput Screening Profiles

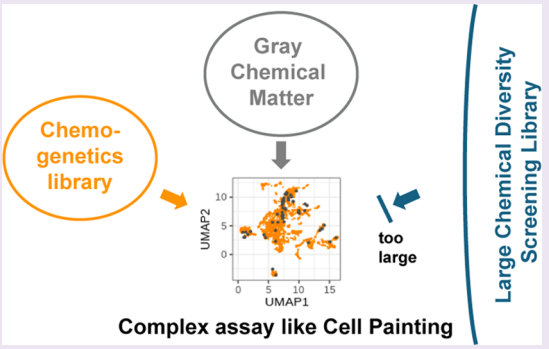
Jason R. Thomas,*,∇ Claude Shelton, IV, Jason Murphy, Scott Brittain, Mark-Anthony Bray, Peter Aspesi, John Concannon, Frederick J. King, Robert J. Ihry, Daniel J. Ho, Martin Henault, Andrea Hadjikyriacou, Marilisa Neri, Frederic D. Sigoillot, Helen T. Pham, Matthew Shum, Louise Barys, Michael D. Jones, Eric J. Martin, Anke Blechschmidt, Sébastien Rieffel, Thomas J. Troxler, Felipa A. Mapa, Jeremy L. Jenkins, Rishi K. Jain, Peter S. Kutchukian, Markus Schirle, and Steffen Renner*,∇

Cite This: ACS Chem. Biol. 2024, 19, 938−952

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Phenotypic assays have become an established approach to drug discovery. Greater disease relevance is often achieved through cellular models with increased complexity and more detailed readouts, such as gene expression or advanced imaging. However, the intricate nature and cost of these assays impose limitations on their screening capacity, often restricting screens to well-characterized small compound sets such as chemogenomics libraries. Here, we outline a cheminformatics approach to identify a small set of compounds with likely novel mechanisms of action (MoAs), expanding the MoA search space for throughput limited phenotypic assays. Our approach is based on mining existing large-scale, phenotypic high-throughput screening (HTS) data. It enables the identification of chemotypes that exhibit selectivity across multiple cell-based assays, which are characterized by persistent and broad structure activity relationships (SAR). We validate the effectiveness of our approach in broad cellular profiling assays (Cell Painting, DRUG-seq, and Promotor Signature Profiling) and chemical proteomics experiments. These experiments revealed that the compounds behave similarly to known chemogenetic libraries, but with a notable bias toward novel protein targets. To foster collaboration and advance research in this area, we have curated a public set of such compounds based on the PubChem BioAssay dataset and made it available for use by the scientific community.



INTRODUCTION

A fundamental tenet of chemical biology is that small molecules can reveal unprecedented insights into biology. As such, phenotypic-based screens are commonly utilized to investigate disease-relevant biology. These screens typically employ two approaches: unbiased high-throughput screening (HTS) of a large and chemically diverse compound collection and focused screening of compounds with established targets and/or mechanisms of actions (MoAs). The unbiased HTS approach allows for the discovery of truly novel chemotypes and MoAs for a specific activity of interest but requires the screening of very large diversity-oriented chemical libraries. The sheer size of these screens can preclude screening of complex, disease-relevant assays, which are often difficult to miniaturize and scale-up. Moreover, the specialized instrumentation and data processing infrastructure required for screens of this scale generally require partnerships with dedicated screening centers or occur within specialized groups.

Screening of a chemogenetic library, a curated collection of compounds with annotated targets and MoAs, is increasingly used as an orthogonal strategy to discover potential disease-modifying targets and underlying MoAs.[1−3] This approach has several key advantages: (1) the smaller scale of these screens allows for the utilization of assay formats not traditionally associated with HTS campaigns, and (2) the integration of

Received: December 5, 2023
Revised: February 28, 2024
Accepted: March 1, 2024
Published: April 2, 2024

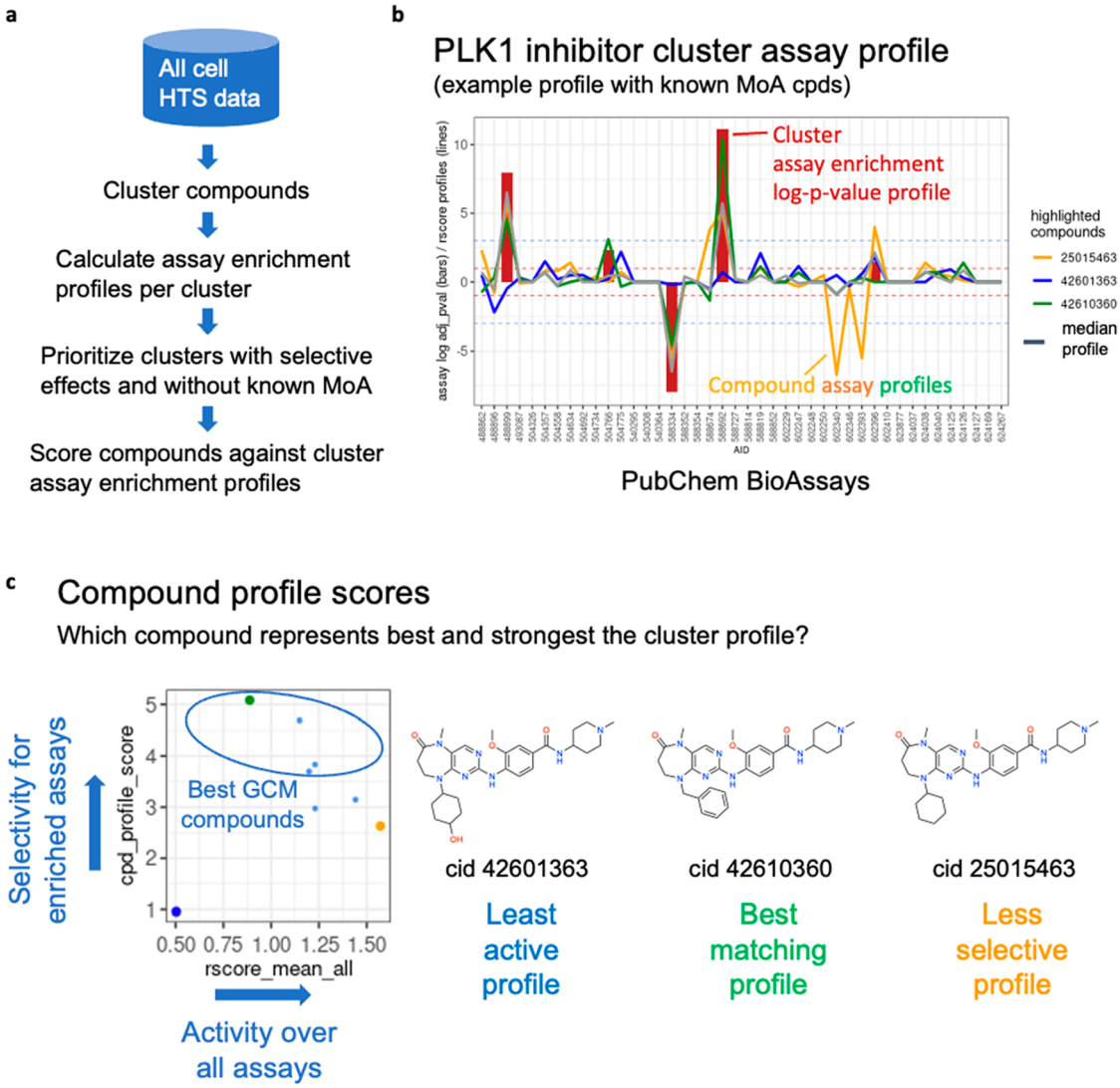


Figure 1. How to calculate Gray Chemical Matter (GCM). (a) Overview of the calculation pipeline. (b) Assay enrichment profile of a GCM cluster (bars) and individual activity profiles of cluster compounds (lines). The bars represent the logged adjusted *p*-values of the assay enrichment calculations. Bars are set to negative values if activities are in the opposite direction as the assay was intended (agonists in antagonist assays, and antagonists in agonist assays). Bars >1 or < −1 (red dashed lines) are significantly enriched. Compounds are considered active in an assay if rscores are >3 or < −3 (blue dashed lines). (c) Profile scores are used to identify compounds that best represent the cluster enrichment profiles. Three compounds are highlighted in the plot and are also shown with their respective assay signatures in panel (b): CID 42610360 (green) with the best matching profile, CID 250154663 (orange) with a less selective profile, and CID 42601363 (blue) with the least active profile. Compounds with the highest profile scores are the most interesting, as they show the most activity and selectivity. Weaker profile scores can be caused either by weaker activity (blue signature in panel (b) with only low rscores) or a lack of selectivity (orange signature in panel (b) with high rscores in enriched but also nonenriched assays).

target annotations within the library enables a rapid transition from screening to hypothesis-driven research. Currently only ca. 10% of the human genome (ca. 2K targets for 20K genes) is covered by such libraries.[4] Target coverage is likely to remain static due to the time and effort required to develop tool compounds for new targets. This limitation calls for the design of alternative and faster approaches of identifying compounds with new and distinct MoAs in order to continue expanding the impact of chemogenomic libraries.

The appeal of phenotypic screens resides in their target-agnostic approach. These screens allow for the discovery of modulators of well-known critical signaling proteins as well as specific yet indirect mechanisms that achieve the same desired effect. When viewed as a whole, cellular HTS data are rich in MoA mechanisms which, if mined effectively, can uncover unbiased insights pointing toward potentially novel MoAs and targets.

Many informatics-based approaches have been proposed to develop screening libraries that are enriched in bioactive compounds based on existing knowledge of bioactive chemotypes, i.e., employing chemogenomics information from target families[5−7] or biology-enriched chemotypes.[8−10] More recently, machine learning models trained on large chemogenomics datasets[11,12] coupled with generative chemistry[13,14] are gaining traction. However, all of these strategies rely on well-characterized bioactive compounds to extrapolate and expand to neighboring MoAs with similar target proteins and target profiles.

https://doi.org/10.1021/acschembio.3c00737
ACS Chem. Biol. 2024, 19, 938−952

By analyzing the activity landscape of compounds from legacy HTS data, distinct fingerprints of chemotype−phenotype associations can emerge. It is widely known that HTS fingerprints are highly correlated between structurally distinct compounds possessing the same target/MoA. In fact, compounds can be clustered solely based on HTS fingerprints, effectively grouping compounds with shared targets or MoAs, regardless of their chemical structure.[15−20] In this study, we present a cheminformatics framework that leverages existing cellular HTS data to identify associations between chemotype and phenotype activity based solely on their phenotypic activity. Through chemical clustering of related HTS fingerprints, we successfully identified groups of structurally related compounds exhibiting persistent and broad structure−activity relationships (SARs). We refer to this phenomenon as "dynamic SAR", in contrast to "flat SAR", which is characterized by minimal changes in compound activity despite structural variations. The key advantage of our framework is that it enriches compounds with cellular activity, potential MoAs, and targets that are not currently represented by existing chemogenetic libraries.

RESULTS AND DISCUSSION

Computational Framework. HTS data are susceptible to assay artifacts.[21,22] Therefore, it is critical for computational HTS mining approaches to avoid inadvertent enrichment of these artifacts. Additionally, certain classes of compounds have unusually high hit rates across a diverse panel of assays, because of their well-established biological impact (e.g., HDAC inhibitors or ATP-competitive pan-kinase inhibitors). On the opposing side of the spectrum is the so-called Dark Chemical Matter (DCM),[23] which are compounds that have shown minimal assay activity, despite being tested in at least 100 biochemical and cellular assays. We aimed to identify a middle ground between the extremes of frequent hitters[24] and DCM, where phenotypic activity can serve as a meaningful measure of modulating a specific target, regardless of the intended assay outcome. Even if the target is unknown, the activity landscape can provide some assurance of the selectivity. Inspired by the DCM terminology, we introduced the term "Gray Chemical Matter" (GCM) to describe compounds in this range.

The GCM workflow consists of the following steps (see Figure 1): (1) obtain a set of cell-based HTS assay datasets, (2) cluster the compounds based on structural similarity and retain only clusters with sufficiently complete assay data matrices to be able to generate assay profiles, (3), calculate an enrichment score for each assay to identify clusters with enriched activity, (4) prioritize clusters with selective profiles and without known MoAs, and (5) score individual compounds within the cluster based on how well they represent the overall cluster profile.

A key step of the GCM pipeline is determining whether a chemical cluster significantly affects a given assay. This is particularly challenging since primary screening data from HTS assays are often generated at a single concentration without replication, resulting in variable assay hit rates and noisy data, making it inherently difficult to assess whether a chemical cluster is over-represented among the active compounds. To address this, we used the Fisher exact test to identify chemical clusters with a hit rate in assays that was significantly higher than that expected by chance. This statistical test compares the number of actives and inactive assay compounds within a chemical cluster against the total number of active and inactive compounds, irrespective of clustering. If the fraction of actives within the cluster is significantly higher than the overall assay hit rate, then the cluster is considered to be enriched for that assay. This approach is inspired by compound set enrichment and scaffold network enrichment methodologies,[25,26] used to identify weak but significant hits in primary HTS data. Using this statistical approach, similar compounds can be treated similarly to replicates of the same compound, thereby increasing the confidence in the chemotype effect on an assay.

In a typical screening project, the assay is designed to identify hits in one prespecified direction, either inhibition or activation. However, to permit an unbiased approach toward detectable MoAs, the data were analyzed without consideration of the desired outcome of the screen, i.e., we allowed for finding agonistic activity in an antagonism screen and vice versa. For this reason, independent statistical tests were performed for both directions of an assay.

Another key step is to score the compounds of a GCM cluster based on how closely they match the cluster assay profile. The cheminformatic framework enables the identification of potentially interesting clusters, but testing entire clusters in future assays is not feasible due to practical limitations. Nonetheless, tests can still be conducted on a single compound from the cluster that has the best alignment with the overall cluster profile. For this purpose, we developed a profile score that quantifies how well the activity profile of an individual compound compares to the assay enrichment profile for all compounds in the same chemical cluster. Highest scores are obtained for compounds with strong effects in enriched assays and weak activities in nonenriched assays. The profile score is calculated as

$$\text{profile score}_{\text{cpd}} = \frac{\sum_{\text{array}\,a=1}^{n} \text{rscore}_{\text{cpd},a} \times \text{assay direction}_{a} \times \text{assay-enriched}_{a}}{\text{mean}(\text{absolute}(\text{rscore}_{\text{cpd,assays}}))}$$

The numerator of the profile score quantifies how well the compound's assay profile matches the assay enrichment profile of the GCM cluster. The denominator normalizes the score to the overall mean absolute activity of the compound over all of the assays. The rscore term represents the number of median absolute deviations that the activity of compound "cpd" measured in assay *a* deviates from the median of that particular assay. The "assay direction" term has values of +1 for assays enriched in the intended direction (i.e., agonists in an assay that was run for agonists and inhibitors in an assay that was run for inhibitors) or −1 for assays enriched in the opposite direction (enrichment of agonists in and inhibitor assay or inhibitors enriched in an agonist assay). The same directionality convention is used for the sign of the rscore activity values. The value of the term "assay enriched" can be either +1 for enriched assays or 0 for assays without enrichment.

The profile score is designed to prioritize compounds with high rscore values for enriched assays while assigning near-zero values for nonenriched assays. This approach allows selection of compounds that have the strongest effects in a specific subset of cellular assays, while exhibiting minimal activity in all other profiled assays.

PubChem Gray Chemical Matter. For the PubChem[27] GCM dataset, we identified 171 cellular HTS assays with >10k compounds tested, totaling ∼1 million unique compounds.

https://doi.org/10.1021/acschembio.3c00737
ACS Chem. Biol. 2024, 19, 938−952

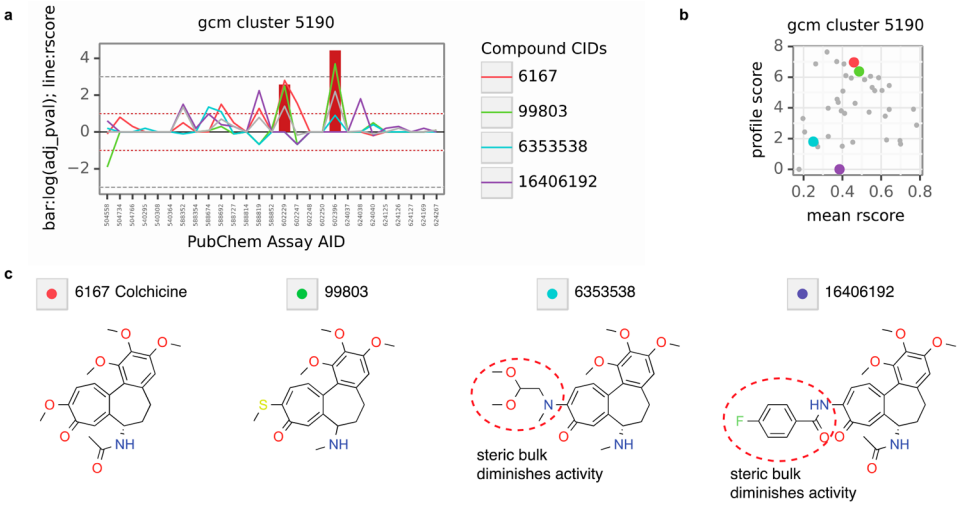


Figure 2. Colchicine SAR on tubulin correlates with the GCM SAR. (a) Assay enrichment profile of the Colchicine GCM cluster. (b) GCM profile scores of the colchicine SAR cluster. (c) Selected colchicine analogues demonstrating consistent SAR reported on tubulin and on GCM profile scores. Colchicine and PubChem CID 99803 are reported active on tubulin. CID 6353538 and CID 16406192 have bulky steric groups that diminish activity on tubulin and on the GCM profile scores.

After clustering and filtering to ensure sufficient data completeness, we obtained 23 000 chemical clusters. Subsequent calculation of the assay enrichment profiles yielded 1956 clusters with significant enrichment in at least one assay. Of those, 1455 clusters matching the following criteria were kept as PubChem GCM candidates: ≥10 assays tested, <20% of tested assays showing enrichment (limited to a maximum of 6 enriched assays), and <200 compounds tested in any one of the assays. The cluster size limit avoids excessively large clusters with potential multiple independent MoAs.

To validate our approach, we did not exclude known chemogenetic compounds from the PubChem GCM dataset, allowing us to investigate their GCM profiles in more detail. Since these compounds often have well-described targets, their ability to match the overall assay profile for a cluster served as a strong indication that the assay activity was likely driven by the assigned target.

Out of the 1455 PubChem GCM clusters, we identified 20 clusters that contained compounds from the Novartis chemogenetic library (Figure S1). Within these clusters, six compounds were associated with the highest-ranking profile scores within their respective clusters, indicating that their activity aligned well with the overall cluster activity (Table S1). This correspondence provides compelling evidence that the annotated targets are likely responsible for the activity observed in the cluster. Notably, we observed clear examples where the assay profile correlated with the known SAR for the respective scaffolds. For example, colchicine and analogs from the same cluster exhibited activity patterns consistent with the established SAR on tubulin,[28] as well as the GCM phenotypic profile score SAR (see Figure 2). However, we acknowledge that the SAR analysis is not exhaustive due to a variety of factors, such as different sources of information (e.g., peer-reviewed manuscripts vs patents), variations in assays, assays conducted by different research laboratories, and the limited availability of inactive compound data. Additionally, while nine compounds did not achieve the top-ranking score, their activity remained consistent with the activity of the PubChem GCM cluster. In only three instances did a chemogenetic library member not correlate with the assay profile, indicating that the profile activity in these cases is likely driven by a different, yet unknown target (see Table S1). These findings underscore the ability of our computational framework to identify compound clusters enriched with specific cellular activity with defined targets.

For a broader MoA assessment, we annotated all PubChem GCM compounds with dose response activities from ChEMBL,[29] focusing on compounds where human target gene information was available. The ChEMBL activities for PubChem GCM compounds spanned a wide range of IC_{50} values, ranging from <1 nM to >100 μM. The threshold of biochemical activity translating to cellular activity is entirely target-dependent, although biochemical potency values of ≤100 nM is generally agreed upon.[30] Of the 750 PubChem GCM clusters with at least one biochemical activity potency value available in ChEMBL, only 47 GCM clusters scored in the range where one could reasonably expect the cellular activity to be attributable to the biochemical target (<100 nM). For 14 of these GCM clusters, the compounds with the best profile score were annotated with ChEMBL targets, and 29 clusters had an annotated target within the best five ranking profile score GCMs. Furthermore, 9 of the 47 GCM clusters had activity annotations for at least three compounds. Of those, four clusters had a Pearson correlation of the ChEMBL potency values with the profile score of >0.5 (see Figure S2). These additional examples emphasize the principles of the GCM workflow. Specifically, they demonstrate that engagement with a specific molecular target is linked with an enriched assay activity and selectivity across a wide range of assays with dynamic SAR.

When dealing with compounds of unknown MoA, particularly those identified through phenotypic screens, the most convincing evidence for compound engagement with a specific cellular target is through demonstration of selective cellular activity within a chemical series with persistent and dynamic SAR. The preservation of dynamic SAR indicates a molecular recognition event such as binding to a defined pocket. However, it is important to note that SAR changes can also affect other physicochemical factors that influence cellular activity such as cell permeability or solubility. Thus, examples

https://doi.org/10.1021/acschembio.3c00737
ACS Chem. Biol. 2024, 19, 938−952

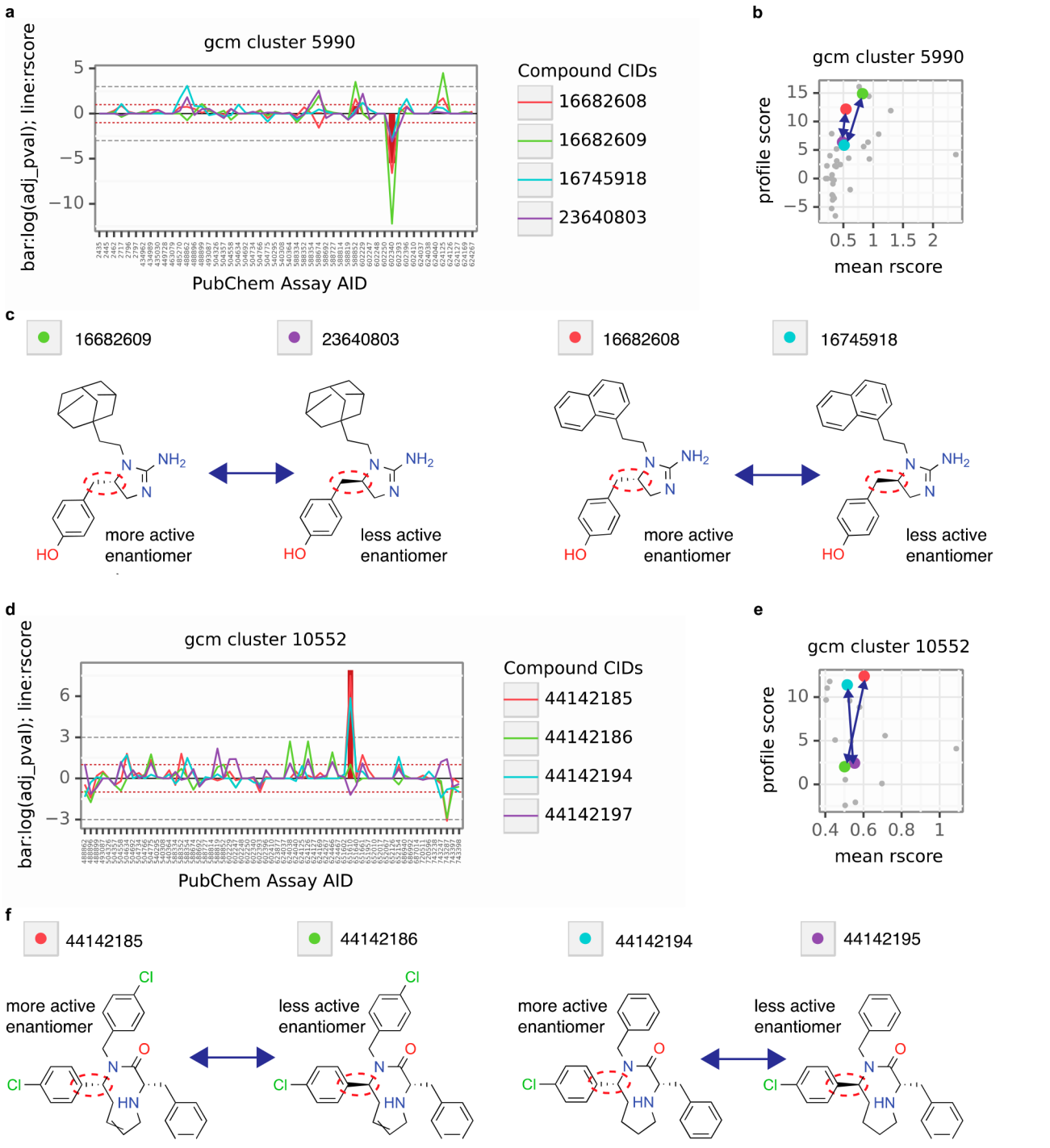


Figure 3. Consistent SAR observed for enantiomer pairs within GCM clusters indicates specific molecular recognition events on an unknown target. (a and d) Assay enrichment profiles of PubChem GCM clusters. (b and e) GCM profile scores of PubChem GCM clusters. (c and f) Both GCM clusters contain two enantiomer pairs which show consistent SAR patterns in their GCM profile scores. Note that cluster 5990 has enriched activity in the opposite direction of an intended assay (negative bar and profile activities in panel (d)). The stereochemically dependent SAR in this example provides clear evidence of the value in considering compound activity in both assay directions.

of enantiomer pairs with significant differences in cellular activity can provide clear and compelling evidence of target specific interactions between a compound and a protein target in cells. In our analysis, we mined the PubChem GCM cluster for examples of enantiomer pairs and discovered two clusters where the enantiomers exhibited striking differences in rscore values (Figure 3). This underscores that, even for compound clusters without any annotated targets, clear evidence of selective and specific target engagement can still be found. We also explored additional examples of dynamic SAR not defined by stereochemistry, specifically looking at clusters of closely related analogues. Their full assay profiles are shown in Figure S3. Additionally, we noticed that clusters with exclusively active compounds, which might be considered potential

https://doi.org/10.1021/acschembio.3c00737
ACS Chem. Biol. 2024, 19, 938−952

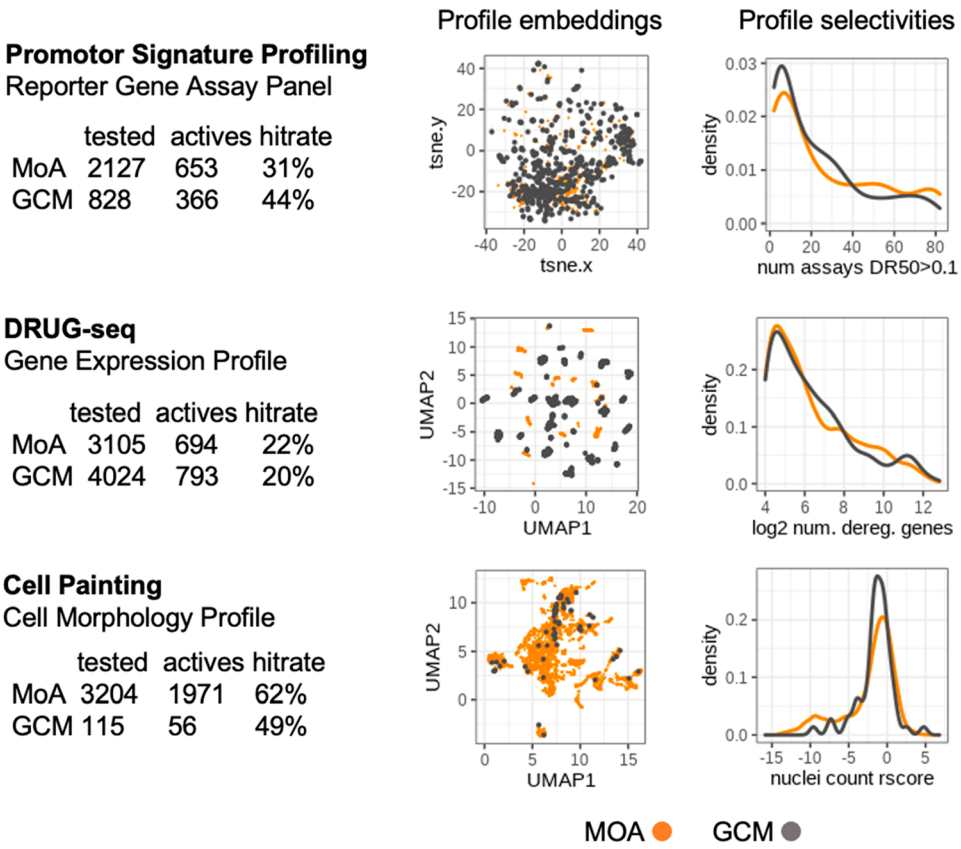


Figure 4. Results of Novartis GCM in Novartis profiling assays in comparison with chemogenetic compounds. GCMs are similar to chemogenomic library compounds in terms of hit rates, phenotype coverage over the profile maps, and selectivity of profiles.

artifact activities, only occur within clusters with <20 cluster members (Figure S4). This indicates that, for many such cases, there were simply not enough analogues tested to identify inactive cluster members as well.

Novartis Gray Chemical Matter. The Novartis cell HTS data were processed with the same pipeline as that for the PubChem GCM. To focus on compounds relevant to mammalian biology, we excluded assay data from non-mammalian cell lines. The resulting Novartis GCM data features 160 assays with >40k compound clusters and consists of >1.5 million compounds.[31]

For the Novartis GCM, 11 000 clusters were identified with at least one assay enriched. After applying similar filtering criteria as the PubChem GCM workflow, this led to 6.8k GCM candidate clusters. To focus on potentially novel MoAs, clusters containing compounds from the Novartis chemogenetic library were removed, as these compounds have well-established targets and MoAs.[1] Additionally, we also applied computational target prediction strategies to remove Novartis GCM clusters with a high likelihood of being driven by a well-described protein target. Clusters were excluded if they had either a high confidence prediction for 10% or medium confidence prediction for 20% of the compounds within the cluster.[32] This procedure yielded a set of 4.8k GCM clusters.

It is interesting to note that there were only 233 compounds that overlapped between the Novartis GCM and the PubChem GCM, representing 90 Novartis GCM clusters and 88 PubChem GCM clusters, respectively. The reason for this finding is most likely attributable to several factors: utilization of different screening collections, different representatives of clusters being present in screening collections, use of different assays with distinct biological contexts, and different subsets of screening collections tested for each of the assays. Given these factors, incorporating GCM from orthogonal sources will increase compound and biological diversity.

Physicochemical Properties of GCM Compounds. To ensure the selection procedure for GCM compounds did not introduce a bias toward unfavorable physicochemical properties, we compared the distributions of these properties, for the Novartis GCM, the PubChem GCM, and the Novartis chemogenetic library (see Figure S5). All three sets exhibited distributions that fell within reasonable ranges for molecular weight, log *P*, TPSA, H-bond donors and acceptors, and the number of rotatable bonds (all descriptors were calculated by RDKit[33]). For example, the mean MW values were 385 ± 101 Da (Novartis GCM), 366 ± 93 Da (PubChem GCM), and 400 ± 143 Da (chemogenetic library). The respective log *P* mean values are 3.8 ± 1.4, 3.7 ± 1.4, and 3.4 ± 2.0; TPSA mean values are 66 ± 35, 64 ± 32, and 85 ± 50; H-bond acceptor mean values are 4.6 ± 2, 4.4 ± 1.9, and 5.2 ± 2.5; H-bond donor mean values are 1.1 ± 1.1, 1.1 ± 1.0, and 1.9 ± 1.8; and number of rotatable bonds mean values are 5.0 ± 3.0, 4.5 ± 2.6, and 5.6 ± 4.0. These findings support the readiness of compounds identified through the GCM workflow for deployment in unbiased screening efforts.

Cellular Profiling Assays Reveal Broad Coverage of Biology Encompassed by GCM Compounds. Given the wide range of cellular HTS assays and diverse activity profiles, we anticipate that the GCM compounds are likely to encompass a broad spectrum of MoAs. To validate this

https://doi.org/10.1021/acschembio.3c00737
ACS Chem. Biol. 2024, 19, 938−952

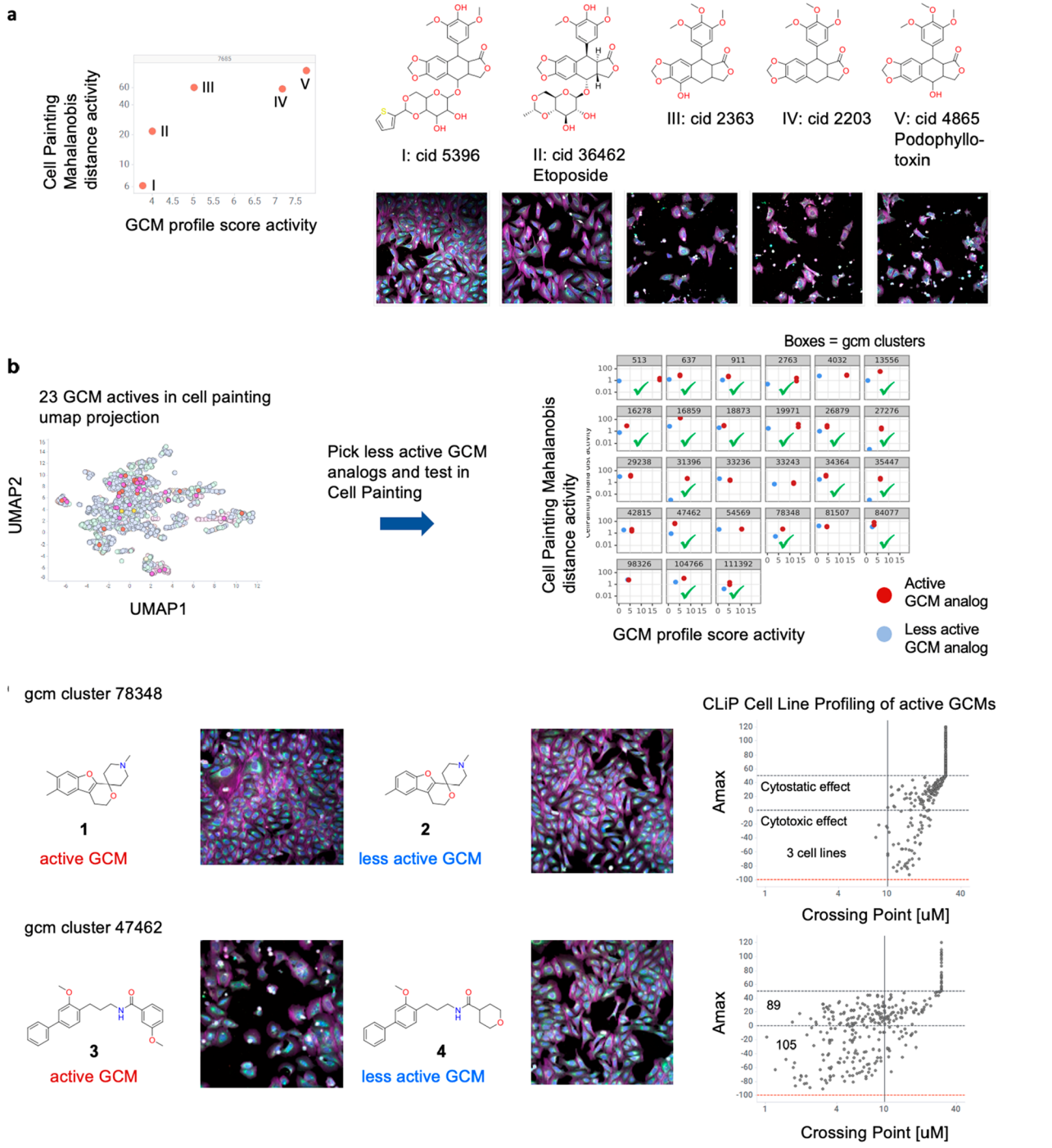


Figure 5. SAR transfer from GCM profiles to Cell Painting and CLiP. (a) Podophyllotoxin PubChem GCM SAR translates to the Cell Painting phenotype strength SAR. (b) 23 pairs of active and less active GCM pairs were tested in Cell Painting. For 19 pairs, the rank of activity was preserved. The active GCMs are shown both with their initial result and the repeated measurement, together with the inactive GCM analogues. In many cases, the activities for active GCM compounds between the initial and repeat experiments are so close that the individual cell painting measurements cannot be distinguished. (c) Cell Painting images of two pairs of active and less active GCMs. Compounds **1** and **2** show only changes in the cell morphology, which is reflected in a very clean profile in the CLiP assay. Compounds **3** and **4** also impact the cell nuclei numbers, which is also reflected in the cytotoxicity observed in 105 out of 300 cell lines profiled in the CLiP assay.

hypothesis, we sought to compare the hit rate and breadth of biological response across multiple profiling platforms between GCM compounds and those from the Novartis chemogenetic library. We selected three distinct platforms for this analysis: Promoter Signature Profiling (PSP),[34] which utilizes a panel of reporter genes and is conducted in HEK293T cells; DRUG-seq,[35,36] a high-throughput transcription profiling assay performed in NGN2 neurons; and Cell Painting,[37−39] a morphological profiling assay applied in U2OS cells. These platforms offer diverse readouts and cellular backgrounds and do not require the compounds to impact cellular proliferation to generate an activity signature. Different numbers of GCM

https://doi.org/10.1021/acschembio.3c00737
ACS Chem. Biol. 2024, 19, 938−952

compounds were selected for each profiling assay based on compound availability, assay capacity, and resources. The selection of GCM clusters was prioritized based on profile score strength, selectivity, and a diverse coverage of enriched assays to increase biological diversity.

In each of the profiling assays, the GCM compounds mirrored the coverage of compounds with well-established MoAs (Figure 4). This finding suggests that, as a collection, the individual GCM clusters have diverse and distinct MoAs. Moreover, the distribution of profiles in the GCM compounds behaved similarly as profiles from the known MoA collection. This similarity was observed in the distribution of affected reporter genes, differentially expressed genes, and nuclei counts. Interestingly, the hit rate for GCM compounds closely matched that of the Novartis chemogenetic library, suggesting that GCM compounds have comparable levels of selectivity to a curated compound collection with defined targets and MoAs. It is important to note that the GCM compounds are primary hits from screening data and have seen no synthetic efforts to enhance their properties. In summary, the GCM compounds perform similarly to compounds with known MoAs across multiple profiling platforms in terms of biological diversity, hit rate, and selectivity of phenotypes. Collectively, these findings suggest that GCM collections are highly promising for enabling biological discoveries. Furthermore, when utilized in combination with known chemogenomics libraries, GCM collections have the potential to significantly expand the exploratory landscape of MoAs for throughput-limited phenotypic-based screens.

SAR Transfer from GCM Profiles to Novel Assays. An important principle guiding the discovery of GCM compounds is that dynamic SAR within and across assays can indicate that a cluster of compounds has a specific target and a significant level of selectivity for its target. The value of these compounds in future assays relies on the ability of this SAR to extend to assays that have not been previously tested. To assess the translatability of SAR, we tested analogues with diverse activity from multiple GCM clusters in the Cell Painting profiling assay. The aim was to evaluate the conservation of SAR within the context of broad morphological responses.

As a first step in validating this approach, we tested five podophyllotoxin analogues that were also observed in the PubChem data. These compounds exhibited a range of activity profile scores similar to those found in the Novartis GCM data. Profiling these analogs in Cell Painting revealed a ranking consistent with their profile score activity, which correlated with their phenotypic strength as quantified by the Mahalanobis distance relative to the neutral control (DMSO) phenotype (Figure 5a).

We selected 23 GCM compounds from the Novartis GCM dataset that had been previously shown activity in the Cell Painting profiling assay, for retesting to evaluate the robustness of their SAR (Figure 4). To provide a basis for comparison, we included structurally similar but less active GCM cluster members (Figure 5b). Out of the 23 pairs of GCM compounds, 19 pairs (83%) of the less active GCM cluster members exhibited a weaker phenotype (indicated by a decreased Mahalanobis distance) or showed no phenotypic change relative to DMSO. While it is very unlikely that compounds selected from a meta-analysis will be as selective as compounds that have gone through rounds of medicinal chemistry optimization, it was reassuring to observe that, in most cases, the dynamic SAR is preserved in an orthogonal assay readout.

We decided to delve deeper into two GCM SAR pairs and conducted a more-detailed examination (Figure 5c). The active GCM compound from cluster 78348 (compound **1**) specifically affected cell morphology, while the active GCM from cluster 47462 (compound **3**) not only impacted cell morphology but also reduced the nuclei count. The activity of each compound correlated with their rscore values, as compounds from the same cluster with lower rscore values failed to produce the same morphological effect. To investigate whether the observed effects were due to a broad cell viability MoA, we conducted a Cell Line Inhibitor Profiling (CLiP) assay to characterize compounds **1** and **3**,[40] assessing cell viability across more than 300 well-characterized CCLE cell lines. As anticipated, compound **3**, with its lower nuclei count, affected the viability of more cell lines than compound **1**. We were pleased to find that compound **1** had a minimal impact on cell viability, affecting only a small number of cell lines at the highest concentration tested. However, we were surprised to discover that compound **3** influenced the viability of approximately one-third of the cell lines tested. Note that the cellular HTS assays used to determine the profile score typically have shorter time points (ranging from hours to overnight), whereas the CLiP assay extended to 72 h. This extended time frame in the CLiP assay might explain the broader impact on cell viability observed for compound **3**. Importantly, since compound **3** did not universally affect cell viability, it suggests the possibility of a distinct MoA that could explain the Cell Painting and CLiP profiles. By comparing the overall morphological and phenotypic outcomes of compounds **1** and **3**, these results underscore that, while GCM compounds may have some influence on cell viability, the computational framework itself is not biased toward general cellular mechanisms that broadly impact cell viability.

To gain a deeper understanding of how the dynamic SAR affects target engagement, we focused on a specific GCM cluster that contained an electrophilic moiety where the presence of the electrophile appears crucial for cluster activity (Figure S6). This suggests that active compounds within this cluster, which possess a Michael acceptor, likely engage their target(s) through covalent labeling of a cysteine residue. To assess the selectivity differences between an active and a less active GCM compound across the proteome, we conducted a live-cell competitive proteome-wide cysteine profiling experiment using an acid-cleavable iodoacetamide probe in HEK293T cells. The results of this experiment showed that the less-active GCM compound (**10**) competed with the labeling of 95 sites, while the active GCM compound (**11**) competed with only 7 sites. While it remains uncertain how representative these stark differences in proteome selectivity are for the entire GCM compound collection, these findings shed light on how dynamic SAR may influence proteome selectivity, which, in turn, may lead to specific and selective phenotypic activity.

Cheminformatic Prediction of Known Target Space. Cheminformatics tools, such as pQSAR models,[12,41] have proven to be accurate in predicting the potency of an unknown compound for binding to a specific target based on the known SAR for that target. In our study, we utilized pQSAR models for 827 targets to compare the hit frequency of the Novartis chemogenetic library with that of the GCM compounds. The analysis revealed that, on average, each compound from the

https://doi.org/10.1021/acschembio.3c00737
ACS Chem. Biol. 2024, 19, 938−952

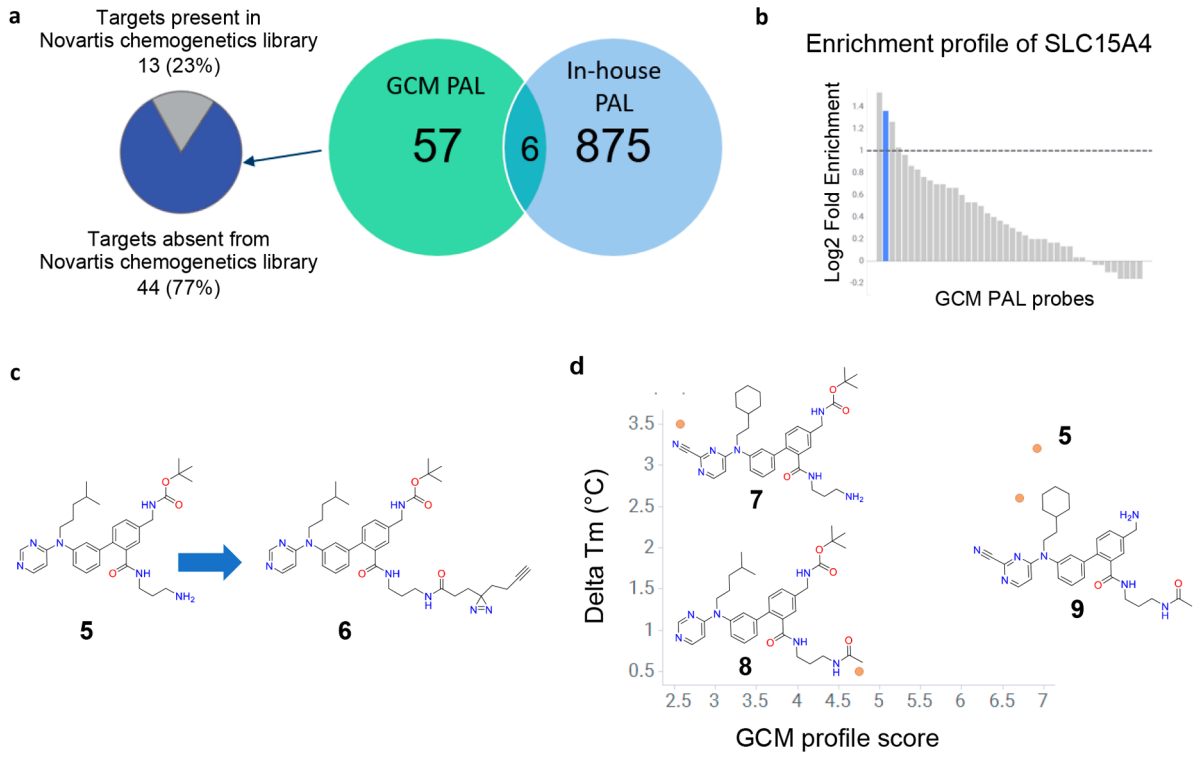


Figure 6. GCM compounds engage protein targets not covered by chemogenomics libraries. (a) Venn diagram comparing analysis of GCM exclusive hits with proteins exhibiting exclusive enrichment with PAL probes synthesized for internal project use. Pie chart depicts the number of known Novartis chemogenomics library members contained within the list of targets exclusively enriched by one GCM PAL probe. (b) Enrichment profile of SLC15A4 across panel GCM PAL probes. Dotted line indicates 2-fold enrichment over DMSO control. (c) While an enrichment of SLC15A4 was observed with GCM PAL probe **6**, the parent compound, **5**, was assayed for direct binding. (d) Scatter plot of delta T_m values was derived from nanoDSF experiments with purified SLC15A4 and GCM profile scores.

Novartis chemogenetic library was predicted to bind to 32 targets, whereas the GCM compounds were predicted to bind to an average of 8 targets. This 4-fold reduction in target prediction suggests that GCM compounds likely bind to targets that are distinct from those represented by current chemogenetic libraries.

Chemical Proteomic Profiling. The relative lack of pQSAR predictions for GCM compounds sparked a hypothesis that these compounds may interact with novel targets. To determine which proteins can bind to GCM and potentially link them to novel phenotypes with protein targets, chemical proteomics screening using photoaffinity labeling (PAL) was performed. PAL probes were synthesized for 57 GCM compounds, which were selected based on compound availability, compatibility with a one-step reaction to produce the PAL probe, and whether there was evidence that modifications could be tolerated at the site for the PAL group based on SAR within the cluster.

HEK293T cells were treated with 1 μM PAL probe for 2 h. After photoirradiation and cell lysis, click chemistry was used to append biotin to the probe-modified proteins. These proteins were then enriched, and the relative abundance was determined using mass spectrometry with isobaric tagging. By profiling the 57 GCM PAL probes, a total of 6879 proteins were identified. Among these proteins, 63 were selectively enriched 3-fold relative to the DMSO control by only one GCM PAL probe. To assess the uniqueness of these enrichments by GCM probes, the enrichment of 54 PAL probes from internal projects (also performed in HEK293T cells and treated with the 1 μM PAL probe) was compared. The PAL probes derived from internal projects have generally progressed through traditional phenotypic screening flowcharts. Project derived probes are carefully selected based on desired activity, followed by a thorough counter screening process and subsequent optimization. Consequently, these probes serve as a valuable point of comparison to the GCM PAL probes. Differences in proteome coverage between GCM-PAL and in-house PAL would emphasize the effectiveness of the GCM workflow in capturing novel modes of action that may not be identified through conventional screening paradigms. While the PAL probes from internal projects led to the identification of more proteins with at least 3-fold enrichment, relative to the DMSO control, there was minimal overlap with the GCM PAL probes (Figure 6A). Furthermore, comparing the list of proteins uniquely enriched by GCM compounds to the annotated targets of the Novartis chemogenetic library revealed that most of these targets have no known ligands (Table S2). These results highlight the potential of GCM compounds to exhibit novel phenotypes by targeting potentially untapped regions of the proteome.

Identification of SLC15A4 Binders from GCM Profiling. Deorphanizing protein function, from single proteins to entire families, is a challenge that has been taken up by the chemical biology community. A recent notable effort in this regard is the RESOLUTE consortium,[42] a precompetitive collaborative initiative between academic and pharmaceutical industry. The main objective of the consortium is to identify ligands and elucidate the function of as many members as

https://doi.org/10.1021/acschembio.3c00737
ACS Chem. Biol. 2024, 19, 938−952

possible within the solute carrier transporter (SLC) superfamily, a group of 446 members.

The PAL-based chemical proteomics experiments provided encouraging insights into the potential of GCM compounds as potential ligand candidates for targets lacking known ligands. In particular, we investigated whether GCM compounds were capable of binding to SLC15A4, an SLC transporter without any reported ligands. Mining the results from the PAL-based screen, we identified three GCM PAL probes that were able to enrich SLC15A4 more than or equal to 2-fold, relative to the DMSO control (Figure 6B). To directly assess compound binding, representative compounds from each of the three GCM clusters were assayed for their ability to increase the thermal stability of SLC15A4 via nanodifference scanning fluorimetry (nanoDSF). Gratifyingly, several compounds from one GCM cluster (compounds **5**, **7**, **8**, **9**) exhibited a positive shift in the T_m values for SLC15A4 (0.5−3.5 °C) (Figures 6C and 6D). These results highlight the potential of GCM compounds as starting points for ligand discovery for novel targets.

DISCUSSION

In this paper, we present a computational framework that provides valuable insights into compounds with selective and specific cellular activity, identified using legacy cellular HTS data. The framework incorporates procedural concepts to capture compounds that function through a specific target or MoA, in an assay and in a target-agnostic manner. The result of this approach is a compound collection that includes both active and inactive members, covering diverse MoAs, and is capable of engaging protein targets not already covered by chemogenetic libraries. It is important to note that while the characterization of GCM selectivity and specificity was performed with a proprietary compound collection, the framework was also able to identify previously unnoticed features within the publicly available PubChem database. We anticipate that these publicly available compounds hold the potential for future drug discovery efforts.

Existing computational approaches for lead generation that utilize HTS fingerprints rely on the similarity principle. This requires having an active compound with a desired phenotypic profile to identify additional compounds with similar HTS profiles or to train machine learning models on known active compounds and their HTS fingerprints. Our approach focuses on an unbiased utilization of HTS fingerprints to reveal potentially novel MoAs. While HTS fingerprints play an important role in our computational workflow, it is important to note that our approach extends far beyond simple clustering based on assay activity. The workflow incorporates multiple crucial aspects, including selectivity, SAR, and assay noise, derived from primary screening data. By considering these factors, we can curate a small and highly promising set of molecules for testing in assays.

Undoubtedly, compounds with potentially novel MoAs could be identified by considering compound profiles individually, rather than as a cluster. We believe that the strength of the GCM approach lies in its ability to demonstrate consistent SAR across various assays. While the assay profile of any individual compound is susceptible to uncertainties in HTS, the confidence in the results increases when more structurally related compounds consistently exhibit similar activity.

In the development of the GCM workflow, the decision to consider both assay directions has been an uncommon strategic choice. Typically, the direction for which the assay has not been specifically designed exhibits lower sensitivity and may be more susceptible to assay artifacts. However, these technical considerations were weighed against the objective of identifying novel MoAs that are not covered by traditional screening approaches. By incorporating data that are typically not considered in screening, the overall goal of the workflow is better supported compared to only considering the intended assay direction. Ultimately, this factor is not prominently featured in the GCM analysis; note that, out of the 1455 PubChem GCM clusters, 509 clusters demonstrate statistically significant assay enrichment in the unintended direction. Among these clusters, only 230 PubChem GCM clusters showed significant enrichment exclusively in the unintended direction. While there is inevitably some level of risk in considering both assay directions, we believe that the data show that this is a balanced risk.

The pQSAR predictions and PAL data led us to believe that GCM compounds have a high likelihood of producing novel MoAs. However, when we assessed the MoA space covered by GCM compounds in various profiling assays, there is noticeable overlap with the MoA space covered by chemogenomic library members. The overlapping MoA space could be attributed to targeting different nodes in the same signaling pathway or engaging with the same target. Since the Novartis GCM data are built on a foundation on pathway screens,[43] this convergence on signaling pathways may be a consequence of the data structure. However, it is worth noting that the ability to modulate a signaling pathway at multiple nodes can be a valuable feature. Also, while it is certainly possible that GCM compounds may directly modulate targets already covered by chemogenomics library compounds, the profiling assay employed does not necessarily distinguish between modalities. One need only consider inhibitors of ribosome translation: cycloheximide,[44] PF-06446846,[45] homoharringtonine,[46] and SRI-41315.[47] All compounds inhibit protein synthesis by preventing ribosome function, yet each one does so through entirely separate and unique mechanisms.

In full transparency, while a significant effort was made to characterize as large of a sample of the Novartis GCM collection as possible, claims of cellular specificity, diversity of MoA, and novelty of protein target engagement could only be corroborated for a subset of the GCM collection. The sheer size of the Novartis GCM collection precludes such a comprehensive study. Additionally, while we identified a novel ligand for SLC15A4, we recognize the need for additional data before claiming a functional ligand for this target. These data are presented to demonstrate how the GCM library enabled PAL, a technology that is not usually associated with screening, to screen and identify ligands for a target of interest within a very small set of tested compounds.

As a concept, GCM fills a gap between chemogenetic libraries with highly defined targets and unbiased large diversity screening libraries. In assays with limited throughput, screening of GCM collections, in addition to chemogenomic libraries, is a promising strategy to enhance the coverage of assayable MoA space without significantly increasing the burden on throughput, as we have demonstrated with profiling assays like DRUG-seq, cell painting and PSP, or PAL screening.

https://doi.org/10.1021/acschembio.3c00737
ACS Chem. Biol. 2024, 19, 938−952

We anticipate that this will be a significant benefit for newly developing assays that are not yet sufficiently miniaturized or in screening applications lacking available automation. We anticipate that the GCM approach, the computational pipeline, and the published PubChem library will foster novel MoA discoveries and the reuse of the screening information that many institutions have collected for novel discoveries.

METHODS

GCM Pipeline. The code used to calculate the GCM compounds together with the PubChem results is published in GitHub: https://github.com/Novartis/GreyChemicalMatter.

Assay Data Preparation. All cell HTS assays were normalized to rscores according to rscore = (activity − median activity)/median absolute deviation of activity. This normalization allows for a general data-driven calling of active compounds that have activities outside the background distribution of the assays in the same manner over all assays.

Compound Clustering. Compounds were encoded by morgan2 fingerprints with RDKit[33] and chemfp[48] and the Tanimoto similarity matrix was calculated. Clustering was calculated with MCL,[49] using a Tanimoto similarity cutoff of 0.5 and a perplexity parameter of 1.8.

Assay Enrichment Profile Calculation for Chemical Clusters. For each chemical cluster, for each assay and assay direction, we calculated whether there were significantly more actives than expected from the background hit rates of the assays found in the chemical cluster. Actives were defined as compounds with rscore values of more than 3 or less than −3, i.e., all compounds with an activity outside the background activity distribution of the assays.

P-values were calculated using the Fisher exact test with alternative "greater" from the scipy.stats package, followed by "fdr_bh" multiple hypothesis correction from statsmodels. Assays with adjusted *p*-values of <0.1 were considered significantly enriched for the respective chemical clusters.

One challenge using observed assay data that was generated for purposes different from calculating GCM cluster profiles is that chemical clusters can have strongly varying amounts of data from the different assays, which makes it difficult to compare compound profiles over multiple assays. Therefore, we wanted to discard assays with very small amounts of data in a cluster compared with assays with more data. For that purpose, we identified the assay with most data points in the chemical cluster and only kept additional assays which had the same number of compounds tested. Such assays are marked as "qualified for profile" in our data.

Assess Chemical Clusters by Their Assay Enrichment Profiles. Chemical clusters were evaluated based on their assay enrichment profiles to determine whether they qualify as GCM. GCM clusters were defined as clusters matching the following criteria:

(1) More than 10 assays tested and qualified for the profile to guarantee a minimum number of data to assess the selectivity of the cluster.

(2) At least one assay enriched to focus on active compounds.

(3) Less than 20% of assays in the cluster enriched and max 5 assays enriched, to prioritize clusters with selective biology and avoid broad toxic and unspecific MoAs or artifact effects of compounds.

(4) Less than 200 compounds with data in any of the assays, to avoid chemical clusters that are too large which might be driven by multiple nonoverlapping MoAs with multiple SAR structures.

Calculate Compound Profile Scores. Compounds profile scores were calculated using the profile score formula (from the main section) to prioritize compounds with strong effects on enriched assays in the enriched assay activity directions and with little effects on other assays. Compounds are only considered active if they have at least one rscore >3 in an enriched assay in the enriched direction; otherwise, they are considered inactive.

PubChem GCM. All PubChem assay data was downloaded from NCBI via "rsync --copy --links --recursive --times --verbose rsync://ftp.ncbi.nlm.nih.gov/pubchem/Bioassay2/Concise/CSV/Data/data/" on January 27, 2021. The dataset was filtered to cell-based assays using metadata from PubChem, retaining 3900 assays as input for the GCM pipeline.

ChEMBL Annotations. Compound clinical phases and target activity annotations were obtained from our in-house integrated version of ChEMBL release 31.

pQSAR Predictions. Affinity predictions for internal assays were predicted by the pretrained pQSAR models available at Novartis. Prediction models were filtered according to standard criteria with >50 compounds in the training set, the pIC_{50} standard deviation was >0.5, correlation of prediction with the experiment from 5-fold cross validation (Q2orig) was >0.3, and the correlation of prediction with the experiment on external test set (R2ext) was > 0.3. As pQSAR predicts affinities for individual assays, assays were aggregated at the target gene level and only the most potent predictions were retained for each compound and target gene. For calling hits, we used the *z*-score normalization (*z*-score = (pIC_{50} − mean pIC_{50})/standard deviation of pIC_{50}) of predicted pIC_{50} values and considered all predictions with a *z*-score of >3 as binders for that target gene. The *z*-scores are normalized using the predicted activities of the 6 million compounds in the Novartis compound database.

pQSAR is a massively multitask regression model covering 9000 assays (about 1/2 cellular), 2 million compounds, and 20 million AC_{50}s. On the very challenging "realistic" test set that was comprised of the singletons and smallest clusters for each assay, the models have a median Pearson correlation with experiment of $r^2 = 0.54$, statistically comparable to 4-concentration AC_{50}s.[41] Since GCMs were derived from the same compound set as the dose response training data for pQSAR, and many of the pQSAR dose response assays cover the same biology as the single concentration assays used for GCM, one can assume there would likely be representatives of GCM clusters with measured activity for the pQSAR MoAs, if the GCM compounds were active on the respective assays. Therefore, we assume the predictions for GCMs are generally within the applicability domains of the multitask models.

iTRACE (Isobaric Tagging and Reactivity-Based Acid Cleavable Enrichment) Covalent Chemical Proteomics. HEK293T cells were seeded at 1×10^6 cells per 15 cm dish and cultured until confluent. Cells were then treated with DMSO or test compound at 50 μM for 1 h in triplicate. Cells were washed and pelleted before resuspension in 50 mM 5% glycerol, 150 mM NaCl, 1.5 mM $MgCl_2$, 0.8% NP-40, and then lysed by probe sonication (with an amplitude of 10, 1 s on/1 s off, for 30s). Lysates were clarified by centrifugation at 1000 rpm for 10 min at 4 °C. One milligram (1 mg) per sample was treated with the cysteine reactive biotin iodoacetamide DADPS probe (dialkoxydiphenylsilane) from Click Chemistry Tools at 500 μM for 1 h at room temperature (RT). Excess biotin probe was removed by cleanup with a cold acetone crash at −20 °C for 1 h. Acetone was removed and pellet was air-dried for 10 min and resuspended in 0.1% Rapigest and 200 mM EPPS. Samples were reduced with 2 mM DTT for 15 min at 65 °C and alkylated with 55 mM iodoacetamide for 1 h in darkness at RT. Each sample was digested overnight with 20 μg LysC/trypsin (Promega) at 37 °C. Samples were diluted to 0.8 mL with 0.1% SDS and incubated with 100 μL of high-capacity ultralink streptavidin agarose (Thermo) for 1 h at RT on rotator. Beads were transferred to a 1.2 μm filter plate and washed a total of 15 times; 5× 0.1%SDS and 5× PBS and 5× distilled water. Peptides were eluted by cleaving the DADPS linker with 300 μL of 2.5% formic acid for 1 h at RT. The eluted peptides were collected by centrifugation and concentrated by speedvac. The eluted DADPS labeled cysteine-containing peptides were resuspended in 100 μL of 50 mM TEAB and 20 μL of each TMTpro (Thermo) isobaric label in acetonitrile was added for 1 h at RT. Sixteen xTMTpro-labeled samples were pooled and fractionated on a Dionex LC with an Xbridge 2.1 × 150 mm C18 column at pH 10. The resulting fractions were concatenated to 15 fractions and dissolved in 20 μL of 2.5% formic acid. Fractions were analyzed by nanoLC-MS/

https://doi.org/10.1021/acschembio.3c00737
ACS Chem. Biol. 2024, 19, 938−952

MS using an Easy-nLC 1200 high-performance liquid chromatography system (Thermo) interfaced with an Orbitrap Eclipse Tribrid Mass Spectrometer (Thermo). A Ionopticks (75 μm × 250 mm) Aurora Ultimate C18 column (at 45 °C) was used to separate iTRACE enriched cysteine peptides at 300 nL/min using a mobile phase A: 2% acetonitrile + 0.1% formic acid in water and a mobile phase B: 98% acetonitrile + 0.1% formic acid in water over a gradient of 3%−45% B over 90 min. TMTpro-labeled peptides were analyzed using SPS-RTS (real time search) on an Orbitrap Eclipse. MS1 scans were acquired from m/z 400−1400 at a mass resolution of 100 000 with AGC set to auto and a charge state of 2−5. SPS-RTS scans were searched using Comet with FDR filtering on; MS2 CID spectra were acquired with isolation window of 0.7 in Turbo mode. DADPS Modified TMTpro-labeled cysteine peptides quantified using SPS with a HCD collision energy of 55% and a resolution of 55k. Raw files were processed using Proteome Discoverer 2.5. Data were searched against a reference human proteome using Mascot.

Photoaffinity-Based Chemical Proteomics. Author: After replacement of normal growth media with Phenol-Red-free Optimem (ThermoFisher P/N 11058021), HEK293T cells cultured in 15 cm dishes were treated with vehicle or GCM PAL probe (1 μM, 2 h, 37 °C), all treatments performed in duplicate. Probe engaged targets where photo cross-linked at 4 °C with a 40 W UV lamp (UVP, Model P/N 95-0043-04). After harvest, cell pellets were resuspended in 250 μL lysis buffer (50 mM HEPES pH 8, 150 mM NaCl, 1.5 mM $MgCl_2$, 5% glycerol) containing 4% SDS, vortexed 30 s, and heated (5 min, 95 °C). Subsequently, a probe sonicator was used to reduce sample viscosity. Copper-catalyzed azide−alkyne cycloaddition (CuAAC) was performed by sequential addition of 650 μL lysis buffer, 20 μL biotin picolyl azide (5 mM in DMSO), 58.8 μL TBTA (1.7 mM in 4:1 t-BuOH:DMSO), 20 μL $CuSO_4$ (50 mM in H_2O and 20 μL TCEP (50 mM in H_2O) to prepared lysates. After 2 h of incubation at 37 °C, samples were precipitated with the addition of 4 mL of cold acetone and incubation at −80 °C for 1 h. Precipitated protein was collected by centrifugation (2000 g) and resolubilized in 1% SDS-PBS (1 mL). After determining the protein concentration (ThermoFisher P/N 22662), normalized total protein amounts (3−5 mg, 1 mL) were added to 50 μL of Neutravidin Agarose Resin (ThermoFisher P/N 29201) and incubated with end-over-end rotation overnight at RT. Samples were washed with 1 mL, three times each: PBS (0.4% NP-40, 1 mM DTT), PBS (1 mM DTT). Afterward, enriched samples were eluted in 80 μL 2× LDS buffer (ThermoFisher P/N 84788), and alkylated with 5 μL iodoaceteamide (1 M in H_2O, 1 h). Detergent was removed from samples using detergent removal spin columns (ThermoFisher P/N 87777) and trypsinized in solution overnight (5 μL, 0.02 μg/μL, ThermoFisher P/N 90057). Samples were labeled with TMT10plex isobaric tags (ThermoFisher P/N 90110), according to manufacturer's instructions. Tagged samples were combined, dried using a vacuum concentrator, and resuspended in 100 μL 0.1% formic acid in H_2O. Samples were fractionated by high-pH reverse-phase chromatography, and quantitative TMT-based proteomic data acquistion was performed as described previously.[50] Acquired MS data was processed using ThermoFisher Proteome Discoverer software. Trypsin cleavage specificity (cleavage at K, R, except if followed by P) allowed for up to two missed cleavages. Cysteine carbidomethylation was set as a fixed modification; methionine and TMT modification of N-termini and lysine residues were set as variable. Summed abundances with the most confident centroid selected from 20 ppm window were used for reporter ion ratio calculation with ANOVA statistical analysis to estimate differential abundance significance. Data were filtered for only high-confidence protein identifications with a <1% FDR cutoff derived from >2 unique quantified peptides.

SLC15A4 Protein Expression. Recombinant human SLC15A4 including a C-terminal cleavable eGFP-TwinStrep-His tag was expressed in HEK293 ExpiF cells via PEI max mediated transient transfection. Cultures were supplemented with 3 mM sodium butyrate and incubated for 3 days at 33 °C.

SLC15A4 Protein Purification. Pellet from 3.6 L of culture was lyzed with dispersion homogenizer in high salt HEPES-based buffer at pH 7.4, followed by wash and clarification from soluble material at 38.4 kg. Target membrane protein was solubilized for 150 min with 1% of DDM/CHS and clarified by ultracentrifugation at 149 kg. Purification occurs via Strep-affinity batch binding, followed by gravity purification and biotin elution. The SLC15A4 containing fraction were pooled and cleaved with HRV 3C enzyme overnight at +4 °C and finally loaded on SEC column for polishing.

The final and highly pure pool was concentrated at 100 kDa cutoff to ∼1 mg mL^{-1}, corresponding to yields of ∼0.25 mg/L of culture.

All buffers contained 0.03% DDM (0.006% CHS), and purification steps were carried out at +4 °C.

This material consistently gave, upon NanoDSF Prometheus analysis, a melting temperature of ∼58 °C, with T_m shifts observed upon specific compound addition.

SLC15A4 Nanodifferential Scanning Fluorimetry (nanoDSF). The nanodifferential scanning fluorimetry (nanoDSF) is based on intrinsic protein fluorescence using aromatic residues (tryptophan, tyrosine). nanoDSF measures the changes in the intrinsic fluorescence intensity ratio (350:330 nm), as a function of temperature.

The Prometheus NT.48 instrument (NanoTemper Technologies) was used to determine the melting temperatures of SLC15A4 in the presence and absence of compounds. The capillaries (high sensitivity) were filled with 10 μL of sample containing 0.2 mg mL^{-1} SLC15A4 diluted in purification buffer (refer to protein purification). A temperature gradient of 1 °C min^{-1} from 25 °C to 85 °C was applied, and the ratio of intrinsic protein fluorescence at 350:330 nm was recorded. Small molecules were added to the final concentration of 50 μM with a DMSO content of 5% (v/v). Protein stability was not affected by DMSO additions up to 6% (v/v). Apo protein was measured in quadruplets, and all measurements containing compounds were performed in duplicate. A control compound was included during every assay run to monitor the assay performance. The protein stabilization upon small molecule addition was recorded as dT_m in °C [T_m(compound) − T_m(apo)]. The nanoDSF data analysis was performed using PR.ThermControl v2.0.4 software (NanoTemper Technologies).

Cell Painting (Morphological Profiling Assay). Morphological profiling was performed as described in the previously published Cell Painting protocol.[37] Briefly, U2OS cells were plated at 400 cells/well in a 1536-well microplate format and incubated with compounds for 24 h at 37 °C. Four replicate plates were generated using four compound concentrations (0.01, 0.1, 1, and 10 μM). The Cell Painting assay labels eight cellular subcompartments with six fluorescent probes acquired in five imaged channels: nuclear DNA (Hoechst 33342: Invitrogen), endoplasmic reticulum (Concanavalin A/Alexa Fluor 488 conjugate: Invitrogen), nucleoli, and cytoplasmic RNA (SYTO 14 green fluorescent nucleic acid stain: Invitrogen), F-actin (phalloidin/Alexa Fluor 568 conjugate: Invitrogen), Golgi and plasma membrane (wheat-germ agglutinin/Alexa Fluor 555 conjugate: Invitrogen), and mitochondria (MitoTracker Deep Red: Invitrogen). Images were captured at 4 sites/well at 20× magnification (0.75NA), using a confocal microscope (Model CV8000, Yokogawa Corporation, Japan).

The morphological signatures were obtained from the fluorescent images using approaches that have also been previously described.[37,51] In brief, CellProfiler software (version 2.2.0,[52]) was used to correct each fluorescent image for uneven illumination, followed by extraction of nuclei counts in addition to ∼2400 single-cell morphological features (e.g., shape, intensity, texture, adjacency, etc.). The nuclei counts were normalized by robust Z-scoring using the median and median absolute deviation (MAD) of the neutral controls of each plate. The single-cell features were summarized into a morphological profile for each treatment by taking the per-feature median for each well. These profiles were then corrected for spatial plate-based heterogeneities using robust local regression, followed by per-feature normalization by robust Z-scoring of each feature against the corresponding neutral control feature on the same plate. The final feature set were selected by removing those features with low variance and high collinearity and/or correlation. For each profile, the

https://doi.org/10.1021/acschembio.3c00737
ACS Chem. Biol. 2024, 19, 938−952

Mahalanobis distance was calculated from the reduced feature set. Finally, the morphological signature was generated by removing the replicate with the lowest correlation to the others, followed by taking the per-feature median across the remaining replicates.

A given compound/concentration signature was considered to express a phenotype (or deemed "active", i.e., distinguishable from the neutral controls), if the following was true: (1) the median pairwise replicate correlation was higher than the 95th percentile of the pairwise neutral control replicate correlation distribution; and (2) the Mahalanobis distance was higher than the 95th percentile of the neutral control Mahalanobis distance distribution. Since the plate edge wells tended to be prone to incubator temperature variations and media evaporation, treatments were considered active using criterion (2) for these wells. The distribution of the phenotype-producing treatments was visualized via UMAPs of the high-dimensional image feature space.[53]

Nuclei counts were score-normalized using median and mad (median absolute deviation) of the neutral controls of each plate. The final value for each treatment was derived from the median over all replicates.

DRUG-seq (Transcriptions Profiling Assay). The DRUG-seq assay was run and analyzed as described in ref 35.

PSP (Promotor Signature Profiling Assay). PSP was run and analyzed as described in this publication [PSP, see ref 34].

Compounds were considered active if they had a DR_{50} value of >0.1 in at least one assay at time points 2 (12 h) or 3 (24 h).

CLiP (Growth Inhibition Assay Across Cancer Cell Line Panel). CLiP[40] was run and analyzed as described in ref 54. Cells in a growth medium were plated into a 1536 well plate (5 μL/well; 250 cells/well) by using a GNF Bottle Valve liquid handler. A Labcyte Echo acoustic transfer instrument was used to transfer 15 nL of compounds in DMSO to each well (final concentrations of 30, 9.5, 3, 1, 0.3, 0.1, 0.03, and 0.01 μM). The cells were then incubated (37 °C, 95% relative humidity, 5% CO_2) for 3 days and 6 h prior to addition of 4 μL of 50% Cell-Titer Glo (Promega) in water using a GNF Bottle Valve liquid handler. Plates were incubated with Cell Titer Glo for 15 min at RT prior to reading luminescence (5 s exposure) on a PerkinElmer ViewLux. For determining GI_{50} values, data were normalized to a day 0 cell count measured using a cell plate copy that was not treated with any compounds, and growth inhibition dose−response curves were calculated using Helios.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acschembio.3c00737.

Additional examples and details on PubChem GCM and Novartis GCM examples; overall properties and MoA coverage; ^{1}H NMR spectra for compounds (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Jason R. Thomas − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*; Email: jason.thomas@novartis.com

Steffen Renner − *Novartis Biomedical Research, Basel 4056, Switzerland*; orcid.org/0000-0002-0720-5629; Email: steffen.renner@novartis.com, steffen.renner@gmail.com

Authors

Claude Shelton, IV − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*

Jason Murphy − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*

Scott Brittain − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*

Mark-Anthony Bray − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*

Peter Aspesi − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*

John Concannon − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*

Frederick J. King − *Novartis Biomedical Research, San Diego, California 92121, United States*

Robert J. Ihry − *Novartis Biomedical Research, San Diego, California 92121, United States*; orcid.org/0000-0003-4519-3136

Daniel J. Ho − *Novartis Biomedical Research, San Diego, California 92121, United States*

Martin Henault − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*

Andrea Hadjikyriacou − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*

Marilisa Neri − *Novartis Biomedical Research, Basel 4056, Switzerland*

Frederic D. Sigoillot − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*

Helen T. Pham − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*

Matthew Shum − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*

Louise Barys − *Novartis Biomedical Research, Basel 4056, Switzerland*

Michael D. Jones − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*

Eric J. Martin − *Novartis Biomedical Research, Emeryville, California 94608, United States*; orcid.org/0000-0001-7040-5108

Anke Blechschmidt − *Novartis Biomedical Research, Basel 4056, Switzerland*

Sébastien Rieffel − *Novartis Biomedical Research, Basel 4056, Switzerland*

Thomas J. Troxler − *Novartis Biomedical Research, Basel 4056, Switzerland*

Felipa A. Mapa − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*

Jeremy L. Jenkins − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*

Rishi K. Jain − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*

Peter S. Kutchukian − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*

Markus Schirle − *Novartis Biomedical Research, Cambridge, Massachusetts 02139, United States*; orcid.org/0000-0003-4933-2623

Complete contact information is available at: https://pubs.acs.org/10.1021/acschembio.3c00737

Author Contributions

[▽]Jason R. Thomas and Steffen Renner had equal contributions to this work.

Funding

No external funding was used for our studies.

Notes

The authors declare the following competing financial interest(s): All authors were or are current employees of Novartis and may own shares.

https://doi.org/10.1021/acschembio.3c00737
ACS Chem. Biol. 2024, 19, 938−952

■ ACKNOWLEDGMENTS

We would like to thank our Novartis colleagues for valuable input and discussions, in particular from project teams that screened GCM compounds. Specifically, we acknowledge J. Davies, P. Farmer, and E. Lounkine for great discussions and input on the concept of GCM, and the hithub team for making the Novartis screening data and metadata easily available for data mining efforts such as GCM.

■ REFERENCES

(1) Canham, S. M.; Wang, Y.; Cornett, A.; Auld, D. S.; Baeschlin, D. K.; Patoor, M.; Skaanderup, P. R.; Honda, A.; Llamas, L.; Wendel, G.; et al. Systematic chemogenetic library assembly. *Cell Chem. Biol.* **2020**, *27* (9), 1124−1129.

(2) Elkins, J. M.; Fedele, V.; Szklarz, M.; Abdul Azeez, K. R.; Salah, E.; Mikolajczyk, J.; Romanov, S.; Sepetov, N.; Huang, X.-P.; Roth, B. L.; et al. Comprehensive characterization of the published kinase inhibitor set. *Nat. Biotechnol.* **2016**, *34* (1), 95−103.

(3) Liu, Y.; Platchek, M.; Kement, B.; Bee, W. T.; Truong, M.; Zeng, X.; Hung, S.; Lin, H.; Morrow, D.; Kallal, L. A.; et al. A novel approach applying a chemical biology strategy in phenotypic screening reveals pathway-selective regulators of histone 3 K27 trimethylation. *Mol. Biosyst.* **2014**, *10* (2), 251−257.

(4) Carter, A. J.; Kraemer, O.; Zwick, M.; Mueller-Fahrnow, A.; Arrowsmith, C. H.; Edwards, A. M. Target 2035: probing the human proteome. *Drug Discovery Today* **2019**, *24* (11), 2111−2115.

(5) Hartenfeller, M.; Renner, S.; Jacoby, E. Reaction-driven de novo design: a keystone for automated design of target family-oriented libraries. *De novo Mol. Des.* **2013**, 245−266.

(6) Renner, S.; Popov, M.; Schuffenhauer, A.; Roth, H.-J.; Breitenstein, W.; Marzinzik, A.; Lewis, I.; Krastel, P.; Nigsch, F.; Jenkins, J.; Jacoby, E. Recent trends and observations in the design of high-quality screening collections. *Future Med. Chem.* **2011**, *3* (6), 751−766.

(7) Schneider, P.; Schneider, G. Privileged structures revisited. *Angew. Chem., Int. Ed.* **2017**, *56* (27), 7971−7974.

(8) Over, B.; Wetzel, S.; Grütter, C.; Nakai, Y.; Renner, S.; Rauh, D.; Waldmann, H. Natural-product-derived fragments for fragment-based ligand discovery. *Nat. Chem.* **2013**, *5* (1), 21−28.

(9) Renner, S.; Van Otterlo, W. A.; Dominguez Seoane, M.; Möcklinghoff, S.; Hofmann, B.; Wetzel, S.; Schuffenhauer, A.; Ertl, P.; Oprea, T. I.; Steinhilber, D.; et al. Bioactivity-guided mapping and navigation of chemical space. *Nat. Chem. Biol.* **2009**, *5* (8), 585−592.

(10) Wetzel, S.; Klein, K.; Renner, S.; Rauh, D.; Oprea, T. I.; Mutzel, P.; Waldmann, H. Interactive exploration of chemical space with Scaffold Hunter. *Nat. Chem. Biol.* **2009**, *5* (8), 581−583.

(11) Heyndrickx, W.; Mervin, L.; Morawietz, T.; Sturm, N.; Friedrich, L.; Zalewski, A.; Pentina, A.; Humbeck, L.; Oldenhof, M.; Niwayama, R.et al. Melloddy: Cross-pharma federated learning at unprecedented scale unlocks benefits in qsar without compromising proprietary information. *J. Chem. Inf. Model.* **2023**..

(12) Martin, E. J.; Zhu, X.-W. Collaborative profile-QSAR: a natural platform for building collaborative models among competing companies. *J. Chem. Inf. Model.* **2021**, *61* (4), 1603−1616.

(13) Godinez, W. J.; Ma, E. J.; Chao, A. T.; Pei, L.; Skewes-Cox, P.; Canham, S. M.; Jenkins, J. L.; Young, J. M.; Martin, E. J.; Guiguemde, W. A. Design of potent antimalarials with generative chemistry. *Nat. Machine Intell.* **2022**, *4* (2), 180−186.

(14) Zhavoronkov, A.; Ivanenkov, Y. A.; Aliper, A.; Veselov, M. S.; Aladinskiy, V. A.; Aladinskaya, A. V.; Terentiev, V. A.; Polykovskiy, D. A.; Kuznetsov, M. D.; Asadulaev, A.; et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* **2019**, *37* (9), 1038−1040.

(15) Helal, K. Y.; Maciejewski, M.; Gregori-Puigjane, E.; Glick, M.; Wassermann, A. M. Public domain HTS fingerprints: design and evaluation of compound bioactivity profiles from PubChem's bioassay repository. *J. Chem. Inf. Model.* **2016**, *56* (2), 390−398.

(16) Petrone, P. M.; Simms, B.; Nigsch, F.; Lounkine, E.; Kutchukian, P.; Cornett, A.; Deng, Z.; Davies, J. W.; Jenkins, J. L.; Glick, M. Rethinking molecular similarity: comparing compounds on the basis of biological activity. *ACS Chem. Biol.* **2012**, *7* (8), 1399−1409.

(17) Petrone, P. M.; Wassermann, A. M.; Lounkine, E.; Kutchukian, P.; Simms, B.; Jenkins, J.; Selzer, P.; Glick, M. Biodiversity of small molecules−a new perspective in screening set selection. *Drug Discovery Today* **2013**, *18* (13−14), 674−680.

(18) Riniker, S.; Wang, Y.; Jenkins, J. L.; Landrum, G. A. Using information from historical high-throughput screens to predict active compounds. *J. Chem. Inf. Model.* **2014**, *54* (7), 1880−1891.

(19) Wassermann, A. M.; Lounkine, E.; Glick, M. Bioturbo similarity searching: combining chemical and biological similarity to discover structurally diverse bioactive molecules. *J. Chem. Inf. Model.* **2013**, *53* (3), 692−703.

(20) Wassermann, A. M.; Lounkine, E.; Urban, L.; Whitebread, S.; Chen, S.; Hughes, K.; Guo, H.; Kutlina, E.; Fekete, A.; Klumpp, M.; Glick, M. A screening pattern recognition method finds new and divergent targets for drugs and natural products. *ACS Chem. Biol.* **2014**, *9* (7), 1622−1631.

(21) Baell, J. B.; Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **2010**, *53* (7), 2719−2740.

(22) Seidler, J.; McGovern, S. L.; Doman, T. N.; Shoichet, B. K. Identification and prediction of promiscuous aggregating inhibitors among known drugs. *J. Med. Chem.* **2003**, *46* (21), 4477−4486.

(23) Wassermann, A. M.; Lounkine, E.; Hoepfner, D.; Le Goff, G.; King, F. J.; Studer, C.; Peltier, J. M.; Grippo, M. L.; Prindle, V.; Tao, J.; et al. Dark chemical matter as a promising starting point for drug lead discovery. *Nat. Chem. Biol.* **2015**, *11* (12), 958−966.

(24) Roche, O.; Schneider, P.; Zuegge, J.; Guba, W.; Kansy, M.; Alanine, A.; Bleicher, K.; Danel, F.; Gutknecht, E.-M.; Rogers-Evans, M.; et al. Development of a virtual screening method for identification of "frequent hitters" in compound libraries. *J. Med. Chem.* **2002**, *45* (1), 137−142.

(25) Varin, T.; Gubler, H.; Parker, C. N.; Zhang, J.-H.; Raman, P.; Ertl, P.; Schuffenhauer, A. Compound set enrichment: a novel approach to analysis of primary HTS data. *J. Chem. Inf. Model.* **2010**, *50* (12), 2067−2078.

(26) Varin, T.; Schuffenhauer, A.; Ertl, P.; Renner, S. Mining for bioactive scaffolds with scaffold networks: improved compound set enrichment from primary screening data. *J. Chem. Inf. Model.* **2011**, *51* (7), 1528−1538.

(27) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* **2021**, *49* (D1), D1388−D1395.

(28) Chen, J.; Liu, T.; Dong, X.; Hu, Y. Recent development and SAR analysis of colchicine binding site inhibitors. *Mini Rev. Med. Chem.* **2009**, *9* (10), 1174−1190.

(29) Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrián-Uhalte, E.; et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **2017**, *45* (D1), D945−D954.

(30) Bunnage, M. E.; Chekler, E. L. P.; Jones, L. H. Target validation using chemical probes. *Nat. Chem. Biol.* **2013**, *9* (4), 195−199.

(31) Schuffenhauer, A.; Schneider, N.; Hintermann, S.; Auld, D.; Blank, J.; Cotesta, S.; Engeloch, C.; Fechner, N.; Gaul, C.; Giovannoni, J. Evolution of Novartis' small molecule screening deck design. *J. Med. Chem.* **2020**, *63* (23), 14425−14447.

(32) Wang, Y.; Cornett, A.; King, F. J.; Mao, Y.; Nigsch, F.; Paris, C. G.; McAllister, G.; Jenkins, J. L. Evidence-based and quantitative prioritization of tool compounds in phenotypic drug discovery. *Cell Chem. Biol.* **2016**, *23* (7), 862−874.

(33) Landrum, G. *RDKit: Open-source cheminformatics, Release 1.1-79*; 2013: 4, https://www.rdkit.org.

https://doi.org/10.1021/acschembio.3c00737
ACS Chem. Biol. 2024, 19, 938−952

(34) King, F. J.; Selinger, D. W.; Mapa, F. A.; Janes, J.; Wu, H.; Smith, T. R.; Wang, Q.-Y.; Niyomrattanakitand, P.; Sipes, D. G.; Brinker, A.; Porter, J. A.; Myer, V. E. Pathway reporter assays reveal small molecule mechanisms of action. *JALA: J. Assoc. Lab. Automation* **2009**, *14* (6), 374−382.
(35) Li, J.; Ho, D. J.; Henault, M.; Yang, C.; Neri, M.; Ge, R.; Renner, S.; Mansur, L.; Lindeman, A.; Kelly, B.; et al. DRUG-seq Provides Unbiased Biological Activity Readouts for Neuroscience Drug Discovery. *ACS Chem. Biol.* **2022**, *17* (6), 1401−1414.
(36) Ye, C.; Ho, D. J.; Neri, M.; Yang, C.; Kulkarni, T.; Randhawa, R.; Henault, M.; Mostacci, N.; Farmer, P.; Renner, S.; et al. DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. *Nat. Commun.* **2018**, *9* (1), 4307.
(37) Bray, M.-A.; Singh, S.; Han, H.; Davis, C. T.; Borgeson, B.; Hartland, C.; Kost-Alimova, M.; Gustafsdottir, S. M.; Gibson, C. C.; Carpenter, A. E. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* **2016**, *11* (9), 1757−1774.
(38) Cimini, B. A.; Chandrasekaran, S. N.; Kost-Alimova, M.; Miller, L.; Goodale, A.; Fritchman, B.; Byrne, P.; Garg, S.; Jamali, N.; Logan, D. J.; et al. Optimizing the Cell Painting assay for image-based profiling. *Nat. Protoc.* **2023**, *18*, 1981.
(39) Reisen, F.; Sauty De Chalon, A.; Pfeifer, M.; Zhang, X.; Gabriel, D.; Selzer, P. Linking phenotypes and modes of action through high-content screen fingerprints. *Assay Drug Dev. Technol.* **2015**, *13* (7), 415−427.
(40) Barretina, J.; Caponigro, G.; Stransky, N.; Venkatesan, K.; Margolin, A. A.; Kim, S.; Wilson, C. J.; Lehár, J.; Kryukov, G. V.; Sonkin, D.; et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **2012**, *483* (7391), 603−607.
(41) Martin, E. J.; Polyakov, V. R.; Zhu, X.-W.; Tian, L.; Mukherjee, P.; Liu, X. All-assay-Max2 pQSAR: activity predictions as accurate as four-concentration IC50s for 8558 Novartis assays. *J. Chem. Inf. Model.* **2019**, *59* (10), 4450−4459.
(42) Superti-Furga, G.; Lackner, D.; Wiedmer, T.; Ingles-Prieto, A.; Barbosa, B.; Girardi, E.; Goldmann, U.; Gürtl, B.; Klavins, K.; Klimek, C.; et al. The RESOLUTE consortium: unlocking SLC transporters for drug discovery. *Nat. Rev. Drug Discovery* **2020**, *19* (7), 429−430.
(43) Fishman, M. C.; Porter, J. A. A new grammar for drug discovery. *Nature* **2005**, *437* (7058), 491−493.
(44) Myasnikov, A. G.; Kundhavai Natchiar, S.; Nebout, M.; Hazemann, I.; Imbert, V.; Khatter, H.; Peyron, J.-F.; Klaholz, B. P. Structure−function insights reveal the human ribosome as a cancer target for antibiotics. *Nat. Commun.* **2016**, *7* (1), 12856.
(45) Lintner, N. G.; McClure, K. F.; Petersen, D.; Londregan, A. T.; Piotrowski, D. W.; Wei, L.; Xiao, J.; Bolt, M.; Loria, P. M.; Maguire, B. Selective stalling of human translation through small-molecule engagement of the ribosome nascent chain. *PLoS Biol.* **2017**, *15* (3), No. e2001882.
(46) Gürel, G.; Blaha, G.; Moore, P. B.; Steitz, T. A. U2504 determines the species specificity of the a-site cleft antibiotics:: The structures of tiamulin, homoharringtonine, and bruceantin bound to the ribosome. *J. Mol. Biol.* **2009**, *389* (1), 146−156.
(47) Coelho, J. P.; Yip, M. C.; Oltion, K.; Taunton, J.; Shao, S. The eRF1 degrader SRI-41315 acts as a molecular glue at the ribosomal decoding center. *Nat. Chem. Biol.* **2024**, 1−8.
(48) Dalke, A. The chemfp project. *J. Cheminform.* **2019**, *11* (1), 1−21.
(49) Van Dongen, S. Graph clustering via a discrete uncoupling process. *SIAM J. Matrix Anal. Appl.* **2008**, *30* (1), 121−141.
(50) Thomas, J. R.; Brittain, S. M.; Lipps, J.; Llamas, L.; Jain, R. K.; Schirle, M. A Photoaffinity Labeling-Based Chemoproteomics Strategy for Unbiased Target Deconvolution of Small Molecule Drug Candidates In *Proteomics for Drug Discovery. Methods in Molecular Biology, vol 1647*; Lazar, I.; Kontoyianni, M.; Lazar, A., Humana Press: New York, NY, 2017. https://doi.org/10.1007/978-1-4939-7201-2_1.org/10.1007/978-1-4939-7201-2_1
(51) Caicedo, J. C.; Cooper, S.; Heigwer, F.; Warchal, S.; Qiu, P.; Molnar, C.; Vasilevich, A. S.; Barry, J. D.; Bansal, H. S.; Kraus, O.; et al. Data-analysis strategies for image-based cell profiling. *Nat. Methods* **2017**, *14* (9), 849−863.
(52) 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. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **2006**, *7*, R100.
(53) McInnes, L.; Healy, J.; Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv 2018. arXiv preprint, arXiv:1802.03426 1802.
(54) Isobe, Y.; Okumura, M.; McGregor, L. M.; Brittain, S. M.; Jones, M. D.; Liang, X.; White, R.; Forrester, W.; McKenna, J. M.; Tallarico, J. A.; et al. Manumycin polyketides act as molecular glues between UBR7 and P53. *Nat. Chem. Biol.* **2020**, *16* (11), 1189−1198.

https://doi.org/10.1021/acschembio.3c00737
ACS Chem. Biol. 2024, 19, 938−952