COMMUNICATION

Cell painting transfer increases screening hit rate

Ethan Cohen[1,2], Maxime Corbe[1,3], Cláudio A. Franco[4,5], Francisca F. Vasconcelos[4], Franck Perez[3,6], Elaine Del Nery[3], Guillaume Bollot[2] and Auguste Genovesio[1,*]

[1]Computational Bioimaging and Bioinformatics, Institut de Biologie de l'Ecole Normale Supérieure, PSL University, Paris, France
[2]Synsight, 4 Rue Pierre Fontaine, 91000 Évry-Courcouronnes, France
[3]Biophenics Laboratory, Institut Curie, PSL Research University, Department of Translational Research, Cell and Tissue Imaging Facility (PICT-IBiSA), Paris, France
[4]Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal
[5]Católica Medical School, Católica Biomedical Research Centre, Universidade Católica Portuguesa, Lisbon, Portugal
[6]Dynamics of Intra-cellular Organisation – UMR144, Institut Curie, PSL Research University, Paris, France
*Corresponding author. E-mail: auguste.genovesio@ens.psl.eu

Abstract

Drug discovery uses high throughput screening to identify compounds that interact with a molecular target or that alter a phenotype favorably. The cautious selection of molecules used for such a screening is instrumental and is tightly related to the hit rate. In this work, we wondered if cell painting, a general-purpose image-based assay, could be used as an efficient proxy for compound selection, thus increasing the success rate of a specific assay. To this end, we considered cell painting images with 30,000 molecules treatments, and selected compounds that produced a visual effect close to the positive control of an assay, by using the Frechet Inception Distance. We then compared the hit rates of such a preselection with what was actually obtained in real screening campaigns. As a result, cell painting would have permitted a significant increase in the success rate and, even for one of the assays, would have allowed to reach 80% of the hits with 10 times fewer compounds to test. We conclude that images of a cell painting assay can be directly used for compound selection prior to screening, and we provide a simple quantitative approach in order to do so.

Impact Statement

This paper proposes a simple method to increase the efficiency of a drug screening campaign. It leverages deep learning and the cell painting assay to select, before screening, a set of compounds that will be more likely to produce an expected effect. We demonstrate the performance of this hit transfer approach on 3 screening campaigns where, in the best case scenario, screening 10 times less compounds could still lead to identification of 80% of the hits. This work should be useful to drug discovery professional and researcher to reduce the cost of large screening campaign and improve the early drug discovery step.

1. Introduction

Target-based high throughput screening is currently the main approach to drug discovery. It consists of identifying active chemical compounds that interact with a preidentified target through automated parallelized experimental tests[1,2]. Another approach that has gained in popularity, especially in seeking first-in class therapeutic drugs, is phenotypic screening, during which one seeks compounds that modulate a

phenotype of interest[3,4]. To this end, a library of compounds is selected in the hope that a part of these, defined as "hits," will reproduce or approach the phenotype obtained with a positive control tool treatment.

One of the main issues that concerns all types of molecular screens is the selection of the compound library to be tested, as this will have an impact on the final hit rate. Given the size of the chemical space, which is in the order of 10^{60}, a random selection of compounds is largely suboptimal. Various methods have been proposed to reduce the attrition rate of a screening, through a better preselection of compounds to test[5]. Furthermore, specific libraries have been designed, such as the Dundee Kinase Inhibitor library to address specific modulation types[6].

Cell painting (CP) is a phenotypic image-based assay developed at the Broad Institute that offers to capture high-level and general information from images of cells under perturbations, through labeling of important organelles[7]. These organelles range from the nuclei, to nucleoli, mitochondria, actin, and tubulin networks. On top of CP's ability to monitor the state of the cells, the images of a large U2OS cell painting screen containing 30 K compounds were made publicly available[8]. The community recently benefited from the release of CP images of 120,000 compound treatment by the JUMP consortium[9].

In this work, we wondered whether CP coupled with transfer learning could provide a good proxy to efficiently preselect active compounds for a specific assay. To this end, we place ourselves in a real case scenario where we evaluate the hit rate we would have obtained on high-content screening campaigns we performed, had we used cell painting prior to screening to select 5, 10, or 15% of the compounds.

2. Results

2.1. *Transferring hits from cell painting through inception*

To evaluate the preselection gain that cell painting could offer in practice, we scanned recently performed in-house Compound Screening campaigns (CPDS at Institut Curie, Biophenics). For each assay, we considered all compounds that were both in the considered screen and were part of the 30 k compound tested in the CP assay introduced in Bray *et al.*[8] publicly available at http://gigadb.org/dataset/100351. We considered only the assays that had a positive control and at least 380 compounds and five hits in common, which ended up in the selection of three screens. For each of these three screens, we then ranked the compounds by decreasing similarity with our positive control using the CP assay. To measure similarity between images of cells perturbed by two compound treatments in the CP assay, we used the Frechet Inception Distance (FID, see Section 4)[10]. In short, computing the FID between two conditions consists in vectorizing all the images from each condition through a pretrained convolutional network (Inception) and computing the Frechet distance between these high-dimensional sample feature distributions[11]. We then examined how many of the hits we actually obtained during the screening campaign would have been ranked in the first 5%, the first 10%, and the first 15% of this list. Furthermore, to assess the significance of these results, we performed a Fisher exact test[12]. Table 1 reports the results we obtained for each screen.

Table 1. *Hit prediction using cell painting on three recent screening campaigns.*

Screens	Compounds also in CP	Hits also in CP	Top 5%	Top 10%	Top 15%
CPDS#1	400	5	2/5 (2.1e-2)	4/5 (4.e-4)	4/5 (2.1e-4)
CPDS#2	380	32	14/32 (3.7e-13)	19/32 (1.2e-13)	19/32 (1.0e-9)
CPDS#3	360	38	7/38 (1.1e-3)	10/38 (1.8e-3)	15/38 (7.5e-5)

Note. The first column lists the compound screening campaign (CPDS). The second column indicates the number of compounds tested in the screen that were also found in the cell painting assay. The third column indicates, among those compounds, how many were selected as hits in the considered screen. The three remaining columns indicate the ratio of hits obtained in the Top 5, 10, and 15% of the ranked list of compounds. The *p*-value of an exact Fisher test is reported in parentheses.

2.2. Robustness to variation of experimental settings

In any primary screen, a dedicated protocol is followed. It typically comprises many parameters such as cell line, cell seeding, incubation time, concentration, or labeling. This protocol varies from screen to screen depending on the end goal and the cell painting assay presented in Bray *et al.* does not escape this rule. The consequence is that the same perturbation in two different assays does not necessarily produce the same phenotypes. In fact, we observed that cell phenotypes can be significantly different even with a similar compound treatment (Figure 1a,b). However, interestingly, these variations in protocols did not prevent the prediction from being highly significant. The reason may be that while a difference in phenotype is obvious for the same compound across screens, it does not prevent the control and the hits to look similar in each individual screen. For instance, Figure 1a,c displays the cell painting images of the positive control condition Brefeldin A for the CPDS#2, and Piperlongumine, a compound correctly predicted as a hit using FID. The images of phenotypes in cell painting expectedly look similar as Piperlongumine was selected by this means. However, while these phenotypes in CP both look dissimilar from the phenotypes in the CPDS#2 (probably due to the difference in incubation time), they seem to again match when considering CPDS#2 only.

2.3. Robustness to the positive control relevance

CPDS#1 and 2 were designed to monitor the trafficking of fluorescently labeled reporter proteins out of the ER using the Retention Using Selective Hooks (RUSH) assay[13]. In such screens, we seek for compounds that reproduce a specific phenotype, precisely defined by the positive control compound
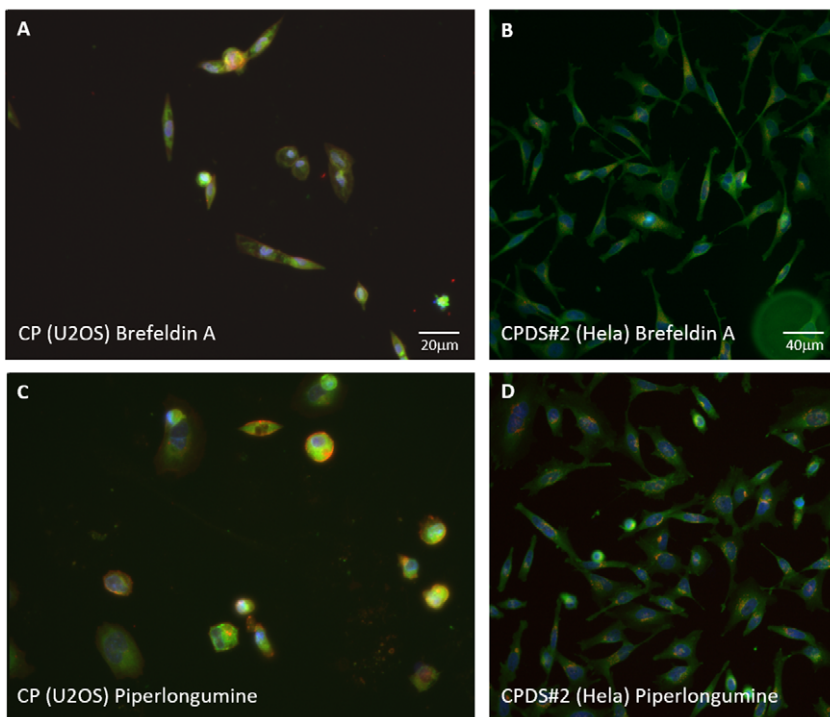

Figure 1. *(a) Effect of Brefeldin A treatment on the cell painting assay (5 μM, U2OS, incubation time: 24 hr). (b) Effect of Brefeldin A treatment on the CPDS#2 (10 μM, Hela, incubation time: 120 min). (c) Effect of Piperlongumine on the cell painting assay, a compound selected as close to Brefeldin A using FID and independently as a hit in CPDS#2. (d) Effect of Piperlongumine on the CPDS#2. Graphs (a,b) produce dissimilar phenotypes, but (a) is close to (c), and (b) is close to (d).*

Brefeldin A (BFA). On the other hand, CPDS#3 is a cell survival screen where the goal is to identify toxic compounds, but no specific positive control was used to this end. Instead, hits were formerly selected based on cell count. We then applied the following strategy: we arbitrarily chose two cytotoxic compounds (Lovastatin and Fluvastatin) as positive controls in cell painting assay in order to select a list of possibly toxic compounds[14]. Interestingly, the results for CPDS#3 displayed in Table 1 remain highly significant, which suggests that—at least for such an obvious phenotype such as cell death—it seems sufficient to grab a selection of candidates that loosely reproduces the effect of a toxic compound, to perform better than a random selection of compounds.

3. Discussion

In this work, we wondered if CP could be used in practice to select an efficient list of compounds to be tested before a screening campaign. To this end, we propose to simply use the Frechet Inception Distance as a metric in a large cell painting assay, to select compounds likely to reproduce the phenotype obtained with a positive control in a specific assay. Using already achieved screening campaigns, we quantitatively demonstrated that this strategy would have allowed, in practice, to drastically reduce the number of molecules to screen, while consistently improving hit rate. Concretely, screening only 10% of compounds using this strategy would allow us to pick up from 25% to up to 80% of the hits, depending on the assay. Overall, it seems to be constantly beneficial to use cell painting for compound selection prior to screening, rather than using a compound library directly.

Furthermore, our results suggest that, while phenotypes produced by a given compound could largely vary from CP to a specific screen, due to the variation in experimental settings, for example, cell model, labeling, compound concentration and incubation time, phenotype similarities of two compound treatments in cell painting was likely reproducible in a specific screen. Finally, our results suggest that, at least when the phenotype of interest is cell death, it is enough to arbitrarily select a few toxic compounds as positive controls, to obtain a compound selection that is more likely to lead to hits than random.

Importantly, the overlap between our compound library and the libraries used in Bray *et al.* reached about 400 compounds at best. In consequence, some highly ranked compounds in the cell painting assay could not be observed in practice in our specific screens. Therefore we anticipate that the hit ratio obtained using the suggested approach could be significantly higher in a real-case scenario, when used for future screens where all compounds close to a positive control in cell painting could be tested. We anticipate that this approach, straightforward to use, will also largely benefit from the soon to be released CP-JUMP dataset by the Broad Institute that is expected to comprise more than 120,000 compound treatments.

4. Methods

4.1. Screening assays

We performed an exhaustive search in the assays that were previously screened in-house. We computed the intersection of compounds used in each screen and the 30 k compounds tested in the CP assay introduced by Bray *et al.*[8]. We selected those assays that comprise the positive control and more than three hits in this intersection. This filtering led to only three assays among 60, mostly because the positive control in our assays was not frequently part of the CP assay. In some other cases, it was due to the discrepancies between the compound libraries leading to a small overlap.

The three compound drug screening campaigns (CPDS) retrieved this way were performed using 1,600 compounds obtained from Prestwick (1,280 off-patent small molecules, mostly approved drugs from FDA, EMA, and other agencies and a set of 320 phytochemical compounds), all tested at 10 μM.

For CPDS#1, A549 cells stably expressing the GFP-ACE2 RUSH reporter[13] were seeded in 384-wp (Viewplate 384, Perkin Elmer) for 24 hr, treated with compounds for 90 min, then subsequently treated with 40 μM of biotin for 60 min.

For CPDS#2 screen, HeLa cells stably expressing the GFP-VEGF RUSH reporter were seeded in 384-wp for 24 hr, treated with compounds for 120 min, then subsequently treated with biotin as previously described.

In both previous screens, RUSH reporters are retained in the endoplasmic reticulum (ER) through a streptavidin-based interaction that is relieved by biotin addition to cells. Brefeldin A (BFA) is added to the screen as a positive control of ER retention.

For CPDS#3 screen, we employed the Ewing sarcoma A673 cells with compounds being incubated for 24 hr, as previously described[14].

Image acquisition was performed after fixation of cells with 4% formaldehyde solution and nuclei staining with 0.2 μg/mL of DAPI using the INCell 2200 automated widefield system (GE Healthcare, Chicago, IL) at a 20× magnification (Nikon 20X/0.45, Plan Apo, CFI/60).

4.2. Cell painting

The cell painting assay is a generic high-content phenotypic assay that does not target a particular process but offers a general observation of cellular states. This assay proposes to multiplex several cell markers allowing simultaneous observation of the nucleus, endoplasmic reticulum, nucleoli, cytoplasmic RNA, actin, Golgi, plasma membrane, and mitochondria[7]. It can be produced at high throughput to observe the morphometric changes that a library of molecules can operate on cells. Large databases of images corresponding to this assay subjected to the effect of multiple molecules, but also to genetic deregulations, have recently been put online by a consortium of pharmaceutical players coordinated by the Broad Institute[9].

4.3. Image-based sample similarity

To measure the similarity between images of cells perturbed by two compound treatments in the CP assay, we used the Frechet Inception Distance (FID)[15]. FID was primarily designed to compare distributions of synthetic images generated by a generative model with real images used to train the model. Briefly, it consists in passing all images through an Inception v3 network pretrained on ImageNet, then computing the squared Wasserstein metric between the two distributions approximated as multivariate Gaussian[11]. It results in the closed-form formula:

$$\text{FID} = \left\|\mu_r - \mu_g\right\|^2 + Tr\left(\Sigma_r - \Sigma_g - 2\left(\Sigma_r \Sigma_g\right)^{1/2}\right)$$ where $N(\mu_r, \Sigma_r)$ and $N\left(\mu_g, \Sigma_g\right)$ are the gaussian fitted to the real and generated data separately.

4.4. Rank and statistical test

After ranking all compounds by decreasing order of similarity with a positive control using FID, we examined the fraction of hits we would have obtained had we decided to screen only the first 5%, the first 10%, or the first 15% of the most similar compounds in CP. After this, we tested the significance of each of these ratios compared to the total ratio of hits. To this end we performed an exact Fisher test that computes a p-value, using the hypergeometric law, which is the exact probability to obtain a ratio equal or more extreme than the observed ratio.

Authorship Contributions. A.G. and E.C. proposed cell painting transfer as a way to increase screening hit rate. E.C., M.C., and C.A.F. performed data preprocessing. E.C. implemented and performed data processing. F.F.V., F.P., E.D.N., and G.B. provided screening data and scientific insights. A.G. and E.C. wrote the manuscript. All authors edited the manuscript.

Data Availability Statement. Cell painting screening data and metadata are publicly available from the cell image library http://www.cellimagelibrary.org/pages/project_20269.

References

1. Macarron R, Banks MN, Bojanic D, *et al.* (2011) Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov* **10**, 188–195.
2. Swinney DC & Anthony J (2011) How were new medicines discovered? *Nat Rev Drug Discov* **10**, 507–519.
3. Zheng W, Thorne N & McKew JC (2013) Phenotypic screens as a renewed approach for drug discovery. *Drug Discov Today* **18**, 1067–1073.
4. Kotz J (2012) Phenotypic screening, take two. *Science-Business eXchange* **5**, 380–380.
5. Waring MJ, Arrowsmith J, Leach AR, *et al.* (2015) An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat Rev Drug Discov* **14**, 475–486.
6. Brenk R, Schipani A, James D, *et al.* (2008) Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *ChemMedChem* **3**, 435–444.
7. Bray M-A, Singh S, Han H, *et al.* (2016) Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat Protoc* **11**, 1757–1774.
8. Bray M-A, Gustafsdottir SM, Rohban MH, *et al.* (2017) A dataset of images and morphological profiles of 30 000 small-molecule treatments using the Cell Painting assay. *Gigascience* **6**, 1–5.
9. Chandrasekaran SN, Cimini BA, Goodale A, *et al.* (2022) Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. *bioRxiv.* https://www.biorxiv.org/content/10.1101/2022.01.05.475090v1.
10. Heusel M, Ramsauer H, Unterthiner T, Nessler B & Hochreiter S (2017) Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Adv Neural Inf Process Syst* **30**, 6629–6640.
11. Szegedy C, Liu W & Jia Y (2015) Going deeper with convolutions. *Proc Estonian Acad Sci Biol Ecol*.
12. Fisher RA (1922) On the interpretation of $\chi 2$ from contingency tables, and the calculation of P. *J R Stat Soc* **85**, 87.
13. Boncompain G & Perez F (2013) Fluorescence-based analysis of trafficking in mammalian cells. *Methods Cell Biol* **118**, 179–194.
14. Buchou C, Laud-Duval K, van der Ent W, *et al.* (2022) Upregulation of the mevalonate pathway through EWSR1-FLI1/EGR2 regulatory axis confers Ewing cells exquisite sensitivity to statins. *Cancers* **14**, 2327.
15. Heusel M, Ramsauer H, Unterthiner T, Nessler B, & Hochreiter S (2017) GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*, pp. 6629–6640. Red Hook, NY: Curran Associates.