RESEARCH ARTICLE

Screening for variable drug responses using human iPSC cohorts

Melpomeni Platani[1], Hao Jiang[1], Lindsay Davidson[2], Santosh Hariharan[3¤], Regis Doyonnas[3], Angus I. Lamond[1], Jason R. Swedlow[1*]

1 Molecular Cell and Developmental Biology, School of Life Sciences, University of Dundee, Dundee, Scotland, United Kingdom, **2** Human Pluripotent Stem Cell Facility, School of Life Sciences, University of Dundee, Dundee, Scotland, United Kingdom, **3** Discovery Sciences, Pfizer, Groton, Connecticut, United States of America

¤Current address: VRTX, 50 Northern Ave, Boston MA, 02210, United States of America.
* jrswedlow@dundee.ac.uk

OPEN ACCESS

Citation: Platani M, Jiang H, Davidson L, Hariharan S, Doyonnas R, Lamond AI, Swedlow JR, et al. (2025) Screening for variable drug responses using human iPSC cohorts. PLoS One 20(5): e0323953. https://doi.org/10.1371/journal.pone.0323953

Editor: Nanako Kawaguchi, Tokyo Women's Medical University, JAPAN

Received: January 10, 2025

Accepted: April 17, 2025

Published: May 30, 2025

Data availability statement: Image data was deposited to the Image Data Resource (https://idr.openmicroscopy.org) under accession number idr0169. The mass spectrometry proteomics data was deposited to the ProteomeXchange Consortium, via the PRIDE partner repository with the dataset identifier PXD050194.

Abstract

We have developed a laboratory-based drug screening platform that uses a cohort of human induced pluripotent stem cell (hiPSC) lines, derived from different donors, to predict variable drug responses of potential clinical relevance. This builds on recent findings that pluripotent hiPSC lines express a broad repertoire of gene transcripts and proteins, whose expression levels reflect the genetic identity of the donor. We demonstrate that a cohort of hiPSC lines from different donors can be screened efficiently in their pluripotent state, using high-throughput Cell Painting assays. Variable phenotypic responses between hiPSC lines were detected with a wide range of clinically approved drugs, in use across multiple disease areas. Furthermore, information on mechanisms of drug-cell interactions underlying the observed variable responses was derived by using quantitative proteomic analysis to compare sets of hiPSC lines that had been stratified objectively, based upon variable response, Cell Painting data. We propose that information derived from comparative drug screening, using curated libraries of hiPSC lines from different donors, can help to improve the delivery of safe new drugs suitable for a broad range of genetic backgrounds and sexual diversity within human populations.

Introduction

Most current drug development pipelines involve an initial laboratory-based research phase, in which disease-relevant molecular targets are identified (usually proteins) and then either small molecules, or biological effectors (e.g., antibodies), are characterised that specifically bind to, or otherwise modulate, the targets. Promising drug candidates are then progressed from animal models to human clinical trials, where they must be evaluated for efficacy and possible toxicity in human patients before they can be certified for use by regulatory authorities [1–3]. Unfortunately,

Funding: This work was supported by the Biotechnology and Biological Sciences Research Council [grant numbers BB/V010948/1, APP3732 to AIL]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: The authors Melpomeni Platani, Angus I. Lamond and Jason R. Swedlow would like to declare a conflict of interest through their affiliation with Tartan Cell Technologies, Ltd. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

these pipelines often have very high failure rates, with 80% or more drug candidates in major disease areas failing to pass clinical trials [4]. The high failure rates at late stages of drug development are a major factor in the high cost of bringing safe and effective new drugs to market.

Another significant limitation with current pre-clinical drug development pipelines is the failure to account for the suitability of drugs for use with diverse patient groups from different sexes and/or genetic backgrounds. Variable drug responses resulting from natural genetic variations in human populations is recognised to contribute to the high failure rates of drug candidates in clinical trials [5]. Fundamentally, drug development pipelines that focus on mapping drug molecules to targets in the laboratory, do not account for normal population-level differences between individuals, which affect the outcome of clinical trials on patients.

There are now many examples of drugs, used in oncology [6], anti-psychotics [7], cholesterol-lowering medications [8] and others [9], where variable efficacy and toxicity are well-documented and included in standard dosing guidelines. In many cases, differential responses result from epistatic interactions and have been mapped to genetic variants affecting the expression levels and/or activity of cell proteins and pathways, for example, metabolic enzymes, membrane transporters and signalling molecules etc. A key gap in the drug development pipeline, therefore, is the lack of assays that report on the impact of population-level phenotypic variation during the pre-clinical phases of drug development.

Several projects have now established cohorts of human induced pluripotent cell (hiPSC) lines from multiple donors using standardised protocols [10,11]. In particular, the Human Induced Pluripotent Stem Cell Initiative, (HipSci, https://www.hipsci.org), has established a library of >700 cell lines, generated from over 300 individuals, including healthy male and female donors of varied ages, as well as patient cohorts with known genetic disorders [12]. The HipSci cell lines were all reprogrammed from human skin fibroblasts, following transduction of the pluripotency transcription factors using a Sendai virus vector [12]. Many of these cell lines have been subjected to mRNA and protein expression analysis and these data are publicly available. Interestingly, comparison of mRNA and protein expression patterns in the HipSci lines showed that mRNA and protein levels in separate hiPSC lines from the same individual are more similar to each other than to lines derived from different donors. This shows that gene expression states of the HipSci lines are strongly defined by the genetic identity of the donor [12,13]. The HipSci cohort therefore provides a tractable system for measuring how human genetic variation affects cellular phenotypes, potentially including variable drug responses.

The ability of hiPSCs to be re-differentiated into different cell types, using standardised differentiation protocols, is frequently used for experimental analyses of tissue-specific responses and disease mechanisms [14]. However, using the undifferentiated, pluripotent state of hiPSCs for drug screening provides some potential advantages. For example, studies that have measured in depth gene transcript and protein expression in hiPSCs and embryonic stem cells (ESCs) [13,15], find that stem cells express a broader repertoire of gene transcripts and proteins than terminally

differentiated, primary cells and tissues, including most tumour-derived, transformed cell lines. For example, hiPSC lines express ~2.5-fold more receptor tyrosine kinases than many primary cells, or tumour cell lines. For the majority of chromosomes, > 50% of their known protein coding genes are expressed in hiPSCs ([13] and references therein). While no single cell type expresses all potential drug targets, this broad molecular expression profile in pluripotent stem cells, including many pathways targeted by clinical therapeutics, makes the use of pluripotent hiPSCs an attractive foundation for compound and drug screening.

In this report, we describe the establishment and characterisation of a laboratory-based drug screening platform that compares drug responses between multiple hiPSC lines in the pluripotent state. The platform combines a Cell Painting assay with advanced analytics and mass-spectrometry (MS)-based proteomic profiling to detect and explain variable drug responses. We validate the use of this screening platform using different clinically-approved drugs that target diverse physiological pathways and disease areas.

Results

Establishing Pluripotency in hiPSC Cohorts in High-Throughput Format

To test whether we could detect and characterize variable drug responses across a cohort of hiPSC lines from different donors, the high throughput Cell Painting assay [16] was first adapted for use with hiPSC lines grown in the pluripotent state. We reasoned that an image-based Cell Painting assay could be used to detect variable responses and stratify cell lines accordingly, followed by using quantitative MS-based proteomics to identify potential mechanisms causing the variable responses (see Fig 1A).

First, we established whether conditions could be established for culturing multiple hiPSC lines in the pluripotent state in 384 multi-well plates, suitable for high content imaging and analysis. Twenty-eight independent hiPSC lines, all derived from different donors (S1 Table, “Run1”), were propagated in mTeSR medium, plated at $3x10^4$ cells/cm^2 in 384 multi-well plates for 72 hrs and then fixed and analysed by immunofluorescence to detect the established markers of pluripotency, i.e., Oct-4, Sox-2, Nanog and TRA-1–81 (S1A Fig). All four markers showed strong, homogeneous staining.

We conclude, first, that multiple hiPSC lines can be cultured successfully in multi-well format suitable for large-scale, compound profiling by image analysis and second, that the panels of hiPSCs can be maintained in a pluripotent state throughout the 48–72 hours required for conducting the screening protocol.

Using cell painting to profile variable drug responses

Next, the set of 28 independent hiPSC lines from different donors, (hereafter, “donors”), were grown in 384 multi-well plates for 24 hrs (see Methods for details), then treated either with compounds from a library of FDA-approved drugs, or with DMSO carrier, each at a single concentration (S2 Table). After incubation for a further 24 hrs in the presence of the drugs, the cells were fixed and stained with Cell Painting markers (see Materials and Methods). Images were recorded in an automated, plate-based imager, then imported into OMERO (9). Images were segmented with respect to the ‘nuclei’, ‘cytoplasm’, ‘Golgi’ and ‘protrusions’, using a custom CellProfiler pipeline [17]. For each drug-donor combination, image features were calculated, using 871 separate feature measurements, from a total of 6,000-8,000 cells per donor-drug pair. After Robust Z’ normalisation [18–20], all measured features were scaled as the number of standard deviations different from DMSO for each of the hiPSC lines [16].

Fig 1B shows a heatmap of the measured features of the 28 donors, comparing treatment with 4 of the drugs from the library, i.e., atorvastatin, simvastatin, rapamycin and afatinib. The heatmap shows that each drug has characteristic feature patterns, but that different donors show varied magnitudes of response, visible as either darker, or lighter, horizontal stripes, respectively (Fig 1B). The feature patterns for atorvastatin and simvastatin are similar, consistent with them being

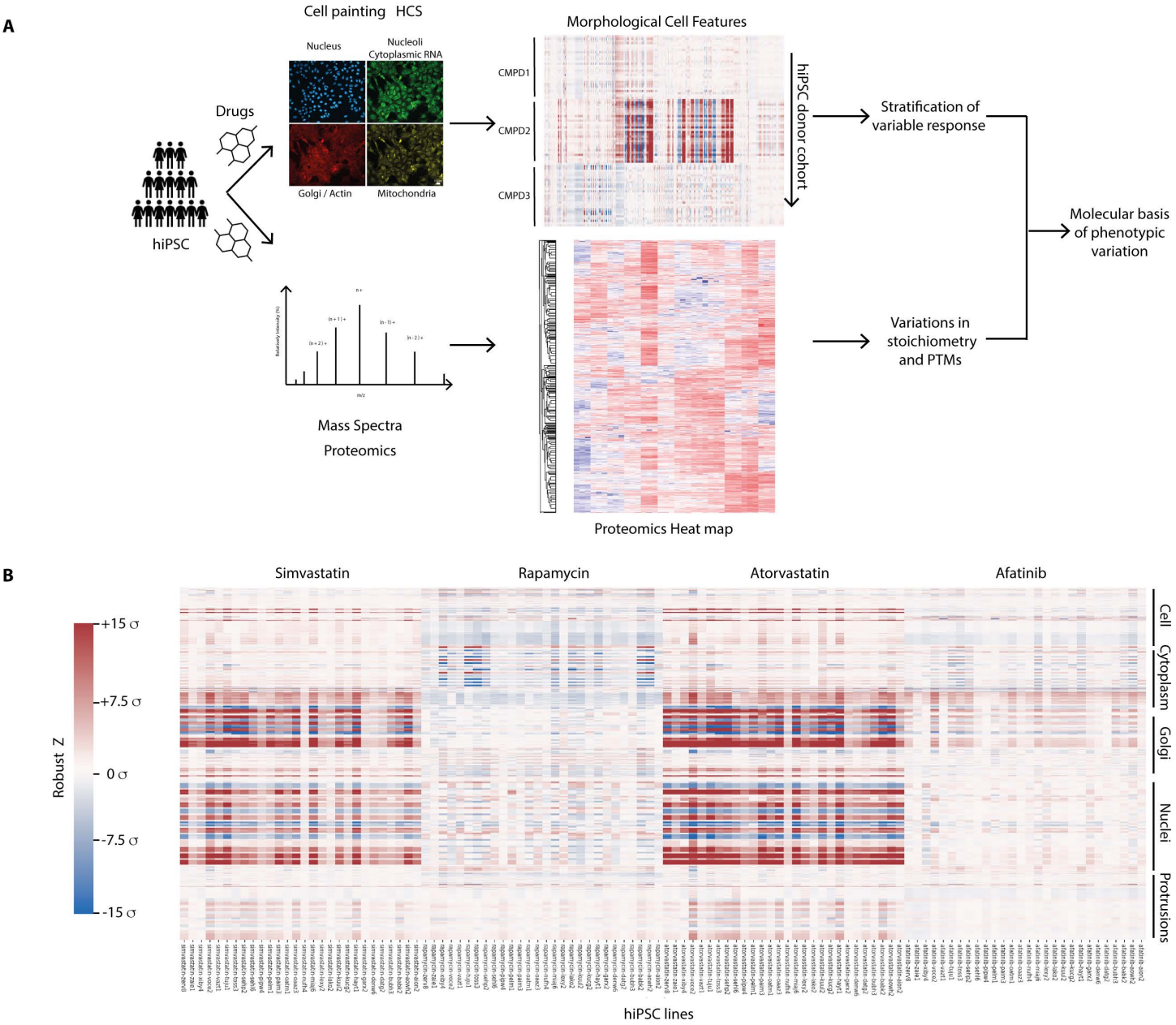

Fig 1. (A) Study overview. Quality controlled hiPSC lines from 43 donors in their pluripotent state were used for Cell Painting and proteomic analysis, by mass spectrometry, following addition of 52 FDA approved drugs. 38.2 million cells were imaged across all donors and 871 morphological traits were quantified per donor/drug combination. Variability of drug response of the hiPSC cohort was studied using both phenotypic screening and proteomics. (B) Heatmap of RobustZ scores from Cell Painting shows variations in the magnitude of drug responses across hiPSC lines. Differential responses to afatinib, atorvastatin, rapamycin and simvastatin are shown. The response to each drug creates a characteristic pattern of features, but with differences in the magnitude of the response. Color bar shows RobustZ scores normalized to DMSO expressed as standard deviations.

https://doi.org/10.1371/journal.pone.0323953.g001

competitive inhibitors of the same protein target, i.e., HMG CoA reductase, the rate-limiting enzyme of the cholesterol biosynthetic pathway. Interestingly, atorvastatin, which has a higher target affinity, also shows a stronger response than simvastatin in this assay.

These data demonstrate that the Cell Painting assay, which was originally developed for analysis of immortalised cancer lines [16], also generates characteristic phenotypic fingerprints in hiPSC lines. Moreover, the data show that Cell Painting fingerprints from hiPSC cohorts reveal variable responses between the lines from different donors. The major variation corresponds to differences in the magnitude of features.

Next, the assay was expanded to profile the hiPSC cohort responses to 52 FDA-approved drugs, which are used to treat a range of different diseases (S2 Table). A subset of the drugs tested are shown for visibility with full data available on Image Data Resource IDR [21]. To remove highly correlated features, we calculated the Spearman correlation coefficient, s, and removed all features where $s > 0.98$, which left 442 filtered features for further analysis.

Fig 2A shows a heatmap of the Cell Painting induction, calculated as the fraction of filtered features with RobustZ $> 2\sigma$ from DMSO (i.e., significant with >95% certainty) [22]. This effectively reduces the filtered features for each drug-donor pair to a single statistic, which may underestimate the complexity of phenotypes. Despite this limitation, Fig 2A shows clearly that, first, hiPSC lines derived from different donors exhibit different levels of induction when treated with the same drugs and second, that examples of such differential response are observed between cell lines for all of the 52 different drugs tested.

In all but one case, the individual cell lines show varied responses and appear to be either hyper- or hyposensitive to specific drugs. The exception was the cell line voce2, which showed a higher induction level to several, but not all drugs. Combining the results shown in Fig 1B (feature heatmap) and Fig 2A (induction heatmap), we conclude that a significant source of variable responses between donor cell lines seen in the Cell Painting assay results from differences in the magnitude of response shown by each line.

We next asked if phenotypes caused by different compounds could be distinguished across the cohort of hiPSC lines. The filtered features from all drug-donor pairs were subjected to embedding with UMAP, which fits a manifold to a high-dimensional dataset and then projects this to a 2D plane [23] (Fig 2B). This showed that compound responses clustered according to their known mechanism of action. For example, statins, everolimus and rapamycin, topoisomerase inhibitors, cytotoxics and microtubule depolymerizers all formed separate, individual clusters. Interestingly, fluphenazine, an antipsychotic that acts by blocking specific dopamine receptors in the brain and that is used to treat schizophrenia and other psychotic disorders, clustered near the statins. While clustering in UMAP is not always a reliable indicator of similarity, we also noted the pattern of induction scores for fluphenazine was similar to that observed for atorvastatin (Fig 2A). These results are consistent with reports that fluphenazine, like statins, affects lipid metabolism in human patients [24].

Besides showing that many drugs with similar modes of action cluster in this assay, the data also reveal cell lines that show variable responses to the same drugs. For example, when treated with rotenone, which disrupts microtubules [25], most hiPSC lines cluster together ("MT Modulators, Fig 2B). An exception is the nufh4 cell line (pink and blue dot), which clusters instead with hiPSC lines treated with fenbendazole. This is a drug which also causes microtubule depolymerisation, but by a different mechanism to rotenone, which inhibits microtubule assembly dynamics [26] (filled circles with trefoil, Fig 2B). The nufh4 line consistently shows a lower induction score when treated with microtubule deploymerizers (Fig 2A), which is consistent with nufh4 having a different response to rotenone than the other hiPSC lines.

To test whether these results were reproducible, a selection of the previously analysed hiPSC lines (S1 Table) were regrown from frozen stocks and analysed in Cell Painting assays after treatment with the same set of FDA-approved drugs, but this time using a different High Content Screening (HCS) imaging system (data collected on a Yokogawa CV7000 instead of InCell 2200). Using UMAP for visualisation, we observed similar clustering of Cell Painting features generated in this repeat set of assays to those collected from the original experiment that was performed 6 months previously, using independent cultures of each hiPSC line and with data acquired using a different type of HCS imaging system (S1B Fig).

To further asses the reproducibility of the assay, we compared the measured phenotypes of hiPSC lines in an additional run of the assay, executed after an interval of 2 yrs 8 months (both datasets were collected on the same InCell 2200

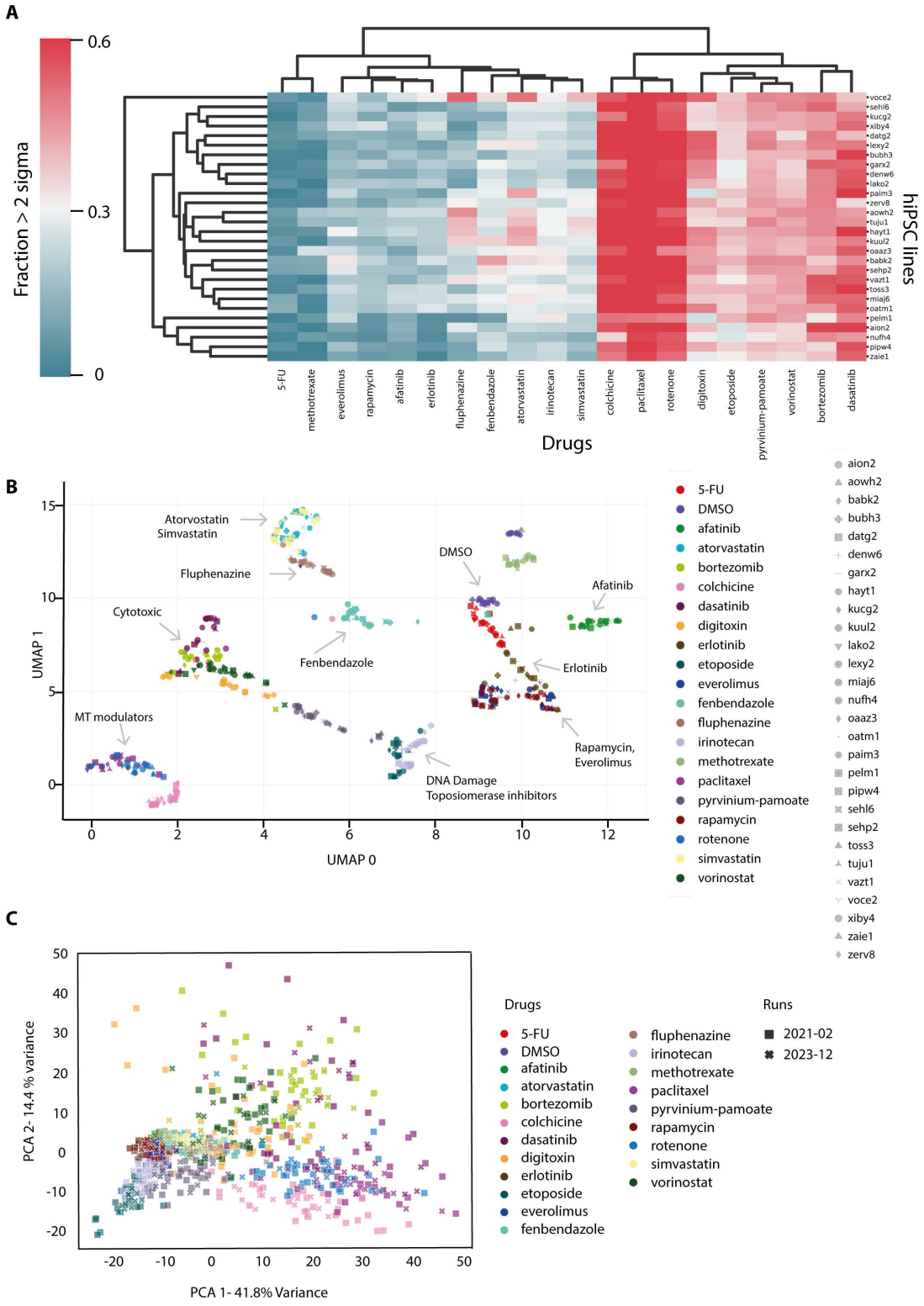

Fig 2. (A) Induction plot of drug responses for each hiPSC line shows variations in response magnitude. Induction is the fraction of features for each combination of drug and hiPSC line that differ in value by > 2σ (standard deviation) away from the DMSO control. (B) UMAP embedding of Cell Painting features measured for 28 hiPSC lines treated with different drugs. Cell lines are represented on the graph with different shapes and drugs with

different colours. Representative separated clusters are indicated by arrows. For visibility, a subset of drugs used in this screen are shown. (C) PCA analysis of Cell Painting features measured for human iPSC lines from phenotypic screens performed in 2021−02 (28 lines) and 2023−12 (18 lines). Runs are represented on the graph with different shapes ("Run 2021-02" solid squares and "Run 2023-12" crosses) and drugs with different colours. The reproducibility of the assay results on different dates is indicated by the overlap of drug treatments in the different assay runs.

https://doi.org/10.1371/journal.pone.0323953.g002

imager) but with a different cohort of hiPSC lines (S1 Table). Fig 2C shows the comparison of the two assays, using Principal Component Analysis (PCA). This further validates the comparison between the two different runs and also avoids the effects of clustering in high-dimensional space that can affect UMAP [27]. PCA analysis shows that similar compounds are indistinguishable in the two assay runs. S1 (C and E) Fig shows a comparison of the two assays and again reveals that phenotypic responses are similar. As expected, PCA and UMAP show different clustering properties and different variations across cell lines, but the specific clustering responses of cell lines treated with different classes of drugs, as well as the variable responses observed to the same drug, between cell lines derived from different donors, are reproducible in this assay format. We conclude this represents intrinsic properties of the respective cell lines, independent of variables specific to individual assay runs, such as specific cell line stocks, prepared media, or the instrumentation used for image acquisition.

While many of the drug treatments resulted in tight clustering of the responses of different cell lines on the UMAP and PCA plots (Fig 2), a subset of the drugs tested, i.e., statins, 5-FU and erlotinib, showed more dispersed clustering of cell line responses. We hypothesised that this behaviour reflected an increased degree of variable response to these drugs between the respective cell lines. To test this hypothesis, we calculated the Euclidean distance of the feature vectors measured for each compound and DMSO, for each donor. S3B Fig shows the result of this calculation, along with the mean and coefficient of variation (CoV) of the Euclidean distance for each drug across donors. This metric is elevated for simvastatin, atorvastatin, erlotinib and fenbendazole, consistent with the evidence from UMAP and the induction plot that response to these drugs varies across the donors [27].

The variable responses shown in Fig 2 and S3B Fig might be due to intrinsic variations within an assay. To directly assess the variations in Cell Painting features in different assays, we calculated the Earth Mover Distance (EMD, see Methods) which has been shown previously to provide a useful measure of differences between phenotypic features in HCS assays [28]. S2A Fig shows the distribution of EMDs for all features from all compound treatments common between these two assay runs. Most show EMDs similar to DMSO, suggesting the differences are in the range of the assay controls. Cytotoxic compounds (e.g., bortezimib, dasatinib, colchicine) show a broader distribution of EMDs, which is likely due to cell death caused by these compounds during incubation. Inspired by the overall similarities in features between the two assays, we calculated induction scores for the drugs and donors that are common across these assays. Comparison of heatmaps of these induction scores shows the variations between drugs, but also between donors within individual drugs, are similar (S2B, C Fig). These data suggest that the variations we observe in drug responses (e.g., Figs 1 and 2) across hiPSC lines generated from different donors are robust and reproducible.

Linking variable responses to protein expression & stoichiometry

Our previous studies on this HipSci cohort have demonstrated that variations in protein expression levels between lines derived from different donors can be controlled by specific genomic loci, i.e., protein Quantitative Trait Loci (pQTLs) [13]. This suggested the hypothesis that a potential source of differential drug responses between hiPSC lines might be due to epistatic interactions, i.e., resulting from genomic variation between donors causing differences in protein expression that can modulate drug-induced phenotypes, for example by affecting the stoichiometry of components of drug response pathways [29]. Importantly, this hypothesis takes into account drug-cell interactions outside of the direct interaction of a drug with its protein target. Furthermore, it can potentially explain why patterns of drug response features measured by

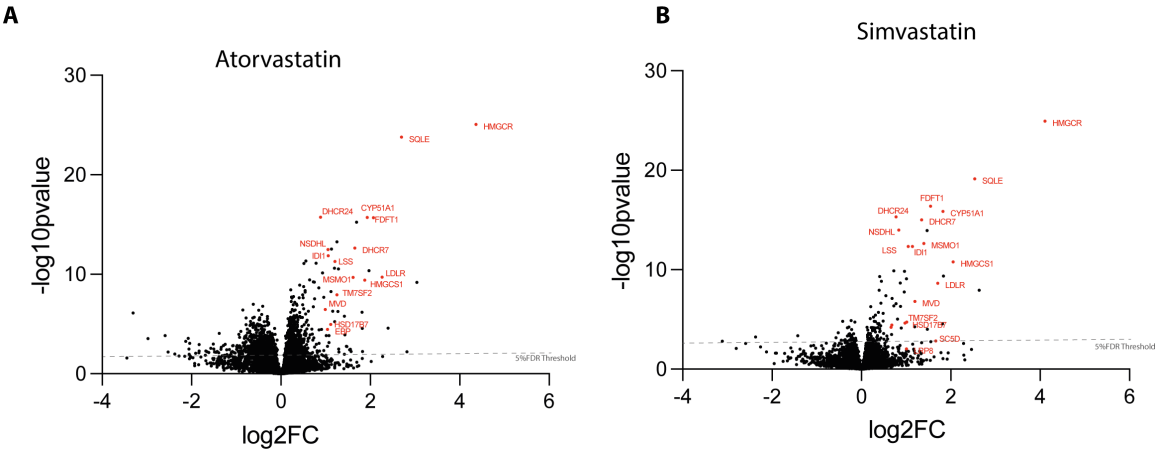

Fig 3. Volcano plots showing differential protein expression relative to DMSO in all four (tuju1, hayt1, denw6 and zaie1), hiPSC lines, treated with either atorvastatin (A) or simvastatin (B). The protein expression fold change (FC) and p value were calculated using Perseus. All sample runs (n = 12), from each drug, were considered as one group, and the DMSO control set (n = 12) was considered as a separate group. Representative proteins that play a role in the HMG-CoA reductase pathway are highlighted in red and show increased expression following drug treatment.

https://doi.org/10.1371/journal.pone.0323953.g003

Cell Painting appear similar across cell lines, while the magnitude of responses between lines differ, as seen in the feature and induction heatmaps (Figs 1B and 2A).

To test this hypothesis, we first used the induction plot (Fig 2A) and a graph of induction values (S3A Fig), to classify cell lines as, respectively, "low" and "high" responders, for two well-characterised drugs, i.e., atorvastatin and simvastatin. These drugs share the same target, but are well-known to have different responses and efficacies in different individuals [30]. Representative cell lines denw6 and zaie1 ("low responders") and hayt1 and tuju1 ("high responders"), were each treated separately with either DMSO (control), or with either atorvastatin, or simvastatin, each at 5 μM, for 24 hrs, as in the Cell Painting assay. Following treatment, cell extracts were prepared and proteomes analysed by LC-MS/MS (see Methods; quantitative proteomics were performed on a small subselection of donors and compounds because the culture of cells required a larger culture format). From the resulting data, differences in protein expression levels between DMSO- and drug-treated lines were visualised in 'volcano plots' (Fig 3). Combining data for treatment of all four cell lines (tuju1, hayt1, denw6, zaie1), with either atorvastatin (Fig 3A), or simvastatin (Fig 3B), showed that many of the proteins whose expression increased after drug treatment are involved in lipid metabolism (red dots in Fig 3A, B). Further, GO term enrichment analysis also showed a significant enrichment for expression of protein factors involved in lipid metabolism (S3 and S4 Tables).

These data suggest that homeostatic mechanisms in the hiPSC lines respond to drug mediated inhibition of cholesterol metabolism by increasing the expression of enzymes required for cholesterol synthesis. While a common type of response to inhibition of cholesterol synthesis is seen across the hiPSC lines, differences were evident in the magnitude of responses between individual cell lines. This is illustrated by the response to either atorvastatin, or simvastatin, as detected by Cell Painting induction (Fig 2A). We therefore tested whether variation in response magnitude might be linked to differences in protein expression between the respective high and low response lines. Volcano plots comparing protein expression in the high and low response lines, after treatment with either atorvastatin (Fig 4A), or simvastatin (S4A Fig), show that multiple proteins involved in the cholesterol synthesis pathway (highlighted in purple), together with some other classes of proteins, have increased expression specifically in the high response lines. These data are consistent with the hypothesis that differential protein expression may explain, at least in part, the differences in the magnitude of drug responses between individual hiPSC lines.

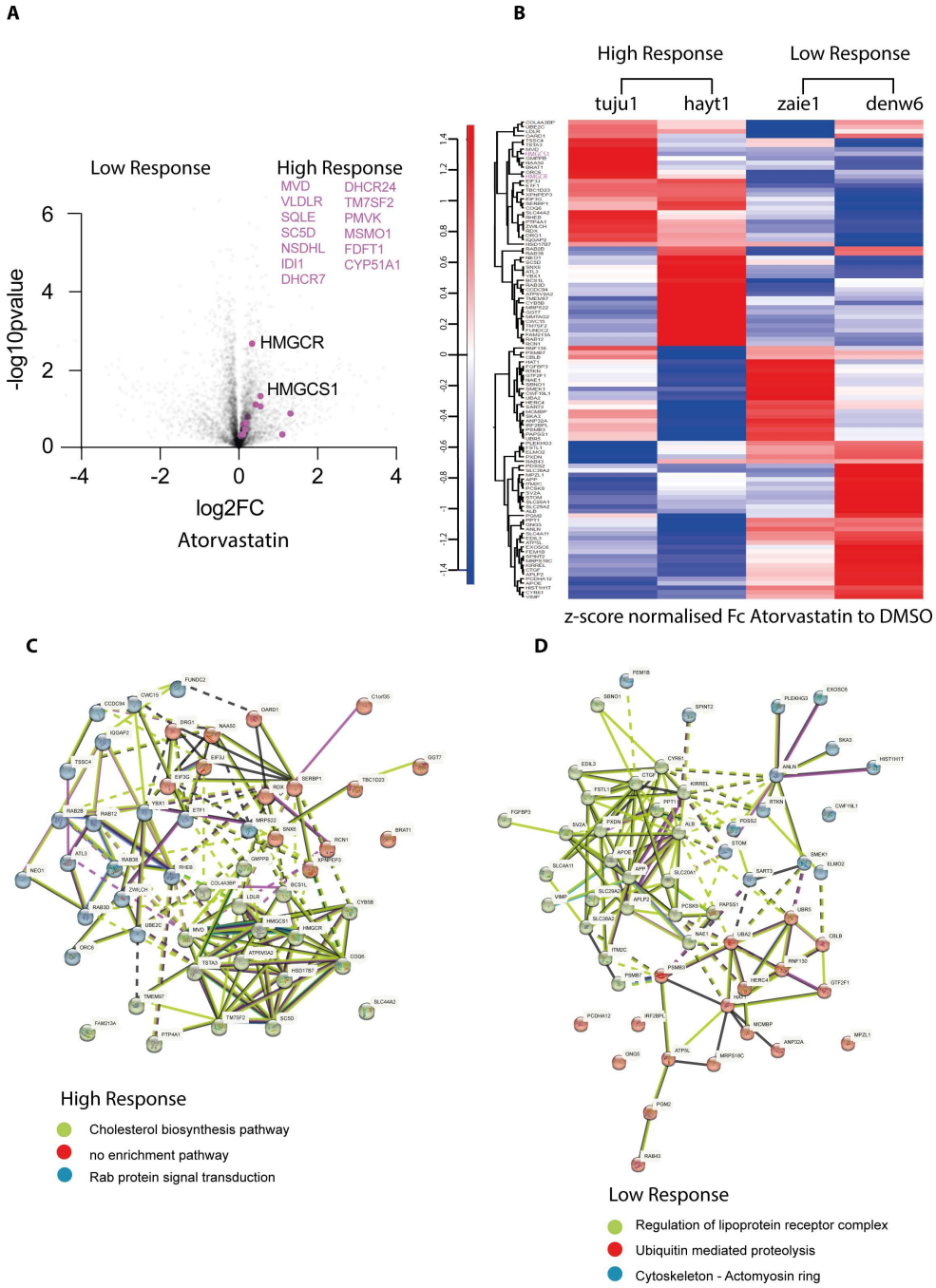

Fig 4. (A) Volcano plot of differential protein expression in four hiPSC lines, stratified by their respective high, or low, response to atorvastatin (see S3A Fig). High response lines: tuju1 and hayt1. Low response lines: denw6 and zaie1. Protein FC and p value for each cell line were calculated as in Fig 3. Sample runs from the drug treatment set (n = 6), were considered as one group and the DMSO control set (n = 6), were considered as a separate group. Members of the cholesterol biosynthesis pathway are represented as purple dots and listed on the right side of the graph. (B) Heat map analysis of proteins showing differential expression ($log2FC_{high\text{-}low} > 0.263$) between the respective high and low response hiPSC lines, following atorvastatin treatment. Heatmap generation steps: the protein fold change (FC_{high}) and p value (P_{high}) for high response cell lines were calculated using sample runs from both high response cell lines to Atorvastatin set (n = 6), considered as one group and the DMSO control set (n = 6) considered as a separate group. The protein fold change (FC_{low}) and p value (P_{low}) for low response cell lines were also calculated. Only proteins with both $P_{high} \leq 0.05$ and $P_{low} \leq 0.05$ were considered further to compare the difference between high and low response. log2FC of these filtered proteins was calculated as ($log2FC_{high\text{-}low} = log2FC_{high} - log2FC_{low}$). Heatmap shows FC between drug and DMSO control of the filtered list z-score normalised. StringDB analysis of filtered proteins shown in (B), in high response lines (C) and low response lines (D).

https://doi.org/10.1371/journal.pone.0323953.g004

Next, we compared the expression of individual proteins in each of the four drug-treated lines (Fig 4B). Proteins were identified whose expression differed significantly in high vs low responders ($|\log_2 FC| > 0.263$ and $p < 0.05$, see Materials and Methods). This reveals clear differences in the responses to the two statin drugs, suggesting that there may be variation in the mechanisms of response to statin treatment between each hiPSC line (see also S4B Fig; S9 Table).

Interestingly, changes in protein expression in the high response lines include increased expression of HMGCR and HMGCS1, which are directly involved in the cholesterol mevalonate synthesis pathway. In contrast, both "low response lines", show increased expression of proteins involved in regulating the stability and recycling of lipoprotein receptor complexes. This includes PCSK9, a protein that provides post translational regulation of LDL receptor and directs it for lysosomal degradation, rather than regular recycling potentially limiting the efficacy of statins. In addition, SCARB1, a cell surface HDL receptor that mediates HDL-cholesteryl ester uptake, as well as ApoE, a plasma protein that participates in the transport of cholesterol and other lipids between cells and has been linked to the effect of statins [31–39]. Lipoprotein receptor complexes play an important role in the regulation of plasma and intracellular cholesterol homeostasis [40]. These differences in protein expression between the respective high and low responder hiPSC lines to statin treatment suggest that variable activity of biosynthetic and/or influx mechanisms may be involved in determining how each hiPSC line responds to statins. In addition to identifying well-established targets of statins, characterising the proteins whose expression increases in the high response lines, using GO enrichment and STRING analysis (https://string-db.org/), identified factors involved in Rab protein signal transduction and in the metabolism of RNA (Fig 4C and S4C Fig; S5 and S7 Tables). These are all known to change expression upon statin treatment [41–45]. In contrast, the low response lines showed increased expression of proteins linked to the cytoskeleton, along with several ubiquitin E2 and E3 enzymes, suggesting that ubiquitin-conjugation levels may affect the degree of statin response (Fig 4D and S4D Fig; S6 and S8 Tables).

These data support the view that a diversity of factors and pathways can affect the cellular response to well-defined drug perturbations and these differences are due, at least in part, to population-level variations in gene and protein expression between the respective lines.

To explore further the observed variable drug responses, we calculated the pairwise correlation matrix for the induction metric (see Fig 2A) for 32 drugs across donors. This provides a measure of the pattern of variable response across a given set of donors. S5 Fig shows a visualisation of these data in a heatmap and identifies sets of compounds that share common patterns of variable response. For example, erlotinib and afatinib, both ErbB family receptor inhibitors and the statins and fluphenazine, behave similarly. Everolimus, etoposide and irinotecan share a similar variable response pattern, which may be linked to a role for the mTor pathway in controlling the DNA damage response [46,47]. In addition, warfarin, omeprazole, niclosamide and metoclopramide, which have very different targets, nonetheless share a similar variable response pattern across hiPSC lines. We hypothesise that this may be linked to a shared off-target effect, e.g., mitotoxicity [48]. This is an interesting point to explore in further studies in the future.

Discussion

We have established, to our knowledge, the first laboratory pipelines for the systematic analysis of population-level variable drug responses, using a cohort of hiPSC lines, derived from multiple, healthy donors. Previous work has described culturing hiPSCs from a single patient in 384-well plate for drug screening, In this study, we have used a cohort of hiPSC lines from many different donors and Cell Painting to garner significantly more information from the screens [12,49]. This pipeline provides a 'genetically-informed', high-throughput assay for identifying and characterising variable drug responses in human cells, using Cell Painting and quantitative proteomics. This analysis provides an initial quantitative measure of variable response across a cohort of hiPSC lines [50].

The choice of screening hiPSC lines specifically in their undifferentiated, pluripotent state leverages the recent discovery that, at least in part, gene expression in hiPSC lines reflects the genetic identity of the donor at both the RNA and protein levels [12,13]. Thus, the undifferentiated hiPSC lines can behave as avatars of the donors, with respect to their proteomes and cellular phenotypes. The corollary is that drug screening assays, using cohorts of pluripotent hiPSC lines from different donors, allows objective stratification based on measurements of phenotypic responses. We hypothesise that these data may reveal important information about variable human drug responses that is germane to clinically relevant differential drug responses between patients. It is important to note that the HipSci cohort covers limited ethnic diversity, with a high concentration of ethnic European donors [51]. In the future, we aim to expand our working cohort to include hiPSC lines from a broader range of donor ethnicity, which may provide an opportunity to detect variable efficacy and toxicity of drugs in different ethnic groups [50,51].

The data presented in this study show that features can be extracted from the analysis of high-throughput, Cell Painting, fluorescence microscopy images and used to stratify differences in the magnitude of drug responses between different hiSPC lines, (illustrated in Figs 1 and 2). The data also show that the stratification of cell lines, based upon their degree of drug response, is reproducible in this assay format. Further, by using quantitative proteomic analysis to compare the stratified sets of hiPSC lines we show that additional information relevant to understanding the mechanisms underlying variable drug response phenotypes can be identified. We have focussed on proteome level analysis here because proteins are the direct targets of most drugs in clinical use and because proteins are also the primary mediators of most disease processes and mechanisms of drug action.

We propose a hypothesis to explain the reproducible, differential response to drug treatment we have detected. This hypothesis postulates that variations between the cell lines, at the genomic and/or epigenomic level, in turn determine the respective proteomes and thereby lead to differences in drug-cell interactions. We envisage that drug-cell response phenotypes can reflect multiple epistatic interactions that are predominantly mediated at the proteome level. This hypothesis is supported by our comparison of proteomes of hiPSC lines that were stratified by Cell Painting as either ‘high’, or ‘low’ responders, respectively, to treatment with either atorvastatin and simvastatin. While we have concentrated here on comparing protein expression levels between cell lines, we note that other comparative analyses, for example comparing protein phosphorylation and other protein post translational modifications (PTMs) and protein-protein interactions, can also provide valuable insights into the mechanisms causing variable drug responses between the different cell lines.

In this study, we have investigated potential mechanisms involved in the variable responses seen to statins in this screening assay. Statins are drugs developed to reduce the level of cholesterol in blood and thus lower the risk of developing cardiac and circulatory diseases, including angina, heart attack and stroke [52]. Separate, but interconnected pathways, control cholesterol homeostasis, including cholesterol influx, mediated by lipoprotein receptors, cholesterol synthesis, regulated by HMG-CoA reductase and the mevalonate pathway and cholesterol efflux, mediated by transport proteins. Each of these pathways respond to multiple inputs that collectively define cholesterol levels [53]. In addition, our data suggest that modulation of protein stability by Ub-mediated proteolysis is also an important factor in cholesterol homeostasis (Fig 4D). Even within the relatively limited sample of different donor lines we have analysed so far, we see evidence of varied regulatory effects. The comparative proteomic analysis supports the view that “high response” lines can be determined by modulation of the endogenous cholesterol synthesis pathway, while “low response lines” may instead reflect primarily regulation of the cholesterol influx and efflux pathways.

In conclusion, we have established and validated a flexible, laboratory-based high-throughput screening platform that can detect and characterise examples of variable drug response for clinically relevant therapeutics. We anticipate that the use of this platform will, in the future, help to improve the success rate of clinical trials by predicting variable responses and providing evidence of biomarkers that may improve the selection of suitable patients most likely to benefit from new drug candidates. We also anticipate that the use of carefully selected hiPSC cohorts, derived from donors of different

sex and genetic backgrounds, can help to improve the development of safe new drugs that are suitable for the needs of modern, diverse human societies.

Materials & methods

Cell culture

Human iPSC lines used in this study were from the HipSci cohort (4). Feeder-free human iPSC lines were cultured in Essential 8 (E8) medium (E8 complete medium supplemented with (50x) E8 supplement ThermoFisher-A1517001) on tissue-culture dishes coated with 10 μg/cm^2 of reduced Growth Factor Basement Membrane matrix (Geltrex, ThermoFisher A1413202 resuspended in basal medium DMEM/F12 Thermo Fisher 21331020). Medium was changed daily.

To passage feeder-free hiPSC lines, cells were washed with PBS and incubated briefly (3–5 min), with 0.5 mM PBS-EDTA solution. After PBS-EDTA solution was removed, cell clusters were resuspended in E8 medium and seeded at ratios of 1:4–1:6, depending on their confluency, on Geltrex-coated tissue culture dishes. Once established, undifferentiated, feeder-free human iPSC lines were expanded and aliquots frozen for future use, prior to transition to TeSR medium for further study. To transition to TeSR medium supplemented with bFGF (Peprotech 100–1813, at a final concentration of 30 ng/ml), Noggin (Peprotech 120-10C, at a final concentration of 10ng/ml), Activin A (Peprotech 120-14P, at a final concentration of 10ng/ml), E8 medium of hiPSC lines was exchanged daily with medium containing increasing amounts of TeSR medium, until human iPSC lines were completely maintained in TeSR medium alone, typically over 1–2 weeks.

To passage human iPSC lines after transition to TeSR medium, cells were washed with PBS and incubated briefly with TrypLE Select (ThermoFisher 12563029), to create a single cell suspension prior to resuspension in TeSR medium supplemented with ROCK kinase inhibitor (Tocris 1254, at a final concentration of 10 μM), to enhance single cell survival, bFGF (Peprotech 100–1813, at a final concentration of 30 ng/ml), Noggin (Peprotech 120-10C, at a final concentration of 10ng/ml), Activin A (Peprotech 120-14P, at a final concentration of 10ng/ml) and seeded at a concentration of 5x10^4 cells/cm^2 on Geltrex-coated tissue culture dishes. 24 hrs after replating Rock kinase inhibitor was removed from the culture medium. Pluripotency of TeSR-transitioned hiPSC lines was verified by immunofluorescence (see below) prior to either cell banking, or experimental use. Cells were grown at 37°C and 5% CO_2. Medium was changed daily.

hiPSC line ID number and details of the cell lines are listed in S1 Table.

The data from the experimental runs are included in this report. Run 2021−02, 52 drugs, 28 donors 1,456 drug-donors pairs). Run 2021−10, 8 donors, 52 drugs (416 drug-donor pairs). Run 2023−12, 18 donors, 52drug (936 drug-donor pairs).

Cell staining

384-well tissue culture plates (CellCarrier-384 Ultra Microplates, Perkin Elmer 6057300) were coated with human plasma fibronectin (Merck Millipore FC010) at a concentration of 5 μg/cm^2. Cells were passaged manually as described above with TrypLE Select, counted and seeded on the fibronectin-coated wells, at a concentration of 3x10^4/cm^2. Cell line plating was in rows, with six wells per condition for each cell line (six technical replicates). Each replicate was sampled at 9 locations and repeated in a separate biological replicate separated by a week (i.e., 2 bioreplicates per assay run).

Cells were incubated for 24 hrs prior to drug treatment, followed by a further 24 hrs incubation before final fixation, staining and high content imaging. Antibody staining of a separate 384-well plate plated for each biological replicate, at 48 hrs, confirmed that pluripotency was maintained at 48 hrs.

Mitotracker (Thermo Fisher M22426), staining was performed in live cells with a 30 min incubation at 37°C prior to fixation. Fixation, permeabilisation and immunostaining were performed at room temperature. 8% Paraformaldehyde in

PBS (PFA, Sigma F8775), was added to an equal volume of cell culture medium in wells of a 384-well plate for a final concentration of 4% Paraformaldehyde and incubated for 15 min. Fixed cells were washed in PBS and permeabilised with 0.1% v/v Triton-X100 (Sigma) for 15 min. Following permeabilisation, cells were washed in TBS and stained with DAPI (Thermo Fisher D1306 at 5 μg/ml), SYTO 14 (Thermo Fisher S7576 at 2.5 μM), Concanavalin A (Thermo Fisher C11252 at 20 μg/ml), Phalloidin Alexa (Thermo Fisher A12381 at 66 nM), and WGA (Thermo Fisher W11262 at 11.8 μg/ml) for 1hr. Staining volume was 30 μl/well. Cell fixation, permeabilisation, staining and washing steps (2x, wash volume 50 μl), were performed using automated liquid handling systems (405 plate washer, BioTek– Tempest, Formulatrix and HighRes Biosolutions Automation systems).

Drug treatment

Drug dosing on 384-well plates was performed using an Echo acoustic liquid handler and corresponding software (Labcyte). Screened compounds were selected from the appropriate chemical library plates containing either Cloud library (Enamine), or an in-house library of publicly available compounds. Drug dosing was performed at final concentrations of; etoposide 1.7 μM, Berberine Chloride 2.68 μM and Fenbendazole 3.34 μM. For all other compounds drug dosing was performed at a final concentration of 5 μM. S2 Table lists the compounds and their final concentrations used in this study. For proteomics analysis drug treatments with either simvastatin (Tocris 1965), or atorvastatin; (Selleckchem S5715) were performed in triplicate in 6-well plates at a final concentration of 5 μM for 24hrs.

Cell staining and analysis of pluripotency markers

All cell lines used in this study were quality controlled for pluripotency prior and during the HCS assays. For each experiment 8 wells of each cell line were individually stained with pluripotency marker antibodies.

All fixation, permeabilisation and immunostaining analyses were performed at room temperature, apart from primary antibody incubation, which was carried out at 4°C overnight. 8% Paraformaldehyde in PBS (PFA, Sigma F8775), was added to an equal volume of cell culture medium in wells of a 384-well plate for a final concentration of 4% Paraformaldehyde and incubated for 15 min. Fixed cells were washed twice in PBS and permeabilised with 0.5% v/v Triton-X100 (Sigma) for 10 min (apart from TRA1–81 staining). Following permeabilization cells were washed twice in PBS, then blocked in 10% normal donkey serum for 1 hr prior to antibody incubations. Pluripotency marker antibodies used in this study are: Nanog, Oct4, Sox2, Tra-1–81 all at 1:500, (Oct4 (C30A3), Sox2 (D6D9), Nanog (D73G4) and TRA181 (TRA-1–81) Cell Signalling StemLight Pluripotency Antibody Kit 9656). All affinity purified donkey secondary antibodies (either Alexa 488 or Alexa 594) were purchased from Jackson Immunoresearch Europe. Hoechst 33342 (Thermo Fisher) was used to stain DNA. Images of cells (for each hiPSC line, 9 fields/well, 16 wells per pluripotency marker), stained with pluripotency marker antibodies were segmented based on DAPI signal and the integrated nuclear intensity of each marker was quantified using CellProfiler. Distributions of integrated intensities of all nuclear pluripotency markers were monotonic and correlated with integrated DAPI staining.

Imaging

Imaging data were collected on an InCell 2200 high content imager, with a 20x/NA 0.45 Plan Fluor lens and a Quad2 multichroic mirror (GE, Issaquah, WA) or a CV7000 high content imager, with a 20x/0.45 lens at epifluorescence mode (Yokogawa). For each assay, we performed 6 technical replicates (each in separate wells), with each replicate sampled at 9 locations, and repeated in a separate biological replicate separated by a week (i.e., 2 bio replicates per assay run). All of these data are then combined into a single well measurement, producing an n-dimensional data vector per well.

Data processing, analysis and visualization of multiple cell lines

Raw images were imported into OMERO Plus (Glencoe Software, Inc., [54]) and then processed using a custom pipeline in CellProfiler [16], which segmented nuclei, cytoplasm, Golgi and cortical protrusions and calculated a range of defined features. All further steps were executed using the Pandas Python library and Numpy libraries [55,56]. Features for each plate were Robust-Z normalised by the median of features for the DMSO control for that plate [16,57].The Robust Z-score converts the population of features derived from individual cells to a distribution of zero mean and unit variance. Robust Z is calculated according to the formula $x'_i = (x_i - \mu)/M$ where x_i is the measured value for each feature, x'_i is the normalised output value and μ and M are the median and median absolute deviation [18–20]. Features with a coefficient of variation >0.50 were removed from further analysis. We further removed all features with |Spearman coefficient| > 0.98. Data were then visualised using Uniform Manifold Approximation and Projection (UMAP [23]). UMAP parameters were n_neighbours = 15, mindist = 0.01 and a cosine metric was used for scaling the distance between points. GO enrichment analysis was carried out using WebGestalt (WEB-based Gene SeT AnaLysis Toolkit) [58,59]. Functional protein-protein interaction analysis was carried out using STRING-DB [60].

Nearest neighbour distances were calculated using Euclidean distance in scikit-learn [61]. The Earth Mover Distance (EMD) was calculated using sci-py [62] for filtered features from individual wells for each drug-donor pair across a pair of assays. The frequency distribution of EMDs for filtered features was calculated and plotted using Seaborn [63].

Proteomics sample preparation

To support parallel MS-based proteomics analysis in this screening platform in a 6-well plate format, we also tested whether the expression of pluripotency markers was maintained when hiPSC lines were plated at $5x10^4$ cells/cm^2, prior to any drug treatment and cultured as previously described. This showed that strong, homogeneous expression of the four pluripotency markers was maintained at this higher cell density (not shown).

All hiPSCs samples were washed twice with 1x PBS, on ice, then centrifuged at 300g to collect the cell pellets. 4% SDS lysis buffer (10 mM TCEP, 100 mM Tris-HCl, pH 7.4) was used to lyse the cells and extract proteins, with 1x Protease and Phosphatase Inhibitor (A32961, Thermo Fisher) added. 1 mL lysis buffer was added to the cell pellets. After 10 cycles of sonication, (30s on/off), using the Bioruptor® Pico, the protein samples were centrifuged at 20,000*xg* for 15 min. The supernatant proteins were denatured at 95°C for 10 min, then alkylated with 40 mM IAA (Iodoacetamide), (final concentration) and incubated in the dark at room temperature for 30 min. The protein concentration was measured using the EZQ™ Protein Quantitation Kit following the manufacturer's instructions.

500 ug proteins from each sample were further processed using the SP3 protocol, as described [64]. In brief, protein samples were mixed with SP3 beads (1:10, protein:beads) and digested with LysC/trypsin mixture (1:50, enzyme:protein). Peptides were eluted from the SP3 beads with 2% DMSO. The peptide concentration was measured using the Pierce™ Quantitative Fluorometric Peptide Assay following the manufacturer's instructions. Peptide samples were stored at −20°C before LC-MS/MS analysis.

LC-MS/MS analysis

For each cell line, three biological replicates per condition were generated. They were combined in one sample, which was analysed in three separate, replicate MS runs. Each hiPSC line was treated three times with either vehicle control, or specified drug.

All peptide samples were analysed by using an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific), equipped with a Dionex UltiMate™ 3000 RSLCnano System, Thermo Fisher Scientific). Peptides from each sample were loaded onto a 75 μm × 2 cm PepMap-C18 pre-column and resolved on a 75 μm × 50 cm PepMap-C18 EASY-Spray temperature-controlled integrated column-emitter (Thermo Fisher Scientific), using either a 160 min or a 280 min multistep gradient from 5% B to 95% B with a constant flow rate of 300 nl/min. The mobile phases are: H_2O incorporating 0.1% FA

(solvent A) and 80% ACN in H_2O incorporating 0.1% FA (solvent B). The details for 160 min gradient are: 0–6 min, 1%−5% B; 6–145 min, 5%−38% B; 145–155 min, 38%−95% B; 155–160 min, 95%B. The details for 280 min gradient are: 0–6 min, 1%−5% B; 6–255 min, 5%−38% B; 255–270 min, 38%−95% B; 270–280 min, 95% B.

For total proteome analysis, the data were acquired under the control of Xcalibur software in data-independent acquisition mode (DIA). For each sample, one run of 280 min method with 97 DIA windows was performed for boosting the identification of proteins [65] and another three replicate runs of 160 min method were performed for the quantification of proteins.

For the 160 min gradient method, the survey scan was acquired in the Orbitrap covering the m/z range from 375 to 1,175 Da with a mass resolution of 60,000 and an automatic gain control (AGC) target of 4.0×10^5 ions. MS/MS was acquired in the Orbitrap covering the m/z range from 200 to 2,000 Da with a mass resolution of 30,000 and an AGC target of 1.0×10^5 and 60 ms maximum injection time. The sequential DIA isolation windows are 10 m/z from 375 to 775 and 40 m/z from 775–1175. Fragmentation was performed using a HCD collision energy of 30%. For the 280 min gradient method, the survey scan was acquired in the Orbitrap covering the m/z range from 355 to 1,150 Da with a mass resolution of 120,000 and an automatic gain control (AGC) target of 4.0×10^5 ions. MS/MS was acquired in the Orbitrap covering the m/z range from 200 to 2,000 Da with a mass resolution of 30,000 and an AGC target of 1.0×10^5 and 60 ms maximum injection time. The sequential DIA isolation windows are 15 m/z from 350 to 410, 10 m/z from 410 to 440, 15 m/z from 440 to 645, 5 m/z from 645 to 775, 10 m/z from 775 to 875, 15 m/z from 875 to 965 and 27 m/z from 965–1150. The fragmentation was performed using a HCD collision energy of 30%.

DIA MS data analysis

For DIA analysis, the raw MS data of different sample runs from both drug and DMSO set were analysed together using DIA-NN v1.8.1 [66]. The in silico spectral library was generated by DIA-NN, using the Homo sapiens database from UniProt (SwissProt October 2021). The data were searched with the following parameters: stable modification of carbamidomethyl (C), N-term M excision, with a maximum of 1 missed tryptic cleavage threshold, match between runs (MBR) option checked. DIA-NN output reports were further processed with an R package, as described previously [67] to perform the MaxLFQ-based protein quantification, with precursors x samples matrix filtered at 1% precursor and protein group FDR. Statistical analyses were performed using Perseus v1.6.15.0 [68,69] In brief, the protein quantification result was imported to Perseus and the copy number of each protein was calculated based on its intensity using a “Proteomic Ruler” [70]. Only proteins with valid quantification values in at least 1/3 of total sample runs were considered for further comparison. Missing values were replaced from a normal distribution. After median normalisation, two sample student’s T tests were performed for p-value generation, using permutation-based FDR for truncation. Heatmap and K-means clustering analysis was performed based on the Euclidean distance, using the default settings in Perseus. In total, 5,895 proteins were quantified using DIA analysis and used for further analysis.

For heat map generation of Fig 4B and S4B Fig: At the first step of differential expression analysis, a two sample-test using Student’s T-test with the default setting in Perseus was performed to determine if the means of two groups (drug and DMSO, for either high or low response cell lines) are significantly different from each other.

The protein fold change ($FC_{high\ or\ low}$) and p value ($P_{high\ or\ low}$) were generated, respectively.

Proteins with $p < 0.05$ were considered further and used to create the results show in Fig 4B and S4B Fig. Z-score normalisation was used for Heat map analysis, according to the protocol in [69]. Data available at PRIDE [71].

Supporting information

S1 Fig. (A) Maintenance of pluripotency in hiPSC lines in HCS assay format. Representative micrographs from six hiPSC lines, plated in 384-well format and stained with pluripotency markers (Oct4, Sox2, Nanog and TRA-181). Scale bar; 100μm. (B) UMAP Embedding of Cell Painting features measured for eight human iPSC lines from phenotypic

screens performed on two different dates (2021−02 and 2021−10 on two different imaging systems. hiPSC lines were treated with drugs as described (see Materials and Methods). Cell lines are represented on the graph with different shapes and drugs with different colours. Arrows point to representative drug clusters. Note the high degree of reproducibility of the assay results generated on different dates. (C) PCA analysis of Cell Painting features measured for hiPSC lines in the 2021−02 assay run. Drugs are represented with different colours. (D) PCA analysis of Cell Painting features measured for hiPSC lines in the 2021−10 run. Drugs are represented with different colours. (E) PCA analysis of Cell Painting features measured for hiPSC lines in the 2023−12 run. (Run "2023-12" −18 hiPSC lines). Drugs are represented with different colours.
(PDF)

S2 Fig. (A) Plot of distributions of Wasserstein distances between each feature calculated from Cell Painting from the two assays shown in S1B-D Fig. The data for these assays were collected in February and October 2021 respectively, on two different imaging systems (InCell 2200 and CV700 high content imagers). The Wasserstein distance provides a metric of similarity of two distribution [28]. Wasserstein distances of features for each drug-donor pair were calculated and the frequency distribution of the medians of the features for each drug was then plotted. Different drugs show different distributions, with cytotoxic drugs (e.g., bortezimib, dasatinib, paclitaxel and colchicine) showing higher Wasserstein distances, i.e., great variation between features in the two different runs. (B) and (C) Induction plots of drugs and donors that were common across the two assays show in S1B-D Fig. Values in the heatmap are calculated induction values using a cutoff of 2 sigma (same is in Fig 2A).
(PDF)

S3 Fig. (A) Bar graph of induction values for hiPSC lines treated with atorvastatin (orange) or with simvastatin (blue). Cell lines used for proteomic analysis are shown in grey boxes. (B) Table of nearest neighbour analysis. Spearman correlation filtered features for all drugs were used to calculate Euclidean distance from DMSO. Induction mean, standard deviation and coefficient of variation (CoV) for each drug are shown.
(PDF)

S4 Fig. (A) Volcano plot of differential protein expression in four hiPSC lines, stratified by their respective high, or low, response to simvastatin (see S3A Fig). High response lines: tuju1 and hayt1. Low response lines: denw6 and zaie1. Protein FC and p value for each cell line were calculated as in Fig 3. Sample runs from drug treatment set (n = 6) were considered as one group and the DMSO control set (n = 6) was considered as a separate group. Members of cholesterol biosynthesis pathway are represented as purple dots and listed on the right side of the graph. (B) Heat map analysis of proteins showing differential expression ($log2FC_{high\text{-}low} > 0.263$) between high and low response hiPSC lines following simvastatin treatment. Heatmap generation steps: The protein fold change (FC_{high}) and p value (P_{high}) for high response cell lines were calculated with sample runs from both high response cell lines to Atorvastatin set (n = 6) considered as one group and the DMSO control set (n = 6) considered as a separate group. The protein fold change (FC_{low}) and p value (P_{low}) for low response cell lines were also calculated. Only proteins with both $P_{high} \leq 0.05$ and $P_{low} \leq 0.05$ were considered further, to compare the difference between high and low response. log2FC of these filtered proteins was calculated as ($log2FC_{high\text{-}low} = log2FC_{high} - log2FC_{low}$). Heatmap shows FC between drug and DMSO of filtered list z-score normalised. StringDB analysis of filtered proteins shown in (B) in high response lines (C) and low response lines (D).
(PDF)

S5 Fig. (A) Pairwise correlation matrix of induction values (see Fig 2A) for all drugs and across all (28) donors. Note the clusters formed based on the patterns of induction across the donor cohort, suggesting similar responses between different drugs.
(PDF)

S1 Table. List of hiPSC lines used in this study.
(PDF)

S2 Table List of drugs, disease targets and concentrations used in this study.
(PDF)

S3 Table GO enrichment analysis of top 5% proteins showing increased expression in all four hiPSC lines used for proteomic analysis following atorvastatin treatment. Note pathways related to lipid metabolism.
(PDF)

S4 Table. GO enrichment analysis of top 5% proteins showing increased expression in all four hiPSC lines used for proteomic analysis following simvastatin treatment.
(PDF)

S5 Table GO Enrichment Analysis of gene sets/pathways and their associated p-values for highly expressed proteins (Fig 4 heat map (B) values >0.2) of high response lines following atorvastatin treatment.
(PDF)

S6 Table. GO Enrichment Analysis of gene sets/pathways and their associated p-values for highly expressed proteins (Fig 4 heat map (B) values >0.2) of low response lines following atorvastatin treatment.
(PDF)

S7 Table. GO Enrichment Analysis of gene sets/pathways and their associated p-values for highly expressed proteins (S4 Fig heat map (B) values >0.2) of high response lines following simvastatin treatment.
(PDF)

S8 Table. GO Enrichment Analysis of gene sets/pathways and their associated p-values for highly expressed proteins (S4 Fig heat map (B) values >0.2) of low response lines following simvastatin treatment.
(PDF)

S9 Table List of proteins used for Heat map analysis of proteins showing differential expression (|log2FC| > 0.263) between high and low response hiPSC lines following atorvastatin treatment (Fig 4B) and simvastatin treatment (S4B Fig).
(XLSX)

Acknowledgments

We thank members of the National Phenotypic Screening Centre and Human Pluripotent Stem Cell Facility at the University of Dundee for invaluable help and advice. We thank University of Dundee's School of Life Sciences for their support of this project.

Author contributions

Conceptualization: Jason R. Swedlow, Melpomeni Platani, Regis Doyonnas, Santosh Hariharan, Angus I. Lamond.

Data curation: Melpomeni Platani.

Formal analysis: Jason R. Swedlow, Melpomeni Platani.

Investigation: Melpomeni Platani, Lindsay Davidson, Hao Jiang.

Methodology: Lindsay Davidson.

Resources: Regis Doyonnas, Santosh Hariharan.

Writing – original draft: Jason R. Swedlow, Melpomeni Platani, Angus I. Lamond.

Writing – review & editing: Jason R. Swedlow, Melpomeni Platani, Angus I. Lamond.

References

1. Loewa A, Feng JJ, Hedtrich S. Human disease models in drug development. Nat Rev Bioeng. 2023;1–15. https://doi.org/10.1038/s44222-023-00063-3 PMID: 37359774
2. Golding H, Khurana S, Zaitseva M. What Is the Predictive Value of Animal Models for Vaccine Efficacy in Humans? The Importance of Bridging Studies and Species-Independent Correlates of Protection. Cold Spring Harb Perspect Biol. 2018;10(4):a028902. https://doi.org/10.1101/cshperspect.a028902 PMID: 28348035
3. Franco R, Cedazo-Minguez A. Successful therapies for Alzheimer's disease: why so many in animal models and none in humans?. Front Pharmacol. 2014;5:146. https://doi.org/10.3389/fphar.2014.00146 PMID: 25009496
4. Dowden H, Munro J. Trends in clinical success rates and therapeutic focus. Nat Rev Drug Discov. 2019;18(7):495–6. https://doi.org/10.1038/d41573-019-00074-z PMID: 31267067
5. Harrison RK. Phase II and phase III failures: 2013-2015. Nat Rev Drug Discov. 2016;15(12):817–8. https://doi.org/10.1038/nrd.2016.184 PMID: 27811931
6. Rakshith HT, Lohita S, Rebello AP, Goudanavar PS, Raghavendra Naveen N. Sex differences in drug effects and/or toxicity in oncology. Curr Res Pharmacol Drug Discov. 2023;4:100152. https://doi.org/10.1016/j.crphar.2022.100152 PMID: 36714036
7. Oh S, Lee TY, Kim M, Kim SH, Lee S, Cho S, et al. Effectiveness of antipsychotic drugs in schizophrenia: a 10-year retrospective study in a Korean tertiary hospital. NPJ Schizophr. 2020;6(1):32. https://doi.org/10.1038/s41537-020-00122-3 PMID: 33214559
8. Wong ND, Zhao Y, Quek RGW, Blumenthal RS, Budoff MJ, Cushman M, et al. Residual atherosclerotic cardiovascular disease risk in statin-treated adults: The Multi-Ethnic Study of Atherosclerosis. J Clin Lipidol. 2017;11(5):1223–33. https://doi.org/10.1016/j.jacl.2017.06.015 PMID: 28754224
9. Tyson RJ, Park CC, Powell JR, Patterson JH, Weiner D, Watkins PB, et al. Precision Dosing Priority Criteria: Drug, Disease, and Patient Population Variables. Front Pharmacol. 2020;11:420. https://doi.org/10.3389/fphar.2020.00420 PMID: 32390828
10. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126(4):663–76. https://doi.org/10.1016/j.cell.2006.07.024 PMID: 16904174
11. Malik N, Rao MS. A review of the methods for human iPSC derivation. Methods Mol Biol. 2013;997:23–33. https://doi.org/10.1007/978-1-62703-348-0_3 PMID: 23546745
12. Kilpinen H, Goncalves A, Leha A, Afzal V, Alasoo K, Ashford S, et al. Common genetic variation drives molecular heterogeneity in human iPSCs. Nature. 2017;546(7658):370–5. https://doi.org/10.1038/nature22403 PMID: 28489815
13. Mirauta BA, Seaton DD, Bensaddek D, Brenes A, Bonder MJ, Kilpinen H, et al. Population-scale proteome variation in human induced pluripotent stem cells. Elife. 2020;9:e57390. https://doi.org/10.7554/eLife.57390 PMID: 32773033
14. Takahashi K, Yamanaka S. Induced pluripotent stem cells in medicine and biology. Development. 2013;140(12):2457–61. https://doi.org/10.1242/dev.092551 PMID: 23715538
15. Brenes A, Bensaddek D, Hukelmann J, Afzal V, Lamond AI. The iPSC proteomic compendium. bioRxiv. 2018;469916. https://doi.org/10.1101/469916
16. Bray M-A, Singh S, Han H, Davis CT, Borgeson B, Hartland C, et al. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc. 2016;11(9):1757–74. https://doi.org/10.1038/nprot.2016.105 PMID: 27560178
17. McQuin C, Goodman A, Chernyshev V, Kamentsky L, Cimini BA, Karhohs KW, et al. CellProfiler 3.0: Next-generation image processing for biology. PLoS Biol. 2018;16(7):e2005970. https://doi.org/10.1371/journal.pbio.2005970 PMID: 29969450
18. Bray MA, Carpenter A. Advanced assay development guidelines for image-based high content screening and analysis. Markossian S, Grossman A, Arkin M, Auld D, Austin C, Baell J, editors. Assay guidance manual. Bethesda (MD); 2017.
19. Birmingham A, Selfors LM, Forster T, Wrobel D, Kennedy CJ, Shanks E, et al. Statistical methods for analysis of high-throughput RNA interference screens. Nat Methods. 2009;6(8):569–75. https://doi.org/10.1038/nmeth.1351 PMID: 19644458
20. Malo N, Hanley JA, Cerquozzi S, Pelletier J, Nadon R. Statistical practice in high-throughput screening data analysis. Nat Biotechnol. 2006;24(2):167–75. https://doi.org/10.1038/nbt1186 PMID: 16465162
21. Williams E, Moore J, Li SW, Rustici G, Tarkowska A, Chessel A, et al. The Image Data Resource: A Bioimage Data Integration and Publication Platform. Nat Methods. 2017;14(8):775–81. https://doi.org/10.1038/nmeth.4326 PMID: 28775673
22. Willis C, Nyffeler J, Harrill J. Phenotypic Profiling of Reference Chemicals across Biologically Diverse Cell Types Using the Cell Painting Assay. SLAS Discov. 2020;25(7):755–69. https://doi.org/10.1177/2472555220928004 PMID: 32546035
23. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv [statML]. 2018.
24. Thomas EA, George RC, Danielson PE, Nelson PA, Warren AJ, Lo D, et al. Antipsychotic drug treatment alters expression of mRNAs encoding lipid metabolism-related proteins. Mol Psychiatry. 2003;8(12):983–93, 950. https://doi.org/10.1038/sj.mp.4001425 PMID: 14647396

25. Srivastava P, Panda D. Rotenone inhibits mammalian cell proliferation by inhibiting microtubule assembly through tubulin binding. FEBS J. 2007;274(18):4788–801. https://doi.org/10.1111/j.1742-4658.2007.06004.x PMID: 17697112
26. Dogra N, Kumar A, Mukhopadhyay T. Fenbendazole acts as a moderate microtubule destabilizing agent and causes cancer cell death by modulating multiple cellular pathways. Sci Rep. 2018;8(1):11926. https://doi.org/10.1038/s41598-018-30158-6 PMID: 30093705
27. Chari T, Pachter L. The spacious Art of Single-Cell Genomics. bioRxiv. 2021. https://doi.org/10.1101/2021.08.25.457696
28. Pearson YE, Kremb S, Butterfoss GL, Xie X, Fahs H, Gunsalus KC. A statistical framework for high-content phenotypic profiling using cellular feature distributions. Commun Biol. 2022;5(1):1409. https://doi.org/10.1038/s42003-022-04343-3 PMID: 36550289
29. Roumeliotis TI, Williams SP, Gonçalves E, Alsinet C, Del Castillo Velasco-Herrera M, Aben N, et al. Genomic Determinants of Protein Abundance Variation in Colorectal Cancer Cells. Cell Rep. 2017;20(9):2201–14. https://doi.org/10.1016/j.celrep.2017.08.010 PMID: 28854368
30. Sirtori CR, Mombelli G, Triolo M, Laaksonen R. Clinical response to statins: mechanism(s) of variable activity and adverse effects. Ann Med. 2012;44(5):419–32. https://doi.org/10.3109/07853890.2011.582135 PMID: 21623698
31. Nozue T. Lipid Lowering Therapy and Circulating PCSK9 Concentration. J Atheroscler Thromb. 2017;24(9):895–907. https://doi.org/10.5551/jat.RV17012 PMID: 28804094
32. Huang Y, Mahley RW. Apolipoprotein E: structure and function in lipid metabolism, neurobiology, and Alzheimer's diseases. Neurobiol Dis. 2014;72 Pt A:3–12. https://doi.org/10.1016/j.nbd.2014.08.025 PMID: 25173806
33. Tudorache IF, Trusca VG, Gafencu AV. Apolipoprotein E - A Multifunctional Protein with Implications in Various Pathologies as a Result of Its Structural Features. Comput Struct Biotechnol J. 2017;15:359–65. https://doi.org/10.1016/j.csbj.2017.05.003 PMID: 28660014
34. Bi Q, Zhou X, Lu Y, Fu W, Wang Y, Wang F, et al. Polymorphisms of the apolipoprotein E gene affect response to atorvastatin therapy in acute ischemic stroke. Front Cardiovasc Med. 2022;9:1024014. https://doi.org/10.3389/fcvm.2022.1024014 PMID: 36426228
35. Zhang L, He S, Li Z, Gan X, Li S, Cheng X, et al. Apolipoprotein E polymorphisms contribute to statin response in Chinese ASCVD patients with dyslipidemia. Lipids Health Dis. 2019;18(1):129. https://doi.org/10.1186/s12944-019-1069-5 PMID: 31153375
36. Cai C, Wen Z, Li L. The relationship between ApoE gene polymorphism and the efficacy of statins controlling hyperlipidemia. Am J Transl Res. 2021;13(6):6772–7. PMID: 34306425
37. Zintzaras E, Kitsios GD, Triposkiadis F, Lau J, Raman G. APOE gene polymorphisms and response to statin therapy. Pharmacogenomics J. 2009;9(4):248–57. https://doi.org/10.1038/tpj.2009.25 PMID: 19529002
38. Shen W-J, Azhar S, Kraemer FB. SR-B1: A Unique Multifunctional Receptor for Cholesterol Influx and Efflux. Annu Rev Physiol. 2018;80:95–116. https://doi.org/10.1146/annurev-physiol-021317-121550 PMID: 29125794
39. Mahley RW. Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. Science. 1988;240(4852):622–30. https://doi.org/10.1126/science.3283935 PMID: 3283935
40. Xia X-D, Peng Z-S, Gu H-M, Wang M, Wang G-Q, Zhang D-W. Regulation of PCSK9 Expression and Function: Mechanisms and Therapeutic Implications. Front Cardiovasc Med. 2021;8:764038. https://doi.org/10.3389/fcvm.2021.764038 PMID: 34782856
41. Istvan ES, Deisenhofer J. Structural mechanism for statin inhibition of HMG-CoA reductase. Science. 2001;292(5519):1160–4. https://doi.org/10.1126/science.1059344 PMID: 11349148
42. Ali BR, Nouvel I, Leung KF, Hume AN, Seabra MC. A novel statin-mediated "prenylation block-and-release" assay provides insight into the membrane targeting mechanisms of small GTPases. Biochem Biophys Res Commun. 2010;397(1):34–41. https://doi.org/10.1016/j.bbrc.2010.05.045 PMID: 20471365
43. Ronzier E, Parks XX, Qudsi H, Lopes CM. Statin-specific inhibition of Rab-GTPase regulates cPKC-mediated IKs internalization. Sci Rep. 2019;9(1):17747. https://doi.org/10.1038/s41598-019-53700-6 PMID: 31780674
44. Brown MS, Goldstein JL. Multivalent feedback regulation of HMG CoA reductase, a control mechanism coordinating isoprenoid synthesis and cell growth. J Lipid Res. 1980;21(5):505–17. PMID: 6995544
45. Grunwald SA, Popp O, Haafke S, Jedraszczak N, Grieben U, Saar K, et al. Statin-induced myopathic changes in primary human muscle cells and reversal by a prostaglandin F2 alpha analogue. Sci Rep. 2020;10(1):2158. https://doi.org/10.1038/s41598-020-58668-2 PMID: 32034223
46. Zhou X, Liu W, Hu X, Dorrance A, Garzon R, Houghton PJ, et al. Regulation of CHK1 by mTOR contributes to the evasion of DNA damage barrier of cancer cells. Sci Rep. 2017;7(1):1535. https://doi.org/10.1038/s41598-017-01729-w PMID: 28484242
47. Amar-Schwartz A, Ben Hur V, Jbara A, Cohen Y, Barnabas GD, Arbib E, et al. S6K1 phosphorylates Cdk1 and MSH6 to regulate DNA repair. Elife. 2022;11:e79128. https://doi.org/10.7554/eLife.79128 PMID: 36189922
48. Seal S, Carreras-Puigvert J, Trapotsi M-A, Yang H, Spjuth O, Bender A. Integrating cell morphology with gene expression and chemical structure to aid mitochondrial toxicity detection. Commun Biol. 2022;5(1):858. https://doi.org/10.1038/s42003-022-03763-5 PMID: 35999457
49. Ryu S, Chu P-H, Malley C, Braisted J, Ormanoglu P, Huang R, et al. Human Pluripotent Stem Cells for High-Throughput Drug Screening and Characterization of Small Molecules. Methods Mol Biol. 2022;2454:811–27. https://doi.org/10.1007/7651_2021_394 PMID: 34128205
50. Naidoo J, Hurrell T, Scholefield J. The generation of human induced pluripotent stem cell lines from individuals of Black African ancestry in South Africa. Stem Cell Res. 2024;81:103534. https://doi.org/10.1016/j.scr.2024.103534 PMID: 39146664
51. Hurrell T, Naidoo J, Ntlhafu T, Scholefield J. An African perspective on genetically diverse human induced pluripotent stem cell lines. Nat Commun. 2024;15(1):8581. https://doi.org/10.1038/s41467-024-52781-w PMID: 39362853

52. Adhyaru BB, Jacobson TA. Safety and efficacy of statin therapy. Nat Rev Cardiol. 2018;15(12):757–69. https://doi.org/10.1038/s41569-018-0098-5 PMID: 30375494

53. Whirl-Carrillo M, Huddart R, Gong L, Sangkuhl K, Thorn CF, Whaley R, et al. An Evidence-Based Framework for Evaluating Pharmacogenomics Knowledge for Personalized Medicine. Clin Pharmacol Ther. 2021;110(3):563–72. https://doi.org/10.1002/cpt.2350 PMID: 34216021

54. Allan C, Burel J-M, Moore J, Blackburn C, Linkert M, Loynton S, et al. OMERO: flexible, model-driven data management for experimental biology. Nat Methods. 2012;9(3):245–53. https://doi.org/10.1038/nmeth.1896 PMID: 22373911

55. McKinney W. Data Structures for Statistical Computing in Python. In: Proceedings of the Python in Science Conference. SciPy; 2010. 56–61. https://doi.org/10.25080/majora-92bf1922-00a

56. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585(7825):357–62. https://doi.org/10.1038/s41586-020-2649-2 PMID: 32939066

57. Chandrasekaran SN, Cimini BA, Goodale A, Miller L, Kost-Alimova M, Jamali N, et al. Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. Nat Methods. 2024;21(6):1114–21. https://doi.org/10.1038/s41592-024-02241-6 PMID: 38594452

58. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 2019;47(W1):W199–205. https://doi.org/10.1093/nar/gkz401 PMID: 31114916

59. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Res. 2017;45(W1):W130–7. https://doi.org/10.1093/nar/gkx356 PMID: 28472511

60. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023;51(D1):D638–46. https://doi.org/10.1093/nar/gkac1000 PMID: 36370105

61. Pedregosa F, Varoquaux G, Gramfort A, Duchesnay E. Scikit-learn: machine learning in python. J Mach Learn Res. 2011;12:2825–30.

62. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261–72. https://doi.org/10.1038/s41592-019-0686-2 PMID: 32015543

63. Waskom M. Seaborn: statistical data visualization. J Open Source Softw. 2021;6(60):3021.

64. Hughes CS, Moggridge S, Müller T, Sorensen PH, Morin GB, Krijgsveld J. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nat Protoc. 2019;14(1):68–85. https://doi.org/10.1038/s41596-018-0082-x PMID: 30464214

65. Doellinger J, Blumenscheit C, Schneider A, Lasch P. Isolation Window Optimization of Data-Independent Acquisition Using Predicted Libraries for Deep and Accurate Proteome Profiling. Anal Chem. 2020;92(18):12185–92. https://doi.org/10.1021/acs.analchem.0c00994 PMID: 32840101

66. Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat Methods. 2020;17(1):41–4. https://doi.org/10.1038/s41592-019-0638-x PMID: 31768060

67. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics. 2014;13(9):2513–26. https://doi.org/10.1074/mcp.M113.031591 PMID: 24942700

68. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote) omics data. Nat Methods. 2016;13(9):731–40. https://doi.org/10.1038/nmeth.3901 PMID: 27348712

69. Tyanova S, Cox J. Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. Methods Mol Biol. 2018;1711:133–48. https://doi.org/10.1007/978-1-4939-7493-1_7 PMID: 29344888

70. Wiśniewski JR, Hein MY, Cox J, Mann M. A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. Mol Cell Proteomics. 2014;13(12):3497–506. https://doi.org/10.1074/mcp.M113.037309 PMID: 25225357

71. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. 2022;50(D1):D543–52. https://doi.org/10.1093/nar/gkab1038 PMID: 34723319