Screening for variable drug responses using human iPSC cohorts

Melpomeni Platani[1], Hao Jiang[1], Lindsay Davidson[2], Santosh Hariharan[3,4], Regis Doyonnas[3], Angus I. Lamond[1], Jason R. Swedlow[1] *

[1] Molecular Cell and Developmental Biology, School of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, Scotland, UK.

[2] Human Pluripotent Stem Cell Facility, School of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, Scotland, UK.

[3] Discovery Sciences, Pfizer, 445 Eastern Point Road, Groton, CT 06340, US Groton CT, USA.

[4] Current address: VRTX, 50 Northern Ave, Boston MA, 02210, USA.

* Corresponding author:

E-mail: jrswedlow@dundee.ac.uk

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.

32 **INTRODUCTION**

33 Most current drug development pipelines involve an initial laboratory-based research
34 phase, in which disease-relevant molecular targets are identified (usually proteins) and
35 then either small molecules, or biological effectors (e.g. antibodies), are characterised
36 that specifically bind to, or otherwise modulate, the targets. Promising drug candidates
37 are then progressed from animal models to human clinical trials, where they must be
38 evaluated for efficacy and possible toxicity in human patients before they can be certified
39 for use by regulatory authorities (*1-3*). Unfortunately, these pipelines often have very high
40 failure rates, with 80% or more drug candidates in major disease areas failing to pass
41 clinical trials (*4*). The high failure rates at late stages of drug development are a major
42 factor in the high cost of bringing safe and effective new drugs to market.

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

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

60 Several projects have now established cohorts of human induced pluripotent cell (hiPSC)
61 lines from multiple donors using standardised protocols (*10, 11*). In particular, the Human
62 Induced Pluripotent Stem Cell Initiative, (HipSci, https://www.hipsci.org), has established
63 a library of >700 cell lines, generated from over 300 individuals, including healthy male
64 and female donors of varied ages, as well as patient cohorts with known genetic
65 disorders (*12*). The HipSci cell lines were all reprogrammed from human skin fibroblasts,
66 following transduction of the pluripotency transcription factors using a Sendai virus vector

67 (*12*). Many of these cell lines have been subjected to mRNA and protein expression
68 analysis and these data are publicly available. Interestingly, comparison of mRNA and
69 protein expression patterns in the HipSci lines showed that mRNA and protein levels in
70 separate hiPSC lines from the same individual are more similar to each other than to
71 lines derived from different donors. This shows that gene expression states of the HipSci
72 lines are strongly defined by the genetic identity of the donor (*12, 13*). The HipSci cohort
73 therefore provides a tractable system for measuring how human genetic variation affects
74 cellular phenotypes, potentially including variable drug responses.

75 The ability of hiPSCs to be re-differentiated into different cell types, using standardised
76 differentiation protocols, is frequently used for experimental analyses of tissue-specific
77 responses and disease mechanisms (*14*). However, using the undifferentiated,
78 pluripotent state of hiPSCs for drug screening provides some potential advantages. For
79 example, studies that have measured in depth gene transcript and protein expression
80 in hiPSCs and embryonic stem cells (ESCs) (*13, 15*), find that stem cells express a
81 broader repertoire of gene transcripts and proteins than terminally differentiated, primary
82 cells and tissues, including most tumour-derived, transformed cell lines. For example,
83 hiPSC lines express ~2.5-fold more receptor tyrosine kinases than many primary cells,
84 or tumour cell lines. For the majority of chromosomes, >50% of their known protein
85 coding genes are expressed in hiPSCs ((*13*) and references therein). While no single
86 cell type expresses all potential drug targets, this broad molecular expression profile in
87 pluripotent stem cells, including many pathways targeted by clinical therapeutics, makes
88 the use of pluripotent hiPSCs an attractive foundation for compound and drug screening.

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

95 **RESULTS**

96 **Establishing Pluripotency in hiPSC Cohorts in High-Throughput Format**

97 To test whether we could detect and characterize variable drug responses across a
98 cohort of hiPSC lines from different donors, the high throughput Cell Painting assay (*16*)
99 was first adapted for use with hiPSC lines grown in the pluripotent state. We reasoned
100 that an image-based Cell Painting assay could be used to detect variable responses and

101 stratify cell lines accordingly, followed by using quantitative MS-based proteomics to
102 identify potential mechanisms causing the variable responses (see Figure 1A).

103 First, we established whether conditions could be established for culturing multiple
104 hiPSC lines in the pluripotent state in 384 multi-well plates, suitable for high content
105 imaging and analysis. Twenty-eight independent hiPSC lines, all derived from different
106 donors (Supp Table 1, "Run1"), were propagated in mTeSR medium, plated at $3x10^4$
107 cells/cm^2 in 384 multi-well plates for 72 hrs and then fixed and analysed by
108 immunofluorescence to detect the established markers of pluripotency, i.e., Oct-4, Sox-
109 2, Nanog and TRA-1-81 (Supp Figure 1A). All four markers showed strong,
110 homogeneous staining.

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

115 **Using Cell Painting to Profile Variable Drug Responses**

116 Next, the set of 28 independent hiPSC lines from different donors, (hereafter, "donors"),
117 were grown in 384 multi-well plates for 24 hrs (see Methods for details), then treated
118 either with compounds from a library of FDA-approved drugs , or with DMSO carrier,
119 each at a single concentration (Supp Table 2). After incubation for a further 24 hrs in the
120 presence of the drugs, the cells were fixed and stained with Cell Painting markers (see
121 Materials and Methods). Images were recorded in an automated, plate-based imager,
122 then imported into OMERO (9). Images were segmented with respect to the 'nuclei',
123 'cytoplasm', 'Golgi' and 'protrusions', using a custom CellProfiler pipeline (*17*). For each
124 drug-donor combination, image features were calculated, using 871 separate feature
125 measurements, from a total of 6,000-8,000 cells per donor-drug pair. After Robust Z'
126 normalisation (*18-20*), all measured features were scaled as the number of standard
127 deviations different from DMSO for each of the hiPSC lines (*16*).

128 Figure 1B shows a heatmap of the measured features of the 28 donors, comparing
129 treatment with 4 of the drugs from the library, i.e., atorvastatin, simvastatin, rapamycin
130 and afatinib. The heatmap shows that each drug has characteristic feature patterns, but
131 that different donors show varied magnitudes of response, visible as either darker, or
132 lighter, horizontal stripes, respectively (Figure 1B). The feature patterns for atorvastatin
133 and simvastatin are similar, consistent with them being competitive inhibitors of the same
134 protein target, i.e., HMG CoA reductase, the rate-limiting enzyme of the cholesterol

135 biosynthetic pathway. Interestingly, atorvastatin, which has a higher target affinity, also
136 shows a stronger response than simvastatin in this assay.

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

142 Next, the assay was expanded to profile the hiPSC cohort responses to 52 FDA-
143 approved drugs, which are used to treat a range of different diseases (Supp Table 2). A
144 subset of the drugs tested are shown for visibility with full data available on Image Data
145 Resource IDR (*21*). To remove highly correlated features, we calculated the Spearman
146 correlation coefficient, s, and removed all features where $s > 0.98$, which left 442 filtered
147 features for further analysis.

148 Figure 2A shows a heatmap of the Cell Painting induction, calculated as the fraction of
149 filtered features with RobustZ > 2σ from DMSO (i.e. significant with >95% certainty) (*22*).
150 This effectively reduces the filtered features for each drug-donor pair to a single statistic,
151 which may underestimate the complexity of phenotypes. Despite this limitation, Figure
152 2A shows clearly that, first, hiPSC lines derived from different donors exhibit different
153 levels of induction when treated with the same drugs and second, that examples of such
154 differential response are observed between cell lines for all of the 52 different drugs
155 tested.

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

163 We next asked if phenotypes caused by different compounds could be distinguished
164 across the cohort of hiPSC lines. The filtered features from all drug-donor pairs were
165 subjected to embedding with UMAP, which fits a manifold to a high-dimensional dataset
166 and then projects this to a 2D plane (*23*) (Figure 2B). This showed that compound
167 responses clustered according to their known mechanism of action. For example, statins,
168 everolimus and rapamycin, topoisomerase inhibitors, cytotoxics and microtubule
169 depolymerizers all formed separate, individual clusters. Interestingly, fluphenazine, an

170 antipsychotic that acts by blocking specific dopamine receptors in the brain and that is
171 used to treat schizophrenia and other psychotic disorders, clustered near the statins.
172 While clustering in UMAP is not always a reliable indicator of similarity, we also noted
173 the pattern of induction scores for fluphenazine was similar to that observed for
174 atorvastatin (Figure. 2A). These results are consistent with reports that fluphenazine,
175 like statins, affects lipid metabolism in human patients(*24*).

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

186 To test whether these results were reproducible, a selection of the previously analysed
187 hiPSC lines (Supp Table 1) were regrown from frozen stocks and analysed in Cell
188 Painting assays after treatment with the same set of FDA-approved drugs, but this time
189 using a different High Content Screening (HCS) imaging system (data collected on a
190 Yokogawa CV7000 instead of InCell 2200). Using UMAP for visualisation, we observed
191 similar clustering of Cell Painting features generated in this repeat set of assays to those
192 collected from the original experiment that was performed 6 months previously, using
193 independent cultures of each hiPSC line and with data acquired using a different type of
194 HCS imaging system (Supp Figure 1B).

195 To further asses the reproducibility of the assay, we compared the measured phenotypes
196 of hiPSC lines in an additional run of the assay, executed after an interval of 2 yrs 8
197 months (both datasets were collected on the same InCell 2200 imager) but with a
198 different cohort of hiPSC lines (Supp Table 1). Figure 2C shows the comparison of the
199 two assays, using Principal Component Analysis (PCA). This further validates the
200 comparison between the two different runs and also avoids the effects of clustering in
201 high-dimensional space that can affect UMAP (*27*). PCA analysis shows that similar
202 compounds are indistinguishable in the two assay runs. Supplemental Figure 1 (C,E)
203 shows a comparison of the two assays and again reveals that phenotypic responses are
204 similar. As expected, PCA and UMAP show different clustering properties and different
205 variations across cell lines, but the specific clustering responses of cell lines treated with

206 different classes of drugs, as well as the variable responses observed to the same drug,
207 between cell lines derived from different donors, are reproducible in this assay format.
208 We conclude this represents intrinsic properties of the respective cell lines, independent
209 of variables specific to individual assay runs, such as specific cell line stocks, prepared
210 media, or the instrumentation used for image acquisition.

211 While many of the drug treatments resulted in tight clustering of the responses of different
212 cell lines on the UMAP and PCA plots (Figure 2), a subset of the drugs tested, i.e. statins,
213 5-FU and erlotinib, showed more dispersed clustering of cell line responses. We
214 hypothesised that this behaviour reflected an increased degree of variable response to
215 these drugs between the respective cell lines. To test this hypothesis, we calculated the
216 Euclidean distance of the feature vectors measured for each compound and DMSO, for
217 each donor. Supplemental Fig 3B shows the result of this calculation, along with the
218 mean and coefficient of variation (CoV) of the Euclidean distance for each drug across
219 donors. This metric is elevated for simvastatin, atorvastatin, erlotinib and fenbendazole,
220 consistent with the evidence from UMAP and the induction plot that response to these
221 drugs varies across the donors (*27*).

222 The variable responses shown in Figure 2 and Supplemental Fig 3B might be due to
223 intrinsic variations within an assay. To directly assess the variations in Cell Painting
224 features in different assays, we calculated the Earth Mover Distance (EMD, see
225 Methods) which has been shown previously to provide a useful measure of differences
226 between phenotypic features in HCS assays (*28*). Supp Fig 2A shows the distribution of
227 EMDs for all features from all compound treatments common between these two assay
228 runs. Most show EMDs similar to DMSO, suggesting the differences are in the range of
229 the assay controls. Cytotoxic compounds (e.g., bortezimib, dasatinib, colchicine) show a
230 broader distribution of EMDs, which is likely due to cell death caused by these
231 compounds during incubation. Inspired by the overall similarities in features between the
232 two assays, we calculated induction scores for the drugs and donors that are common
233 across these assays. Comparison of heatmaps of these induction scores shows the
234 variations between drugs, but also between donors within individual drugs, are similar
235 (Supp Fig 2B,C). These data suggest that the variations we observe in drug responses
236 (e.g., Figures 1 and 2) across hiPSC lines generated from different donors are robust
237 and reproducible.

238 **Linking Variable Responses to Protein Expression & Stoichiometry**

239 Our previous studies on this HipSci cohort have demonstrated that variations in protein
240 expression levels between lines derived from different donors can be controlled by

241 specific genomic loci, i.e., protein Quantitative Trait Loci (pQTLs) (*13*). This suggested
242 the hypothesis that a potential source of differential drug responses between hiPSC lines
243 might be due to epistatic interactions, i.e., resulting from genomic variation between
244 donors causing differences in protein expression that can modulate drug-induced
245 phenotypes, for example by affecting the stoichiometry of components of drug response
246 pathways (*29*). Importantly, this hypothesis takes into account drug-cell interactions
247 outside of the direct interaction of a drug with its protein target. Furthermore, it can
248 potentially explain why patterns of drug response features measured by Cell Painting
249 appear similar across cell lines, while the magnitude of responses between lines differ,
250 as seen in the feature and induction heatmaps (Figures 1B and 2A).

251 To test this hypothesis, we first used the induction plot (Figure 2A) and a graph of
252 induction values (Supp Figure 3A), to classify cell lines as, respectively, “low” and “high”
253 responders, for two well-characterised drugs, i.e., atorvastatin and simvastatin. These
254 drugs share the same target, but are well-known to have different responses and
255 efficacies in different individuals (*30*). Representative cell lines denw6 and zaie1 (“low
256 responders”) and hayt1 and tuju1 (“high responders”), were each treated separately with
257 either DMSO (control), or with either atorvastatin, or simvastatin, each at 5 μM, for 24
258 hrs, as in the Cell Painting assay. Following treatment, cell extracts were prepared and
259 proteomes analysed by LC-MS/MS (see Methods). From the resulting data, differences
260 in protein expression levels between DMSO- and drug-treated lines were visualised in
261 ‘volcano plots’ (Figure 3). Combining data for treatment of all four cell lines (tuju1, hayt1,
262 denw6, zaie1), with either atorvastatin (Figure 3A), or simvastatin (Figure 3B), showed
263 that many of the proteins whose expression increased after drug treatment are involved
264 in lipid metabolism (red dots in Figure 3A, B). Further, GO term enrichment analysis also
265 showed a significant enrichment for expression of protein factors involved in lipid
266 metabolism (Supp Table 3 and 4).

267 These data suggest that homeostatic mechanisms in the hiPSC lines respond to drug
268 mediated inhibition of cholesterol metabolism by increasing the expression of enzymes
269 required for cholesterol synthesis. While a common type of response to inhibition of
270 cholesterol synthesis is seen across the hiPSC lines, differences were evident in the
271 magnitude of responses between individual cell lines. This is illustrated by the response
272 to either atorvastatin, or simvastatin, as detected by Cell Painting induction (Figure 2A).
273 We therefore tested whether variation in response magnitude might be linked to
274 differences in protein expression between the respective high and low response lines.
275 Volcano plots comparing protein expression in the high and low response lines, after
276 treatment with either atorvastatin (Figure 4A), or simvastatin (Supp Figure. 4A), show

277 that multiple proteins involved in the cholesterol synthesis pathway (highlighted in
278 purple), together with some other classes of proteins, have increased expression
279 specifically in the high response lines. These data are consistent with the hypothesis that
280 differential protein expression may explain, at least in part, the differences in the
281 magnitude of drug responses between individual hiPSC lines.

282 Next, we compared the expression of individual proteins in each of the four drug-treated
283 lines (Figure 4B). Proteins were identified whose expression differed significantly in high
284 vs low responders ($|\log_2 FC| > 0.263$ and $p < 0.05$, see Materials and Methods). This
285 reveals clear differences in the responses to the two statin drugs, suggesting that there
286 may be variation in the mechanisms of response to statin treatment between each hiPSC
287 line (see also Supp Figure 4B, Supp Table 9).

288 Interestingly, changes in protein expression in the high response lines include increased
289 expression of HMGCR and HMGCS1, which are directly involved in the cholesterol
290 mevalonate synthesis pathway. In contrast, both “low response lines”, show increased
291 expression of proteins involved in regulating the stability and recycling of lipoprotein
292 receptor complexes. This includes PCSK9, a protein that provides post translational
293 regulation of LDL receptor and directs it for lysosomal degradation, rather than regular
294 recycling potentially limiting the efficacy of statins. In addition, SCARB1, a cell surface
295 HDL receptor that mediates HDL-cholesteryl ester uptake, as well as ApoE, a plasma
296 protein that participates in the transport of cholesterol and other lipids between cells and
297 has been linked to the effect of statins (*31-39*). Lipoprotein receptor complexes play an
298 important role in the regulation of plasma and intracellular cholesterol homeostasis(*40*).
299 These differences in protein expression between the respective high and low responder
300 hiPSC lines to statin treatment suggest that variable activity of biosynthetic and/or influx
301 mechanisms may be involved in determining how each hiPSC line responds to statins.
302 In addition to identifying well-established targets of statins, characterising the proteins
303 whose expression increases in the high response lines, using GO enrichment and
304 STRING analysis (https://string-db.org/), identified factors involved in Rab protein signal
305 transduction and in the metabolism of RNA (Figure 4C and Supp Figure 4C; Supp Table
306 5 and 7). These are all known to change expression upon statin treatment (*41-45*). In
307 contrast, the low response lines showed increased expression of proteins linked to the
308 cytoskeleton, along with several ubiquitin E2 and E3 enzymes, suggesting that ubiquitin-
309 conjugation levels may affect the degree of statin response (Figure 4D and Supp Figure
310 4D; Supp Table 6 and 8).

311 These data support the view that a diversity of factors and pathways can affect the
312 cellular response to well-defined drug perturbations and these differences are due, at

313 least in part, to population-level variations in gene and protein expression between the
314 respective lines.

315 To explore further the observed variable drug responses, we calculated the pairwise
316 correlation matrix for the induction metric (see Figure 2A) for 32 drugs across donors.
317 This provides a measure of the pattern of variable response across a given set of donors.
318 Supp Figure 5 shows a visualisation of these data in a heatmap and identifies sets of
319 compounds that share common patterns of variable response. For example, erlotinib and
320 afatinib, both ErbB family receptor inhibitors and the statins and fluphenazine, behave
321 similarly. Everolimus, etoposide and irinotecan share a similar variable response pattern,
322 which may be linked to a role for the mTor pathway in controlling the DNA damage
323 response (*46, 47*). In addition, warfarin, omeprazole, niclosamide and metoclopramide,
324 which have very different targets, nonetheless share a similar variable response pattern
325 across hiPSC lines. We hypothesise that this may be linked to a shared off-target effect,
326 e.g., mitotoxicity (*48*). This is an interesting point to explore in further studies in the future.

327

328 **DISCUSSION**

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

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

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

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

377 In this study, we have investigated potential mechanisms involved in the variable
378 responses seen to statins in this screening assay. Statins are drugs developed to reduce
379 the level of cholesterol in blood and thus lower the risk of developing cardiac and
380 circulatory diseases, including angina, heart attack and stroke (*52*). Separate, but
381 interconnected pathways, control cholesterol homeostasis, including cholesterol influx,
382 mediated by lipoprotein receptors, cholesterol synthesis, regulated by HMG-CoA
383 reductase and the mevalonate pathway and cholesterol efflux, mediated by transport
384 proteins. Each of these pathways respond to multiple inputs that collectively define
385 cholesterol levels (*53*). In addition, our data suggest that modulation of protein stability
386 by Ub-mediated proteolysis is also an important factor in cholesterol homeostasis (Figure
387 4D). Even within the relatively limited sample of different donor lines we have analysed
388 so far, we see evidence of varied regulatory effects. The comparative proteomic analysis
389 supports the view that "high response" lines can be determined by modulation of the
390 endogenous cholesterol synthesis pathway, while "low response lines" may instead
391 reflect primarily regulation of the cholesterol influx and efflux pathways.

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

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

401 **Acknowledgements**

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

406 **Declaration of Conflicting Interest**

407 The authors Melpomeni Platani, Angus I. Lamond and Jason R. Swedlow would like to
408 declare a conflict of interest through their affiliation with Tartan Cell Technologies, Ltd.

409

410

411 **MATERIALS & METHODS**

412 **Cell Culture**

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

419 To passage feeder-free hiPSC lines, cells were washed with PBS and incubated briefly
420 (3-5 min), with 0.5 mM PBS-EDTA solution. After PBS-EDTA solution was removed, cell
421 clusters were resuspended in E8 medium and seeded at ratios of 1:4 to 1:6, depending
422 on their confluency, on Geltrex-coated tissue culture dishes. Once established,
423 undifferentiated, feeder-free human iPSC lines were expanded and aliquots frozen for
424 future use, prior to transition to TeSR medium for further study. To transition to TeSR
425 medium supplemented with bFGF (Peprotech 100-1813, at a final concentration of 30
426 ng/ml), Noggin (Peprotech 120-10C, at a final concentration of 10ng/ml), Activin A
427 (Peprotech 120-14P, at a final concentration of 10ng/ml), E8 medium of hiPSC lines was
428 exchanged daily with medium containing increasing amounts of TeSR medium, until
429 human iPSC lines were completely maintained in TeSR medium alone, typically over 1-
430 2 weeks.

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

442 hiPSC line ID number and details of the cell lines are listed in Supp Table 1.

443 The data from the experimental runs are included in this report. Run 2021-02, 52 drugs,
444 28 donors 1,456 drug-donors pairs). Run 2021-10, 8 donors, 52 drugs (416 drug-donor
445 pairs). Run 2023-12, 18 donors, 52drug (936 drug-donor pairs).

446 **Cell Staining**

447 384-well tissue culture plates (CellCarrier-384 Ultra Microplates, Perkin Elmer 6057300)
448 were coated with human plasma fibronectin (Merck Millipore FC010) at a concentration
449 of 5 µg/cm^2. Cells were passaged as described above with TrypLE Select, counted and
450 seeded on the fibronectin-coated wells, at a concentration of 3x10^4 /cm^2. Cell line plating
451 was in rows, with six wells per condition for each cell line (six technical replicates). Each
452 replicate was sampled at 9 locations and repeated in a separate biological replicate
453 separated by a week (i.e. 2 bioreplicates per assay run).

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

457 Mitotracker (Thermo Fisher M22426), staining was performed in live cells with a 30 min
458 incubation at 37°C prior to fixation. Fixation, permeabilisation and immunostaining were
459 performed at room temperature. 8% Paraformaldehyde in PBS (PFA, Sigma F8775),
460 was added to an equal volume of cell culture medium in wells of a 384-well plate for a
461 final concentration of 4% Paraformaldehyde and incubated for 15 min. Fixed cells were
462 washed in PBS and permeabilised with 0.1% v/v Triton-X100 (Sigma) for 15 min.
463 Following permeabilisation, cells were washed in TBS and stained with DAPI (Thermo
464 Fisher D1306 at 5 µg/ml), SYTO 14 (Thermo Fisher S7576 at 2.5 µM), Concanavalin A
465 (Thermo Fisher C11252 at 20 µg/ml), Phalloidin Alexa (Thermo Fisher A12381 at 66
466 nM), and WGA (Thermo Fisher W11262 at 11.8 µg/ml) for 1hr. Staining volume was 30
467 µl/well. Cell fixation, permeabilisation, staining and washing steps (2x, wash volume 50
468 µl) were performed using automated liquid handling systems (405 plate washer, BioTek–
469 Tempest, Formulatrix).

470 **Drug treatment**

471 Drug dosing on 384-well plates was performed using an Echo acoustic liquid handler
472 and corresponding software (Labcyte). Screened compounds were selected from the
473 appropriate chemical library plates containing either Cloud library (Enamine), or an in-
474 house library of publicly available compounds. Drug dosing was performed at final
475 concentrations of; etoposide 1.7 µM, Berberine Chloride 2.68 µM and Fenbendazole
476 3.34 µM. For all other compounds drug dosing was performed at a final concentration of
477 5 µM. Supp Table 2 lists the compounds and their final concentrations used in this study.
478 For proteomics analysis drug treatments with either simvastatin (Tocris 1965), or
479 atorvastatin; (Selleckchem S5715) were performed in triplicate in 6-well plates at a final
480 concentration of 5 µM for 24hrs.

481 **Cell staining and analysis of pluripotency markers**

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

485 All fixation, permeabilisation and immunostaining analyses were performed at room
486 temperature, apart from primary antibody incubation, which was carried out at 4°C
487 overnight. 8% Paraformaldehyde in PBS (PFA, Sigma F8775), was added to an equal
488 volume of cell culture medium in wells of a 384-well plate for a final concentration of 4%
489 Paraformaldehyde and incubated for 15 min. Fixed cells were washed twice in PBS and
490 permeabilised with 0.5% v/v Triton-X100 (Sigma) for 10 min (apart from TRA1-81
491 staining). Following permeabilization cells were washed twice in PBS, then blocked in
492 10% normal donkey serum for 1 hr prior to antibody incubations. Pluripotency marker
493 antibodies used in this study are: Nanog, Oct4, Sox2, Tra-1-81 all at 1:500, (Oct4
494 (C30A3), Sox2 (D6D9), Nanog (D73G4) and TRA181 (TRA-1-81) Cell Signalling
495 StemLight Pluripotency Antibody Kit 9656). All affinity purified donkey secondary
496 antibodies (either Alexa 488 or Alexa 594) were purchased from Jackson
497 Immunoresearch Europe. Hoechst 33342 (Thermo Fisher) was used to stain DNA.
498 Images of cells (for each hiPSC line, 9 fields/well, 16 wells per pluripotency marker),
499 stained with pluripotency marker antibodies were segmented based on DAPI signal and
500 the integrated nuclear intensity of each marker was quantified using CellProfiler.
501 Distributions of integrated intensities of all nuclear pluripotency markers were monotonic
502 and correlated with integrated DAPI staining.

503 **Imaging**

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

511 **Data Processing, Analysis and Visualization of Multiple Cell Lines**

512 Raw images were imported into OMERO Plus (Glencoe Software, Inc., (*54*)) and then
513 processed using a custom pipeline in CellProfiler (*16*), which segmented nuclei,
514 cytoplasm, Golgi and cortical protrusions and calculated a range of defined features. All
515 further steps were executed using the Pandas Python library (*55*). Features for each

516 plate were Robust-Z normalised by the median of features for the DMSO control for that
517 plate (*16, 56*).The Robust Z-score converts the population of features derived from
518 individual cells to a distribution of zero mean and unit variance. Robust Z is calculated
519 according to the formula $x'_i = (x_i - \mu)/M$ where x_i is the measured value for each feature,

520 x'_i is the normalised output value and μ and M are the median and median absolute
521 deviation (*18-20*). Features with a coefficient of variation >0.50 were removed from
522 further analysis. We further removed all features with |Spearman coefficient| > 0.98. Data
523 were then visualised using Uniform Manifold Approximation and Projection (UMAP (*23*)).
524 UMAP parameters were n_neighbours = 15, mindist = 0.01 and a cosine metric was
525 used for scaling the distance between points. GO enrichment analysis was carried out
526 using WebGestalt (WEB-based Gene SeT AnaLysis Toolkit) (*57, 58*). Functional protein-
527 protein interaction analysis was carried out using STRING-DB (*59*).

528 Nearest neighbour distances were calculated using Euclidean distance in scikit-learn
529 (*60*). The Earth Mover Distance (EMD) was calculated using sci-py (*61*) for filtered features from
530 individual wells for each drug-donor pair across a pair of assays. The frequency distribution of
531 EMDs for filtered features was calculated and plotted using Seaborn (*62*).

532 **Proteomics sample preparation**

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

539 All hiPSCs samples were washed twice with 1x PBS, on ice, then centrifuged at 300g to
540 collect the cell pellets. 4% SDS lysis buffer (10 mM TCEP, 100 mM Tris-HCl, pH 7.4)
541 was used to lyse the cells and extract proteins, with 1x Protease and Phosphatase
542 Inhibitor (A32961, Thermo Fisher) added. 1 mL lysis buffer was added to the cell pellets.
543 After 10 cycles of sonication, (30s on/off), using the Bioruptor® Pico, the protein samples
544 were centrifuged at 20,000*xg* for 15 min. The supernatant proteins were denatured at
545 95°C for 10 min, then alkylated with 40 mM IAA (Iodoacetamide), (final concentration)
546 and incubated in the dark at room temperature for 30 min. The protein concentration was
547 measured using the EZQ™ Protein Quantitation Kit following the manufacturer's
548 instructions.

549 500 ug proteins from each sample were further processed using the SP3 protocol, as
550 described (*63*). In brief, protein samples were mixed with SP3 beads (1:10,

551 protein:beads) and digested with LysC/trypsin mixture (1:50, enzyme:protein). Peptides
552 were eluted from the SP3 beads with 2% DMSO. The peptide concentration was
553 measured using the Pierce™ Quantitative Fluorometric Peptide Assay following the
554 manufacturer's instructions. Peptide samples were stored at -20℃ before LC-MS/MS
555 analysis.

556 **LC-MS/MS analysis**

557 For each cell line, three biological replicates per condition were generated. They were
558 combined in one sample, which was analysed in three separate, replicate MS runs. Each
559 hiPSC line was treated three times with either vehicle control, or specified drug.
560
561 All peptide samples were analysed by using an Orbitrap Fusion Tribrid mass
562 spectrometer (Thermo Fisher Scientific), equipped with a Dionex UltiMate™ 3000
563 RSLCnano System, Thermo Fisher Scientific). Peptides from each sample were loaded
564 onto a 75 µm × 2 cm PepMap-C18 pre-column and resolved on a 75 µm × 50 cm
565 PepMap-C18 EASY-Spray temperature-controlled integrated column-emitter (Thermo
566 Fisher Scientific), using either a 160 min or a 280 min multistep gradient from 5% B to
567 95% B with a constant flow rate of 300 nl/min. The mobile phases are: H_2O incorporating
568 0.1% FA (solvent A) and 80% ACN in H_2O incorporating 0.1% FA (solvent B). The details
569 for 160 min gradient are: 0-6 min, 1%-5% B; 6-145 min, 5%-38% B; 145-155 min, 38%-
570 95% B; 155-160 min, 95%B. The details for 280 min gradient are: 0-6 min, 1%-5% B; 6-
571 255 min, 5%-38% B; 255-270 min, 38%-95% B; 270-280 min, 95% B.
572 For total proteome analysis, the data were acquired under the control of Xcalibur
573 software in data-independent acquisition mode (DIA). For each sample, one run of 280
574 min method with 97 DIA windows was performed for boosting the identification of
575 proteins (*64*) and another three replicate runs of 160 min method were performed for the
576 quantification of proteins.
577 For the 160 min gradient method, the survey scan was acquired in the Orbitrap covering
578 the m/z range from 375 to 1,175 Da with a mass resolution of 60,000 and an automatic
579 gain control (AGC) target of 4.0×10^5 ions. MS/MS was acquired in the Orbitrap covering
580 the m/z range from 200 to 2,000 Da with a mass resolution of 30,000 and an AGC target
581 of 1.0×10^5 and 60 ms maximum injection time. The sequential DIA isolation windows
582 are 10 m/z from 375 to 775 and 40 m/z from 775-1175. Fragmentation was performed
583 using a HCD collision energy of 30%. For the 280 min gradient method, the survey scan
584 was acquired in the Orbitrap covering the m/z range from 355 to 1,150 Da with a mass
585 resolution of 120,000 and an automatic gain control (AGC) target of 4.0×10^5 ions.
586 MS/MS was acquired in the Orbitrap covering the m/z range from 200 to 2,000 Da with

587 a mass resolution of 30,000 and an AGC target of 1.0×10^5 and 60 ms maximum injection
588 time. The sequential DIA isolation windows are 15 m/z from 350 to 410, 10 m/z from 410
589 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
590 from 875 to 965 and 27 m/z from 965-1150. The fragmentation was performed using a
591 HCD collision energy of 30%.

592 **DIA MS data analysis**

593 For DIA analysis, the raw MS data of different sample runs from both drug and DMSO
594 set were analysed together using DIA-NN v1.8.1(*65*). The in silico spectral library was
595 generated by DIA-NN, using the Homo sapiens database from UniProt (SwissProt
596 October 2021). The data were searched with the following parameters: stable
597 modification of carbamidomethyl (C), N-term M excision, with a maximum of 1 missed
598 tryptic cleavage threshold, match between runs (MBR) option checked. DIA-NN output
599 reports were further processed with an R package, as described previously (*66*) to
600 perform the MaxLFQ-based protein quantification, with precursors x samples matrix
601 filtered at 1% precursor and protein group FDR. Statistical analyses were performed
602 using Perseus v1.6.15.0 (*67, 68*) In brief, the protein quantification result was imported
603 to Perseus and the copy number of each protein was calculated based on its intensity
604 using a "Proteomic Ruler" (*69*). Only proteins with valid quantification values in at least
605 1/3 of total sample runs were considered for further comparison. Missing values were
606 replaced from a normal distribution. After median normalisation, two sample student's T
607 tests were performed for p-value generation, using permutation-based FDR for
608 truncation. Heatmap and K-means clustering analysis was performed based on the
609 Euclidean distance, using the default settings in Perseus. In total, 5,895 proteins were
610 quantified using DIA analysis and used for further analysis.
611 For heat map generation of Figure 4B and Supp Figure 4B: At the first step of differential
612 expression analysis, a two sample-test using Student's T-test with the default setting in
613 Perseus was performed to determine if the means of two groups (drug and DMSO, for
614 either high or low response cell lines) are significantly different from each other.
615 The protein fold change ($FC_{high\ or\ low}$) and p value ($P_{high\ or\ low}$) were generated, respectively.
616 Proteins with $p<0.05$ were considered further and used to create the results show in
617 Figure 4B and Supp Figure 4B. Z-score normalisation was used for Heat map analysis,
618 according to the protocol in (*68*).
619

620 **Data availability**

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium, via the PRIDE (*70*) partner repository with the dataset identifier PXD050194. Cell painting data will be deposited in the Image Data Resource (IDR).

Code availability

All data analysis code used methods published by others (*16*) and tools from pandas (*71*), numpy (*72*), scikit-learn (*73*), SciPy (*61*).

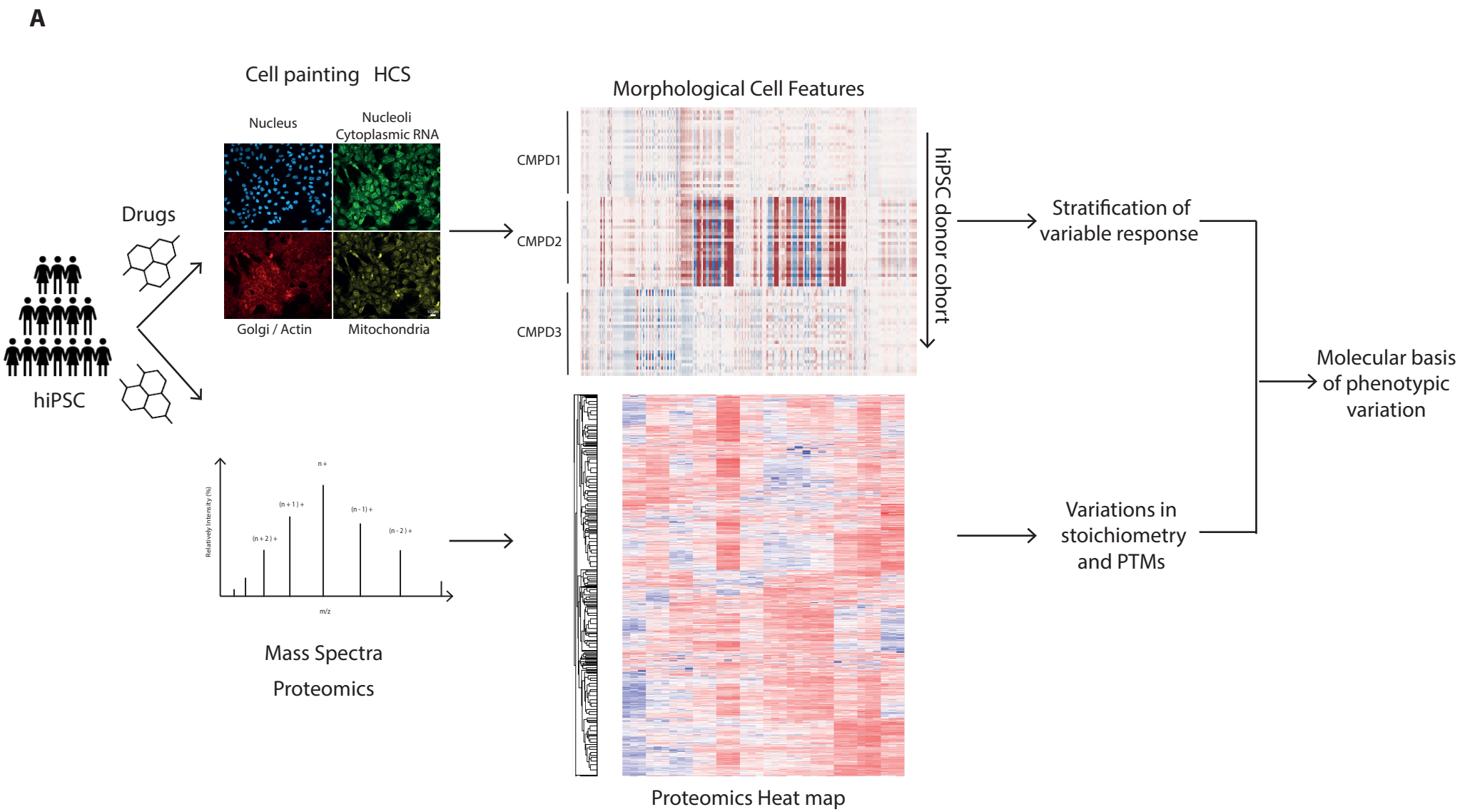

A
Cell painting HCS
Nucleus
Nucleoli
Cytoplasmic RNA
Golgi / Actin
Mitochondria
Drugs
hiPSC
Morphological Cell Features
CMPD1
CMPD2
CMPD3
hiPSC donor cohort
Stratification of variable response
Molecular basis of phenotypic variation
Relatively Intensity (%)
m/z
Mass Spectra
Proteomics
Variations in stoichiometry and PTMs
Proteomics Heat map

Figure 1A

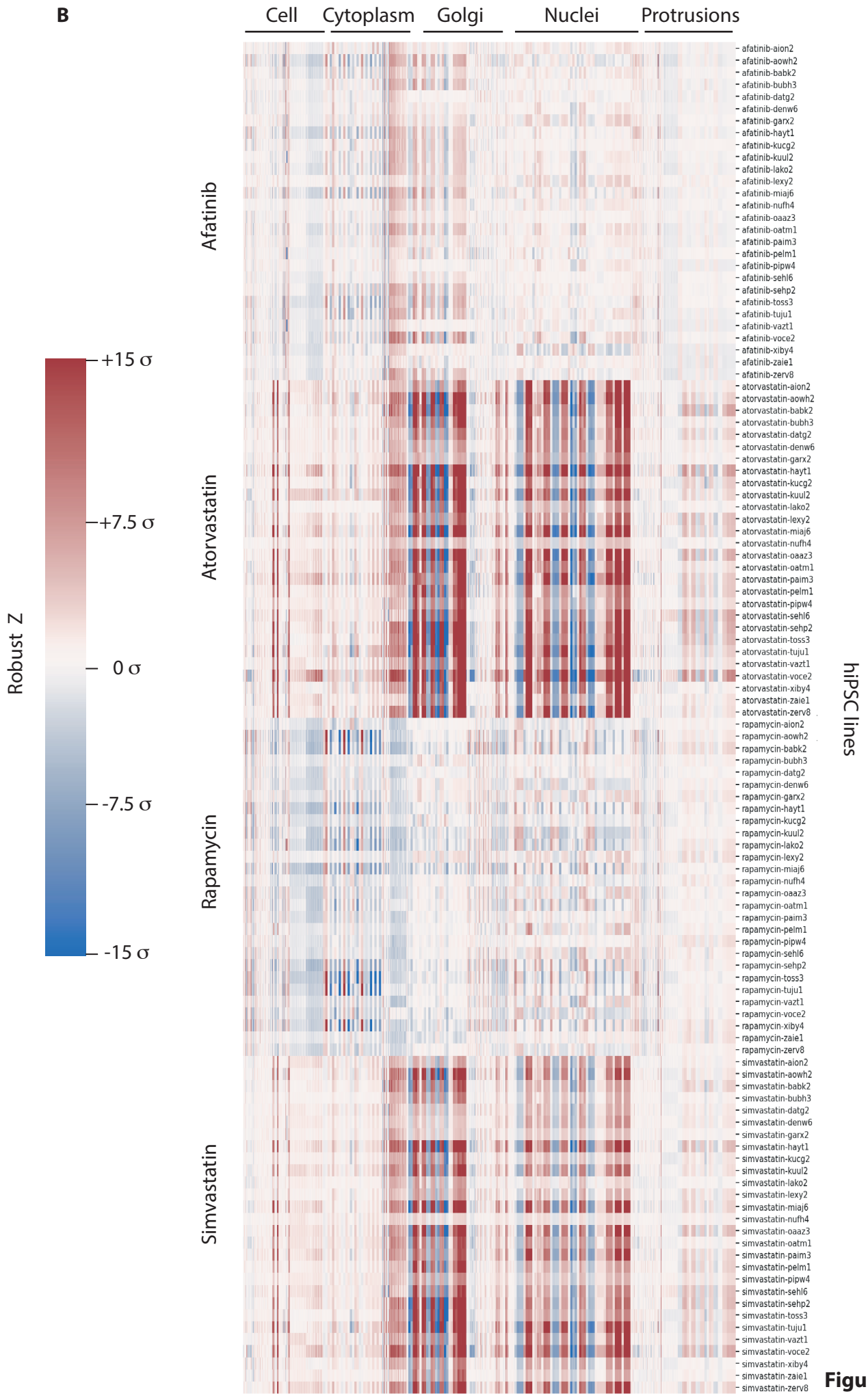

B
Cell
Cytoplasm
Golgi
Nuclei
Protrusions
Afatinib
Atorvastatin
Rapamycin
Simvastatin
hiPSC lines
Robust Z
+15 σ
+7.5 σ
0 σ
-7.5 σ
-15 σ
afatinib-aion2
afatinib-aowh2
afatinib-babk2
afatinib-bubh3
afatinib-datg2
afatinib-denw6
afatinib-garx2
afatinib-hayt1
afatinib-kucg2
afatinib-kuul2
afatinib-lako2
afatinib-lexy2
afatinib-miaj6
afatinib-nufh4
afatinib-oaaz3
afatinib-oatm1
afatinib-paim3
afatinib-pelm1
afatinib-pipw4
afatinib-sehl6
afatinib-sehp2
afatinib-toss3
afatinib-tuju1
afatinib-vazt1
afatinib-voce2
afatinib-xiby4
afatinib-zaie1
afatinib-zerv8
atorvastatin-aion2
atorvastatin-aowh2
atorvastatin-babk2
atorvastatin-bubh3
atorvastatin-datg2
atorvastatin-denw6
atorvastatin-garx2
atorvastatin-hayt1
atorvastatin-kucg2
atorvastatin-kuul2
atorvastatin-lako2
atorvastatin-lexy2
atorvastatin-miaj6
atorvastatin-nufh4
atorvastatin-oaaz3
atorvastatin-oatm1
atorvastatin-paim3
atorvastatin-pelm1
atorvastatin-pipw4
atorvastatin-sehl6
atorvastatin-sehp2
atorvastatin-toss3
atorvastatin-tuju1
atorvastatin-vazt1
atorvastatin-voce2
atorvastatin-xiby4
atorvastatin-zaie1
atorvastatin-zerv8
rapamycin-aion2
rapamycin-aowh2
rapamycin-babk2
rapamycin-bubh3
rapamycin-datg2
rapamycin-denw6
rapamycin-garx2
rapamycin-hayt1
rapamycin-kucg2
rapamycin-kuul2
rapamycin-lako2
rapamycin-lexy2
rapamycin-miaj6
rapamycin-nufh4
rapamycin-oaaz3
rapamycin-oatm1
rapamycin-paim3
rapamycin-pelm1
rapamycin-pipw4
rapamycin-sehl6
rapamycin-sehp2
rapamycin-toss3
rapamycin-tuju1
rapamycin-vazt1
rapamycin-voce2
rapamycin-xiby4
rapamycin-zaie1
rapamycin-zerv8
simvastatin-aion2
simvastatin-aowh2
simvastatin-babk2
simvastatin-bubh3
simvastatin-datg2
simvastatin-denw6
simvastatin-garx2
simvastatin-hayt1
simvastatin-kucg2
simvastatin-kuul2
simvastatin-lako2
simvastatin-lexy2
simvastatin-miaj6
simvastatin-nufh4
simvastatin-oaaz3
simvastatin-oatm1
simvastatin-paim3
simvastatin-pelm1
simvastatin-pipw4
simvastatin-sehl6
simvastatin-sehp2
simvastatin-toss3
simvastatin-tuju1
simvastatin-vazt1
simvastatin-voce2
simvastatin-xiby4
simvastatin-zaie1
simvastatin-zerv8

Figure 1B

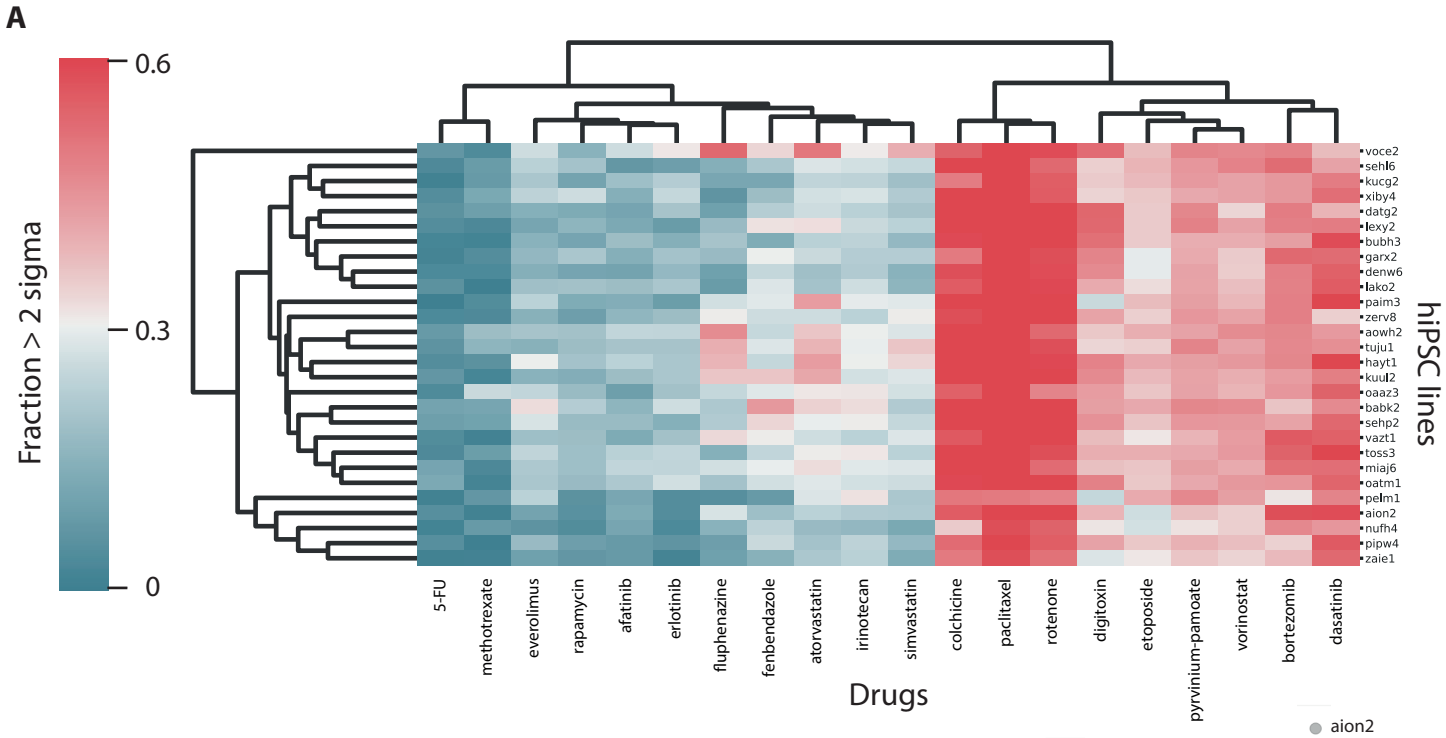
630

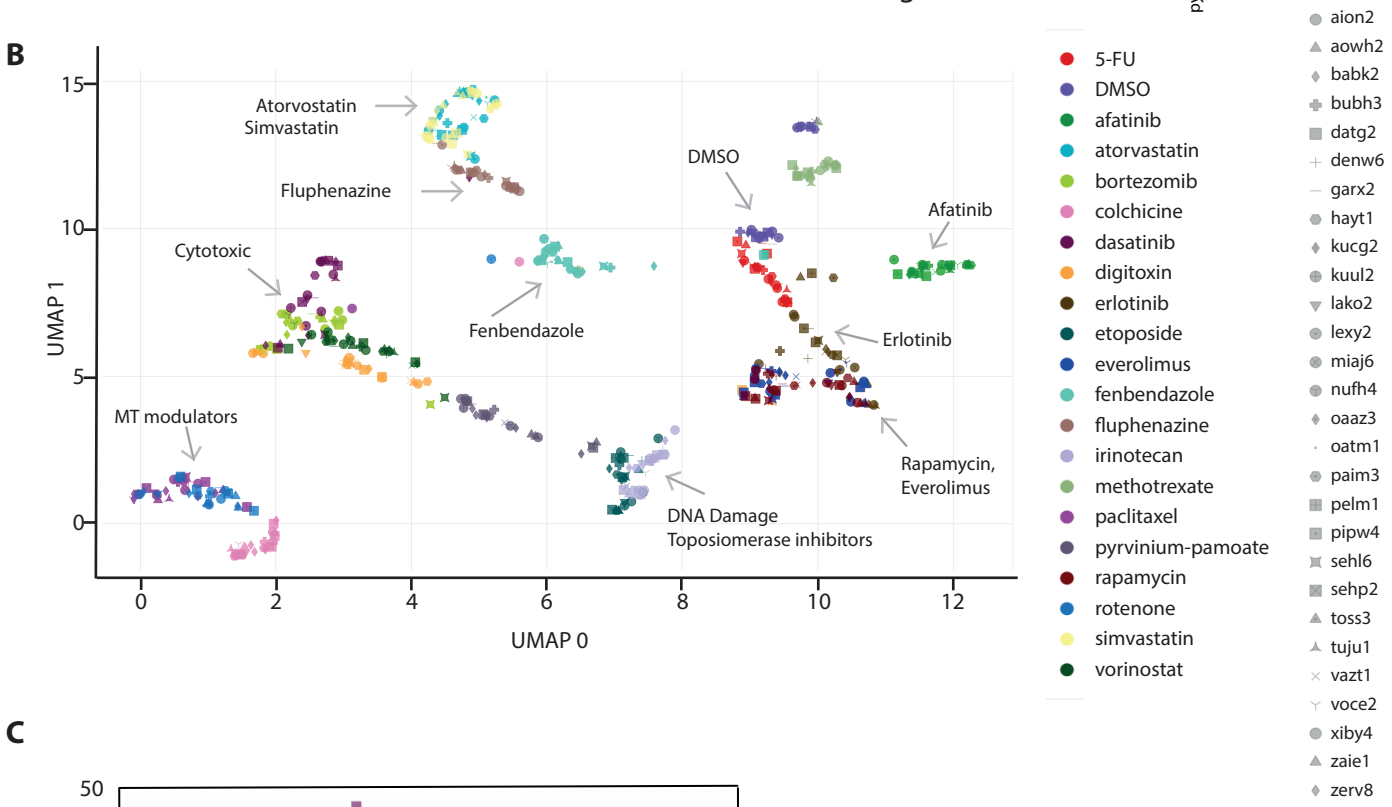
631 Figure 1

632 (A) Study overview. Quality controlled hiPSC lines from 43 donors in their pluripotent
633 state were used for Cell Painting and proteomic analysis, by mass spectrometry,
634 following addition of 52 FDA approved drugs.-38.2 million cells were imaged across all
635 donors and 871 morphological traits were quantified per donor/drug combination.
636 Variability of drug response of the hiPSC cohort was studied using both phenotypic
637 screening and proteomics.

638 (B) Heatmap of RobustZ scores from Cell Painting shows variations in the magnitude of
639 drug responses across hiPSC lines. Differential responses to afatinib, atorvastatin,
640 rapamycin and simvastatin are shown. The response to each drug creates a
641 characteristic pattern of features, but with differences in the magnitude of the response.
642 Color bar shows RobustZ scores normalized to DMSO expressed as standard
643 deviations.

644


A
Fraction > 2 sigma
0.6
0.3
0
hiPSC lines
voce2
sehl6
kucg2
xiby4
datg2
lexy2
bubh3
garx2
denw6
lako2
paim3
zerv8
aowh2
tuju1
hayt1
kuul2
oaaz3
babk2
sehp2
vazt1
toss3
miaj6
oatm1
pelm1
aion2
nufh4
pipw4
zaie1
5-FU
methotrexate
everolimus
rapamycin
afatinib
erlotinib
fluphenazine
fenbendazole
atorvastatin
irinotecan
simvastatin
colchicine
paclitaxel
rotenone
digitoxin
etoposide
pyrvinium-pamoate
vorinostat
bortezomib
dasatinib
Drugs


B
UMAP 1
UMAP 0
Atorvostatin
Simvastatin
Fluphenazine
Cytotoxic
MT modulators
Fenbendazole
DMSO
Afatinib
Erlotinib
Rapamycin,
Everolimus
DNA Damage
Toposiomerase inhibitors
5-FU
DMSO
afatinib
atorvastatin
bortezomib
colchicine
dasatinib
digitoxin
erlotinib
etoposide
everolimus
fenbendazole
fluphenazine
irinotecan
methotrexate
paclitaxel
pyrvinium-pamoate
rapamycin
rotenone
simvastatin
vorinostat
aion2
aowh2
babk2
bubh3
datg2
denw6
garx2
hayt1
kucg2
kuul2
lako2
lexy2
miaj6
nufh4
oaaz3
oatm1
paim3
pelm1
pipw4
sehl6
sehp2
toss3
tuju1
vazt1
voce2
xiby4
zaie1
zerv8

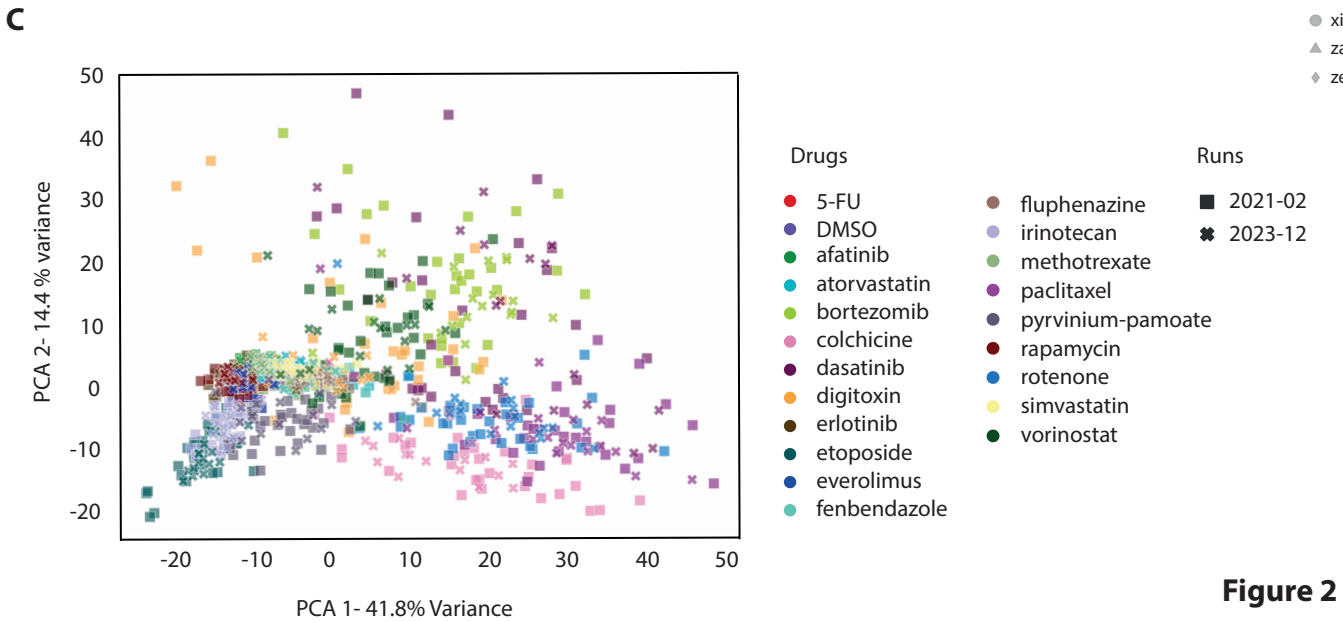

C
PCA 2- 14.4 % variance
PCA 1- 41.8% Variance
Drugs
5-FU
DMSO
afatinib
atorvastatin
bortezomib
colchicine
dasatinib
digitoxin
erlotinib
etoposide
everolimus
fenbendazole
fluphenazine
irinotecan
methotrexate
paclitaxel
pyrvinium-pamoate
rapamycin
rotenone
simvastatin
vorinostat
Runs
2021-02
2023-12

Figure 2

646 Figure 2

647 (A) Induction plot of drug responses for each hiPSC line shows variations in response
648 magnitude. Induction is the fraction of features for each combination of drug and hiPSC
649 line that differ in value by > 2σ (standard deviation) away from the DMSO control.

650 (B) UMAP embedding of Cell Painting features measured for 28 hiPSC lines treated with
651 different drugs. Cell lines are represented on the graph with different shapes and drugs
652 with different colours. Representative separated clusters are indicated by arrows. For
653 visibility, a subset of drugs used in this screen are shown.

654 (C) PCA analysis of Cell Painting features measured for human iPSC lines from
655 phenotypic screens performed in 2021-02 (28 lines) and 2023-12 (18 lines). Runs are
656 represented on the graph with different shapes ("Run 2021-02" solid squares and "Run
657 2023-12" crosses) and drugs with different colours. The reproducibility of the assay
658 results on different dates is indicated by the overlap of drug treatments in the different
659 assay runs.

660

661

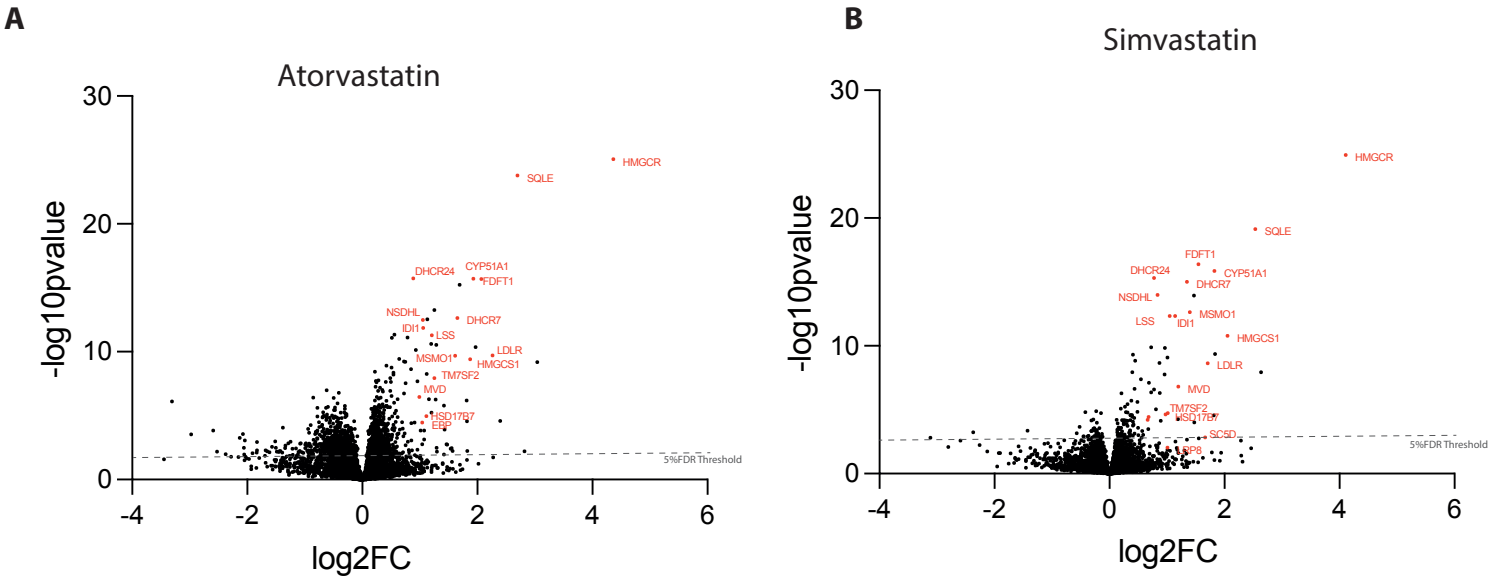

A
Atorvastatin
-log10pvalue
log2FC
HMGCR
SQLE
DHCR24
CYP51A1
FDFT1
NSDHL
DHCR7
IDI1
LSS
LDLR
MSMO1
HMGCS1
TM7SF2
MVD
HSD17B7
EBP
5%FDR Threshold
B
Simvastatin
-log10pvalue
log2FC
HMGCR
SQLE
FDFT1
DHCR24
CYP51A1
DHCR7
NSDHL
MSMO1
LSS
IDI1
HMGCS1
LDLR
MVD
TM7SF2
HSD17B7
SC5D
EBP
5%FDR Threshold

Figure 3

663 Figure 3

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

671

672

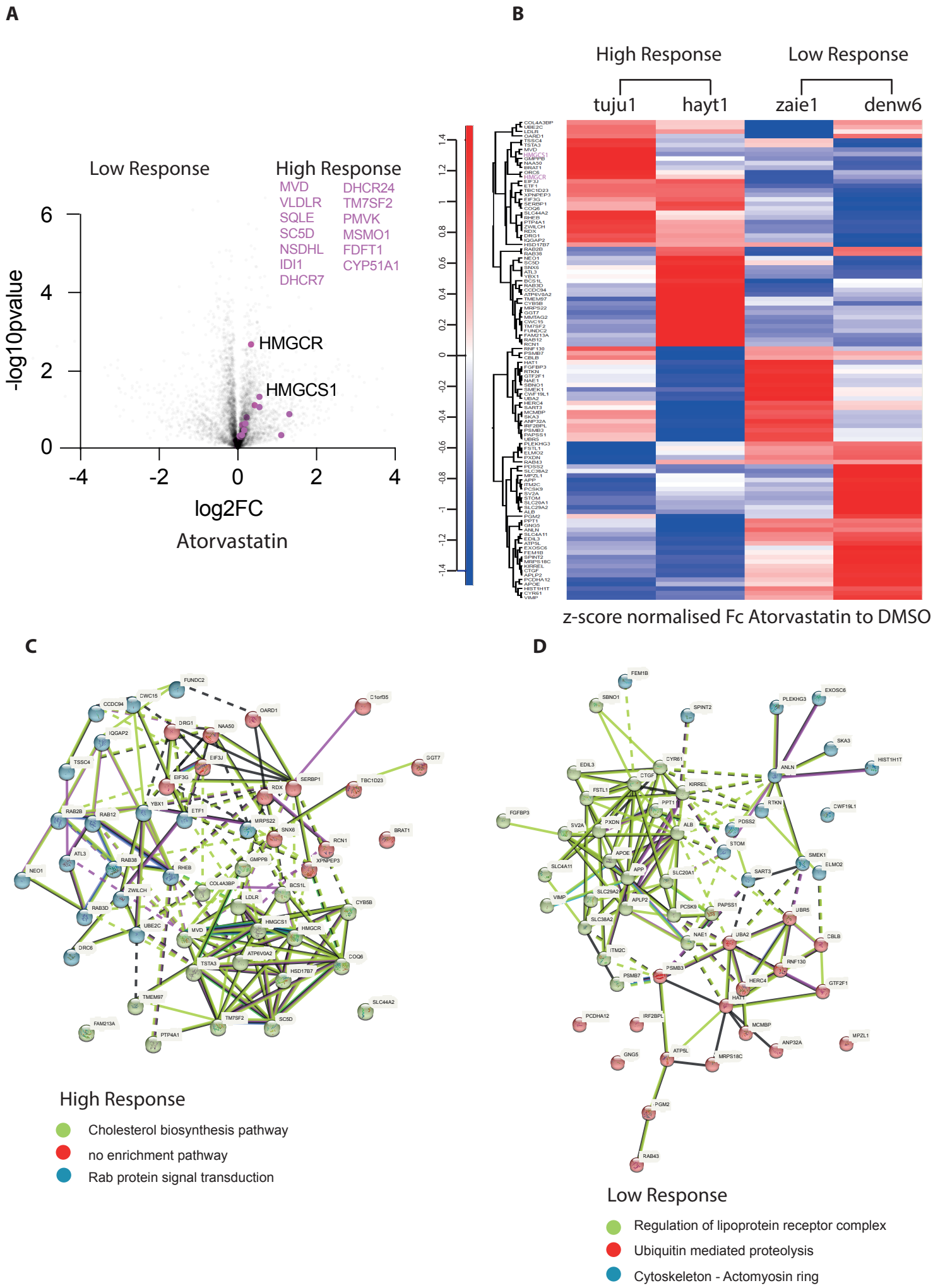

A
Low Response
High Response
MVD
VLDLR
SQLE
SC5D
NSDHL
IDI1
DHCR7
DHCR24
TM7SF2
PMVK
MSMO1
FDFT1
CYP51A1
HMGCR
HMGCS1
-log10pvalue
log2FC
Atorvastatin
B
High Response
Low Response
tuju1
hayt1
zaie1
denw6
z-score normalised Fc Atorvastatin to DMSO
C
High Response
Cholesterol biosynthesis pathway
no enrichment pathway
Rab protein signal transduction
D
Low Response
Regulation of lipoprotein receptor complex
Ubiquitin mediated proteolysis
Cytoskeleton - Actomyosin ring

Figure 4

674 Figure 4

675 (A) Volcano plot of differential protein expression in four hiPSC lines, stratified by their
676 respective high, or low, response to atorvastatin (see Supp Fig 3A). High response
677 lines: tuju1 and hayt1. Low response lines: denw6 and zaie1. Protein FC and p value
678 for each cell line were calculated as in Fig. 3. Sample runs from the drug treatment set
679 (n=6), were considered as one group and the DMSO control set (n=6), were
680 considered as a separate group. Members of the cholesterol biosynthesis pathway are
681 represented as purple dots and listed on the right side of the graph.

682 (B) Heat map analysis of proteins showing differential expression ($log2FC_{high-low}$ >0.263)
683 between the respective high and low response hiPSC lines, following atorvastatin
684 treatment. Heatmap generation steps: the protein fold change (FC_{high}) and p value
685 (P_{high}) for high response cell lines were calculated using sample runs from both high
686 response cell lines to Atorvastatin set (n=6), considered as one group and the DMSO
687 control set (n=6) considered as a separate group. The protein fold change (FC_{low}) and
688 p value (P_{low}) for low response cell lines were also calculated. Only proteins with both
689 $P_{high} \leq 0.05$ and $P_{low} \leq 0.05$ were considered further to compare the difference between
690 high and low response. log2FC of these filtered proteins was calculated as ($log2FC_{high-}$
691 $_{low}$= $log2FC_{high}$- $log2FC_{low}$). Heatmap shows FC between drug and DMSO control of the
692 filtered list z-score normalised. StringDB analysis of filtered proteins shown in (B), in
693 high response lines (C) and low response lines (D).
694

695

696 **References**

697 1. A. Loewa, J. J. Feng, S. Hedtrich, Human disease models in drug development.
698 *Nat Rev Bioeng*, 1-15 (2023).
699 2. H. Golding, S. Khurana, M. Zaitseva, What Is the Predictive Value of Animal
700 Models for Vaccine Efficacy in Humans? The Importance of Bridging Studies
701 and Species-Independent Correlates of Protection. *Cold Spring Harb Perspect*
702 *Biol* **10**, (2018).
703 3. R. Franco, A. Cedazo-Minguez, Successful therapies for Alzheimer's disease:
704 why so many in animal models and none in humans? *Front Pharmacol* **5**, 146
705 (2014).
706 4. H. Dowden, J. Munro, Trends in clinical success rates and therapeutic focus.
707 *Nat Rev Drug Discov* **18**, 495-496 (2019).
708 5. R. K. Harrison, Phase II and phase III failures: 2013-2015. *Nat Rev Drug Discov*
709 **15**, 817-818 (2016).
710 6. H. T. Rakshith, S. Lohita, A. P. Rebello, P. S. Goudanavar, N. Raghavendra
711 Naveen, Sex differences in drug effects and/or toxicity in oncology. *Curr Res*
712 *Pharmacol Drug Discov* **4**, 100152 (2023).
713 7. S. Oh *et al.*, Effectiveness of antipsychotic drugs in schizophrenia: a 10-year
714 retrospective study in a Korean tertiary hospital. *NPJ Schizophr* **6**, 32 (2020).
715 8. N. D. Wong *et al.*, Residual atherosclerotic cardiovascular disease risk in statin-
716 treated adults: The Multi-Ethnic Study of Atherosclerosis. *J Clin Lipidol* **11**,
717 1223-1233 (2017).
718 9. R. J. Tyson *et al.*, Precision Dosing Priority Criteria: Drug, Disease, and Patient
719 Population Variables. *Front Pharmacol* **11**, 420 (2020).
720 10. K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse
721 embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676
722 (2006).
723 11. N. Malik, M. S. Rao, A review of the methods for human iPSC derivation.
724 *Methods Mol Biol* **997**, 23-33 (2013).
725 12. H. Kilpinen *et al.*, Common genetic variation drives molecular heterogeneity in
726 human iPSCs. *Nature* **546**, 370-375 (2017).
727 13. B. A. Mirauta *et al.*, Population-scale proteome variation in human induced
728 pluripotent stem cells. *Elife* **9**, (2020).
729 14. K. Takahashi, S. Yamanaka, Induced pluripotent stem cells in medicine and
730 biology. *Development* **140**, 2457-2461 (2013).
731 15. A. Brenes, Bensaddek, D., Hukelmann, J., Afzal, V. & Lamond, A. I. , The iPSC
732 proteomic compendium. *bioRxiv 469916 (2018) doi:10.1101/469916.*, (2018).
733 16. M. A. Bray *et al.*, Cell Painting, a high-content image-based assay for
734 morphological profiling using multiplexed fluorescent dyes. *Nat Protoc* **11**, 1757-
735 1774 (2016).
736 17. C. McQuin *et al.*, CellProfiler 3.0: Next-generation image processing for biology.
737 *PLoS Biol* **16**, e2005970 (2018).
738 18. M. A. Bray, A. Carpenter, in *Assay Guidance Manual,* S. Markossian *et al.*, Eds.
739 (Bethesda (MD), 2017).
740 19. A. Birmingham *et al.*, Statistical methods for analysis of high-throughput RNA
741 interference screens. *Nat Methods* **6**, 569-575 (2009).
742 20. N. Malo, J. A. Hanley, S. Cerquozzi, J. Pelletier, R. Nadon, Statistical practice in
743 high-throughput screening data analysis. *Nat Biotechnol* **24**, 167-175 (2006).
744 21. E. Williams *et al.*, The Image Data Resource: A Bioimage Data Integration and
745 Publication Platform. *Nat Methods* **14**, 775-781 (2017).
746 22. C. Willis, J. Nyffeler, J. Harrill, Phenotypic Profiling of Reference Chemicals
747 across Biologically Diverse Cell Types Using the Cell Painting Assay. *SLAS*
748 *Discov* **25**, 755-769 (2020).

23. L. McInnes, Healy, J. & Melville, J., UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. . *arXiv [stat.ML]* (2018).
24. E. A. Thomas *et al.*, Antipsychotic drug treatment alters expression of mRNAs encoding lipid metabolism-related proteins. *Mol Psychiatry* **8**, 983-993, 950 (2003).
25. P. Srivastava, D. Panda, Rotenone inhibits mammalian cell proliferation by inhibiting microtubule assembly through tubulin binding. *FEBS J* **274**, 4788-4801 (2007).
26. N. Dogra, A. Kumar, T. Mukhopadhyay, Fenbendazole acts as a moderate microtubule destabilizing agent and causes cancer cell death by modulating multiple cellular pathways. *Sci Rep* **8**, 11926 (2018).
27. P. L. Chari T, The spacious Art of Single-Cell Genomics. *bioRxiv* (2021).
28. Y. E. Pearson *et al.*, A statistical framework for high-content phenotypic profiling using cellular feature distributions. *Commun Biol* **5**, 1409 (2022).
29. T. I. Roumeliotis *et al.*, Genomic Determinants of Protein Abundance Variation in Colorectal Cancer Cells. *Cell Rep* **20**, 2201-2214 (2017).
30. C. R. Sirtori, G. Mombelli, M. Triolo, R. Laaksonen, Clinical response to statins: mechanism(s) of variable activity and adverse effects. *Ann Med* **44**, 419-432 (2012).
31. T. Nozue, Lipid Lowering Therapy and Circulating PCSK9 Concentration. *J Atheroscler Thromb* **24**, 895-907 (2017).
32. Y. Huang, R. W. Mahley, Apolipoprotein E: structure and function in lipid metabolism, neurobiology, and Alzheimer's diseases. *Neurobiol Dis* **72 Pt A**, 3-12 (2014).
33. I. F. Tudorache, V. G. Trusca, A. V. Gafencu, Apolipoprotein E - A Multifunctional Protein with Implications in Various Pathologies as a Result of Its Structural Features. *Comput Struct Biotechnol J* **15**, 359-365 (2017).
34. Q. Bi *et al.*, Polymorphisms of the apolipoprotein E gene affect response to atorvastatin therapy in acute ischemic stroke. *Front Cardiovasc Med* **9**, 1024014 (2022).
35. L. Zhang *et al.*, Apolipoprotein E polymorphisms contribute to statin response in Chinese ASCVD patients with dyslipidemia. *Lipids Health Dis* **18**, 129 (2019).
36. C. Cai, Z. Wen, L. Li, The relationship between ApoE gene polymorphism and the efficacy of statins controlling hyperlipidemia. *Am J Transl Res* **13**, 6772-6777 (2021).
37. E. Zintzaras, G. D. Kitsios, F. Triposkiadis, J. Lau, G. Raman, APOE gene polymorphisms and response to statin therapy. *Pharmacogenomics J* **9**, 248-257 (2009).
38. W. J. Shen, S. Azhar, F. B. Kraemer, SR-B1: A Unique Multifunctional Receptor for Cholesterol Influx and Efflux. *Annu Rev Physiol* **80**, 95-116 (2018).
39. R. W. Mahley, Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. *Science* **240**, 622-630 (1988).
40. X. D. Xia *et al.*, Regulation of PCSK9 Expression and Function: Mechanisms and Therapeutic Implications. *Front Cardiovasc Med* **8**, 764038 (2021).
41. E. S. Istvan, J. Deisenhofer, Structural mechanism for statin inhibition of HMG-CoA reductase. *Science* **292**, 1160-1164 (2001).
42. B. R. Ali, I. Nouvel, K. F. Leung, A. N. Hume, M. C. Seabra, A novel statin-mediated "prenylation block-and-release" assay provides insight into the membrane targeting mechanisms of small GTPases. *Biochem Biophys Res Commun* **397**, 34-41 (2010).
43. E. Ronzier, X. X. Parks, H. Qudsi, C. M. Lopes, Statin-specific inhibition of Rab-GTPase regulates cPKC-mediated IKs internalization. *Sci Rep* **9**, 17747 (2019).
44. M. S. Brown, J. L. Goldstein, Multivalent feedback regulation of HMG CoA reductase, a control mechanism coordinating isoprenoid synthesis and cell growth. *J Lipid Res* **21**, 505-517 (1980).

45. S. A. Grunwald *et al.*, Statin-induced myopathic changes in primary human muscle cells and reversal by a prostaglandin F2 alpha analogue. *Sci Rep* **10**, 2158 (2020).
46. X. Zhou *et al.*, Regulation of CHK1 by mTOR contributes to the evasion of DNA damage barrier of cancer cells. *Sci Rep* **7**, 1535 (2017).
47. A. Amar-Schwartz *et al.*, S6K1 phosphorylates Cdk1 and MSH6 to regulate DNA repair. *Elife* **11**, (2022).
48. S. Seal *et al.*, Integrating cell morphology with gene expression and chemical structure to aid mitochondrial toxicity detection. *Commun Biol* **5**, 858 (2022).
49. S. Ryu *et al.*, Human Pluripotent Stem Cells for High-Throughput Drug Screening and Characterization of Small Molecules. *Methods Mol Biol* **2454**, 811-827 (2022).
50. J. Naidoo, T. Hurrell, J. Scholefield, The generation of human induced pluripotent stem cell lines from individuals of Black African ancestry in South Africa. *Stem Cell Res* **81**, 103534 (2024).
51. T. Hurrell, J. Naidoo, T. Ntlhafu, J. Scholefield, An African perspective on genetically diverse human induced pluripotent stem cell lines. *Nat Commun* **15**, 8581 (2024).
52. B. B. Adhyaru, T. A. Jacobson, Safety and efficacy of statin therapy. *Nat Rev Cardiol* **15**, 757-769 (2018).
53. M. Whirl-Carrillo *et al.*, An Evidence-Based Framework for Evaluating Pharmacogenomics Knowledge for Personalized Medicine. *Clin Pharmacol Ther* **110**, 563-572 (2021).
54. C. Allan *et al.*, OMERO: flexible, model-driven data management for experimental biology. *Nat Methods* **9**, 245-253 (2012).
55. M. W, Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference* **56-61**, doi:10.25080/Majora-25092bf21922-25000a. (2010).
56. S. N. Chandrasekaran *et al.*, Three million images and morphological profiles of cells treated with matched chemical and genetic perturbations. *Nat Methods*, (2024).
57. Y. Liao, J. Wang, E. J. Jaehnig, Z. Shi, B. Zhang, WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* **47**, W199-W205 (2019).
58. J. Wang, S. Vasaikar, Z. Shi, M. Greer, B. Zhang, WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res* **45**, W130-W137 (2017).
59. D. Szklarczyk *et al.*, The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* **51**, D638-D646 (2023).
60. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michael, E. Duchesnay, SciKit-learn: Machine Learning in Python. *JMLR* **12**, 2825-2830 (2011).
61. P. Virtanen *et al.*, SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261-272 (2020).
62. M. Waskom, seaborn: statistical data visualization. *JOSS* **6**, 3021 (2021).
63. C. S. Hughes *et al.*, Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc* **14**, 68-85 (2019).
64. J. Doellinger, C. Blumenscheit, A. Schneider, P. Lasch, Isolation Window Optimization of Data-Independent Acquisition Using Predicted Libraries for Deep and Accurate Proteome Profiling. *Anal Chem* **92**, 12185-12192 (2020).
65. V. Demichev, C. B. Messner, S. I. Vernardis, K. S. Lilley, M. Ralser, DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods* **17**, 41-44 (2020).

857 66. J. Cox *et al.*, Accurate proteome-wide label-free quantification by delayed
858 normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell*
859 *Proteomics* **13**, 2513-2526 (2014).
860 67. S. Tyanova *et al.*, The Perseus computational platform for comprehensive
861 analysis of (prote)omics data. *Nat Methods* **13**, 731-740 (2016).
862 68. S. Tyanova, J. Cox, Perseus: A Bioinformatics Platform for Integrative Analysis
863 of Proteomics Data in Cancer Research. *Methods Mol Biol* **1711**, 133-148
864 (2018).
865 69. J. R. Wisniewski, M. Y. Hein, J. Cox, M. Mann, A "proteomic ruler" for protein
866 copy number and concentration estimation without spike-in standards. *Mol Cell*
867 *Proteomics* **13**, 3497-3506 (2014).
868 70. Y. Perez-Riverol *et al.*, The PRIDE database resources in 2022: a hub for mass
869 spectrometry-based proteomics evidences. *Nucleic Acids Res* **50**, D543-D552
870 (2022).
871 71. T. p. d. team, pandas-dev/pandas: Pandas (v2.2.3). *Zenodo*, (2024).
872 72. C. R. Harris *et al.*, Array programming with NumPy. *Nature* **585**, 357-362
873 (2020).
874 73. F. Pedregosa, Varoquaux, G.,Gramfort, A., Duchesnay, E, Scikit-learn: Machine
875 Learning in Python. *The Journal of Machine Learning Research* **12**, 2825-2830
876 (2011).
877