Science

RESEARCH ARTICLES

Cite as: B. DeMeo *et al*., *Science* 10.1126/science.adi8577 (2025).

Active learning framework leveraging transcriptomics identifies modulators of disease phenotypes

Benjamin DeMeo[1]†, Charlotte Nesbitt[1]†, Samuel A. Miller[1]†, Daniel B. Burkhardt[1]†, Inna Lipchina[1]†, Doris Fu[1], Peter Holderreith[1], David Kim[1], Sergey Kolchenko[1], Artur Szalata[2,3], Ishan Gupta[1], Christine Kerr[1], Thomas Pfefer[1], Raziel Rojas-Rodriguez[1], Sunil Kuppassani[1], Laurens Kruidenier[1], Parul B. Doshi[1], Mahdi Zamanighomi[1], James J. Collins[4,5,6,7], Alex K. Shalek[4,6,8,9]*, Fabian J. Theis[10,2,3]*, Mauricio Cortes[1]*

[1]Cellarity Inc, Somerville, MA, USA. [2]Computational Health Center, Institute of Computational Biology, Helmholtz-Munich, Neuherberg, Germany. [3]TUM School of Computation, Information and Technology, Technical University of Munich, Garching, Germany. [4]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [6]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [7]Wyss Institute, Harvard University, Boston, MA, USA. [8]Department of Chemistry and Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA. [9]Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA, USA. [10]Helmholtz AI, Helmholtz-Munich, Neuherberg, Germany.

***Corresponding author. Email: mcortes@cellarity.com (M.C.); fabian.theis@helmholtz-munich.de (F.J.T.); shalek@mit.edu (A.K.S.)**

†These authors contributed equally to this work.

Phenotypic drug screening remains constrained by the vastness of chemical space and technical challenges scaling experimental workflows. To overcome these barriers, computational methods have been developed to prioritize compounds, but they rely on either single-task models lacking generalizability or heuristic-based genomic proxies that resist optimization. We designed an active deep-learning framework that leverages omics to enable scalable, optimizable identification of compounds that induce complex phenotypes. Our generalizable algorithm outperformed state-of-the-art models on classical recall, translating to a 13-17x increase in phenotypic hit-rate across two hematological discovery campaigns. Combining this algorithm with a lab-in-the-loop signature refinement step, we achieved an additional two-fold increase in hit-rate and molecular insights. In sum, our framework enables efficient phenotypic hit identification campaigns, with broad potential to accelerate drug discovery.

Despite increased spending in therapeutics R&D over the past 20 years (*1*), overall clinical trial success rates have remained stagnant (*2–4*). Several causes have been suggested (*5*), but a recurring theme is the sub-optimality of the single target drug discovery model (*6–9*). Reinforcing this point, genome-wide association studies have shown that diseases tend to be driven by variants in multiple genes (*10*). Moreover, cellular systems often consist of several molecules acting redundantly, resulting in compensatory mechanisms (*11*, *12*). Finally, systematic analyses of clinically-utilized drugs reveal widespread polypharmacology (*13*, *14*). Although single target-based discovery has been the dominant paradigm, retrospective examinations suggest that more than 65% of all approved medicines were actually discovered via phenotypic observations (*9*).

Phenotypic drug discovery, by contrast, aims to modulate a disease-linked behavior in a faithful model. By focusing on cellular attributes, phenotypic screening can directly consider the net effect of on- and off-target molecular responses (polypharmacology), optimizing for a desired outcome. However, this paradigm is constrained by an inherent tradeoff between readout complexity and scalability (*8*): higher-resolution assays measuring complex information-rich phenotypes in disease-relevant systems, such as a molecular signature of the disease process (*15*), can enhance clinical translation but have lower throughput and greater cost; conversely, simpler phenotypic measurement and model combinations support cost-effective, high-throughput screens (HTS), but have reduced clinical translatability (*5*, *16*). To resolve this conundrum, small molecule prioritization strategies are needed that enable efficient screens for clinically-effective agents.

In target-based discovery, virtual screening has improved productivity (*17*). Seeking to translate this approach to phenotypic discovery, several groups have proposed frameworks that leverage AI to rank compounds using task-specific models, yielding novel antibiotics (*18*, *19*) and senolytics (*20*). While these models improved hit-rate compared to traditional screening, they require retraining with large datasets for each new target phenotype.

To overcome task-specificity, researchers have proposed using omics signatures generated in response to chemical perturbations as proxies for multiple phenotypic outcomes. Here, compounds are prioritized based on the probability

that they will induce an omics signature (for example, a gene expression profile) associated with a desired phenotype. An initial implementation showed promise for phenotypic screening in mice (*21*). However, a rigorous evaluation directly testing compounds top-ranked by such a model in a drug discovery setting, with random compound selection as a baseline, is needed to demonstrate impact and drive broader industry adoption.

Key framework features also require refinement. Illustratively, current gene expression prioritization approaches use suboptimal heuristics—such as scoring compounds via statistical tests originally designed for other bioinformatics applications like gene set enrichment (*22–24*)—that cannot be improved via experimental feedback. Relatedly, the success of omics-based prediction depends on the input signature being an accurate representation of the target phenotype. Current approaches infer gene expression signatures from observational associations, which may not translate to the in vitro assay used to model disease, hindering success (*25*); further, the number of available experimental omics signatures is limited, though models that predict virtual signatures directly from chemical structures may help assuage this concern once vetted (*21*, *26*, *27*).

To improve the productivity of omics-based phenotypic drug discovery, we developed a closed-loop active reinforcement learning (ARL) framework. First, we trained a deep-learning architecture, DrugReflector, to predict small molecule modulators of complex cellular phenotypes using compound-induced transcriptomic signatures. In comprehensive benchmarking of compound ranking algorithms enabled by a 1.2M cell dataset spanning 88 chemical perturbations of 10 diverse cancer and primary cell lines, DrugReflector achieved state-of-the-art performance. We next systematically evaluated its omics-based predictions in two campaigns, aiming to induce the differentiation of megakaryocytes and erythrocyte progenitors, relevant for treating anemia and thrombocytopenias (*28–30*). Compared to random compound selection, DrugReflector drove a 13-17X increase in hit-rate. Toward broader utility, it also effectively prioritized compounds associated with disease etiology in two external cancer datasets.

Finally, using our experimental omics data for closed-loop feedback, we realized iterative improvements in screening efficacy. Integrating paired phenotypic and transcriptomic measurements, we refined our input signature via active learning, demonstrating a further two-fold increase in hit-rate and gained insights into factors driving model performance. Our analyses also identified a previously unappreciated pathway to induce the megakaryocyte lineage, further showcasing the translational potential of our approach. Collectively, our lab-in-the loop framework enables greater productivity in drug discovery, empowering the use of more representative and translatable cellular models.

Results

A closed-loop predictive framework to enable phenotypic drug discovery

The core of our phenotypic drug discovery framework is a closed-loop ARL process that nominates compounds to modulate a phenotype of interest (Fig. 1, fig. S1, and methods). Here, we chose to use single-cell transcriptomics as a generalizable omics proxy of cell state, because this data modality is widely available for various tissues in perturbational contexts, and in health and disease (*22*, *31*). First, we analyzed clinical datasets to identify a transcriptional signature for a cellular transition of interest, which was then calibrated to a phenotypic assay to ensure signature induction was positively associated with that phenotype (Step 1). Second, we developed a deep-learning model, DrugReflector, described below, to predict compounds with high probability of inducing that target signature (Step 2). Third, we experimentally screened compounds for phenotypic activity, validating results in multiple donors. These hits are the primary output of our framework and can be used for downstream development (Step 3). Finally, we introduced active signature learning as a closed-loop feedback mechanism using joint transcriptional and phenotypic measurements of hit and non-hit compounds (Step 4). This refined our input signature and improved hit-rate. By comparing our framework to current phenotypic discovery paradigms, we highlighted its enabling developments (fig. S1).

A deep-learning framework for compound prioritization

DrugReflector is an ensemble of three multi-layer perceptron (MLP) classifiers trained to match omics signatures to instigating compounds. As training data, we used a subset of the Connectivity Map (CMap) (*22*), chosen to ensure compound tractability, signature reliability, and broad biological coverage (Fig. 2A and methods). These data were split into three sets evenly dividing perturbation replicates; each model in the ensemble was trained on two of the three. During training, each individual perturbation signature was treated as an independent input—no averaging across replicates was performed. The model was trained for 50 epochs, with recall on unseen compounds evaluated at each epoch (fig. S2A). The checkpoint corresponding to the epoch with the highest recall was selected for downstream prediction.

To evaluate how DrugReflector's performance depended on training set composition, we conducted two down-sampling analyses. In one, we reduced replicate depth by randomly retaining only a subset of perturbation replicates across the dataset. In the other, we reduced biological breadth by limiting the number of cell lines used per compound. As described in supplementary note 1 and shown in fig. S2B, model performance improved with both increasing

replicate depth and cell line diversity, particularly at lower levels of coverage. These results highlight the importance of both technical redundancy and biological diversity in building training data for robust, generalizable predictive models.

Benchmarking DrugReflector against existing methods

To evaluate the performance of our model, we benchmarked DrugReflector against four approaches for matching gene signatures to compounds, using top 1% compound recall as a measure of performance. For each compound, the recall score is 1 if the model ranks the compound with the correct label relative to the top 1% of all predicted compounds based on its transcriptional signature; otherwise, the score is 0. This score is averaged across observations for that compound in the dataset, and then averaged across compounds. The comparison models included two classical baseline methods: a k-nearest neighbor (kNN) classifier, and a logistic regression model. Additionally, we benchmarked against two published in silico statistical methods for connecting omics signatures to disease phenotypes (*21*, *24*): the SigCom LINCS implementation of gene set enrichment analysis (*32*) and Dr. Insight (*23*).

Our benchmarking covered three independent perturbation data sets—Touchstone (*22*), sciPlex3 (*33*), and our own ("Cellarity")—generated for this study. First, we evaluated and compared our model on the CMap Touchstone dataset, comprising 1,000 compounds tested in nine cell lines. Our results show that DrugReflector outperformed all four algorithms, surpassing Dr. Insight by 1,545% (i.e., a 15+-fold improvement) and SigCom by 15% in average recall over cell lines (Fig. 2B). Second, we compared the five algorithms on the sciPlex3 dataset of 188 compounds measured in three CMap cancer cell lines (*22*). DrugReflector again outperformed all algorithms (Dr. Insight and SigCom by 66% and 108% in average recall, respectively)

To evaluate the generalizability of our model to cellular contexts not represented in our training data, such as non-cancer lines, we generated a new scRNA-seq dataset that included six immortalized lines (A549, A375, H1AE, HEK293T, HEP2G, and PC3) and four primary lines: CD8+ T-cells (CD8+), CD34+ hematopoietic progenitors (CD34+), primary adipocytes (PAD), and bronchial epithelial cells (HBEC). Eighty-eight compounds from CMap were tested across the 10 cell types in duplicate with library-matched controls, resulting in 1.26 M cells from 1,737 scRNA-seq samples (fig. S2C). On cancer cell lines present in the training data, we found that DrugReflector outperformed all algorithms, achieving a 323% increase in average recall compared with Dr. Insight and a 73% increase compared with SigCom (Fig. 2B). On primary cancer cell lines outside of the training data, DrugReflector again achieved the highest recall (by 194% and 30% compared to Dr. Insight and SigCom, respectively), although recall was lower overall here compared to cancer lines (Fig. 2, B and C, and fig. S2D).

A limitation of the landmark gene assay is its restriction to 978 genes, or ~5% of protein-coding genes. To remedy this, the CMap authors applied computational inference to estimate the expression of an additional 11,350 genes (*22*). Incorporating these inferred genes into the training data improved DrugReflector's recall of held-out CMap signatures, but reduced recall on external datasets, including sciPlex and Cellarity signatures (fig. S2E and supplementary note 2). This suggests that gene inference can introduce dataset-specific effects that reduce generalizability, outweighing the gains realized by the additional features. We therefore chose not to incorporate inferred genes into DrugReflector.

Finally, we similarly whether virtual signatures inferred from compound structures could expand the space of predictable compounds. Using TranSiGen (*26*), we generated virtual signatures for both CMap and our in-house perturbations, and evaluated compound recall using either DrugReflector or k-nearest neighbor regression. Compared to measured signatures, virtual signatures yielded lower recall—particularly outside of CMap, the source of TranSiGen's training data (fig. S2F and supplementary note 3).

Developing a complex phenotypic assay with a high clinical translatability

Hematopoiesis is an essential developmental process, and aberrant hematopoiesis can lead to numerous proliferative disorders (*34–37*) and cytopenias (*38–40*). Given the prevalence of anemias and platelet deficiencies (*28–30*), we aimed to modulate the lineage commitment of megakaryocytes and erythroid progenitors by applying our framework.

We selected primary human CD34+ hematopoietic stem and progenitor cells (HSPCs) for our screens due to their high clinical relevance to a range of hematologic disorders (*38*, *39*) and the relative abundance of public human scRNA-seq data available to associate gene signatures to hematopoietic processes. Despite their translational value (*41*), CD34+ primary cells are rarely used for phenotypic screening because of the cost and logistical difficulty of sourcing, culturing, and expanding them ex vivo (*42*, *43*). These constraints provided an opportunity to demonstrate the potential of our framework to transform drug discovery by enabling more efficient and scalable phenotypic exploration in clinically meaningful cell types.

To characterize the cell states involved with the megakaryocyte and erythroid lineages, we analyzed a CITE-seq (*44*) dataset we previously generated (*45*), consisting of joint RNA and protein expression data spanning the transcriptome and 134 surface protein makers for primary HSPCs from four healthy donors sampled at five time points over a 10-day time course (Fig. 3A). Combining these data with literature knowledge, we identified progenitor and early lineage-

committed cell states, including cells at a range of differentiation stages along the megakaryocyte (Mk), erythroid (Ery), eosinophil/basophil/mast (EBM), monocyte (Mono), and neutrophil (Neu) lineage trajectories (Fig. 3A and fig. S3A), observing consistency in cellular differentiation across all four donors (fig. S3B).

We confirmed that RNA-derived cell types expressed the expected surface markers for both the Mk and Ery lineages (Fig. 3, B to D). This enabled us to define a flow cytometry gating strategy to detect both (fig. S4, A to C). For each, we also identified positive controls and established our flow assay's dynamic range to facilitate the identification of phenotypically active compounds (fig. S4D and methods).

To establish a hit threshold for each of the two cell-type assays, we first filtered out compounds that lead to low cell viability or for which we measured an insufficient number of cells. We then calculated a significance cutoff relative to DMSO treatment, considering the variation of DMSO samples within and across plates.

Generating transcriptional signatures of megakaryopoiesis and erythropoiesis

To nominate compounds for screening, we identified cell state transitions associated with early differentiation from the bipotential megakaryocyte erythroid progenitor (MEP) into the lineage-committed megakaryocyte progenitor cells (MPC) and erythroid progenitor cells (Ery). MEPs represent an optimal intervention point, as transcriptional and metabolic changes in these cells are associated with commitment to differentiation into either lineage (*46*, *47*). The transcriptional changes underlying these transitions were used as input to DrugReflector.

To quantify these changes, we developed the v-score: an estimate of the difference in the log-normalized count means between the two populations, normalized by the square root of the sum of the variances of the log-transformed counts in each condition. Formally, we defined the v-score between two cell states x and y as follows:

$$\text{v-score}\left(x \to y\right) = \frac{E\left(\log\left(1+y\right)\right) - E\left(\log\left(1+x\right)\right)}{\sqrt{Var\left(\log\left(1+x\right)\right) + Var\left(\log\left(1+y\right)\right)}}$$

Like the CMap level 4 Z-score signatures used to train the model, v-scores quantify differences in expression in units of standard deviation. Unlike the *t*-statistic, the v-score is independent in expectation of the number of cells in each group, enabling well-calibrated comparisons between populations with differing cell count.

DrugReflector identifies inducers of the Mk and Ery lineages

We calculated v-scores between the MEP and MPC population and used them as inputs into DrugReflector to obtain a prioritized list of compounds for screening (table S1). We selected top-ranked compounds from the model's output to assess their ability to induce the phenotype of interest.

To experimentally determine which compounds induced our target phenotype, we treated CD34+ cells with each model-nominated compound under HSPC maintenance conditions and evaluated the induction of CD41a+ CD71- CD42b+ Mk population by flow cytometry. We tested 107 compounds that were both ranked within the top 1,000 by DrugReflector and available in our chemical inventory. These compounds represented an approximately uniform sampling across the rank range, minimizing selection bias induced by inventory constraints (fig. S5). Each compound was tested at three doses (100 nM, 1 μM, 10 μM), and the dose at which compounds were maximally inductive but not cytotoxic was recorded.

To evaluate the impact of ML-based compound selection for phenotypic discovery, we compared to brute-force compound screening, in which all compounds in a library are tested for activity. This is an absolute measure not subject to variability in the implementation of a particular alternative algorithm. Furthermore, brute-force screening of compound libraries is the dominant screening paradigm in the pharmaceutical industry (*8*). To estimate the hit-rate of brute-force screening of our compound inventory, we tested a random selection of 87 compounds.

Among our 107 highly-ranked DrugReflector-nominated compounds, we identified 21 above our six standard deviations hit threshold, resulting in a 19.6% hit-rate (95% CI: 13.2-28.1%, Wilson score interval; Fig. 3E). Two compounds were highly active, inducing more than a four-fold increase in Mk progenitors. By contrast, we identified only one compound from our random selection that passed our hit threshold, resulting in a 1.1% hit-rate (95% CI: 0.2-6.2%, Wilson score interval). These results demonstrate that our ML-based prioritization approach was more effective at identifying hit compounds than random selection by approximately 17X (p=1.29e-10, one-sided binomial test). To confirm that these hits validated in multiple donors, we re-tested 17 hit compounds in two additional donors at the dose at which we observed maximal induction of the Mk lineage. While two compounds did not pass our viability or cell count criteria, 13 out of the remaining 15 hit compounds validated in both donors, demonstrating the robustness in our assay and biological translation of our chemical perturbations across different donors (Fig. 3F).

Next, we sought to demonstrate the generalizability of our framework by biasing the MEP population toward Ery progenitor cells. As previously described, v-scores were calculated between the MEP and Ery progenitor population, and used as input to DrugReflector to obtain a prioritized list of high-ranking compounds to test (table S2).

In our DrugReflector-nominated compound screen, after removing samples with too few viable cells, we observed 13 out of 81 compounds passing our 6-standard deviation above the DMSO hit cutoff, representing a 16% hit-rate (95% CI: 9.6-25.5%; Fig. 3G). In our randomly selected compound set, we observed only 1 out of 85 compounds inducing Ery progenitors above our cutoff, representing a 1.2% hit-rate (95% CI: 0.2-6.4%; Fig. 3G). Again, our transcriptomics-based compound prioritization significantly increased our success rate in inducing the desired phenotype by approximately 13X (p<1.12e-8, one-sided binomial test). Once again, we validated these results in multiple donors. Out of 10 hit compounds passing our quality control filter in our validation experiment, 8 increased Ery progenitors in at least one donor, and 5 did so in both (Fig. 3H). These results provide further support for the capacity of our machine learning model to increase phenotypic hit-rate across multiple experimental settings.

DrugReflector recovers standards of care and modulators of disease-relevant pathways

To assess the broader utility of DrugReflector across diverse disease contexts, we evaluated two datasets with distinct etiologies and therapeutic standards: B cell acute lymphoblastic leukemia (B-ALL) (*48*) and breast cancer, using a multi-tumor single-cell atlas containing both estrogen receptor-positive (ER+) and triple-negative breast cancer (TNBC) samples (*49*). For each condition, we constructed v-score signatures representing the transition from malignant to healthy-like states and ranked compounds by their predicted ability to revert the disease phenotype (table S3).

B-ALL is characterized by an accumulation of B cell progenitors, and is marked by dysregulation of tyrosine kinase signaling, with a fraction of B-ALL cases caused by a translocation resulting in the formation of the BCR-ABL fusion protein (the Philadelphia chromosome) (*50*). DrugReflector prioritized ABL inhibitors, including Ponatinib, a clinical standard-of-care, which ranked in the top 1% (78th of 9,597 compounds; fig. S6, A and B) and has shown superior efficacy in the clinic compared to other ABL inhibitors (*51*). Similarly, we observed a prioritization of interventions targeting MAPK, specifically p38 MAPK (fig. S6, A and B), consistent with the importance of MAPK signaling in B-ALL (*52*, *53*). In contrast, compounds targeting unrelated pathways, such as estrogen receptor signaling or general cell cycle control, were not prioritized.

Analysis using transitions from ER+ and TNBC resulted in distinct classes of prioritized compounds. While both transitions prioritized ER inhibitors, the ER+ transition consistently ranked them higher (p=0.031, Wilcoxon signed-rank test; fig. S6A), in accordance with the genomic shift observed in ER+ tumors (*54*). Similarly, while both transitions prioritized cell cycle inhibitors that target mitotic cells through microtubule inhibition, the triple-negative transition showed moderately stronger prioritization (p=0.12, Wilcoxon signed-rank test), consistent with the overt dysregulation of cell cycle (*55*) and poor prognosis of TNBC (*56*). For additional context, we performed cell-cycle classification on each population used to generate our transitions, identifying increased cell cycling in TNBC compared to other transitions evaluated (fig. S6C). Finally, we observed that JAK/STAT inhibitors, such as ruxolitinib, were prioritized by both B-ALL and breast cancer transitions (fig. S6, A and B). This finding is consistent with the central role of JAK/STAT signaling in B cell development and breast tissue specification and its dysregulation in cancer (*57*), highlighting the model's ability to similarly prioritize shared biology observed in both diseases. DrugReflector's disease- and subtype-specific predictions suggest its potential for transferability to phenotypic discovery efforts beyond the training domain, and its generalizability for compound prioritization in diverse therapeutic settings.

Paired transcriptional and phenotypic measurements enable closed-loop active signature learning

Inspired by the field of ARL, which has been used to predict compound-target binding, optimize molecular properties, and characterize structure-activity relationships (*58*, *59*), we postulated that paired phenotypic and transcriptional measurements could be leveraged to refine our phenotype-associated input omics signature. The core premise of ARL is the optimization of a policy by selectively acquiring data points to maximize a reward signal. The policy guides an agent's actions within an environment, with the resulting rewards and changes in state used to update the policy. In our context, the policy consists of DrugReflector and its input signature, the actions included the paired phenotypic and transcriptional measurements for a selection of compounds, and the reward is the hit-rate. The general framework is to identify a learned signature from the paired transcriptional and phenotypic data and use this to update our original signature. The difference between the learned signature and the original signature is a gradient, and the size of the step we take in our policy update is a tunable step size parameter (Fig. 4A).

To implement this paradigm, we performed a scRNA-seq time course on 12 hit and eight non-hit compounds at four time points (Day 1, 2, 5 and 7) with a paired phenotypic readout at Day 7. We observed cells from all major expected cell types (fig. S7A) and saw a strong correlation between the abundance of Mk cells as determined by our scRNA-seq and the phenotypic measurements (fig. S7B). To test our hypothesis that gene expression changes differentiate hit from non-hit compounds, we performed differential expression analysis between each compound and the DMSO negative control

at Day 1 in the HSPC population. We observed a clear increase in expression of Mk marker genes and transcription factors previously associated with Mk maturation among hit compounds. Several of these genes were significantly associated with the Mk induction phenotype (Fig. 4B, adj. p<0.01, Pearson correlation test).

We reasoned that one factor leading to the prediction of inactive compounds could be compounds having a cell-type specific effect in CD34+ cells that differs from the impact measured in the CMap dataset. 43% of compounds in CMap were previously reported to exhibit cell-type specific effects in the cancer cell lines (*22*), and CD34+ HSPCs were absent from the training data. To test this explicitly, for each compound we calculated the distance between the 24-hour signatures in our follow-up experiment and the 10 most similar signatures for the same compound in LINCS. We found that, on average, CD34+ signatures of non-hit compounds were 11% further from their closest neighbors in CMap compared with hit compounds (fig. S7C, p=0.037, paired *t* test). Although this distance to CMap could only be calculated using L1000 landmark genes, most cell-type-specific perturbation effects fell outside the landmark gene set based on our benchmarking dataset (fig. S7D).

To refine the input signature, we used pseudobulked expression from each perturbation and the Day 7 phenotypic outcome to score each gene by its association with phenotype. We then applied this learned signature to update our original signature as described above. To ensure a comprehensive evaluation of all possible ways to apply our active signature learning framework, we tested all combinations of cell types in the Mk lineage and readout days at each step size between 0 and 1 in increments of 0.05 (fig. S8A). We determined that the maximal improvement in hit compound recall, calculated as the average precision (AP), could be obtained using DE analysis on the Day 1 HSPC population with a learning rate of 0.7. Consistent with this finding, Day 1 HSPCs showed the largest number of differentially expressed genes overall (fig. S8B; p=0.025, one-sided Mann-Whitney U test), the greatest number of genes whose change in expression was correlated with our MK phenotype (fig. S8C; $p<10^{-30}$ against all other groups, Mann-Whitney U test), and the highest correlation with matched LINCS signatures (fig. S8D; p=0.021, one-sided Mann-Whitney U test). Furthermore, we investigated the gene signature changes associated for various gene classes in our refined signature, demonstrating a significant enrichment in transcription factors associated with megakaryopoiesis compared to MK marker genes and all other genes (Fig. 4C and fig. S8, E and F; p<1e-3, Mann-Whitney U test).

To validate our ARL framework, we used the refined signature as input to DrugReflector to rank all compounds (table S4) and tested 96 newly-prioritized compounds that were not previously screened. Results from our phenotypic campaign identified 22 new hits out of the 85 compounds that passed quality control (Fig. 4D). To compare the two signatures directly, we compared the ranks of all hit compounds from the original and new screen, demonstrating that the refined signature was better overall at prioritizing hit compounds (Fig. 4E; p=1e-4, Wilcoxon signed rank test). To quantify the resulting improvement in screening efficiency, we plotted phenotypic hit-rate as a function of rank threshold, showing a roughly 2-fold improvement in hit-rate for the top 100 compounds, converging after ~500 compounds (Fig. 4F). The median rank of true hits improved from 463 to 138, a 3.4-fold gain. To confirm this improvement was due to meaningful biological signal in the learned signature, we repeated the procedure using randomized versions of the learned component prior to interpolation. These "scrambled" refinements consistently resulted in worse hit prioritization (fig. S8G; p<1e-4, bootstrap over 10,000 random seeds), confirming that the learned signature contributes functional information relevant to the target phenotype.

Characterizing mechanisms of chemically-induced megakaryocyte lineage commitment

To understand variation in phenotype induction across compounds, we examined cell-type specific differential expression profiles relative to DMSO for each condition and time point (fig. S9A). Given the importance of the 24-hour timepoint in our signature refinement studies, we focused on HSPCs at 24 hours post perturbation. There, we observed five major clusters of perturbations separated by the first two principal components (PCs): an inactive cluster that does not induce Mk, a single compound that inhibits Mk differentiation, one cluster of highly active Mk differentiation compounds, and two clusters of moderately active compounds separated along PC2 (Fig. 5A).

To determine what drives this variation, we performed GSEA to identify biological processes associated with these PCs (fig. S9, B and C). Genes with high PC1 loadings were enriched for antigen processing and JAK/STAT signaling, consistent with known stages of megakaryopoiesis (*60–62*). This cluster of active compounds corresponded to tyrosine kinase inhibitors, indicating that inhibition of one or multiple kinases can partially drive the induction of megakaryocyte progenitors, which has been previously observed for kinase inhibitors (*63*, *64*). In contrast, genes with high PC2 loadings were enriched for lipid and cholesterol biosynthesis, which has recently been recognized as an important process in Mk maturation and platelet formation (*65*, *66*), but has not been previously implicated in early Mk lineage commitment.

To further examine the drivers of chemically-induced megakaryopoiesis, we identified a pseudotime trajectory in the Mk lineage for all cells in the transcriptional validation experiment (Fig. 5B and fig. S10A). We then calculated the

expression of genes in rolling pseudotime windows to identify expression patterns that differ across groups of compounds (fig. S10B). Induction of Mk differentiation and development gene sets were consistent across pseudotime for all classes of compounds (Fig. 5, C and D). This suggested a single program of Mk differentiation regardless of chemical perturbation. However, positive regulators of Mk differentiation and cell cycling genes, known to play a role in Mk differentiation (*47*), were differentially induced by strongly active compounds, especially in MEP and MPC cells. This suggested that up-regulation of these genes is important for the activity of hit compounds. Finally, we saw that compounds that modulate lipid metabolism do this at all time points and in cell types in a pattern consistent with other compounds' activity, but notably stronger (Fig. 5E).

To investigate the mechanisms underlying these transcriptional groupings, we compiled IC50 binding assays from ChEMBL (*67*) and compared all transcriptionally-profiled compounds with less than 20 inhibitory targets under 1 μM. Most high-affinity targets identified for strong and moderate inducers function as kinases, while all moderate inducer (lipid) compounds bound HMGCR with high affinity (Fig. 5F). Further, outside of the lipid class of hits, very few targets were bound at high affinity by more than one compound. This suggests that while the strong and moderate classes may encompass several functional targets, the lipid class likely functions through a single target, HMGCR. CRISPR knockout of HMGCR (fig. S11A) confirmed directional induction of megakaryocyte progenitors on Day 7 (fig. S11B) but was inferior to compounds, likely owing to its essential role in numerous biological processes, as evidenced by a reduction in the fraction of edited cells over time (fig. S11A).

Given the higher target promiscuity of the remaining hit compounds, we sought to deconvolute the likely target of one strong megakaryocyte inducer. BRD-K28392481 targets multiple receptor tyrosine kinases (RTKs), including KDR (VEGFR2), as well as FGFR1-4. To dissect the contribution of this group of RTKs to the phenotypic activity of this compound, we tested four additional tool compounds: two selective for KDR over FGFR proteins and three selective for FGFR proteins over KDR (fig. S12A). The two KDR-selective compounds induced Mk progenitors in a dose response while the FGFR inhibitors were inactive (Fig. 5G). To explore whether KDR was sufficient to induce the increase of Mk progenitors observed with our chemical perturbations, we performed loss of function studies using CRISPR. Targeting KDR using single and dual guides resulted in editing efficiency greater than 60% for single guides and >70% for dual guides, with no apparent loss of edited cells over the course of the assay (fig. S12B). Phenotypic assessment of lineage differentiation detected a moderate increase in Mk progenitors, indicating that inhibition of KDR was not sufficient to phenocopy the chemical effect fully. Consistent with these findings, loss of function of FLT3 resulted in induction of Mk cells, albeit to a lesser degree than the small molecules (fig. S12C). Together these studies shed light on the diverse mechanisms influencing Mk specification and identify HMGCR as a putative target for induction of the Mk lineage. Our studies on KDR, meanwhile, suggest that inhibition of multiple tyrosine kinases is likely driving the robust induction of megakaryopoiesis with our chemical perturbations.

Discussion

Here, we present a closed-loop ARL framework for prediction of disease phenotypes to improve phenotypic discovery. Our generalizable approach links disease biology, chemistry, and phenotypic activity using omics-level data. As we show through extensive benchmarking, our workflow enabled an order of magnitude improvement in hit-rate compared to brute-force screening. In addition, we used paired omics and phenotypic data to refine our original signature through active learning. A single cycle of active signature learning yielded an additional 2-fold improvement in hit-rate and provided a deeper understanding of the biological process under study. As such, our lab-in-the-loop architecture exemplifies the promise of AI-guided drug discovery (*68*).

A feature of our framework is its modular nature and the ability to optimize each component independently. As shown through our studies, the input signatures are critical for the prioritization of compounds and the output post-signature refinement. In the hematopoietic dataset, we leveraged temporal shifts to identify a cell state capturing the transition from an MEP to the earliest cell state that defines an MPC. In other datasets, different features, such as genetic drivers, could be leveraged to select a cell state transition and corresponding gene signature.

During signature refinement, we used a temporal perturbation dataset to elucidate what parameters in the input signature led to increased phenotypic hit-rate. Future work is needed to unravel the relative significance of this finding across different contexts. Relatedly, as more omics perturbations signatures become available, there will be a need to consider how to weight the incremental value of each in prediction and, similarly, to develop robust heuristics for blending existing and learned signatures. More broadly, identifying disease signatures is a rapidly developing field, and the most recent methods focus on differential expression between cell states (*24*), as we do here. In a complex disease, a meaningful signature could be cell type/state specific, temporally restricted, and/or heterogeneous (somatic mutations). Therefore, future efforts should consider how to select context-appropriate strategies for signature identification, such as prioritizing genes based on fate mapping (*69*), paired single-cell genotyping (*70*), transfer learning of signatures

between contexts informed by epigenetic or spatial atlas priors, or by causal inference of regulatory relationships between genes (*71*, *72*).

To maximize the current paradigm, there is also an opportunity to improve the reference perturbation dataset. CMap's L1000 assay is noisy due to its bead-based methodology (*73*), and constrained to 978 genes. Moreover, almost all of CMap's data come from cancer cell lines, limiting generalizability to some primary cell types. To provide a better basis for predictions, datasets tailored to desired therapeutic areas need to be built, considering the value of different data modalities, alone and in combinations.

To examine the vast space of drug-like compounds (*74*, *75*), future models must be able to make predictions directly from chemical structure. While recent approaches have been proposed to predict transcriptional signatures from chemical structures (*21*, *26*), analyses from our internal benchmarks presented here, and a recent Kaggle competition to predict the cell-type specific impact of perturbations (*76*), suggest that current methods, while promising, require further refinement. Existing approaches also work best when applied to cell types present in—or similar to—the training set (*77*), so new algorithms are needed to support generalization.

While our framework prioritizes compounds for zero-shot phenotypic discovery, efficiency could be improved by selecting compounds that maximize learning. Reinforcement learning offers a formal approach to this via acquisition functions that balance exploration and exploitation to maximize a reward, such as hit-rate (*68*). Early screens might favor exploring diverse omic profiles to link omic readout to phenotype, while follow-up screens could focus on exploiting these insights to predict hits more accurately. Similarly, emerging methods designed for causal gene inference (*78*–*80*) could enhance learning efficiency at each step by improving the identification of molecular drivers.

Finally, our approach prosecutes the cell as the target, considering disease states to be informed by one or more dysregulated pathways rather than by a single gene. Our framework capitalizes on this paradigm, putting disease biology into a transcriptional context and addressing the polypharmacology often exhibited by small molecules through use of omics to explicitly consider shifts across multiple pathways at once. This enabled us to identify two distinct groups of molecules that increased Mk lineage commitment. With a combination of chemogenetic and CRISPR studies, we examined the mechanisms informing one of the "strong inducers," demonstrating that KDR alone was insufficient, and the observed phenotypic activity was likely driven by inhibition of multiple tyrosine kinases, including FLT3 and LYN. Furthermore, our omics approach to targeting cell states is well-suited to identify biological insights. We explored a second class of Mk-inducers, which all shared 3-hydroxy-3-methylglutaryl coenzyme A (HMGCR) as a common annotated target. Through follow-up experimentation, we confirmed that this class of molecules was inducing Mk through modulation of the cholesterol synthesis pathway. Interestingly, the Mk-bias of *CALR*-mutant hematopoietic stem cells has been associated with cholesterol biosynthesis pathways (*81*). Similarly, dysregulation of cholesterol homeostasis by deletion of *ABCG4* results in increased megakaryopoiesis in mice (*82*), suggesting a biological basis for our findings.

In sum, the framework presented here has broad utility for drug discovery across disease settings. It connects underlying disease biology to chemical perturbations through omics as a common language, yielding marked improvement in phenotypic hit-rate by focusing on the cell as the target. Our study provides direct evidence that modulating cholesterol biosynthesis in human hematopoietic progenitors is sufficient to induce megakaryocyte fate. In addition, we show that inhibition of multiple tyrosine kinases is required to promote Mk lineage bias. This framework has allowed us to identify druggable nodes in sickle cell disease (*83*) and myelofibrosis. Owing to a surge in publicly available single-cell datasets across diseases (*31*) and single-cell perturbation signatures (*84*, *85*), it is now possible to leverage existing atlases to derive an initial target signature for dozens of indications and apply this paradigm.

Materials and methods are available in the supplementary materials.

REFERENCES AND NOTES

1. D. Austin, T. Hayford, "Research and development in the pharmaceutical industry," Publication No. 57025 (US Congressional Budget Office, 2021); https://www.cbo.gov/publication/57126.
2. C. H. Wong, K. W. Siah, A. W. Lo, Estimation of clinical trial success rates and related parameters. *Biostatistics* **20**, 273–286 (2019). doi:10.1093/biostatistics/kxx069 Medline
3. M. Hay, D. W. Thomas, J. L. Craighead, C. Economides, J. Rosenthal, Clinical development success rates for investigational drugs. *Nat. Biotechnol.* **32**, 40–51 (2014). doi:10.1038/nbt.2786 Medline
4. D. Thomas, D. Chancellor, A. Micklus, S. LaFever, M. Hay, S. Chaudhuri, R. Bowden, A. W. Lo, "Clinical development success rates and contributing factors 2011–2020" (BIO, Informa Pharma Intelligence, and QLS Advisors, 2021); https://www.bio.org/clinical-development-success-rates-and-contributing-factors-2011-2020.
5. J. W. Scannell, A. Blanckley, H. Boldon, B. Warrington, Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* **11**, 191–200 (2012). doi:10.1038/nrd3681 Medline
6. W. Zheng, N. Thorne, J. C. McKew, Phenotypic screens as a renewed approach for drug discovery. *Drug Discov. Today* **18**, 1067–1073 (2013). doi:10.1016/j.drudis.2013.07.001 Medline
7. D. B. Kell, Finding novel pharmaceuticals in the systems biology era using multiple effective drug targets, phenotypic screening and knowledge of transporters: Where drug discovery went wrong and how to fix it. *FEBS J.* **280**, 5957–5980 (2013). doi:10.1111/febs.12268 Medline
8. J. G. Moffat, F. Vincent, J. A. Lee, J. Eder, M. Prunotto, Opportunities and challenges in phenotypic drug discovery: An industry perspective. *Nat. Rev. Drug Discov.* **16**, 531–543 (2017). doi:10.1038/nrd.2017.111 Medline
9. A. Sadri, Is target-based drug discovery efficient? Discovery and "off-target" mechanisms of all drugs. *J. Med. Chem.* **66**, 12651–12677 (2023).

doi:10.1021/acs.jmedchem.2c01737 Medline
10. T. A. Manolio, F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. C. Mackay, S. A. McCarroll, P. M. Visscher, Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009). doi:10.1038/nature08494 Medline
11. J. L. Hartman 4th, B. Garvik, L. Hartwell, Principles for the buffering of genetic variation. *Science* **291**, 1001–1004 (2001). doi:10.1126/science.1056072 Medline
12. A. Rossi, Z. Kontarakis, C. Gerri, H. Nolte, S. Hölper, M. Krüger, D. Y. R. Stainier, Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* **524**, 230–233 (2015). doi:10.1038/nature14580 Medline
13. M. J. Keiser, V. Setola, J. J. Irwin, C. Laggner, A. I. Abbas, S. J. Hufeisen, N. H. Jensen, M. B. Kuijer, R. C. Matos, T. B. Tran, R. Whaley, R. A. Glennon, J. Hert, K. L. H. Thomas, D. D. Edwards, B. K. Shoichet, B. L. Roth, Predicting new molecular targets for known drugs. *Nature* **462**, 175–181 (2009). doi:10.1038/nature08506 Medline
14. S. Klaeger, S. Heinzlmeir, M. Wilhelm, H. Polzer, B. Vick, P.-A. Koenig, M. Reinecke, B. Ruprecht, S. Petzoldt, C. Meng, J. Zecha, K. Reiter, H. Qiao, D. Helm, H. Koch, M. Schoof, G. Canevari, E. Casale, S. R. Depaolini, A. Feuchtinger, Z. Wu, T. Schmidt, L. Rueckert, W. Becker, J. Huenges, A.-K. Garz, B.-O. Gohlke, D. P. Zolg, G. Kayser, T. Vooder, R. Preissner, H. Hahne, N. Tõnisson, K. Kramer, K. Götze, F. Bassermann, J. Schlegl, H.-C. Ehrlich, S. Aiche, A. Walch, P. A. Greif, S. Schneider, E. R. Felder, J. Ruland, G. Médard, I. Jeremias, K. Spiekermann, B. Kuster, The target landscape of clinical kinase drugs. *Science* **358**, eaan4368 (2017). doi:10.1126/science.aan4368 Medline
15. C. V. Theodoris, P. Zhou, L. Liu, Y. Zhang, T. Nishino, Y. Huang, A. Kostina, S. S. Ranade, C. A. Gifford, V. Uspenskiy, A. Malashicheva, S. Ding, D. Srivastava, Network-based screen in iPSC-derived cells reveals therapeutic candidate for heart valve disease. *Science* **371**, eabd0724 (2021). doi:10.1126/science.abd0724 Medline
16. D. Lowe, "The brute rorce bias," *Science*, 12 March 2012; https://www.science.org/content/blog-post/brute-force-bias.
17. W. P. Walters, M. T. Stahl, M. A. Murcko, Virtual screening—An overview. *Drug Discov. Today* **3**, 160–178 (1998). doi:10.1016/S1359-6446(97)01163-X
18. J. M. Stokes, K. Yang, K. Swanson, W. Jin, A. Cubillos-Ruiz, N. M. Donghia, C. R. MacNair, S. French, L. A. Carfrae, Z. Bloom-Ackermann, V. M. Tran, A. Chiappino-Pepe, A. H. Badran, I. W. Andrews, E. J. Chory, G. M. Church, E. D. Brown, T. S. Jaakkola, R. Barzilay, J. J. Collins, A deep learning approach to antibiotic discovery. *Cell* **180**, 688–702.e13 (2020). doi:10.1016/j.cell.2020.01.021 Medline
19. G. Liu, D. B. Catacutan, K. Rathod, K. Swanson, W. Jin, J. C. Mohammed, A. Chiappino-Pepe, S. A. Syed, M. Fragis, K. Rachwalski, J. Magolan, M. G. Surette, B. K. Coombes, T. Jaakkola, R. Barzilay, J. J. Collins, J. M. Stokes, Deep learning-guided discovery of an antibiotic targeting *Acinetobacter baumannii*. *Nat. Chem. Biol.* **19**, 1342–1350 (2023). doi:10.1038/s41589-023-01349-8 Medline
20. F. Wong, S. Omori, N. M. Donghia, E. J. Zheng, J. J. Collins, Discovering small-molecule senolytics with deep neural networks. *Nat. Aging* **3**, 734–750 (2023). doi:10.1038/s43587-023-00415-z Medline
21. J. Zhu, J. Wang, X. Wang, M. Gao, B. Guo, M. Gao, J. Liu, Y. Yu, L. Wang, W. Kong, Y. An, Z. Liu, X. Sun, Z. Huang, H. Zhou, N. Zhang, R. Zheng, Z. Xie, Prediction of drug efficacy from transcriptional profiles with deep learning. *Nat. Biotechnol.* **39**, 1444–1452 (2021). doi:10.1038/s41587-021-00946-z Medline
22. A. Subramanian, R. Narayan, S. M. Corsello, D. D. Peck, T. E. Natoli, X. Lu, J. Gould, J. F. Davis, A. A. Tubelli, J. K. Asiedu, D. L. Lahr, J. E. Hirschman, Z. Liu, M. Donahue, B. Julian, M. Khan, D. Wadden, I. C. Smith, D. Lam, A. Liberzon, C. Toder, M. Bagul, M. Orzechowski, O. M. Enache, F. Piccioni, S. A. Johnson, N. J. Lyons, A. H. Berger, A. F. Shamji, A. N. Brooks, A. Vrcic, C. Flynn, J. Rosains, D. Y. Takeda, R. Hu, D. Davison, J. Lamb, K. Ardlie, L. Hogstrom, P. Greenside, N. S. Gray, P. A. Clemons, S. Silver, X. Wu, W.-N. Zhao, W. Read-Button, X. Wu, S. J. Haggarty, L. V. Ronco, J. S. Boehm, S. L. Schreiber, J. G. Doench, J. A. Bittker, D. E. Root, B. Wong, T. R. Golub, A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452.e17 (2017). doi:10.1016/j.cell.2017.10.049 Medline
23. J. Chan, X. Wang, J. A. Turner, N. E. Baldwin, J. Gu, Breaking the paradigm: Dr Insight empowers signature-free, enhanced drug repurposing. *Bioinformatics* **35**, 2818–2826 (2019). doi:10.1093/bioinformatics/btz006 Medline
24. B. He, Y. Xiao, H. Liang, Q. Huang, Y. Du, Y. Li, D. Garmire, D. Sun, L. X. Garmire, ASGARD is a single-cell guided pipeline to aid repurposing of drugs. *Nat. Commun.* **14**, 993 (2023). doi:10.1038/s41467-023-36637-3 Medline
25. S. Raghavan, P. S. Winter, A. W. Navia, H. L. Williams, A. DenAdel, K. E. Lowder, J. Galvez-Reyes, R. L. Kalekar, N. Mulugeta, K. S. Kapner, M. S. Raghavan, A. A. Borah, N. Liu, S. A. Väyrynen, A. D. Costa, R. W. S. Ng, J. Wang, E. K. Hill, D. Y. Ragon, L. K. Brais, A. M. Jaeger, L. F. Spurr, Y. Y. Li, A. D. Cherniack, M. A. Booker, E. F. Cohen, M. Y. Tolstorukov, I. Wakiro, A. Rotem, B. E. Johnson, J. M. McFarland, E. T. Sicinska, T. E. Jacks, R. J. Sullivan, G. I. Shapiro, T. E. Clancy, K. Perez, D. A. Rubinson, K. Ng, J. M. Cleary, L. Crawford, S. R. Manalis, J. A. Nowak, B. M. Wolpin, W. C. Hahn, A. J. Aguirre, A. K. Shalek, Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. *Cell* **184**, 6119–6137.e26 (2021). doi:10.1016/j.cell.2021.11.017 Medline
26. X. Tong, N. Qu, X. Kong, S. Ni, J. Zhou, K. Wang, L. Zhang, Y. Wen, J. Shi, S. Zhang, X. Li, M. Zheng, Deep representation learning of chemical-induced transcriptional profile for phenotype-based drug discovery. *Nat. Commun.* **15**, 5378 (2024). doi:10.1038/s41467-024-49620-3 Medline
27. X. Qi, L. Zhao, C. Tian, Y. Li, Z.-L. Chen, P. Huo, R. Chen, X. Liu, B. Wan, S. Yang, Y. Zhao, Predicting transcriptional responses to novel chemical perturbations using deep generative model for drug discovery. *Nat. Commun.* **15**, 9256 (2024). doi:10.1038/s41467-024-53457-1 Medline
28. N. J. Kassebaum; GBD 2013 Anemia Collaborators, The global burden of anemia. *Hematol. Oncol. Clin. North Am.* **30**, 247–308 (2016). doi:10.1016/j.hoc.2015.11.002 Medline
29. "Thrombocytopenia in pregnancy," Practice Bulletin Number 207 (The American College of Obstetricians and Gynecologists, 2019); https://www.acog.org/clinical/clinical-guidance/practice-bulletin/articles/2019/03/thrombocytopenia-in-pregnancy.
30. C. Neunert, D. R. Terrell, D. M. Arnold, G. Buchanan, D. B. Cines, N. Cooper, A. Cuker, J. M. Despotovic, J. N. George, R. F. Grace, T. Kühne, D. J. Kuter, W. Lim, K. R. McCrae, B. Pruitt, H. Shimanek, S. K. Vesely, American Society of Hematology 2019 guidelines for immune thrombocytopenia. *Blood Adv.* **3**, 3829–3866 (2019). doi:10.1182/bloodadvances.2019000966 Medline
31. J. E. Rood, A. Maartens, A. Hupalowska, S. A. Teichmann, A. Regev, Impact of the Human Cell Atlas on medicine. *Nat. Med.* **28**, 2486–2496 (2022). doi:10.1038/s41591-022-02104-7 Medline
32. J. E. Evangelista, D. J. B. Clarke, Z. Xie, A. Lachmann, M. Jeon, K. Chen, K. M. Jagodnik, S. L. Jenkins, M. V. Kuleshov, M. L. Wojciechowicz, S. C. Schürer, M. Medvedovic, A. Ma'ayan, SigCom LINCS: Data and metadata search engine for a million gene expression signatures. *Nucleic Acids Res.* **50**, W697–W709 (2022). doi:10.1093/nar/gkac328 Medline
33. S. R. Srivatsan, J. L. McFaline-Figueroa, V. Ramani, L. Saunders, J. Cao, J. Packer, H. A. Pliner, D. L. Jackson, R. M. Daza, L. Christiansen, F. Zhang, F. Steemers, J. Shendure, C. Trapnell, Massively multiplex chemical transcriptomics at single-cell resolution. *Science* **367**, 45–51 (2020). doi:10.1126/science.aax6234 Medline
34. J. Wechsler, M. Greene, M. A. McDevitt, J. Anastasi, J. E. Karp, M. M. Le Beau, J. D. Crispino, Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nat. Genet.* **32**, 148–152 (2002). doi:10.1038/ng955 Medline
35. P. Giannoni, C. Marini, G. Cutrona, K. Todoerti, A. Neri, A. Ibatici, G. Sambuceti, S. Pigozzi, M. Mora, M. Ferrarini, F. Fais, D. de Totero, A high percentage of CD16+ monocytes correlates with the extent of bone erosion in chronic lymphocytic leukemia patients: The impact of leukemic B cells in monocyte differentiation and osteoclast maturation. *Cancers* **14**, 5979 (2022). doi:10.3390/cancers14235979 Medline
36. A. Dhawan, E. Padron, Abnormal monocyte differentiation and function in chronic myelomonocytic leukemia. *Curr. Opin. Hematol.* **29**, 20–26 (2022). doi:10.1097/MOH.0000000000000689 Medline
37. R. A. Voit, L. Tao, F. Yu, L. D. Cato, B. Cohen, T. J. Fleming, M. Antoszewski, X. Liao, C. Fiorini, S. K. Nandakumar, L. Wahlster, K. Teichert, A. Regev, V. G. Sankaran, A genetic disorder reveals a hematopoietic stem cell regulatory network co-opted in leukemia. *Nat. Immunol.* **24**, 69–83 (2023). doi:10.1038/s41590-022-01370-4 Medline
38. J. G. Drachman, G. P. Jarvik, M. G. Mehaffey, Autosomal dominant thrombocytopenia: Incomplete megakaryocyte differentiation and linkage to

human chromosome 10. *Blood* **96**, 118–125 (2000). doi:10.1182/blood.V96.1.118 Medline
39. N. Wang, C. LaVasseur, R. Riaz, J. Papoin, L. Blanc, A. Narla, Targeting of Calbindin 1 rescues erythropoiesis in a human model of Diamond Blackfan anemia. *Blood Cells Mol. Dis.* **102**, 102759 (2023). doi:10.1016/j.bcmd.2023.102759 Medline
40. P. Germino-Watnick, M. Hinds, A. Le, R. Chu, X. Liu, N. Uchida, Hematopoietic Stem cell gene-addition/editing therapy in sickle cell disease. *Cells* **11**, 1843 (2022). doi:10.3390/cells11111843 Medline
41. D. S. Krause, M. J. Fackler, C. I. Civin, W. S. May, CD34: Structure, biology, and clinical utility. *Blood* **87**, 1–13 (1996). doi:10.1182/blood.V87.1.1.1 Medline
42. S. Kumar, H. Geiger, HSC niche biology and HSC expansion ex vivo. *Trends Mol. Med.* **23**, 799–819 (2017). doi:10.1016/j.molmed.2017.07.003 Medline
43. J. H. Cheah, J. A. Bittker, "'So you want to run a high-throughput screen: Do you know how much that costs?'; Costs of high throughput screens and how to fund them" in *High Throughput Screening Methods: Evolution and Refinement*, J. A. Bittker, N. T. Ross, Eds. (Royal Society of Chemistry, 2016), chap. 17, pp. 372–389.
44. M. Stoeckius, C. Hafemeister, W. Stephenson, B. Houck-Loomis, P. K. Chattopadhyay, H. Swerdlow, R. Satija, P. Smibert, Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017). doi:10.1038/nmeth.4380 Medline
45. D. Burkhardt, M. Luecken, A. Benz, P. Holderrieth, J. Bloom, C. Lance, A. Chow, R. Holbrook, Open Problems - Multimodal Single-Cell Integration (Kaggle, 2022); https://kaggle.com/competitions/open-problems-multimodal.
46. T. P. McDonald, P. S. Sullivan, Megakaryocytic and erythrocytic cell lines share a common precursor cell. *Exp. Hematol.* **21**, 1316–1320 (1993). Medline
47. Y. C. Lu, C. Sanada, J. Xavier-Ferrucio, L. Wang, P. X. Zhang, H. L. Grimes, M. Venkatasubramanian, K. Chetal, B. Aronow, N. Salomonis, D. S. Krause, The molecular signature of megakaryocyte-erythroid progenitors reveals a role for the cell cycle in fate specification. *Cell Rep.* **25**, 2083–2093.e4 (2018). doi:10.1016/j.celrep.2018.10.084 Medline
48. J. K. Bailur, S. S. McCachren, K. Pendleton, J. C. Vasquez, H. S. Lim, A. Duffy, D. B. Doxie, A. Kaushal, C. Foster, D. DeRyckere, S. Castellino, M. L. Kemp, P. Qiu, M. V. Dhodapkar, K. M. Dhodapkar, Risk-associated alterations in marrow T cells in pediatric leukemia. *JCI Insight* **5**, e140179 (2020). doi:10.1172/jci.insight.140179 Medline
49. B. Pal, Y. Chen, F. Vaillant, B. D. Capaldo, R. Joyce, X. Song, V. L. Bryant, J. S. Penington, L. Di Stefano, N. Tubau Ribera, S. Wilcox, G. B. Mann, A. T. Papenfuss, G. J. Lindeman, G. K. Smyth, J. E. Visvader; kConFab, A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. *EMBO J.* **40**, e107333 (2021). doi:10.15252/embj.2020107333 Medline
50. M. Passet, R. Kim, E. Clappier, Genetic subtypes of B-cell acute lymphoblastic leukemia in adults. *Blood* **145**, 1451–1463 (2025). doi:10.1182/blood.2023022919 Medline
51. E. Jabbour, H. M. Kantarjian, I. Aldoss, P. Montesinos, J. T. Leonard, D. Gómez-Almaguer, M. R. Baer, C. Gambacorti-Passerini, J. McCloskey, Y. Minami, C. Papayannidis, V. Rocha, P. Rousselot, P. Vachhani, E. S. Wang, B. Wang, M. Hennessy, A. Vorog, N. Patel, T. Yeh, J.-M. Ribera, Ponatinib vs imatinib in frontline Philadelphia chromosome-positive acute lymphoblastic leukemia: A randomized clinical trial. *JAMA* **331**, 1814–1823 (2024). doi:10.1001/jama.2024.4783 Medline
52. D. Khiem, J. G. Cyster, J. J. Schwarz, B. L. Black, A p38 MAPK-MEF2C pathway regulates B-cell proliferation. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17067–17072 (2008). doi:10.1073/pnas.0804868105 Medline
53. J. D. Mclaurin, O. D. Weiner, Multiple sources of signal amplification within the B-cell Ras/MAPK pathway. *Mol. Biol. Cell* **30**, 1610–1620 (2019). doi:10.1091/mbc.E18-09-0560 Medline
54. D. Chi, H. Singhal, L. Li, T. Xiao, W. Liu, M. Pun, R. Jeselsohn, H. He, E. Lim, R. Vadhi, P. Rao, H. Long, J. Garber, M. Brown, Estrogen receptor signaling is reprogrammed during breast tumorigenesis. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 11437–11443 (2019). doi:10.1073/pnas.1819155116 Medline
55. K. L. Thu, I. Soria-Bretones, T. W. Mak, D. W. Cescon, Targeting the cell cycle in breast cancer: Towards the next phase. *Cell Cycle* **17**, 1871–1885 (2018). doi:10.1080/15384101.2018.1502567 Medline
56. M. Kesireddy, L. Elsayed, V. K. Shostrom, P. Agarwal, S. Asif, A. Yellala, J. Krishnamurthy, Overall survival and prognostic factors in metastatic triple-negative breast cancer: A National Cancer Database analysis. *Cancers* **16**, 1791 (2024). doi:10.3390/cancers16101791 Medline
57. C. Xue, Q. Yao, X. Gu, Q. Shi, X. Yuan, Q. Chu, Z. Bao, J. Lu, L. Li, Evolving cognition of the JAK-STAT signaling pathway: Autoimmune disorders and cancer. *Signal Transduct. Target. Ther.* **8**, 204 (2023). doi:10.1038/s41392-023-01468-7 Medline
58. D. Reker, G. Schneider, Active-learning strategies in computer-assisted drug discovery. *Drug Discov. Today* **20**, 458–465 (2015). doi:10.1016/j.drudis.2014.12.004 Medline
59. L. Wang, Z. Zhou, X. Yang, S. Shi, X. Zeng, D. Cao, The present state and challenges of active learning in drug discovery. *Drug Discov. Today* **29**, 103985 (2024). doi:10.1016/j.drudis.2024.103985 Medline
60. B. Woods, W. Chen, S. Chiu, C. Marinaccio, C. Fu, L. Gu, M. Bulic, Q. Yang, A. Zouak, S. Jia, P. K. Suraneni, K. Xu, R. L. Levine, J. D. Crispino, Q. J. Wen, Activation of JAK/STAT signaling in megakaryocytes sustains myeloproliferation *in vivo*. *Clin. Cancer Res.* **25**, 5901–5912 (2019). doi:10.1158/1078-0432.CCR-18-4089 Medline
61. A. Zufferey, E. R. Speck, K. R. Machlus, R. Aslam, L. Guo, M. J. McVey, M. Kim, R. Kapur, E. Boilard, J. E. Italiano Jr., J. W. Semple, Mature murine megakaryocytes present antigen-MHC class I molecules to T cells and transfer them to platelets. *Blood Adv.* **1**, 1773–1785 (2017). doi:10.1182/bloodadvances.2017007021 Medline
62. Z. Fang, X. Liu, G. Peltz, GSEApy: A comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, btac757 (2023). doi:10.1093/bioinformatics/btac757 Medline
63. A. Mazharian, C. Ghevaert, L. Zhang, S. Massberg, S. P. Watson, Dasatinib enhances megakaryocyte differentiation but inhibits platelet formation. *Blood* **117**, 5198–5206 (2011). doi:10.1182/blood-2010-12-326850 Medline
64. K. Suknuntha, Y. J. Choi, H. S. Jung, A. Majumder, S. Shah, I. Slukvin, E. A. Ranheim, Megakaryocytic expansion in gilteritinib-treated acute myeloid leukemia patients is associated with AXL inhibition. *Front. Oncol.* **10**, 585151 (2020). doi:10.3389/fonc.2020.585151 Medline
65. K. L. Kelly, W. J. Reagan, G. E. Sonnenberg, M. Clasquin, K. Hales, S. Asano, P. A. Amor, S. Carvajal-Gonzalez, N. Shirai, M. D. Matthews, K. W. Li, M. K. Hellerstein, N. B. Vera, T. T. Ross, G. Cappon, A. Bergman, C. Buckeridge, Z. Sun, E. Z. Qejvanaj, T. Schmahai, D. Beebe, J. A. Pfefferkorn, W. P. Esler, De novo lipogenesis is essential for platelet production in humans. *Nat. Metab.* **2**, 1163–1178 (2020). doi:10.1038/s42255-020-00272-9 Medline
66. B. de Jonckheere, F. Kollotzek, P. Münzer, V. Göb, M. Fischer, K. Mott, C. Coman, N. N. Troppmair, M.-C. Manke, M. Zdanyte, T. Harm, M. Sigle, D. Kopczynski, A. Bileck, C. Gerner, N. Hoffmann, D. Heinzmann, A. Assinger, M. Gawaz, D. Stegner, H. Schulze, O. Borst, R. Ahrends, Critical shifts in lipid metabolism promote megakaryocyte differentiation and proplatelet formation. *Nat. Cardiovasc. Res.* **2**, 835–852 (2023). doi:10.1038/s44161-023-00325-8 Medline
67. B. Zdrazil, E. Felix, F. Hunter, E. J. Manners, J. Blackshaw, S. Corbett, M. de Veij, H. Ioannidis, D. M. Lopez, J. F. Mosquera, M. P. Magarinos, N. Bosc, R. Arcila, T. Kizilören, A. Gaulton, A. P. Bento, M. F. Adasme, P. Monecke, G. A. Landrum, A. R. Leach, The ChEMBL Database in 2023: A drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* **52**, D1180–D1192 (2024). doi:10.1093/nar/gkad1004 Medline
68. H. Wang, T. Fu, Y. Du, W. Gao, K. Huang, Z. Liu, P. Chandak, S. Liu, P. Van Katwyk, A. Deac, A. Anandkumar, K. Bergen, C. P. Gomes, S. Ho, P. Kohli, J. Lasenby, J. Leskovec, T.-Y. Liu, A. Manrai, D. Marks, B. Ramsundar, L. Song, J. Sun, J. Tang, P. Veličković, M. Welling, L. Zhang, C. W. Coley, Y. Bengio, M. Zitnik, Scientific discovery in the age of artificial intelligence. *Nature* **620**, 47–60 (2023). doi:10.1038/s41586-023-06221-2 Medline
69. M. Lange, V. Bergen, M. Klein, M. Setty, B. Reuter, M. Bakhti, H. Lickert, M. Ansari, J. Schniering, H. B. Schiller, D. Pe'er, F. J. Theis, CellRank for directed single-cell fate mapping. *Nat. Methods* **19**, 159–170 (2022). doi:10.1038/s41592-021-01346-6 Medline
70. A. S. Nam, K.-T. Kim, R. Chaligne, F. Izzo, C. Ang, J. Taylor, R. M. Myers, G. Abu-Zeinah, R. Brand, N. D. Omans, A. Alonso, C. Sheridan, M. Mariani, X. Dai, E. Harrington, A. Pastore, J. R. Cubillos-Ruiz, W. Tam, R. Hoffman, R. Rabadan, J. M. Scandura, O. Abdel-Wahab, P. Smibert, D. A. Landau, Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature* **571**, 355–360 (2019). doi:10.1038/s41586-019-1367-0 Medline

71. J. S. Fleck, S. M. J. Jansen, D. Wollny, F. Zenk, M. Seimiya, A. Jain, R. Okamoto, M. Santel, Z. He, J. G. Camp, B. Treutlein, Inferring and perturbing cell fate regulomes in human brain organoids. *Nature* **621**, 365–372 (2023). doi:10.1038/s41586-022-05279-8 Medline
72. K. Kamimoto, B. Stringa, C. M. Hoffmann, K. Jindal, L. Solnica-Krezel, S. A. Morris, Dissecting cell identity via network inference and in silico gene perturbation. *Nature* **614**, 742–751 (2023). doi:10.1038/s41586-022-05688-9 Medline
73. Y. Qiu, T. Lu, H. Lim, L. Xie, A Bayesian approach to accurate and robust signature detection on LINCS L1000 data. *Bioinformatics* **36**, 2787–2795 (2020). doi:10.1093/bioinformatics/btaa064 Medline
74. R. S. Bohacek, C. McMartin, W. C. Guida, The art and practice of structure-based drug design: A molecular modeling perspective. *Med. Res. Rev.* **16**, 3–50 (1996). doi:10.1002/(SICI)1098-1128(199601)16:1<3:AID-MED1>3.0.CO;2-6 Medline
75. O. O. Grygorenko, D. S. Radchenko, I. Dziuba, A. Chuprina, K. E. Gubina, Y. S. Moroz, Generating multibillion chemical space of readily accessible screening compounds. *iScience* **23**, 101681 (2020). doi:10.1016/j.isci.2020.101681 Medline
76. A. Szałata, A. Benz, R. Cannoodt, M. Cortes, J. Fong, S. Kuppasani, R. Lieberman, T. Liu, J. A. Mas-Rosario, R. Meinl, J. Nourisa, J. Tumiel, T. M. Tunjic, M. Wang, N. Weber, H. Zhao, B. Anchang, F. J. Theis, M. D. Luecken, D. B. Burkhardt, A benchmark for prediction of transcriptomic responses to chemical perturbations across cell types. *Adv. Neural Inf. Process. Syst.* **37**, 20566–20616 (2024).
77. L. Hetzel, S. Boehm, N. Kilbertus, S. Günnemann, M. Lotfollahi, F. Theis, Predicting cellular responses to novel drug perturbations at a single-cell resolution. *Adv. Neural Inf. Process. Syst.* **35**, 26711–26722 (2022).
78. A. Tejada-Lapuerta, P. Bertin, S. Bauer, H. Aliee, Y. Bengio, F. J. Theis, Causal machine learning for single-cell genomics. *Nat. Genet.* **57**, 797–808 (2025). doi:10.1038/s41588-025-02124-2 Medline
79. R. Lopez, J.-C. Hütter, J. K. Pritchard, A. Regev, Large-scale differentiable causal discovery of factor graphs. arXiv:2206.07824 [stat.ML] (2022).
80. J. Jiang, S. Chen, T. Tsou, C. S. McGinnis, T. Khazaei, Q. Zhu, J. H. Park, I.-M. Strazhnik, J. Vielmetter, Y. Gong, J. Hanna, E. D. Chow, D. A. Sivak, Z. J. Gartner, M. Thomson, D-SPIN constructs gene regulatory network models from multiplexed scRNA-seq data revealing organizing principles of cellular perturbation response. bioRxiv 2023.04.19.537364 [Preprint] (2023); https://doi.org/10.1101/2023.04.19.537364.
81. D. Prins, H. J. Park, S. Watcham, J. Li, M. Vacca, H. P. Bastos, A. Gerbaulet, A. Vidal-Puig, B. Göttgens, A. R. Green, The stem/progenitor landscape is reshaped in a mouse model of essential thrombocythemia and causes excess megakaryocyte production. *Sci. Adv.* **6**, eabd3139 (2020). doi:10.1126/sciadv.abd3139 Medline
82. A. J. Murphy, N. Bijl, L. Yvan-Charvet, C. B. Welch, N. Bhagwat, A. Reheman, Y. Wang, J. A. Shaw, R. L. Levine, H. Ni, A. R. Tall, N. Wang, Cholesterol efflux in megakaryocyte progenitors suppresses platelet production and thrombocytosis. *Nat. Med.* **19**, 586–594 (2013). doi:10.1038/nm.3150 Medline
83. M. Cortes, A. J. Monti, S. Sun, O. Lynch, Y. Xia, S. Lin, M. Malamas, S. Steelman, S. Krishnamoorthy, M. Stewart, E. Tozzo, Identification of small molecules that induce therapeutic levels of fetal hemoglobin for treatment of sickle cell disease by pairing machine learning with high-resolution single cell RNA sequencing maps of adult and fetal human erythropoiesis. *Blood* **138** (suppl. 1), 2022 (2021). doi:10.1182/blood-2021-152728
84. Z. Wei, D. Si, B. Duan, Y. Gao, Q. Yu, Z. Zhang, L. Guo, Q. Liu, PerturBase: A comprehensive database for single-cell perturbation data analysis and visualization. *Nucleic Acids Res.* **53**, D1099–D1111 (2025). doi:10.1093/nar/gkae858 Medline
85. J. Zhang, A. A. Ubas, R. de Borja, V. Svensson, N. Thomas, N. Thakar, I. Lai, A. Winters, U. Khan, M. G. Jones, J. D. Thompson, V. Tran, J. Pangallo, E. Papalexi, A. Sapre, H. Nguyen, O. Sanderson, M. Nigos, O. Kaplan, S. Schroeder, B. Hariadi, S. Marrujo, C. C. A. Salvino, G. G. Olivares, R. Koehler, G. Geiss, A. Rosenberg, C. Roco, D. Merico, N. Alidoust, H. Goodarzi, J. Yu, *Tahoe-100M*: A giga-scale single-cell perturbation atlas for context-dependent gene function and cellular modeling. bioRxiv 2025.02.20.639398 [Preprint] (2025); https://doi.org/10.1101/2025.02.20.639398.
86. Cellarity Inc, V-score signatures for cellular state changes, Zenodo (2025); https://doi.org/10.5281/zenodo.16921906.
87. Cellarity Inc, DrugReflector source code, Zenodo (2025); https://doi.org/10.5281/zenodo.16921928.
88. Cellarity Inc, DrugReflector model checkpoints, Zenodo (2025); https://doi.org/10.5281/zenodo.16912445.
89. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: An imperative style, high-performance deep learning library. arXiv:1912.01703 [cs.LG] (2019).
90. Daylight Chemical Information Systems, Inc., Daylight Theory: SMARTS - A Language for Describing Molecular Patterns; https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html.
91. A. Schuffenhauer, N. Schneider, S. Hintermann, D. Auld, J. Blank, S. Cotesta, C. Engeloch, N. Fechner, C. Gaul, J. Giovannoni, J. Jansen, J. Joslin, P. Krastel, E. Lounkine, J. Manchester, L. G. Monovich, A. P. Pelliccioli, M. Schwarze, M. D. Shultz, N. Stiefl, D. K. Baeschlin, Evolution of Novartis' small molecule screening deck design. *J. Med. Chem.* **63**, 14425–14447 (2020). doi:10.1021/acs.jmedchem.0c01332 Medline
92. R. Brenk, A. Schipani, D. James, A. Krasowski, I. H. Gilbert, J. Frearson, P. G. Wyatt, Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *ChemMedChem* **3**, 435–444 (2008). doi:10.1002/cmdc.200700139 Medline
93. T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, "Focal loss for dense object detection," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy (2017), pp. 2999–3007.
94. S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167 [cs.LG] (2015).
95. I. Loshchilov, F. Hutter, SGDR: Stochastic Gradient Descent with Warm Restarts. arXiv:1608.03983 [cs.LG] (2017).
96. T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A next-generation hyperparameter optimization framework. arXiv:1907.10902 [cs.LG] (2019).
97. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018). doi:10.1186/s13059-017-1382-0 Medline
98. K. Petrova, Symphonypy, GitHub (2023); https://github.com/potulabe/symphonypy.
99. S. A. Miller, R. A. Policastro, S. Sriramkumar, T. Lai, T. D. Huntington, C. A. Ladaika, D. Kim, C. Hao, G. E. Zentner, H. M. O'Hagan, LSD1 and aberrant DNA methylation mediate persistence of enteroendocrine progenitors that support *BRAF*-mutant colorectal cancer. *Cancer Res.* **81**, 3791–3805 (2021). doi:10.1158/0008-5472.CAN-20-3562 Medline
100. C. De Donno, S. Hediyeh-Zadeh, A. A. Moinfar, M. Wagenstetter, L. Zappia, M. Lotfollahi, F. J. Theis, Population-level integration of single-cell datasets enables multi-scale analysis across samples. *Nat. Methods* **20**, 1683–1692 (2023). doi:10.1038/s41592-023-02035-2 Medline
101. C. Knox, M. Wilson, C. M. Klinger, M. Franklin, E. Oler, A. Wilson, A. Pon, J. Cox, N. E. L. Chin, S. A. Strawbridge, M. Garcia-Patino, R. Kruger, A. Sivakumaran, S. Sanford, R. Doshi, N. Khetarpal, O. Fatokun, D. Doucet, A. Zubkowski, D. Y. Rayat, H. Jackson, K. Harford, A. Anjum, M. Zakir, F. Wang, S. Tian, B. Lee, J. Liigand, H. Peters, R. Q. R. Wang, T. Nguyen, D. So, M. Sharp, R. da Silva, C. Gabriel, J. Scantlebury, M. Jasinski, D. Ackerman, T. Jewison, T. Sajed, V. Gautam, D. S. Wishart, DrugBank 6.0: The DrugBank Knowledgebase for 2024. *Nucleic Acids Res.* **52**, D1265–D1275 (2024). doi:10.1093/nar/gkad976 Medline
102. H. Li, R. Zhang, Y. Min, D. Ma, D. Zhao, J. Zeng, A knowledge-guided pre-training framework for improving molecular representation learning. *Nat. Commun.* **14**, 7568 (2023). doi:10.1038/s41467-023-43214-1 Medline
103. D. W. Brandt, Core system model: Understanding the impact of reliability on high-throughput screening systems. *Drug Discov. Today* **3**, 61–68 (1998). doi:10.1016/S1359-6446(97)01136-7

ACKNOWLEDGMENTS

We thank past and current members of Cellarity who have contributed to building our platform and provided valuable discussions and ideas throughout the years. In addition, we would like to thank team members who contributed to generation of public datasets throughout the years, including the NeurIPS 2022 competition. Finally, we would like to thank the founding members of Cellarity,

Nick Plugis and Avak Kahvejian, for their original vision. **Funding:** All work in this study was funded by Cellarity, Inc. **Author contributions:** Conceptualization: BD, SAM, DBB, AKS, FJT, MC; Methodology: BD, DBB, IL, SAM, CN, IG, SK, MZ, MC; Investigation: BD, CN, SAM, DBB, DF, PH, DK, AS, TP, RRR, CK; Visualization: BD, DBB, IL, SAM, DF, CN, AKS, MC; Project administration: DBB, IL, MZ, LK, MC; Supervision: LK, PD, MZ, AKS, FJT, MC; Writing – original draft: DBB, AKS, BD, DK, SAM, MC, IL, CN; Writing – revised draft: BD, SAM, CN, AKS, MC; Writing – review & editing: BD, SAM, DBB, CN, IL, MZ, PD, JC, AKS, FJT, MC. **Competing interests:** All authors are employees or compensated consultants to, and have equity interest in Cellarity, Inc. B.D. previously worked as a paid consultant for Dragonfly Therapeutics, A.S. consulted for Exvivo Labs Inc. J.C. is an academic co-founder and board member of Cellarity. A.K.S. reports compensation for consulting and/or scientific advisory board membership from Honeycomb Biotechnologies, Cellarity, Ochre Bio, Relation Therapeutics, Bio-Rad Laboratories, Quotient Therapeutics, Wolf Greenfield, Bio-Rad Laboratories, IntrECate Biotherapeutics, Third Rock Ventures, Parabilis Medicines, JnJ, Pfizer, Santa Ana, Danaher, Passkey Therapeutics, Sail Biosciences and Dahlia Biosciences unrelated to this work. A.K.S. reports research support from Break Through Cancer, Wellcome Leap, NIH, the Bill & Melinda Gates Foundation, Foundation MIT, the Moore Foundation, the Chan Zuckerberg Initiative, Coca-Cola, the UK Medical Research Council, and Becton, Dickinson, and Company. F.J.T. consults for Immunai, Singularity Bio, CytoReason, Cellarity and Curie Bio Operations and has ownership interest in Dermagnostix and Cellarity. **Data and materials availability:** Perturbational transcriptomic data from cancer and primary cell lines, used to benchmark DrugReflector, are available on GEO at accession GSE306429. The multi-modal single-cell time course data used to obtain the megakaryocyte and erythroid induction signatures is available on GEO at accession GSE305370. The perturbational single-cell RNA-seq time course of megakaryocyte differentiation, used to transcriptionally validate predictions and refine the input signature, is available on GEO at accession GSE305979. V-scores from the hematopoiesis signatures and from B-ALL and breast cancer transitions can be downloaded from Zenodo (*86*). Code to run the DrugReflector algorithm and to reproduce the major results of this manuscript are available on Zenodo (*87*). DrugReflector model checkpoints are available on Zenodo as described in the documentation (*88*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adi8577
Materials and Methods
Supplementary Text
Figs. S1 to S12
Tables S1 to S7
References (*89*–*103*)
MDAR Reproducibility Checklist

Submitted 22 July 2024; resubmitted 2 June 2025
Accepted 29 August 2025
Published online 23 October 2025
10.1126/science.adi8577

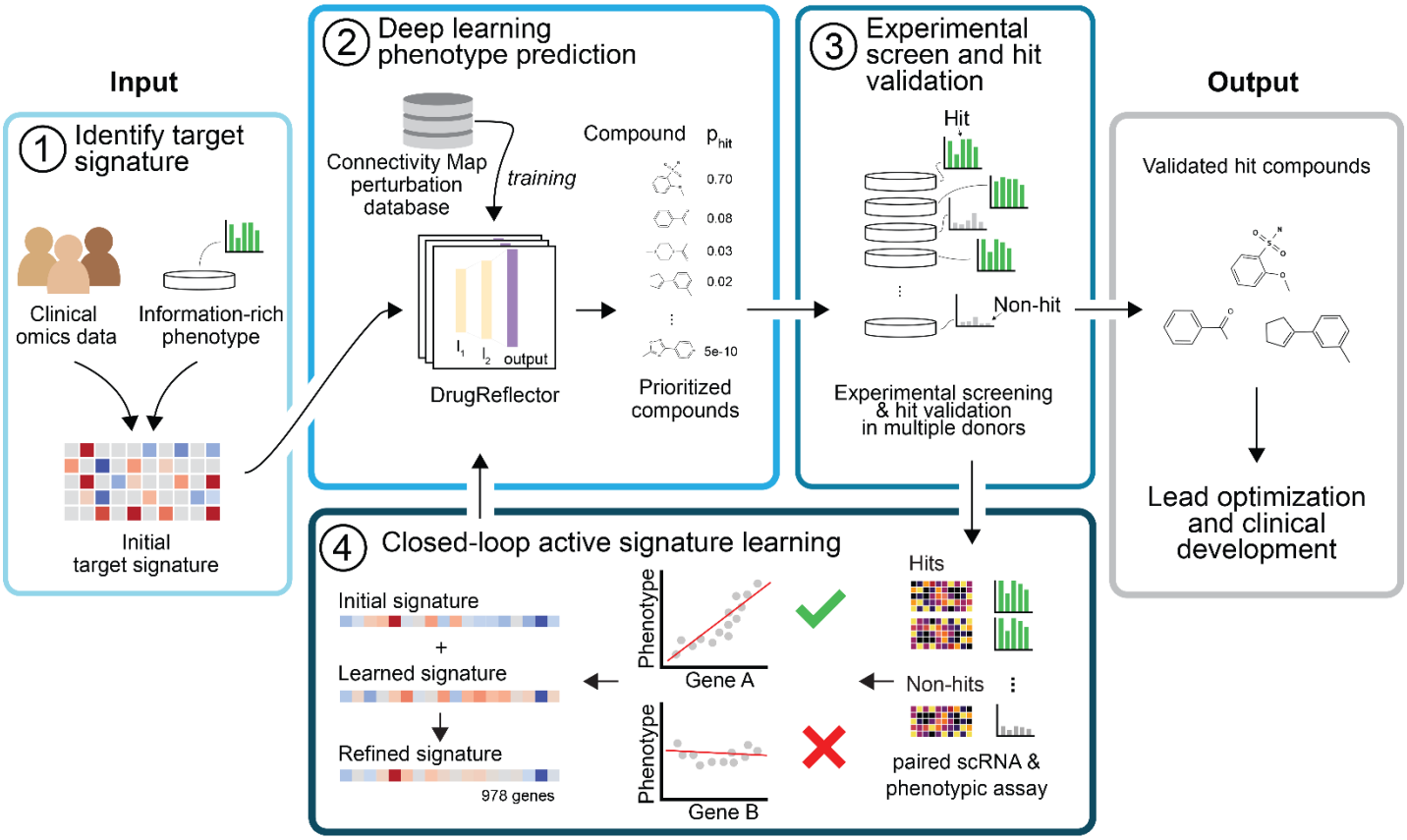
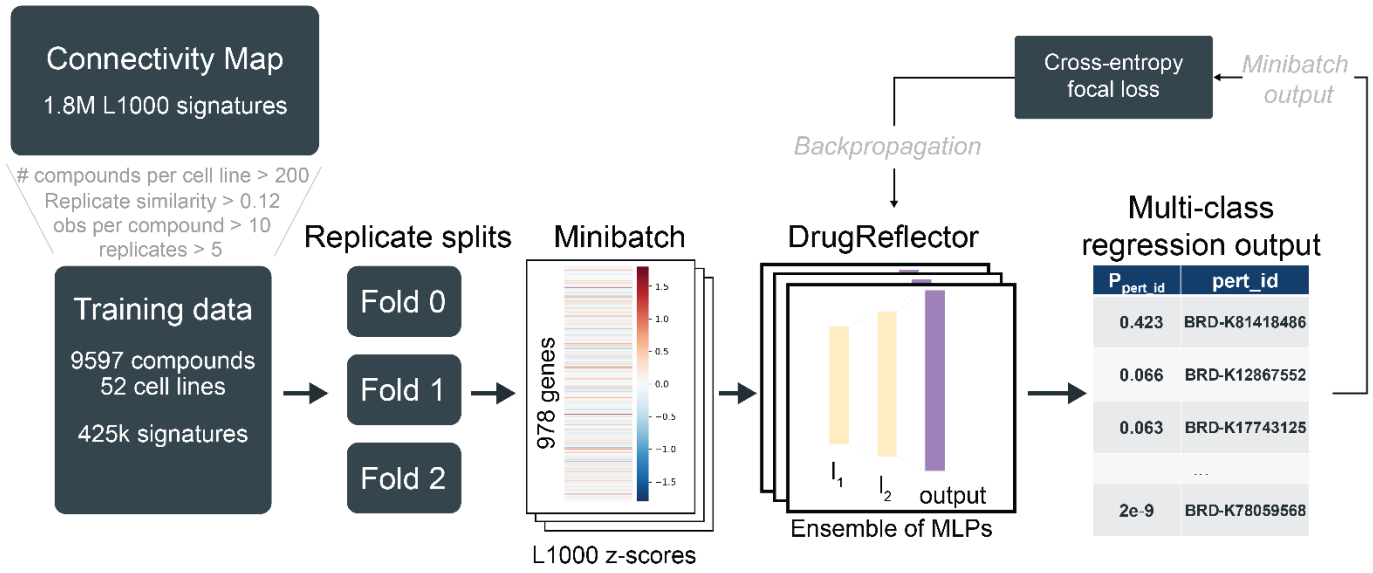
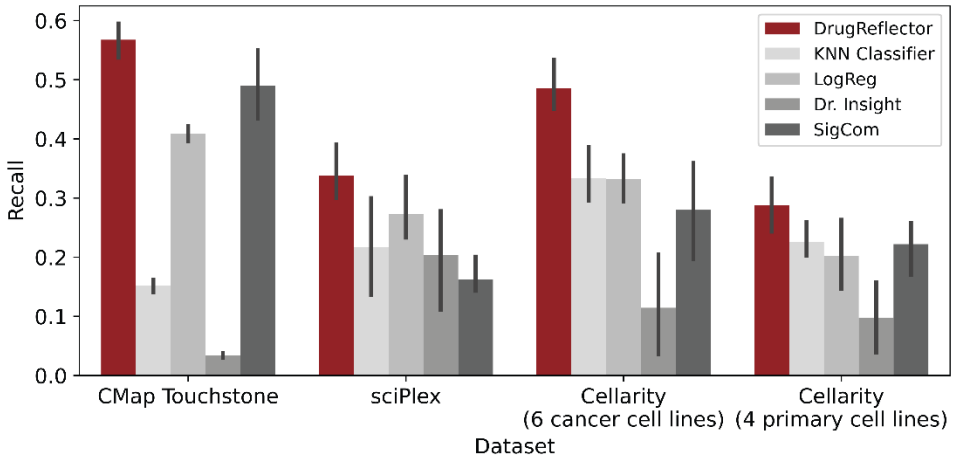


Fig. 1. A modular and generalizable framework to enable phenotypic discovery using omics-level deep learning models. (1) First, a target signature is identified based on clinical data and/or data from an information-rich clinically translatable phenotypic assay (here, a transcriptomic signature from single-cell RNA-seq data). (2) To determine compounds for screening, our deep learning model DrugReflector trained on perturbation signatures (here, the LINCS Connectivity Map) predicts which compounds will likely induce the target signature. (3) Compounds are then experimentally screened, and hit compounds that induce the desired phenotype are identified and validated in multiple donors. Validated hits are the output of this discovery stage and may be used for downstream pre-clinical development. (4) The input signature is actively refined by the lab-in-the-loop use of paired omic and phenotypic measurements, improving the prioritization of active compounds.

A



B



Dataset	CMap Touchstone	sciPlex 3	Cellarity Cancer	Cellarity Primary
Modality	L1000 (bulk)	snRNA	scRNA	scRNA
Cancer cell lines	10	3	6	0
Primary cell lines	0	0	0	4
Perturbation #	1,000	188	88	88

C

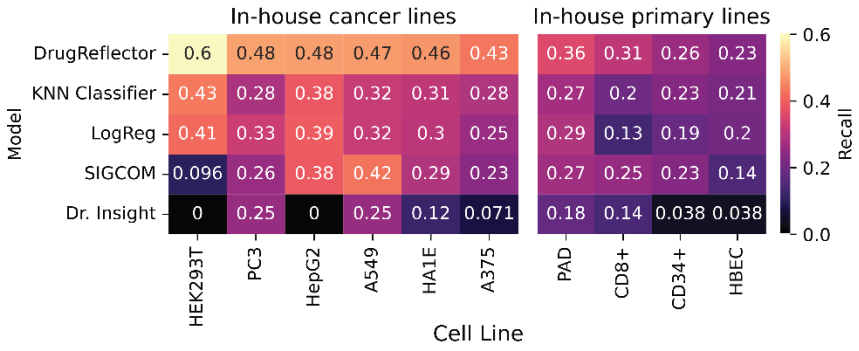


Fig. 2. A deep learning approach to phenotypic virtual screening. (**A**) A schematic representation of the model training regime. The full CMAP dataset was filtered using custom quality control metrics (methods) and partitioned into equal replicate splits for training. Each model in the ensemble was trained on 2 out of 3 folds and validated on the held-out fraction. The model was trained on a multi-class regression task where the goal was to match the correct perturbation labels to batches of input CMAP signatures, which are z-score representations of differential expression across each plate (Level 4 signatures). Full details of model training can be found in the methods section. (**B**) Performance of each algorithm on each benchmarking dataset, as measured by average top-1% compound recall over all cell lines. Error bars denote standard deviation across cell lines. snRNA: single-nucleus RNA-seq; scRNA: single-cell RNA-seq. (**C**) Heat map of recall for each algorithm in each cell line of our internal perturbational transcriptomic screen, with 6 cancer cell lines and 4 primary cell lines.

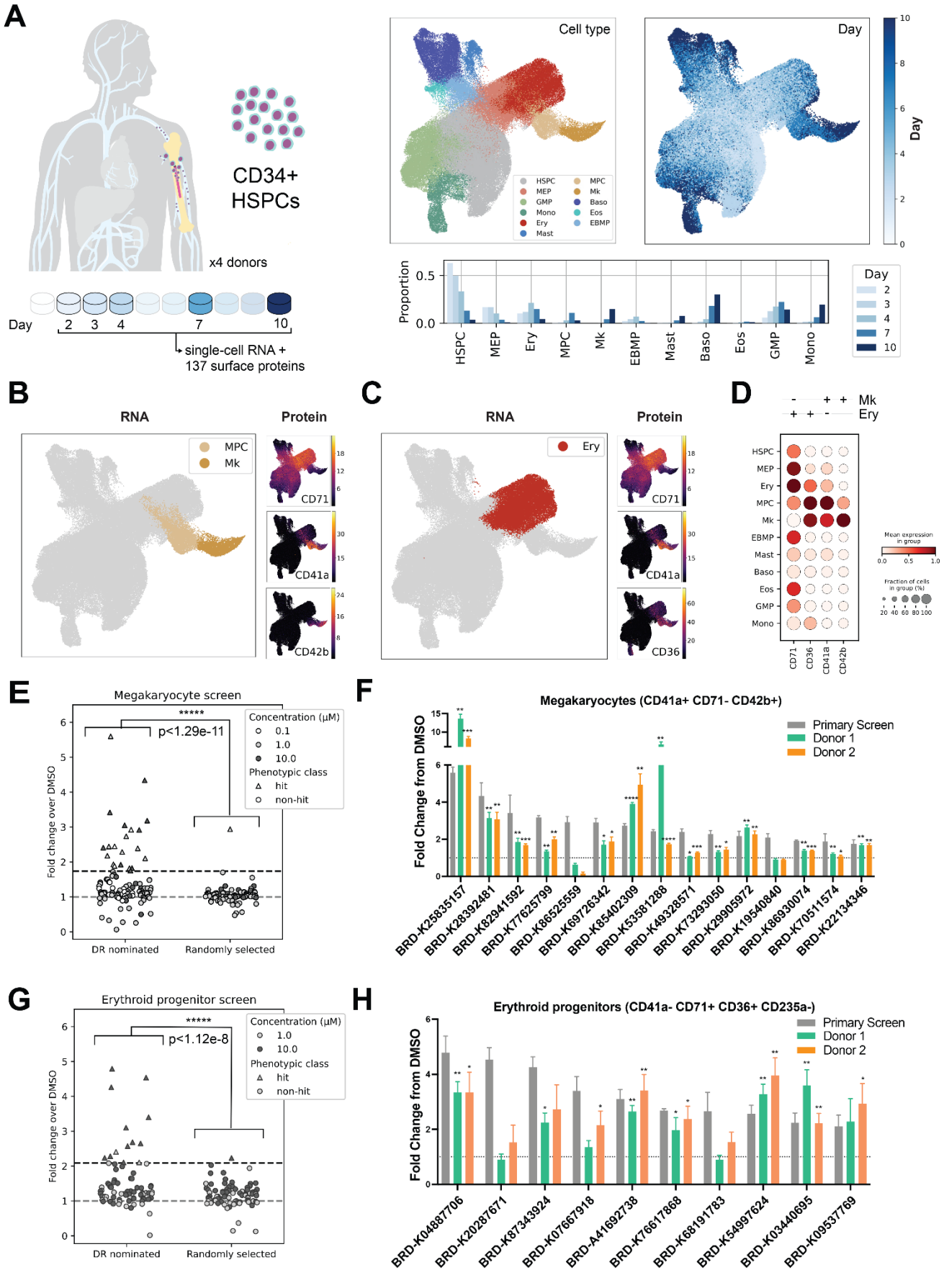
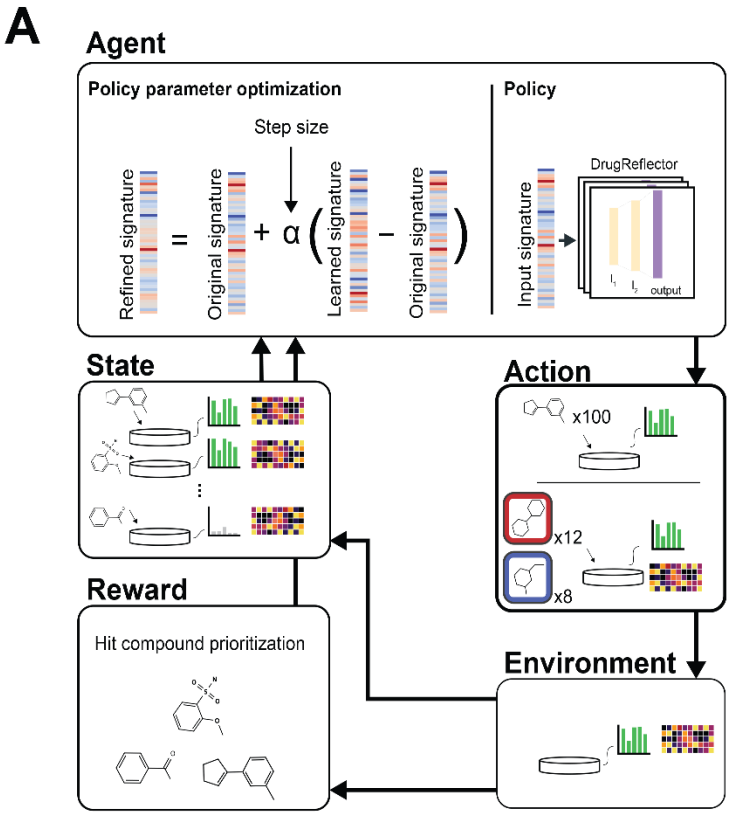
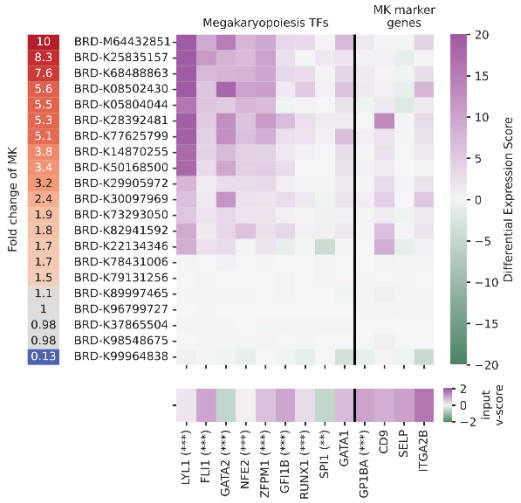
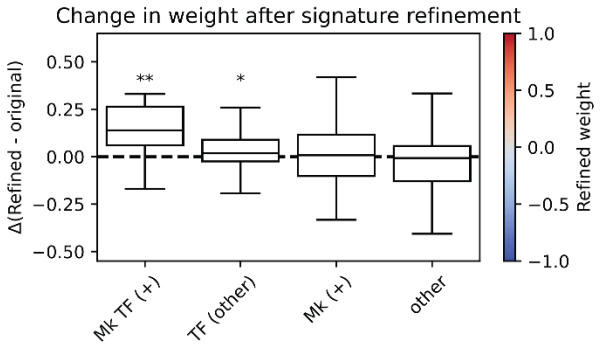
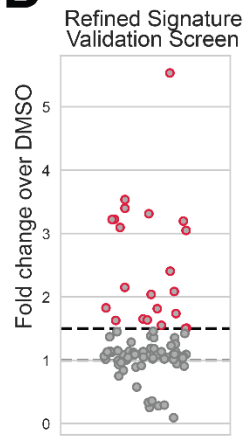


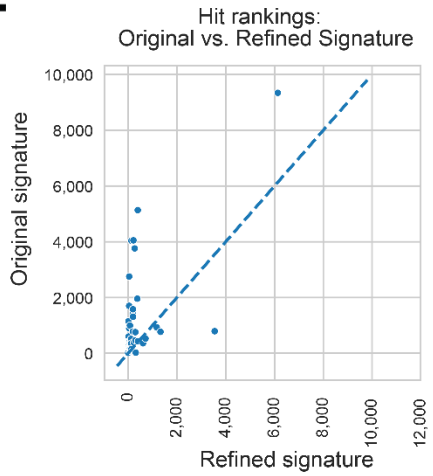
Fig. 3. A single-cell multi-omics guided phenotypic assay capturing multi-lineage hematopoietic differentiation in human primary cells. (**A**) We obtained primary CD34+ hematopoietic stem and progenitor cells (HSPCs) from 4 healthy donors and performed CITE-seq (single-cell RNA-seq + antibody derived tag surface protein marker measurements) at 5 time points over a 10-day time course in vitro. UMAP embeddings are shown for all cells from all donors and time points. Below, the proportion of cells assigned to each type across days. (**B** and **C**) UMAP embeddings calculated from RNA. Larger plots show cell types associated with the Mk (B) or Ery (C) lineage. Smaller plots show the expression of surface markers used for negative or positive identification of cell populations via FACS. (**D**) The marker panel for each flow cytometry phenotypic assay and the average expression of each surface marker in each cell type based on the CITE-seq surface marker measurements. (**E**) Result of experimental validation of compounds to induce Mk differentiation measured with flow cytometry following a 7-day in vitro differentiation in the presence of each compound. Triangles represent hits and circles represent non-hits. The color is the dose at which the compound maximally induced Mk abundance. The gray dashed line denotes a fold-change of 1 relative to DMSO (no change). The black dashed line represents the hit significance cutoff for Mk. The asterisks and p-value show the results of a one-sided binomial test between the hit-rates for each method (methods). (**F**) Validation of Mk hit compounds in 2 additional donors. Gray bars denote the maximal observed change in Mk abundance from the screen in (E) for each compound. Error bars denote standard deviation across technical triplicates. (**G**) Same as for (E), but for the Ery discovery campaign. (**H**) Same as (F) for validation of Ery hit compounds in multiple donors. Abbreviations: HSPC - Hematopoietic Stem and Progenitor Cell; MEP – Megakaryocyte Erythroid Progenitor; GMP – Granulocyte Macrophage Progenitor; Mono – Monocyte; Ery – erythrocyte; MPC – Megakaryocyte Progenitor; Baso – Basophil; Eos – Eosinophil; EBMP – Eosinophil/Basophil/Mast Progenitor.











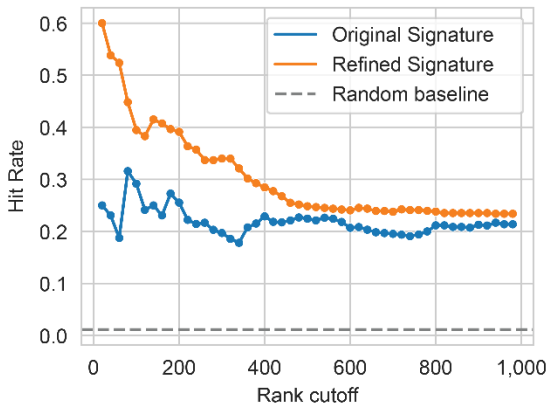


Fig. 4. Increased phenotypic hit-rate through active signature learning using paired transcriptional and phenotypic readouts. (**A**) A visual diagram of the reinforcement learning framework for active signature learning. In reinforcement learning (RL), an agent takes actions to interact with an environment, which yields rewards and changes in state that are fed back into the agent to update the policy. The goal is to learn a policy that maximizes the reward signal. In active signature learning, the agent is the combination of a policy, the output ranking of compounds from DrugReflector given an input signature, and the policy update process. The action is the selection of top compounds from DrugReflector for phenotypic screening, followed by paired phenotypic and transcriptional measurement of the most informative compounds, meaning the compounds with high rank, balanced between hits and non-hits. The reward is the prioritization of hit compounds, and the state is the paired phenotypic and transcriptional measurements. The agent takes the hit compounds for the phenotypic screen and the paired data to learn an updated signature that maximizes the prioritization of hit compounds. (**B**) A heatmap showing the differential expression of genes that play a role in Mk differentiation for each compound perturbation at 24 hours. Left, the observed change in Mk from the paired phenotypic assay. Below, the input v-score from the original prediction for each gene is shown. Asterisks indicate the significance of each gene's correlation with Mk fold change, as measured with a Pearson correlation test. (**C**) Box plot of gene signature changes induced by signature refinement for various gene classes: transcription factors that induce megakaryopoiesis (MK TF (+)), other transcription factors (TF (other)), megakaryocyte markers (MK (+)), and all other genes (other). Asterisks indicate significance relative to the "other" gene set based on Mann-Whitney U tests across 1,000 random 100-gene subsamples, with p-values averaged to assess robustness. (**D**) Result of experimental validation of compounds predicted by DrugReflector using the refined signature as input. Each dot is a compound. The light gray dashed line denotes a fold-change of 1 relative to DMSO (no change).The black dashed line indicates the hit-calling threshold (methods). (**E**) Comparison of hit compound DrugReflector ranks using the original signature vs. the refined signature as input. The refined signature ranked hits lower overall ($p<10^{-4}$, Wilcoxon signed rank test), indicating stronger prioritization. (**F**) Hit-rates for the original signature (blue line) vs. the refined signature (orange line) at various rank thresholds. For each rank threshold, we record the proportion of hits among all tested compounds ranked lower than that threshold. For comparison, the hit-rate from our random baseline is also indicated (gray dashed line; 87 randomly-chosen compounds).

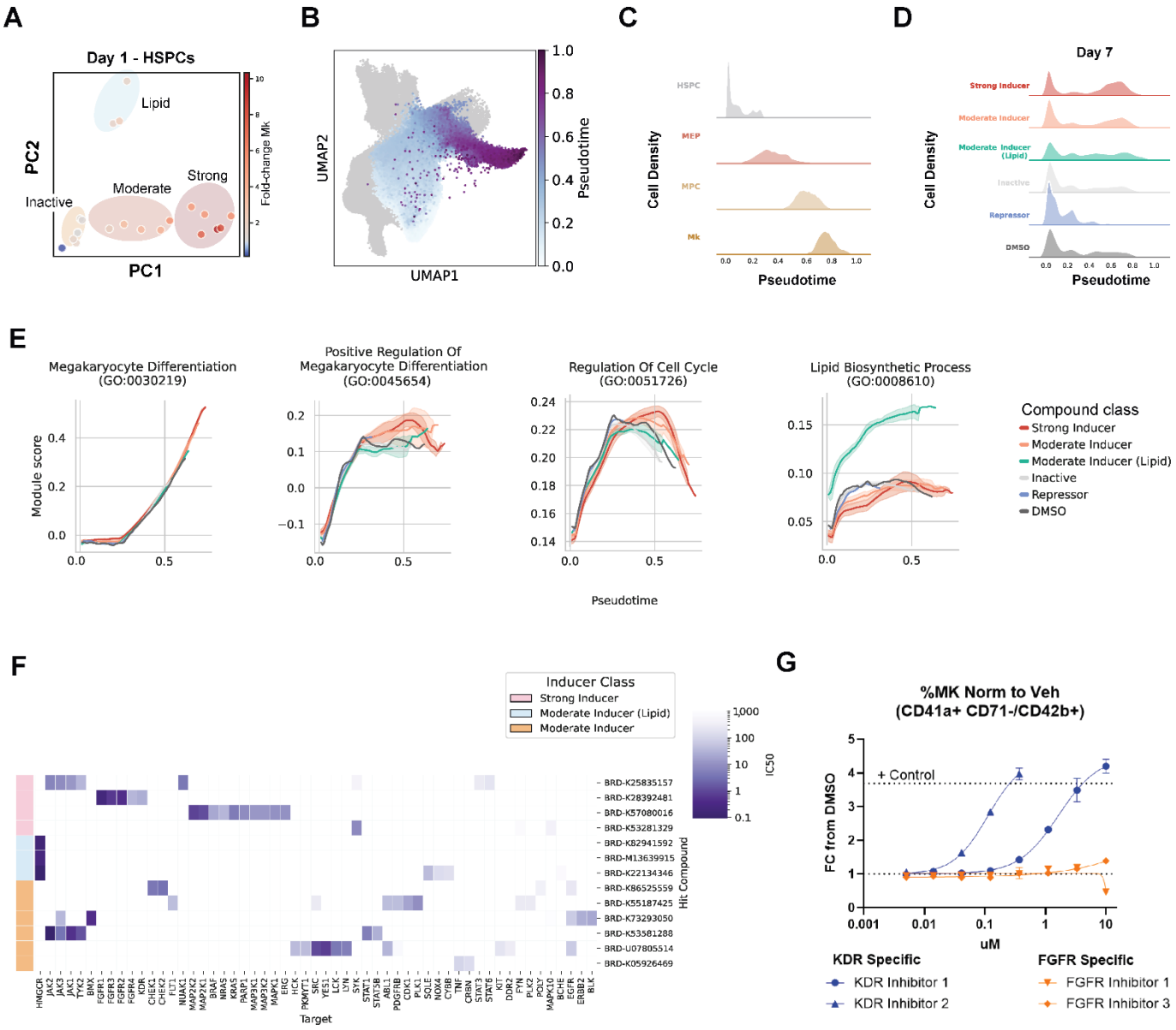


Fig. 5. Unraveling the mechanisms of chemically-induced megakaryopoiesis. (**A**) Principal component analysis of DES signatures derived from predicted Mk-inducers relative to DMSO in HSPCs at 24 hours. Compounds in PC space are colored by fold change in Mk relative to DMSO. (**B**) UMAP-embedding of diffusion pseudotime calculated across HSPC, MEP, MPC, and Mk. (**C**) Ridge plot depicting density of cells per type at various stages of pseudotime. (**D**) The density of cells at Day 7 per compound class. (**E**) Rolling-window average expression of genes associated with GO Biological Processes per compound class across pseudotime. Error bands show the standard deviation of expression across cells treated with compounds in the same class. (**F**) Target inhibition IC50s for megakaryopoiesis inducers from ChEMBL. Hit compounds are grouped by inducer class as annotated in panel (A). Heat map color indicates IC50 on a log-scale. (**G**) Dose response curves for compounds with selective binding affinities to KDR (blue) and FGFR (yellow).