

Drug Repurposing Using Deep Embeddings of Gene Expression Profiles

Yoni Donner,^{*,†} Stéphane Kazmierczak,^{*,‡} and Kristen Fortney^{*,‡}

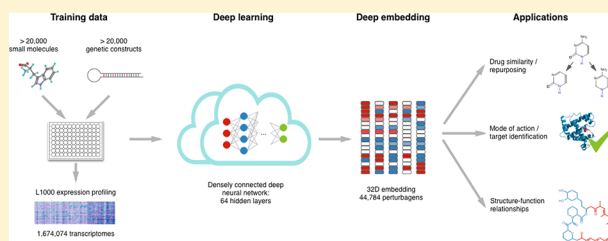
[†]Quantified Mind

[‡]BioAge Labs, Richmond, California 94804, United States

S Supporting Information

ABSTRACT: Computational drug repositioning requires assessment of the functional similarities among compounds. Here, we report a new method for measuring compound functional similarity based on gene expression data. This approach takes advantage of deep neural networks to learn an embedding that substantially denoises expression data, making replicates of the same compound more similar. Our method uses unlabeled data in the sense that it only requires compounds to be labeled by identity rather than detailed pharmacological information, which is often unavailable and costly to obtain. Similarity in the learned embedding space accurately predicted pharmacological similarities despite the lack of any such labels during training and achieved substantially improved performance in comparison with previous similarity measures applied to gene expression measurements. Our method could identify drugs with shared therapeutic and biological targets even when the compounds were structurally dissimilar, thereby revealing previously unreported functional relationships between compounds. Thus, our approach provides an improved engine for drug repurposing based on expression data, which we have made available through the online tool DeepCodex (<http://deepcodex.org>).

KEYWORDS: deep neural network, metric learning, L1000, drug embedding, drug repositioning



INTRODUCTION

In silico drug repositioning¹ relies on having a useful notion of compound similarity. For example, two compounds are more likely to have the same pharmacological mode of action if they are structurally similar, i.e., if they share many chemical substructures. Structural similarity has the advantage of being easily measurable for any pair of compounds. However, when used alone, it is of limited utility: compounds that share a mode of action can be chemically diverse,^{2,3} and small changes in structure can have dramatic effects on function.^{4,5} Other notions of compound similarity that have proven useful in computational repurposing efforts include side-effect similarity and target similarity, but this information is available for only a small proportion of compounds. One promising, relatively new strategy for identifying similar drugs involves the use of gene expression data.^{6–8} In this paradigm, compounds are considered similar if they impact cellular gene expression in the same way (we will refer to this as “functional similarity”). However, standard measures of similarity between expression profiles may only poorly predict pharmacological similarities between compounds. Indeed, even expression profiles corresponding to biological replicates can vary greatly.⁹

Deep learning¹⁰ has been increasingly used in recent years to learn complex functions, typically using large amounts of labeled data. Some examples of applications of deep learning to drug discovery include: representing small molecules using graph convolutions,^{11–13} sequence-to-sequence models,¹⁴ 3D

convolutions with wave transform,¹⁵ generating molecular structures with a defined set of parameters,^{16,17} and predicting drug–target interactions.¹⁸

In this study, we applied deep learning to develop a metric of compound functional similarity to inform in silico drug repositioning. We developed a method for embedding L1000 expression profiles into the unit hypersphere $\{x \in \mathbb{R}^n \mid \|x\|_2 = 1\}$ using a deep neural network. In this space, a similarity between -1 and 1 is obtained by the dot product of two embedded profiles. (In this paper, we use the term “embedding” as commonly used in machine learning to mean a mapping from some input space (here, the space of gene expression profiles) into a space where useful structure exists.)

Our network is trained on a large data set of cellular perturbations, with no labels other than the identities of the perturbagens applied to each sample. Large data sets of gene expression profiles of chemically or genetically perturbed cells are now available,^{19,20} facilitating handling of the problem of

Special Issue: Deep Learning for Drug Discovery and Biomarker Development

Received: March 16, 2018

Revised: June 25, 2018

Accepted: July 12, 2018

Published: July 12, 2018

functional similarity as a deep learning problem. Specifically, we used the LINCS data set,²¹ the largest public resource of gene expression data from drug-treated cells, which comprises more than 1.5 million expression profiles from more than 40,000 perturbagens (compounds or genetic manipulations). LINCS data were gathered using the L1000 platform (Genometry, <http://genometry.com/>), which measures the expression of only 978 genes, yet can capture most of the variance of whole-genome profiles at a much lower cost.^{22–24} Deep learning has previously been applied to the LINCS data to improve the imputation accuracy of whole-genome expression from L1000 profiles,²⁵ predict pharmacological properties of drugs,²⁶ compute L1000 signatures,²⁷ and map L1000 profiles to binary barcodes that improve prediction of compound structure and target information.⁹

The training loss function we used encourages high similarity between embeddings of profiles associated with the same perturbagen and low similarity between embeddings of profiles associated with different perturbagens. We predicted that the similarity in the trained embedding space would not only separate same-perturbagen pairs from different-perturbagen pairs but would also correspond to pharmacological similarity between different perturbagens. In other words, despite not using any labels related to pharmacological similarities between perturbagens, we expected to find that the embeddings of pharmacologically similar perturbagens would be closer in the trained embedding space than those of pharmacologically dissimilar perturbagens.

Our approach is similar to that of Filzen et al., who developed the aforementioned perturbation barcodes, in that both approaches use deep metric learning to embed L1000 expression profiles in a metric space based on a contrastive criterion. However, those authors created general-purpose representations of gene expression profiles, whereas we focused specifically on *in silico* drug repositioning specifically. So for example, while perturbation barcodes are trained to group samples by biological replicates, our embeddings are trained to group samples by perturbagen identity in order to capture functional similarities between drugs. Furthermore, the implementation of our method differs in several important ways, including the neural network architecture, the loss function, and the training procedure. In addition, we used the latest developments in deep learning, including self-normalizing networks²⁸ and dense connectivity,²⁹ to train much deeper networks. Our results demonstrate that our embeddings outperform the perturbation barcodes.

We validated our deep-learned metric of functional similarity in several independent ways. First, when perturbagens were held out during training, the metric successfully clustered profiles from the held-out set, both by perturbagen and by biological replicate. More importantly for drug repositioning applications, the metric was capable of capturing pharmacological similarity even among compound pairs with low structural similarity. Our learned-expression-based metric was capable of complementing structural similarity measures, and a simple approach combining the strengths of both approaches yielded further improvements in results.

We made both our trained neural network as well as our drug results available to researchers through the online tool DeepCodex (<http://deepcodex.org>).

MATERIALS AND METHODS

Data. To train the deep neural network, we used data from the NIH Library of Integrated Network-Based Cellular Signatures Program (LINCS, <http://www.lincsproject.org>). Phase I and II level 4 (z-scores) data for a total of 1,674,074 samples corresponding to 44,784 perturbagens were retrieved from GEO³⁰ (accession numbers GSE92742 and GSE70138). After removal of 106,499 control samples (as indicated by the `sm_pert_type` field), 1,567,575 samples corresponding to 44,693 perturbagens remained. Our analysis only included the 978 L1000 landmark genes and did not use imputations.

To evaluate the embedding performance, we used functional labels from several sources. Therapeutic target classifications for 1005 drugs were obtained from the Anatomical Therapeutic Chemical Classification System (ATC) provided by DrugCentral.³¹ There are 14, 79, 182, and 444 classes for ATC levels 1, 2, 3, and 4, respectively. Biological protein target annotations for 135 target classes and 411 drugs were obtained from ChEMBL.³²

To compute structural similarities, we used extended-connectivity fingerprints³³ (ECFPs) and Molecular ACCess System (MACCS) keys.³⁴ Tanimoto coefficients were used to compute similarity.³⁵

Deep Neural Network. The inputs to our deep neural network are vectors of standardized L1000 expression profiles. Because L1000 measures 978 landmark genes, the vectors are 978-dimensional. Standardization is performed per gene by subtracting the mean and dividing by the standard deviation. Means and variances are estimated over the entire training set.

The calculation of the training loss for a single training example is illustrated in Figure 1. Our embedding network is deep, self-normalizing,²⁸ and densely connected.²⁹ To our knowledge, the densely connected architecture has previously only been used in convolutional networks. Here we successfully use dense connections with no convolutions to train deep embedding networks and observed no performance degradation even when training networks with more than 100 layers. Our implementation is memory-efficient, similar to the memory-efficient implementation described for convolutional networks,³⁶ but simpler because of the lack of batch normalization. A final fully connected layer computes the unnormalized embedding, followed by an L_2 -normalization layer $\frac{1}{\|x\|_2}x$, where x is the unnormalized embedding, to project the embedding on the unit hypersphere.

The network is trained with a modified softmax cross-entropy loss over n classes, where classes are perturbagen identities. For ease of notation, the following describes the loss for a single sample. The loss for an entire batch is defined similarly. We use L_2 -normalized weights with no bias term to obtain the cosines of the angles between the embedding vector and the class weights: $\cos \theta_i = (W_i^T x) / (\|W_i\|_2 \|x\|_2)$, where i is the class index. The loss for a single training sample given $\theta = \{\cos \theta_i\}_{1 \leq i \leq n}$, where n is the number of classes (perturbagens), is

$$l(\theta) = -\log \frac{\exp(\alpha(\cos \theta_i - m))}{\exp(\alpha(\cos \theta_i - m)) + \sum_{j \neq i} \exp(\alpha \cos \theta_j)} \quad (1)$$

where i is the label (perturbagen identity), $\alpha > 0$ is a trainable scaling parameter, and $m \geq 0$ is a nontrainable margin parameter. During training, m is gradually increased from an

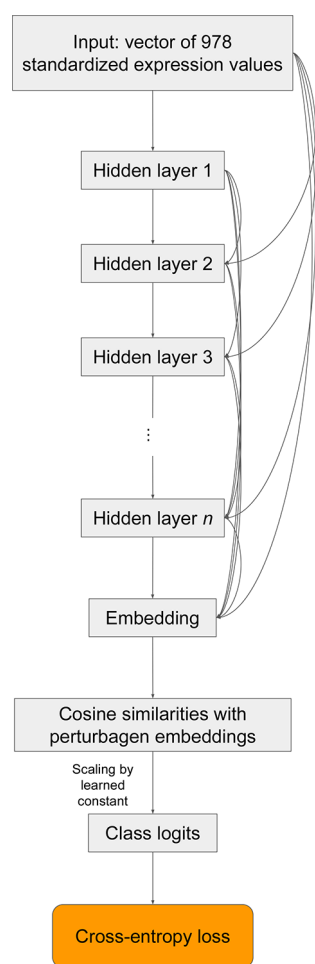


Figure 1. Embedding network architecture and training loss for a single expression profile. Inputs are standardized L1000 profiles and are processed by a densely connected neural network. The output embeddings are used to predict the class (perturbagen identity) of the input by softmax where logits are cosine similarities between the profile embeddings and learned class embeddings, scaled by a learned constant. The prediction cross-entropy loss is used to train the network.

initial value of 0, up to some maximum value. Inclusion of the margin forces the embeddings to be more discriminative and serves as a regularizer. A similar margin has been proposed for convolutional neural networks.³⁷

For the main experiments described here, we used 64 hidden layers, a growth rate of 16, and an embedding size of 32. We trained the network for 8000 steps (about 2 h using a single NVIDIA K80 GPU) with a batch size of 8192, adding Gaussian noise with a standard deviation of 0.3 to the input. The margin m was linearly increased at a rate of 0.0002 per step up to a maximum 0.25. These values were chosen without a hyperparameter search and are therefore unlikely to be optimal. However, as we show in the [Supporting Information](#), we found that predictive accuracy remained high over a wide range of values for all hyperparameters; therefore, we conclude that our method is not sensitive to exact values of hyperparameters. Nonetheless, it is likely that a full hyperparameter search would lead to even higher accuracy, and we leave this exploration to future work.

Computation of Perturbagen-Level Embeddings. In some of our evaluations, as described below, perturbagen-level

rather than sample-level embeddings were required. Because the input of our embedding neural network is a single L1000 expression profile, the output is an embedding of a single expression profile. To embed perturbagens, for which multiple profiles are available, many methods can be considered. For simplicity, to compute an embedding for a specific perturbagen, we simply average over all profile embeddings for the expression profiles of that perturbagen. There may be better ways than simple averaging to aggregate profile-level embeddings into perturbagen-level embeddings, but this simple approach provided good results, so we chose to leave exploration of other methods of aggregation for future work.

Performance Evaluation. Because the goal of the method is to rank perturbagens *in silico* based on functional similarity to a specific query perturbagen, we based our evaluation on similarity queries, defined as follows. A perturbagen-level similarity query is performed by computing the similarities between a single query perturbagen and many perturbagens from a set of candidates that includes both positives and negatives. Here, positives are candidate perturbagens belonging to the same functional group as the query perturbagen, where groups are defined based on multiple criteria as detailed below, and negatives are perturbagens that are not in the same group.

In our evaluations, we also use profile-level similarity queries, which differ from perturbagen-level similarity queries in that they use single expression profiles, rather than perturbagens, for both for the query and the candidates. We use two types of profile queries: same-perturbagen and biological-replicate. For same-perturbagen queries, the positives are expression profiles in which the same perturbagen was applied as in the query profile, and negatives are profiles in which a different perturbagen was applied. Biological-replicate queries are similar, but a profile is considered positive only if it is a biological replicate of the query profile, sharing the same perturbagen, cell line, dose, and time. When performance is optimal, all positives will be ranked above all negatives.

For each query, similarity scores are computed for all the candidates. Let $\{p_i\}_{1 \leq i \leq P}$ and $\{n_j\}_{1 \leq j \leq N}$ denote the scores for the positive and negative candidates, where P and N are the total numbers of positive and negative candidates for that specific query. For each $1 \leq i \leq P$, the rank quantile of the similarity score of the i th positive p_i within all negative scores is computed as follows $q_i = \frac{1}{N} \sum_{j=1}^N \frac{1}{2}(\text{sgn}(n_j - p_i) + 1)$, where $\text{sgn } x$ is the sign function.

We aggregate the quantiles q_i over all queries using the following metrics: the median rank as well as the top-0.001, top-0.01, and top-0.1 recall. Here we use “recall” to refer to the fraction of positives retrieved (i.e., the number of positives retrieved divided by the total number of positives). More precisely, the top- x recall $r(x)$, defined for $x \in [0, 1]$, means the fraction of positives ranked higher (lower quantile) than the x quantile $r(x) = \frac{1}{Q} \sum_{i=1}^Q \frac{1}{2}(\text{sgn}(x - q_i) + 1)$, where Q is the total number of positives over all queries, and the vector q of size Q is the concatenation of all the quantile vectors from the individual queries.

Throughout this paper, we refer to ranks as quantiles (relative rank, out of the total number of negatives). For example, rank 689 of 10,000 is the 0.0689 quantile. The reason for this is that the number of negatives changes between experiments, but assuming that the negatives are drawn independently from an identical distribution, the quantile of the positive is expected to remain the same. In this manner,

rank 689 of 10,000 is considered the same as rank 6890 of 100,000.

Results are visualized in two ways: (1) Recall by quantile. For each quantile $q \in [0, 1]$ (x -axis), the recall $r(x)$ is the y -axis value. As explained above, this recall is the fraction of queries in which the positive is ranked above q . For example, for $q = 0.05$, the x -axis value is 0.05, and the fraction of queries in which the positive is ranked in the top 5% is the y -axis value. (2) Receiver operating characteristic (ROC) curve, where the true positive rate (y -axis) is plotted against the false positive rate (x -axis). Because the ROC curves are generally very similar to the quantile-recall curves, we only show the quantile-recall curves in the main paper and include all of the corresponding ROC curves in the [Supporting Information](#).

Besides the visualizations, the results for each experiment are summarized in a table with the following summary statistics: median rank quantile; top-0.001 recall; top-0.01 recall; top-0.1 recall; area under the ROC Curve (AUC); and Matthews correlation coefficient (MCC). To calculate the MCC, we convert the continuous similarity scores to binary by thresholding at the median.

Experiments. We evaluated our method with two types of experiments. First, we evaluated the generalization ability of the trained networks using cross-validation at the perturbagen level. We refer to this form of evaluation as internal, because it only relies on the training labels and cross-validation, with no use of any external labels not available during training. Second, we evaluated the generalization ability of the similarity measure in the trained embedding space to identify functionally similar drugs. We refer to this as external evaluation, because it uses external labels not available during training.

Internal Evaluation. For the internal evaluation, we evaluated how well the trained embedding network clusters expression profiles from held-out perturbagens. For this purpose, the set of perturbagens was split into two groups, 80% training and 20% test. For each split, we trained an embedding network using only the training perturbagens and used it to embed all profiles corresponding to the held-out perturbagens. Then, each held-out profile was used as a query, with same-perturbagen held-out profiles as positives and different-perturbagen held-out profiles as negatives. Because the number of samples is very large (more than 1.5 million), evaluating the ranks using all samples as candidates requires considerable computational resources. Instead, we estimated the ranks using a random subset of 10,000 candidates. Ranks were aggregated over all queries and over all splits to training and test sets, and the aggregated distribution of ranks was used to calculate the summary statistics shown in the figures and tables.

As a second internal evaluation, we repeated the process described above but in this case only considering biological replicates (samples with the same perturbagen, cell line, dose and time) as positives, rather than all samples with the same perturbagen. Although the network was trained to separate samples by perturbagen, we expected it to also effectively separate samples by biological replicates.

External Evaluation. For the external evaluation, to evaluate the identification of functional similarity between perturbagens, we used sets of compounds that share some function, as described below. For each set and each compound within the set (referred to as the query compound), we ranked all candidate compounds by similarity to the query compound. Compounds within the same set as the query are positives, and

compounds outside the set are negatives. Results are summarized as described above for the internal evaluation. Sets are defined by shared ATC group (https://www.whocc.no/atc/structure_and_principles/) and shared protein target as labeled in ChEMBL.³² Separates sets are defined for each ATC level. For this evaluation, we also computed Cohen's kappa for multiclass classification restricted to the five largest sets.

We also evaluated the methods by their ability to identify structurally similar compounds. For this evaluation, we did not use fixed sets; instead, for each query compound, the positives and negatives were determined by the structural similarity of all other compounds to the query compound. Structural similarity is defined as the Tanimoto coefficient for ECFPs.³³ Specifically, a Tanimoto coefficient above 0.6 is considered positive, and a Tanimoto coefficient below 0.1 is considered negative. Analogous results using MACCS³⁴ are reported in the [Supporting Information](#).

Direct Training vs Holdout Embeddings. Because this is an external evaluation, based on labels not available during training, cross-validation is not necessary. Instead, we trained the embedding network once, on the full data set, and used this network to embed all the expression profiles. We computed the perturbagen embeddings from the individual profile embeddings by averaging as explained above. We refer to this method as “direct training”, because embeddings were computed directly on the same data set that the network was trained on. Although the direct training method does not result in overestimated estimates of performance, because performance is evaluated using labels not available during training, computing the embeddings by applying the trained network on the training set directly may not be ideal for other reasons (see [Supporting Information](#) for a more thorough discussion). We have also experimented with a different approach, based on cross-validation, where multiple networks are trained on subsets of perturbagens, and similarities are computed on the held-out perturbagens. We refer to this second approach as “holdout”. The holdout approach generates unbiased similarities but requires much greater computational resources because of the need to train many networks on different subsets of the full data. Because the results obtained using both approaches are similar (see [Supporting Information](#)), we used the direct training approach in the main experiments, because of its lower complexity, in terms of both computation and implementation.

Baselines. We compared the results for our embeddings with two baselines: the 978 landmark gene z -scores (“ z -scores”) and perturbation barcodes⁹ (“barcode”). For the z -scores, we found that cosine distance performed better than Euclidean distance, so we used the cosine distance for results reported for z -scores. For completeness, and to facilitate comparisons with the experiments described in ref 9, we also provide results for z -scores using Euclidean distance in the [Supporting Information](#). For perturbation barcodes, we used Euclidean distance, because this is the metric suggested in the original paper.⁹ L1000 profiles were encoded as barcodes using the code and model available at <https://github.com/matudor/siamese>.

To compare our embeddings with structural similarity, we used ECFPs and MACCS keys with the Tanimoto coefficient measure of similarity.³⁵

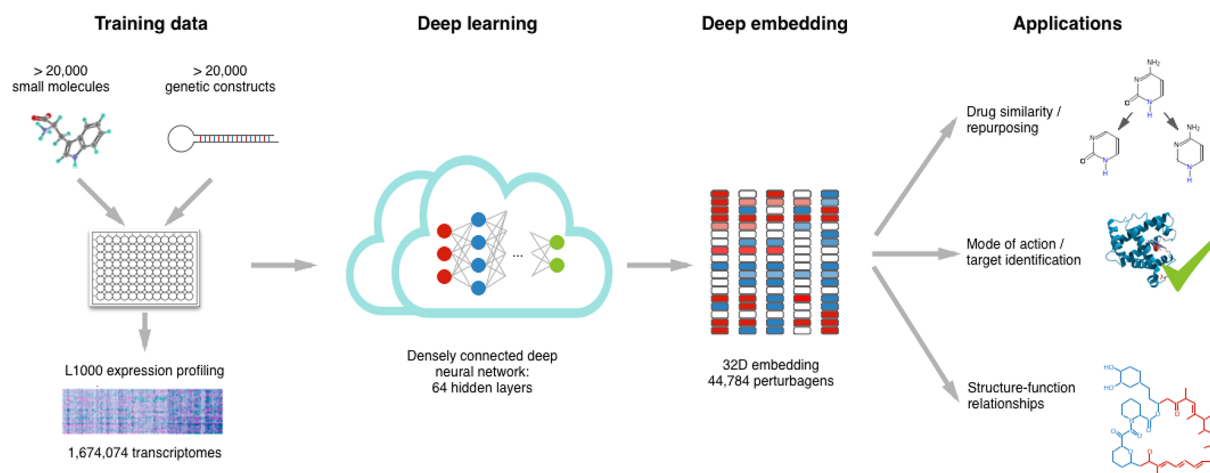


Figure 2. High-level overview of our approach.

RESULTS

Deep Embedding Decreases Noise in Gene Expression Data. We trained a deep neural network to embed L1000 expression profiles in a metric space where similarity, calculated as the dot product between embeddings, clusters expression profiles by perturbagen (Figure 2). Our internal evaluation is based on cross-validation: We train the network on 80% of the perturbagens, leaving out the remaining 20%. For each profile in the held-out subset, we compute its similarity to all other profiles from the same perturbagen (positives) as well as to profiles from other perturbagens (negatives) that are also part of the held-out subset. This way, the query is entirely contained within the held-out subset, with no overlap with the training set. The ranks of each positive among the negatives were computed, and the distribution of ranks was used for evaluation (Materials and Methods). Perfect performance is obtained in the situation in which all positives are ranked above all negatives.

We evaluated performance separately for two disjoint groups of perturbagens, the first comprising small molecules and the second comprising direct manipulations of target gene expression through CRISPR or expression of shRNA or cDNA. In addition, we evaluated performance over all perturbagens. For this experiment, we split the set of perturbagens 41 times into two subsets: training (80% of all perturbagens) and test (20%) and aggregated the distribution of ranks over all test sets.

For all perturbagen groups, the embeddings performed much better than the baselines (Figure 3a and Table 1a; ROC curves in Supporting Figure S1; Euclidean distance in Supporting Figures S5 and S8). Performance on genetic manipulations is slightly better than on small molecules, but even when the evaluation was restricted to small molecules, the embeddings ranked the positive within the top 1% in almost half (48%) of the queries, and the AUC was above 0.9. There was very little variance in AUC and MCC values between different random splits to training and test perturbagens (95% CI for small molecules, genetic manipulations, and all perturbagens, respectively: AUC [0.900, 0.916], [0.934, 0.938], [0.899, 0.913]; MCC [0.639, 0.677], [0.729, 0.738], [0.637, 0.672]). These results suggest that the embeddings effectively represent the effects of the perturbagen with some invariance to other sources of variation, including the variance between biological replicates as well as between cell lines,

doses, post-treatment measurement delays, and other factors. Furthermore, because the evaluation only included perturbagens entirely held out during training, these results demonstrate that the embeddings generalized well to perturbagens not included in the training set.

We also performed another evaluation, on the basis of biological replicates (i.e., profiles sharing the same perturbagen, cell line, dose, and time) rather than perturbagen identity alone. In this evaluation, positives were profiles that were biological replicates of the query profile, and negatives were selected from among the profiles that were not. Identifying biological replicates should be easier than identifying same-perturbagen profiles that vary in terms of cell line or dose. As expected, all methods performed better on this evaluation, but the embeddings still performed much better than the baselines (Figure 3b and Table 1b; ROC curves in Supporting Figure S2; Euclidean distance in Supporting Figures S7 and S8). The between-split variance in AUC and MCC values was very low (95% CI for small molecules, genetic manipulations, and all perturbagens, respectively: AUC [0.965, 0.976], [0.945, 0.947], [0.963, 0.974]; MCC [0.822, 0.861], [0.756, 0.761], [0.812, 0.849]). The difference in performance between the biological-replicate queries and the same-perturbagen queries was larger for the baselines than the embeddings. This may also reflect a level of invariance to sources of variation other than the perturbagen in the learned embedding space.

Deep Embedding Recovers Functionally Similar Sets of Compounds. As an external validation of our results, we tested whether our deep embeddings could improve the ability of gene expression data to identify pairs of similar compounds. For this purpose, we used three different criteria to build test queries for similar compounds (see Materials and Methods for more details): (1) High structural similarity, defined by the Tanimoto coefficient T_c : for MACCS, $T_c > 0.9$ are positives, and $T_c < 0.5$ are negatives, and for ECFPs, $T_c > 0.6$ are positives, and $T_c < 0.1$ are negatives. (2) Shared protein target, defined by ChEMBL. (3) Shared pharmacological or therapeutic properties, defined by ATC.

There are four relevant ATC levels, ranging from the most general (level 1, main anatomical group) to the most specific (level 4, chemical/therapeutic/pharmacological subgroup). In the main text, we provide results for ATC level 2 (therapeutic subgroup), because it adds more new information to the evaluation on top of the biological protein targets defined by

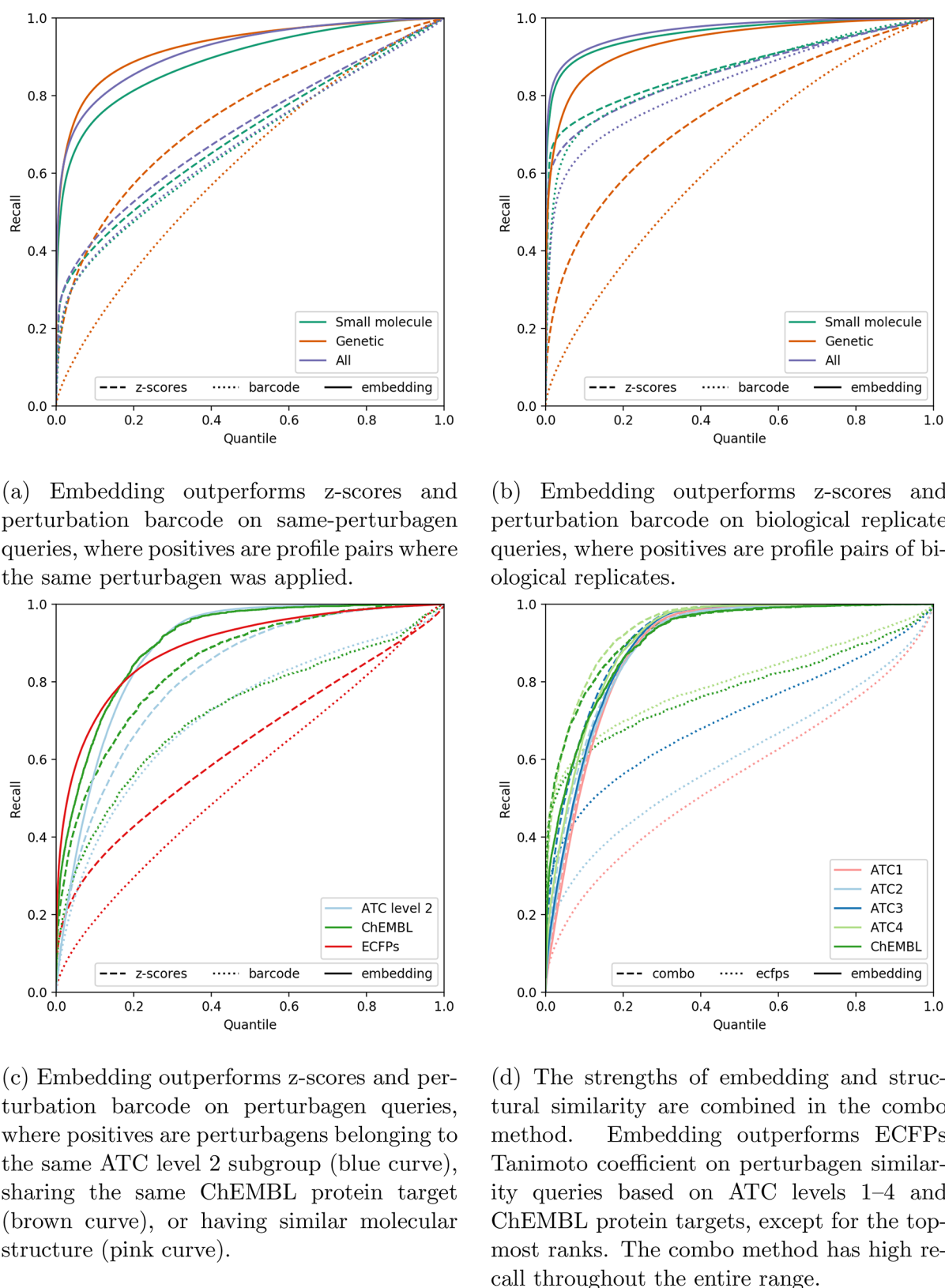


Figure 3. Deep embedding outperforms baselines on four tasks of ranking profiles (top row)/perturbagens (bottom row) by similarity to a query profile/perturbagen. The curves show, for each rank quantile q (x -axis), the fraction of positives ranked better than q (y -axis).

ChEMBL, whereas in the [Supporting Information](#), we include results for all levels. Additionally, MACCS and ECFPs are similar, and the queries they define have significant overlap. In the main text, we present results for ECFPs; the results for

MACCS are very similar and are included in the [Supporting Information](#).

Our embeddings performed better than the baselines on all query types (Figure 3c and Table 2a; full results including

Table 1. Embedding and Baseline Performance on Ranking Expression Profiles by Similarity to a Query Profile^a

perturbagens ^(b)	method	median	top-0.001	top-0.01	top-0.1	AUC	MCC
small molecules	z-scores	0.1960	0.051	0.265	0.412	0.682	0.235
	z-scores-euc	0.4932	0.033	0.091	0.188	0.513	0.012
	barcode	0.2316	0.039	0.174	0.383	0.679	0.252
	embedding	0.0120	0.295	0.480	0.738	0.904	0.649
genetic	z-scores	0.1415	0.074	0.169	0.436	0.736	0.353
	z-scores-euc	0.4057	0.027	0.068	0.216	0.481	−0.038
	barcode	0.3328	0.010	0.041	0.209	0.607	0.154
	embedding	0.0064	0.335	0.548	0.821	0.936	0.734
all	z-scores	0.1698	0.064	0.265	0.430	0.696	0.262
	z-scores-euc	0.5242	0.036	0.086	0.174	0.517	0.016
	barcode	0.2226	0.050	0.189	0.387	0.673	0.248
	embedding	0.0056	0.360	0.557	0.782	0.902	0.648
perturbagens ^(c)	method	median	top-0.001	top-0.01	top-0.1	AUC	MCC
small molecules	z-scores	0.0062	0.103	0.627	0.746	0.855	0.537
	z-scores-euc	0.4405	0.078	0.215	0.318	0.521	0.029
	barcode	0.0184	0.089	0.381	0.714	0.883	0.614
	embedding	0.0005	0.577	0.745	0.903	0.968	0.833
genetic	z-scores	0.1315	0.078	0.180	0.451	0.742	0.361
	z-scores-euc	0.4076	0.030	0.073	0.223	0.477	−0.045
	barcode	0.3126	0.012	0.047	0.226	0.620	0.170
	embedding	0.0051	0.357	0.571	0.845	0.946	0.758
all	z-scores	0.0045	0.166	0.590	0.716	0.843	0.516
	z-scores-euc	0.4367	0.097	0.199	0.302	0.525	0.023
	barcode	0.0229	0.104	0.377	0.657	0.854	0.552
	embedding	0.0001	0.629	0.786	0.916	0.965	0.823

^aFor z-scores, results using Euclidean distance (z-scores-euc) are reported for completeness. ^bEmbedding outperforms baselines on profile queries, where positives are profile pairs where the same perturbagen was applied. ^cEmbedding outperforms baselines on profile queries where positives are pairs of biological replicates.

ATC levels 1–4 and MACCS shown in Supporting Figure S9 and Supporting Table S1; ROC curves in Supporting Figures S3 and S10). The gap between the embeddings and baselines was largest for queries of structural similarity. Structurally similar compounds (the positives for each query) tend to have correlated expression profiles, but the correlations are weak.² Indeed, the raw z-scores performed barely better than chance on structural similarity queries, whereas the embeddings perform much better. One possible explanation for this result is that the embedding is trained to cluster together profiles corresponding to the same compound, which is equivalent to the identity of the chemical structure. The greater similarities in embedding space between structurally similar compounds relative to structurally dissimilar compounds demonstrates good generalization of the training objective.

Given that the embeddings capture some structural information, in the next section we explore the relationship between structural similarity and the similarity in embedding space.

To explore how performance varied across compound classes, we computed the confusion matrix for the five ATC level 2 subgroups containing at least 50 drugs. Performance varied greatly across classes (Table 3), ranging from 32.7 to 68.5% (chance performance being 20%). Because ATC level 2 subgroups are therapeutic subgroups, this variation may be due to large differences in mechanism of action between drugs within the same subgroup. For this reason, we also computed the confusion matrices for ATC level 4 (chemical subgroup) and ChEMBL protein targets, both of which are likely to be more homogeneous, but comprise smaller subgroups, making the accuracy estimates noisier. We set the size threshold to 15,

leaving three subgroups in both cases. In both cases, accuracy was very high for the more specific subgroups and moderately high for less specific subgroups (Table 4).

We also explored how performance varies across cell lines, incubation times, and doses. Of the seven most common cell lines in the data set, performance in human kidney epithelial immortalized cells (HA1E) was much worse than the rest, while the differences between the other six cell lines were small (Tables S4 and S5). To analyze performance by incubation time, we included the two time points that had sufficient data: 6 h and 24 h. Performance was slightly better overall at 24 h compared to 6 h, but the differences were very small (Table S2). We observed much larger differences as a function of dose. We binned doses to three groups to provide sufficient data for comparison: less than 1 μ M, between 1 and 10 μ M, and at least 10 μ M. Performance was best in the middle group and worst in the lowest dose group (Table S3).

Gene Expression Embeddings Complement Structural Similarity. We investigated how deep gene expression embeddings can complement structural similarity in functional similarity queries. Our method aims to improve recall for drug repositioning, i.e., increase the number of hits. As such, it is meant to complement, rather than replace, existing methods. In particular, structural similarity is readily available for all compounds and is commonly used for functional similarity queries.³⁵ Although high structural similarity usually indicates common pharmacological activity, the converse is not true. For example, Maggiora et al.³⁵ found that among groups of compounds with the same activity, the median MACCS Tanimoto similarity within a group varied from ~0.3 to ~0.75.

Table 2. Performance on Ranking Drugs by Similarity to a Query Drug^a

query type ^(b)	method	median	top-0.001	top-0.01	top-0.1	AUC	MCC	κ (five-class)
ATC level 2	z-scores	0.1142	0.022	0.103	0.466	0.785	0.412	0.268
	barcode	0.1712	0.018	0.083	0.376	0.662	0.208	0.120
	embedding	0.0824	0.013	0.096	0.570	0.882	0.563	0.351
ChEMBL target	z-scores	0.0733	0.101	0.218	0.558	0.826	0.438	0.484
	barcode	0.1555	0.069	0.153	0.411	0.680	0.215	0.211
	embedding	0.0513	0.090	0.288	0.658	0.906	0.552	0.620
structure (ECFPs)	z-scores	0.2888	0.073	0.149	0.327	0.530	0.025	
	barcode	0.4196	0.012	0.044	0.185	0.443	−0.100	
	embedding	0.0296	0.181	0.360	0.700	0.940	0.753	0.940
sets ^(c)	method	median	top-0.001	top-0.01	top-0.1	AUC	MCC	
ATC level 1	combo	0.0809	0.024	0.108	0.579	0.891		0.564
	embedding	0.0857	0.014	0.089	0.560	0.889		0.569
	ECFPs	0.3936	0.028	0.078	0.249	0.558		0.059
ATC level 2	combo	0.0651	0.056	0.173	0.629	0.896		0.562
	embedding	0.0773	0.022	0.121	0.590	0.889		0.565
	ECFPs	0.3071	0.081	0.159	0.326	0.607		0.110
ATC level 3	combo	0.0503	0.107	0.248	0.683	0.913		0.573
	embedding	0.0746	0.033	0.144	0.605	0.896		0.573
	ECFPs	0.1250	0.164	0.284	0.472	0.710		0.227
ATC level 4	combo	0.0210	0.209	0.404	0.784	0.935		0.570
	embedding	0.0604	0.049	0.197	0.665	0.908		0.573
	ECFPs	0.0204	0.316	0.460	0.627	0.801		0.328
ChEMBL target	combo	0.0175	0.210	0.451	0.766	0.928		0.548
	embedding	0.0459	0.106	0.317	0.677	0.913		0.556
	ECFPs	0.0256	0.270	0.444	0.605	0.765		0.276

^aSimilarity is defined by shared ATC subgroups, shared ChEMBL protein targets, or similar molecular structure as defined by high ECFPs Tanimoto coefficient. ^bEmbedding outperforms z-scores and perturbation barcodes on ranking queries with positives defined by the shared ATC level 2 subgroup, shared ChEMBL protein target, or high ECFPs Tanimoto coefficient. Cohen's κ was computed for the five largest classes and could not be computed for the ECFP-based queries, because it is not based on a fixed class assignment (Materials and Methods). ^cThe combo method combines the strength of structural similarity and deep embedding. While embedding tends to perform better than ECFPs on top-0.1 accuracy, AUC, and MCC for all subgroups, ECFPs have a lower median rank and higher top-0.001 and top-0.01 accuracy on the two most specific groupings: ATC level 4 and ChEMBL, because ECFPs have very high recall at the topmost ranks, especially for highly specific subgroups. The combo method combines the high top-0.1 accuracy, AUC, and MCC achieved by deep embedding and the low median quantile, high top-0.001 accuracy, and top-0.01 accuracy achieved by ECFPs on the most specific subgroups. Its performance is very good across all of the metrics and all of the groups.

Table 3. Confusion Matrix for Classifying Drugs to One of the Five Largest ATC Level 2 Subgroups^a

#	subgroup	<i>n</i>	1	2	3	4	5	accuracy
1	antibacterials for systemic use	72	33	0	24	10	5	45.8%
2	antineoplastic agents	73	0	50	14	4	5	68.5%
3	ophthalmologicals	86	16	4	37	24	5	43.0%
4	psychoanaleptics	55	6	1	12	18	18	32.7%
5	psycholeptics	53	6	1	8	11	27	50.9%

^aColumn titles for the classes are the numbers matching the rows.

Using our own trained embeddings, we ranked all compounds by similarity to metformin and explored the 10 compounds most similar to metformin in the embedding space. There are no compounds in the data set with high structural similarity to metformin (the highest ECFPs is 0.26), so it is not surprising that all 10 compounds are structurally different from it (Table 5). The top hit, allantoin, is ranked relatively high by structural similarity ($T_c = 0.095$, ranked 33rd out of more than 20,000), but a value below 0.1 would normally not be considered similar. At least four of our top five six hits were previously reported to have some functional similarity to metformin. For example, in a recent study that identified novel candidate metformin and rapamycin mimetics,³⁸ our top hit, allantoin, exhibited the highest similarity to metformin. A recent study³⁹ showed that allantoin acts as a

caloric restriction mimetic, similar to metformin. Metformin and allantoin appear to work through similar mechanisms.⁴⁰ Metformin and sulindac (ranked second) have anticancer properties that may be related to altering cellular responses to oxidative stress,^{41,42} and both drugs have similar abilities to suppress the formation of colorectal aberrant crypt foci (ACF).⁴³ Metformin and todrilazine (ranked fifth) are both hydroxyl scavengers,^{44,45} and both increase *Ins1* promoter activity in MIN6 cells.⁴⁶ Metformin, morantel (ranked sixth), and metrifonate (ranked ninth) inhibit AChE.^{47–49} Like homatropine (ranked eighth), metformin may have anticholinergic activity at high concentrations.⁵⁰ For the remaining four compounds, we did not find a known shared mechanism with metformin.

Table 4. Confusion Matrices for Classifying Drugs to One of the Three Largest ATC Level 4 Subgroups and to One of the Three Largest ChEMBL Protein Target Subgroups^a

source	#	subgroup	<i>n</i>	1	2	3	accuracy
ATC level 4	1	anti-inflammatory preparations, nonsteroids for topical use	21	20	1	0	95.2%
	2	other antineoplastic agents	18	4	12	2	66.7%
	3	protein kinase inhibitors	22	0	0	22	100%
ChEMBL target	1	bacterial penicillin-binding protein	15	11	2	2	73.3%
	2	DNA	22	6	14	2	63.6%
	3	glucocorticoid receptors	19	1	0	18	94.7%

^aColumn titles for the classes are the numbers matching the rows.

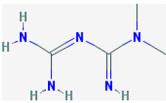
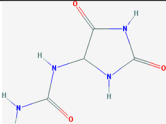
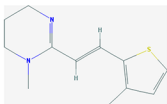
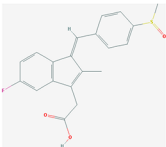
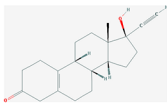
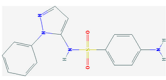
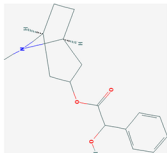
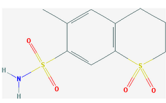

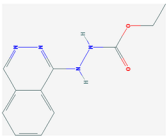
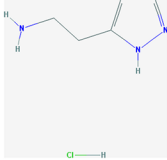
To determine to what extent our embeddings complement structural similarity, we first evaluated the ability of structural similarity alone to identify drugs with the same known protein target or ATC classification (Figure 3d, dotted lines; Table 1b;

ROC curves in Supporting Figure S4). Here, we present results using ECFPs; the performance of the MACCS keys was worse than that of ECFPs, and those results are provided in the Supporting Information. Structural similarity performed best on the lowest level sets, ATC level 4 (chemical/therapeutic/pharmacological subgroup) and ChEMBL protein targets, but its performance degraded rapidly when moving to more general ATC classifications, and for ATC levels 1 and 2 (anatomical/therapeutic subgroups), it was not much better than chance.

Even for ATC level 4 and ChEMBL targets, where structural similarity performed at its best, recall was excellent at the top ranks (top-0.001 recall of 0.316 for ATC level 4 and 0.27 for ChEMBL targets) but increased very slowly below the top 0.1 ranks. The distribution of ranks below the top 0.1 was very close to chance (Supporting Figure S13). The same phenomenon was observed using MACCS (Supporting Figure S14).

In comparison with structural similarity, recall using our trained embeddings was not as high at the top ranks (Figure 3d, solid lines). However, with the exception of the top ranks, the embeddings performed much better and addressed both of the structural similarity shortcomings described above. First,

Table 5. Top 10 Compounds, Ranked by Similarity to Metformin in Embedding Space^a

#	Compound	<i>T_c</i>	Structure	#	Compound	<i>T_c</i>	Structure
	Metformin (biguanide hypoglycemic agent)	1.0					
1	Allantoin (urea hydantoin)	0.095		6	Morantel (antinematodal agent)	0.032	
2	Sulindac (nonsteroidal anti-inflammatory agent)	0.021		7	Noretynodrel (synthetic progestational hormone)	0.011	
3	Sulfaphenazole (sulfonamide anti-infective agent)	0.037		8	Homatropine (anticholinergic; antagonist at muscarinic receptors)	0.070	
4	Meticrane (sulphonamide-derivative; thiazide-like diuretic activity)	0.032		9	Metrifonate (organochloro-phosphate cholinesterase inhibitor)	0.026	
5	Todralazine (antihypertensive agent with both central and peripheral action)	0.054		10	Betazole (histamine H ₂ agonist)	0.027	

^a# indicates the rank. *T_c* is the Tanimoto coefficient using ECFPs. The compounds are chemically diverse, and all of them have relatively low structural similarity to metformin, demonstrating that our method generates suggestions for drug repurposing that would not be captured using structural-similarity-based methods. Brief compound descriptions taken from DrugBank⁵¹ and PubChem.⁵²

the performance degradation was very small at higher levels of classification, with good performance even at ATC level 1: median rank of 0.086 and AUC of 0.889 (Table 2b). Second, embedding recall kept increasing rapidly beyond the few top ranks, outperforming structural similarity starting at quantile 0.0647 for ChEMBL as well as quantiles 0.0867, 0.0544, 0.0202, and 0.0057 for ATC levels 4, 3, 2, and 1, respectively.

These results suggested that the strengths of structural similarity and gene expression embedding could be combined. Accordingly, we explored the simplest possible way of combining the two: a similarity measure that computes an unweighted average of embedding similarity and the ECFPs Tanimoto coefficient. We refer to this similarity measure as “combo”.

The combo method outperformed both structural similarity and the embeddings (Figure 3d, dashed lines; Table 2b; ROC curves in Supporting Figure S4; MACCS results in Supporting Figures S11 and S12 and Supporting Table S6). The improvement of the combo method is most evident in the top ranks, especially for ATC level 4 and ChEMBL targets, exactly where structural similarity performed at its best. The combo method median quantile and top- q accuracy for $q \in [0.001, 0.01, 0.1]$ are significantly better than those of the embeddings ($p < 10^{-1000}$, one-tailed paired-sample t test using bootstrapping with $n = 10,000$ samples), demonstrating that our deep embeddings complement structural similarity and that a combination of both is more effective than either by itself. We chose simple averaging for its simplicity, leaving to future work the exploration of more advanced methods of combining the strengths of structural similarity and gene expression embeddings.

DISCUSSION

Profiling the modes of action of small molecules using high-throughput technologies such as transcriptomics, proteomics, and metabolomics represents a powerful approach to accelerating drug discovery. One of the most promising applications for these large data sets is computational drug repositioning, the automated discovery of a compound's potential targets and disease indications. Accurate assessment of the functional similarities among compounds is essential for this purpose.

In this study, we applied techniques from deep learning to the largest public data set of chemically and genetically perturbed cellular transcriptomes to build a map of compound functional similarity. Our novel method significantly decreases noise in the measurement data and outperforms previous approaches in several ways. In particular, it is better at recovering replicates of the same compound as well as recovering drugs known to be similar in terms of their pharmacology, protein targets, or structure. Therefore, our method should provide a useful and general tool for prioritizing new compounds for specific disease indications or pharmacological behaviors.

In our method, similarity between gene expression profiles is computed using embeddings: a deep neural network embeds input gene expression profiles as vectors in the unit hypersphere, and the similarity between two profiles is defined as the dot product between their embeddings. To train the network, we used a definition of similarity that is highly specific (i.e., based only on perturbation identity) and binary: Two profiles are defined as similar if they measure gene expression after treatment with the same perturbation (small molecule,

ligand, or genetic perturbation), and otherwise, they are defined as dissimilar. This simple definition of similarity allows the network to be trained using large data sets such as LINCS.

Our trained network generalized well to expression profiles of cells treated by perturbagens not included during training. Using the same definition of similarity or a related definition based on biological replicates, our network successfully identified similar profiles within the held-out set. In addition, despite not being directly trained to do so, the network successfully identified biological replicates in the held-out data. More importantly, despite the narrow and binary definition of similarity used in training, we showed that the learned embedding similarity generalizes to additional definitions of similarity based on structure and function. Similarity in embedding space effectively identifies compounds with similar chemical structures, shared protein targets, and shared therapeutic groups.

From the standpoint of optimization, we have identified several opportunities for improvement of this method. For example, in this study, we used only the L1000 expression profiles and the identity of the perturbation in each sample. However, the LINCS data set also contains additional metadata, including cell line, dose, and time point. The use of these metadata could lead to better performance, but such experiments are outside the scope of the current study.

In addition, we used simple averaging to calculate perturbation-level embeddings from profile-level embeddings. However, it is possible that averaging of all profile embeddings corresponding to a perturbation is not the best way to calculate the embedding for that perturbation. One possible shortcoming of the averaging approach is that in some profiles, the treatment may have no effect, e.g., because the dose is too low. If such profiles could be identified and removed prior to calculating the average, perturbation embeddings could better represent their functions. This possibility should be investigated in future work.

Another issue that we have not resolved here is how to best combine embedding similarity with structural similarity. As we have shown, the two approaches are largely complementary and can be easily combined. Even a method as simple as averaging the ECFPs Tanimoto coefficient with the embedding similarity improves on either method alone, and less trivial methods of combining the two are likely to work even better. For example, it is possible to estimate the probability of two compounds sharing a function (such as protein target) based on their structural and embedding similarities. If the probabilities were estimated using a training set that does not include the query compound of interest, the candidates could be sorted according to a model that combines the estimated probabilities, such as Naïve Bayes.

In terms of deep learning, the main contribution of this study is a novel architecture that combines several recent developments, including self-normalizing activations, dense connections, and softmax margin. Using this architecture, we trained neural networks that are much deeper than those previously used to analyze gene expression data.^{25,26,53} For example, in image recognition tasks, the best networks often include hundreds of layers, and performance improves with the number of layers.⁵⁴ The network used in our experiments includes 64 hidden layers, but when we compared networks with different depths and approximately the same number of parameters, we found very little, if any, improvement beyond four hidden layers. Why is there a large gap in accuracy

between very deep and relatively shallow networks in computer vision but not in the task we studied here? One possible explanation is that image recognition requires deep hierarchical representations, because objects are made of parts that are themselves objects, and deep networks are capable of learning robust representations at each level of the hierarchy. By contrast, if the structure of gene expression profiles is less hierarchical than that of images, it would be adequately represented with far fewer layers than are required in computer vision.

■ ASSOCIATED CONTENT

■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.molpharmaceut.8b00284.

Neural network performance remains high over a wide range of hyperparameter values; ROC curves corresponding to the quantile-recall curves in the main text; direct training vs holdout embeddings; additional results: using MACCS instead of ECFPs, using Euclidean distance for z-scores, and using all ATC levels; distribution of ranks for queries using structural similarity (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: yonidonner@gmail.com (Y.D.)

*E-mail: stephane@bioagelabs.com (S.K.)

*E-mail: kristen@bioagelabs.com (K.F.)

ORCID

Yoni Donner: 0000-0002-2940-4215

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Carlos De la Guardia, Chris Patil, Kia Winslow, and C.J. Winslow. This work was supported by BioAge Labs (YD, SK, KF). YD consulted for BioAge Labs for this project.

■ REFERENCES

- (1) Vanhaelen, Q.; Mamoshina, P.; Aliper, A. M.; Artemov, A.; Lezhnina, K.; Ozerov, I.; Labat, I.; Zhavoronkov, A. Design of efficient computational workflows for in silico drug repurposing. *Drug Discovery Today* **2017**, *22*, 210–222.
- (2) Chen, B.; Greenside, P.; Paik, H.; Sirota, M.; Hadley, D.; Butte, A. J. Relating Chemical Structure to Cellular Response: An Integrative Analysis of Gene Expression, Bioactivity, and Structural Data Across 11,000 Compounds. *CPT: Pharmacometrics Syst. Pharmacol.* **2015**, *4*, 576–584.
- (3) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do Structurally Similar Molecules Have Similar Biological Activity? *J. Med. Chem.* **2002**, *45*, 4350–4358.
- (4) Eckert, H.; Bajorath, J. Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discovery Today* **2007**, *12*, 225–233.
- (5) Kubinyi, H. In *3D QSAR in Drug Design: Ligand-Protein Interactions and Molecular Similarity*; Kubinyi, H., Folkers, G., Martin, Y. C., Eds.; Springer Netherlands: Dordrecht, The Netherlands, 1998; pp 225–252.
- (6) Cheng, J.; Yang, L.; Kumar, V.; Agarwal, P. Systematic evaluation of connectivity map for disease indications. *Genome Med.* **2014**, *6*, 95.
- (7) Iorio, F.; Bosotti, R.; Scacheri, E.; Belcastro, V.; Mithbaokar, P.; Ferriero, R.; Murino, L.; Tagliaferri, R.; Brunetti-Pierri, N.; Isacchi, A.; di Bernardo, D. Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 14621–14626.
- (8) Lamb, J.; et al. The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. *Science* **2006**, *313*, 1929–1935.
- (9) Filzen, T. M.; Kutchukian, P. S.; Hermes, J. D.; Li, J.; Tudor, M. Representing high throughput expression profiles via perturbation barcodes reveals compound targets. *PLoS Comput. Biol.* **2017**, *13*, e1005335.
- (10) LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.
- (11) Duvenaud, D.; Maclaurin, D.; Aguilera-Iparraguirre, J.; Gómez-Bombarelli, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R. P. Convolutional Networks on Graphs for Learning Molecular Fingerprints. In *Proceedings of the 28th International Conference on Neural Information Processing Systems—Volume 2*, Montreal, Canada, December 7–12, 2015; Cortes, C., Lee, D. D., Sugiyama, M., Garnett, R., Eds.; 2015; pp 2224–2232.
- (12) Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular graph convolutions: moving beyond fingerprints. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 595–608.
- (13) Altae-Tran, H.; Ramsundar, B.; Pappu, A. S.; Pande, V. Low Data Drug Discovery with One-Shot Learning. *ACS Cent. Sci.* **2017**, *3*, 283–293.
- (14) Xu, Z.; Wang, S.; Zhu, F.; Huang, J. Seq2seq Fingerprint: An Unsupervised Deep Molecular Embedding for Drug Discovery. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, Boston, Massachusetts, USA, August 20–23, 2017; pp 285–294.
- (15) Kuzminykh, D.; Polykovskiy, D.; Kadurin, A.; Zhebrak, A.; Baskov, I.; Nikolenko, S.; Shayakhmetov, R.; Zhavoronkov, A. 3D Molecular Representations Based on the Wave Transform for Convolutional Neural Networks. *Mol. Pharmaceutics* **2018**; Article ASAP.
- (16) Segler, M. H. S.; Kogej, T.; Tyrchan, C.; Waller, M. P. Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. *ACS Cent. Sci.* **2018**, *4*, 120–131.
- (17) Kadurin, A.; Aliper, A.; Kazennov, A.; Mamoshina, P.; Vanhaelen, Q.; Khrabrov, K.; Zhavoronkov, A. The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget* **2017**, *8*, 10883–10890.
- (18) Wen, M.; Zhang, Z.; Niu, S.; Sha, H.; Yang, R.; Yun, Y.; Lu, H. Deep-Learning-Based Drug-Target Interaction Prediction. *J. Proteome Res.* **2017**, *16*, 1401–1409.
- (19) Lamb, J.; et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **2006**, *313*, 1929–1935.
- (20) Subramanian, A.; Narayan, R.; Corsello, S. M.; Peck, D. D.; Natoli, T. E.; Lu, X.; Gould, J.; Davis, J. F.; Tubelli, A. A.; Asiedu, J. K.; et al. A Next Generation Connectivity Map: L1000 Platform And The First 1,000,000 Profiles. *Cell* **2017**, *171*, 1437–1452.
- (21) Keenan, A. B.; Jenkins, S. L.; Jagodnik, K. M.; Koplev, S.; He, E.; Torre, D.; Wang, Z.; Dohlman, A. B.; Silverstein, M. C.; Lachmann, A. The Library of Integrated Network-Based Cellular Signatures NIH Program: System-Level Cataloging of Human Cells Response to Perturbations. *Cell Syst.* **2018**, *6*, 13–24.
- (22) Amit, I.; et al. Unbiased Reconstruction of a Mammalian Transcriptional Network Mediating Pathogen Responses. *Science* **2009**, *326*, 257–263.
- (23) Donner, Y.; Feng, T.; Benoist, C.; Koller, D. Imputing gene expression from selectively reduced probe sets. *Nat. Methods* **2012**, *9*, 1120–1125.
- (24) Duan, Q.; Flynn, C.; Niepel, M.; Hafner, M.; Muhlich, J. L.; Fernandez, N. F.; Rouillard, A. D.; Tan, C. M.; Chen, E. Y.; Golub, T. R.; et al. LINCS Canvas Browser: interactive web app to query,

browse and interrogate LINCS L1000 gene expression signatures. *Nucleic Acids Res.* **2014**, *42*, W449–W460.

(25) Chen, Y.; Li, Y.; Narayan, R.; Subramanian, A.; Xie, X. Gene expression inference with deep learning. *Bioinformatics* **2016**, *32*, 1832–1839.

(26) Aliper, A.; Plis, S.; Artemov, A.; Ulloa, A.; Mamoshina, P.; Zhavoronkov, A. Deep Learning Applications for Predicting Pharmacological Properties of Drugs and Drug Repurposing Using Transcriptomic Data. *Mol. Pharmaceutics* **2016**, *13*, 2524–2530.

(27) Duan, Q.; Reid, S. P.; Clark, N. R.; Wang, Z.; Fernandez, N. F.; Rouillard, A. D.; Readhead, B.; Tritsch, S. R.; Hodos, R.; Hafner, M.; et al. L1000CDS2: LINCS L1000 characteristic direction signatures search engine. *npj Syst Biol Appl* **2016**, *2*, 16015.

(28) Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S. Self-Normalizing Neural Networks. 2017, arXiv.org e-Print archive. <https://arxiv.org/abs/1706.02515>.

(29) Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K. Q. Densely Connected Convolutional Networks. 2017, arXiv.org e-Print archive. <https://arxiv.org/abs/1608.06993>.

(30) Barrett, T.; Wilhite, S. E.; Ledoux, P.; Evangelista, C.; Kim, I. F.; Tomashevsky, M.; Marshall, K. A.; Phillippy, K. H.; Sherman, P. M.; Holko, M.; et al. NCBI GEO: archive for functional genomics data sets-update. *Nucleic Acids Res.* **2012**, *41*, D991–D995.

(31) Ursu, O.; Holmes, J.; Knockel, J.; Bologa, C. G.; Yang, J. J.; Mathias, S. L.; Nelson, S. J.; Oprea, T. I. DrugCentral: online drug compendium. *Nucleic Acids Res.* **2017**, *45*, D932–D939.

(32) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, D1100–D1107.

(33) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754.

(34) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273–1280.

(35) Maggiora, G.; Vogt, M.; Stumpfe, D.; Bajorath, J. Molecular Similarity in Medicinal Chemistry. *J. Med. Chem.* **2014**, *57*, 3186–3204.

(36) Pleiss, G.; Chen, D.; Huang, G.; Li, T.; van der Maaten, L.; Weinberger, K. Q. Memory-Efficient Implementation of DenseNets. 2017, arXiv.org e-Print archive. <https://arxiv.org/abs/1707.06990>

(37) Liang, X.; Wang, X.; Lei, Z.; Liao, S.; Li, S. Soft-Margin Softmax for Deep Classification **2017**, 10635, 413–421.

(38) Aliper, A.; Jellen, L.; Cortese, F.; Artemov, A.; Karpinsky-Semper, D.; Moskalev, A.; Swick, A. G.; Zhavoronkov, A. Towards natural mimetics of metformin and rapamycin. *Aging* **2017**, *9*, 2245–2268.

(39) Calvert, S.; Tacutu, R.; Sharifi, S.; Teixeira, R.; Ghosh, P.; de Magalhães, J. P. A network pharmacology approach reveals new candidate caloric restriction mimetics in *C. elegans*. *Aging Cell* **2016**, *15*, 256–266.

(40) Lin, K. C.; Yeh, L. R.; Chen, L. J.; Wen, Y. J.; Cheng, K. C.; Cheng, J. T. Plasma glucose-lowering action of allantoin is induced by activation of imidazoline I-2 receptors in streptozotocin-induced diabetic rats. *Horm. Metab. Res.* **2012**, *44*, 41–46.

(41) Marchetti, M.; Resnick, L.; Gamliel, E.; Kesaraju, S.; Weissbach, H.; Binnering, D. Sulindac Enhances the Killing of Cancer Cells Exposed to Oxidative Stress. *PLoS One* **2009**, *4*, e5804.

(42) Sena, P.; Mancini, S.; Benincasa, M.; Mariani, F.; Palumbo, C.; Roncucci, L. Metformin Induces Apoptosis and Alters Cellular Responses to Oxidative Stress in Ht29 Colon Cancer Cells: Preliminary Findings. *Int. J. Mol. Sci.* **2018**, *19*, 1478.

(43) Higurashi, T.; Takahashi, H.; Endo, H.; Hosono, K.; Yamada, E.; Ohkubo, H.; Sakai, E.; Uchiyama, T.; Hata, Y.; Fujisawa, N.; et al. Metformin efficacy and safety for colorectal polyps: a double-blind randomized controlled trial. *BMC Cancer* **2012**, *12*, 118.

(44) Hajihashemi, S.; Geuns, J. M. Radical scavenging activity of steviol glycosides, steviol glucuronide, hydroxytyrosol, metformin,

aspirin and leaf extract of *Stevia rebaudiana*. *Free Radicals Antioxid.* **2013**, *3*, s34–s39.

(45) Dimri, M.; Joshi, J.; Chakrabarti, R.; Sehgal, N.; Sureshbabu, A.; Prem Kumar, I. Todralazine Protects Zebrafish from Lethal Effects of Ionizing Radiation: Role of Hematopoietic Cell Expansion. *Zebrafish* **2015**, *12*, 33–47. 25517940[pmid].

(46) Szabat, M.; Modi, H.; Ramracheya, R.; Girbinger, V.; Chan, F.; Lee, J. T. C.; Piske, M.; Kamal, S.; Carol Yang, Y. H.; Welling, A.; Rorsman, P.; Johnson, J. D. High-content screening identifies a role for Na⁺ channels in insulin production. *R. Soc. Open Sci.* **2015**, *2*, 150306.

(47) Markowicz-Piasecka, M.; Sikora, J.; Szydłowska, A.; Skupień, A.; Mikiciuk-Olasik, E.; Huttunen, K. M. Metformin - a Future Therapy for Neurodegenerative Diseases. *Pharm. Res.* **2017**, *34*, 2614–2627.

(48) Vardanyan, R.; Hruby, V. *Synthesis of Best-Seller Drugs*; Academic Press: Boston, MA, USA, 2016; pp 749–764.

(49) Cummings, J. L.; Ringman, J. M. Metrifonate (Trichlorfon): a review of the pharmacology, pharmacokinetics and clinical experience with a new acetylcholinesterase inhibitor for Alzheimer's disease. *Expert Opin. Invest. Drugs* **1999**, *8*, 463–471.

(50) Chew, M. L.; Mulsant, B. H.; Pollock, B. G.; Lehman, M. E.; Greenspan, A.; Mahmoud, R. A.; Kirshner, M. A.; Sorisio, D. A.; Bies, R. R.; Gharabawi, G. Anticholinergic Activity of 107 Medications Commonly Used by Older Adults. *J. Am. Geriatr. Soc.* **2008**, *56*, 1333–1341.

(51) Wishart, D. S.; et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082.

(52) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. PubChem Substance and Compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213.

(53) Putin, E.; Mamoshina, P.; Aliper, A.; Korzinkin, M.; Moskalev, A.; Kolosov, A.; Ostrovskiy, A.; Cantor, C.; Vijg, J.; Zhavoronkov, A. Deep biomarkers of human aging: Application of deep neural networks to biomarker development. *Aging* **2017**, *8*, 1021–1033.

(54) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) **2016**, 770–778.