



Predicting the Effects of Drug Combinations Using Probabilistic Matrix Factorization

Ron Nafshi and Timothy R. Lezon*

Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA, United States

Drug development is costly and time-consuming, and developing novel practical strategies for creating more effective treatments is imperative. One possible solution is to prescribe drugs in combination. Synergistic drug combinations could allow lower doses of each constituent drug, reducing adverse reactions and drug resistance. However, it is not feasible to sufficiently test every combination of drugs for a given illness to determine promising synergistic combinations. Since there is a finite amount of time and resources available for finding synergistic combinations, a model that can identify synergistic combinations from a limited subset of all available combinations could accelerate development of therapeutics. By applying recommender algorithms, such as the low-rank matrix completion algorithm Probabilistic Matrix Factorization (PMF), it may be possible to identify synergistic combinations from partial information of the drug interactions. Here, we use PMF to predict the efficacy of two-drug combinations using the NCI ALMANAC, a robust collection of pairwise drug combinations of 104 FDA-approved anticancer drugs against 60 common cancer cell lines. We find that PMF is able to predict drug combination efficacy with high accuracy from a limited set of combinations and is robust to changes in the individual training data. Moreover, we propose a new PMF-guided experimental design to detect all synergistic combinations without testing every combination.

OPEN ACCESS

Edited by:

Rui Li,
Pfizer, United States

Reviewed by:

JunLin Xu,
Hunan University, China
Rahul Kumar,
Indian Institute of Science, India

*Correspondence:

Timothy R. Lezon
lezon@pitt.edu

Keywords: combination therapies, phenotypic screening, drug discovery, matrix factorization, active learning, experimental design

Specialty section:

This article was submitted to
Drug Discovery in Bioinformatics,
a section of the journal
Frontiers in Bioinformatics

Received: 12 May 2021

Accepted: 30 July 2021

Published: 13 August 2021

Citation:

Nafshi R and Lezon TR (2021)
Predicting the Effects of Drug
Combinations Using Probabilistic
Matrix Factorization.
Front. Bioinform. 1:708815.
doi: 10.3389/fbinf.2021.708815

INTRODUCTION

Complex diseases are increasingly recognized as emerging not from single molecules, but from systemic dysfunction of biological processes. From a systems view, pharmacologically treating complex disease requires engaging multiple components of the dysregulated pathways through polypharmacology or combination therapies. Already combination therapies represent the standard of care for an array of diseases, including cancer (Mokhtari et al., 2017) bacterial infection (Mulani et al., 2019), HIV (Ghosn et al., 2018), neurological and behavior disorders (Ortiz-Orendain et al., 2017). They are also the focus of increasing attention in the search for therapeutics to treat other complex diseases, such as AD (Cummings et al., 2019) and NAFLD (Singh et al., 2017). Unfortunately, discovering effective combination therapies requires either serendipitous discovery in the clinic or laborious searches in pre-clinical models.

Phenotypic approaches have been more successful than target-based approaches in bringing new first-in-class drugs to the clinic (Swinney 2013, 2014), and next-generation *in vitro* disease models

promise to boost the power of phenotypic screens by enhancing their clinical relevance. Screening in patient-derived 3D organoids (Lou and Leung, 2018; Takahashi, 2019) and biomimetic tissue chips that contain multiple interacting cell types in physiological geometries (Mittal et al., 2019; Gough et al., 2021), has proven useful for personalizing therapies for cancer and other complex diseases. Inter-organ side effects have been studied *in vitro* using multiple linked tissue chips (Skardal et al., 2017), paving the way for complete human-on-a-chip disease models. The price of this biological fidelity is a loss of throughput. Although sophisticated disease models more closely resemble their clinical counterparts than do their 2D monoculture predecessors, they are far more expensive to develop and maintain. Large-scale screens and even moderate combination screens are not yet feasible in these systems.

An addressable bottleneck in the phenotypic discovery pipeline is the low hit rate of screens. The overwhelming majority of compounds tested in traditional phenotypic screens are inactive, and the problem is exacerbated in combination screens, where the number of possible combinations scales exponentially with the size of the compound library. Fortunately, the availability of large screening data sets in human cell lines has spurred computational methods for predicting drug efficacies and synergies for complex diseases (Menden et al., 2019; Adam et al., 2020). Deep learning models using Graph Convolutional Networks are a powerful and well explored approach for relating complex relationships between inputs and targets, and many successful models have been developed for predicting values in the drug-interaction network (Sun et al., 2020). Other successful approaches include those that incorporate information on transcriptomic or proteomic profiling (Dawson and Carragher, 2014; Huang et al., 2019; Diaz et al., 2020), compound chemical structures (Sidorov et al., 2019), or drug targets (Yang et al., 2020; Rao et al., 2019; Iwata et al., 2015). Network-based approaches have been applied to create robust methods of identifying therapeutically effective drug combinations (Cheng, Kovács, and Barabási 2019).

Combination screens are amenable to statistical prediction methods that use screening data without supplementing it with details about chemical structures, targets, or OMICS profiles. One particularly powerful approach employs a higher-order factorization machine to predict dose-response surfaces for unique drug combinations using only data from the same screen (Julkunen et al., 2020). The advantages of such methods are that they guard against bias introduced from orthogonal databases, they can be used with unannotated compound libraries, and their predictions are consistent with the data generated in the screen of interest. For methods like this to be effective, the computational and experimental components of the screen must be properly synchronized, so that each informs the other (Stern et al., 2016). A number of studies have addressed the right way to mix computational and experimental work to find effective combination therapies (Calzolari et al., 2008; Gerlee et al., 2013; Kashif et al., 2015; Weiss et al., 2015; Silva et al., 2016; Matlock et al., 2017; Ianevski et al., 2019).

Here we introduce a method for predicting the effects of drug combinations using as input only the effects of other drug

combinations. Mathematically, the problem is identical to filling in the missing values of a symmetric matrix containing the effects of drug combinations. Each row and each column of the matrix corresponds to a drug, and the matrix elements are the effects of the drug combinations. If only a subset of the matrix elements are known, the rest can be inferred by decomposing and reconstructing the partial matrix under certain assumptions (Lezon et al., 2006; Lezon and Bahar, 2010). For the current application, we use Probabilistic Matrix Factorization (PMF) (Salakhutdinov and Mnih, 2007), a collaborative filtering algorithm that has proven successful in other problems of the same class (for an introduction to collaborative filtering algorithms, see Aggarwal, 2016).

The PMF algorithm was first developed to recommend movies to Netflix users based on the movies viewed by other users. The core assumption of PMF is that attitudes or preferences that lead to each user's score for a movie are shared by other users with similar taste. PMF recommends that viewers watch movies that similar viewers enjoyed. The method has since been applied to predict values from other large, sparse and imbalanced data sets. Biomedical applications of PMF include predicting diseases associated with transcription patterns (Ha et al., 2020; Mao, Wang, and Zhang 2019), recommending novel indications for drug repurposing (Meng et al., 2021; Yang et al., 2014), and predicting novel targets from drugs (Cobanoglu et al., 2013; Cobanoglu et al., 2015; Li et al., 2020). In the present context, PMF is used to "recommend" drug combinations based on the known effects of similar combinations. We train our model on phenotypic screening data from the NCI ALMANAC (Holbeck et al., 2017), a robust collection of pairwise drug combinations of 104 FDA approved anticancer drugs against 60 common cancer cell lines. We find that knowing the effects of only 50% of drug combinations allows us to classify the effects of the missing combinations as efficacious with 95% accuracy, and we demonstrate how our method can be incorporated into optimal experimental design.

METHODS

NCI ALMANAC

The NCI ALMANAC is a novel, easy-to-use resource created to help researchers identify new combination therapies. The NCI ALMANAC database (Holbeck et al., 2017) is a collection of pairwise combinations of 104 FDA approved anticancer drugs against the NCI-60, a set of 60 common human tumor cancer cell lines collected by the National Cancer Institute. A total of 5,232 drug-drug pairs were evaluated in each of the cell lines; 304,549 experiments were performed to test each drug at either 9 or 15 combination dose points, for a total of 2,809,671 dose combinations. At each dose combination, the percent cell growth after 2 days was measured and recorded, and the efficacy of the combination calculated as the percent of growth inhibition. A combination that has no effect on cell growth compared to control has zero efficacy; a combination that completely halts cell growth has efficacy 100. See the NCI ALMANAC (Holbeck et al., 2017) for details. For each cell line, the combination

efficacies are arranged into a symmetric matrix, $\mathbf{M}_{104 \times 104}$, where each row and column represent a drug, and each element represents the efficacy of a unique drug-drug combination on that cell line. For purposes of PMF (see below), diagonal elements are ignored. The data is then normalized to mean-zero and unit variance for input into the PMF algorithm.

The synergy of each combination is reported by the NCI ALMANAC as a “ComboScore” that measures the difference between the recorded growth rate after testing and the growth rate expected by Bliss Independence (Bliss, 1939). A positive ComboScore indicates a synergistic combination, whereas a negative ComboScore indicates an antagonistic combination. When applying PMF to predict synergies instead of efficacies, we populate the input matrix \mathbf{M} with ComboScores and normalize as described above.

PMF

Probabilistic Matrix Factorization (PMF) is a collaborative filtering algorithm that factors the low-rank input matrix $\mathbf{M}_{n \times m}$ into the product of two low-rank matrices, $\mathbf{A}_{n \times d}$ and $\mathbf{B}_{m \times d}$ such that $\mathbf{M}_{ij} = \mathbf{A}_i \mathbf{B}_j^T$. Thus, PMF reduces to estimating the two matrices \mathbf{A} and \mathbf{B} . The core assumptions of this are that the values of \mathbf{M} are independent, normally distributed and share a common variance σ^2 . Thus, the conditional probability of entries of \mathbf{M} can be expressed as

$$p(\mathbf{M}|\mathbf{A}, \mathbf{B}, \sigma^2) = \prod_{i=1}^n \prod_{j=1}^m N(\mathbf{M}_{ij}|\mathbf{A}_i \mathbf{B}_j^T, \sigma^2)^{I_{ij}}$$

where I_{ij} is the indicator function equal to 1 if \mathbf{M}_{ij} is known and 0 otherwise (Salakhutdinov and Mnih, 2007).

To solve for the matrices \mathbf{A} and \mathbf{B} , we place a zero-mean spherical Gaussian prior on each, such that $p(\mathbf{A}|\sigma_A) = \prod_{i=1}^n N(\mathbf{A}_i|0, \sigma_A^2 \mathbf{I})$ and $p(\mathbf{B}|\sigma_B) = \prod_{j=1}^m N(\mathbf{B}_j|0, \sigma_B^2 \mathbf{I})$. We can then derive the full posterior distribution of \mathbf{A} and \mathbf{B} as $p(\mathbf{A}, \mathbf{B}|\mathbf{M}, \sigma^2, \sigma_A^2, \sigma_B^2) \propto p(\mathbf{M}|\mathbf{A}, \mathbf{B}, \sigma^2)p(\mathbf{A}|\sigma_A^2)p(\mathbf{B}|\sigma_B^2)$. Maximizing the log-posterior is equivalent to minimizing the sum-of-squared-errors objective function: $L(\mathbf{A}, \mathbf{B}) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^m I_{ij} (\mathbf{M}_{ij} - \mathbf{A}_i \mathbf{B}_j^T)^2 + \frac{\lambda}{2} \sum_{i=1}^n \|\mathbf{A}_i\|_{Fro}^2 + \frac{\lambda}{2} \sum_{j=1}^m \|\mathbf{B}_j\|_{Fro}^2$, where λ is the regularization rate hyperparameter. We then construct a stochastic gradient descent update scheme by differentiating the loss function in terms of \mathbf{A} and \mathbf{B} , such that

$$\frac{\partial}{\partial \mathbf{A}_i} L(\mathbf{A}, \mathbf{B}) = \sum_{j=1}^m I_{ij} (\mathbf{M}_{ij} - \mathbf{A}_i \mathbf{B}_j^T) \mathbf{B}_j + \lambda \mathbf{A}_i$$

$$\frac{\partial}{\partial \mathbf{B}_j} L(\mathbf{A}, \mathbf{B}) = \sum_{i=1}^n I_{ij} (\mathbf{M}_{ij} - \mathbf{A}_i \mathbf{B}_j^T) \mathbf{A}_i + \lambda \mathbf{B}_j$$

Algorithmically, we randomly initialize \mathbf{A} and \mathbf{B} from Gaussian distributions and iteratively update them by descending along these gradients until a minimum of $L(\mathbf{A}, \mathbf{B})$ is reached. Applying these rules simultaneously to both \mathbf{A} and \mathbf{B} guarantees convergence of the algorithm to a local minimum. However, the stochastic nature of the initial conditions implies that each run of PMF may not necessarily converge to the global minimum, or even the same local minimum. This requires that PMF be run multiple times on different random initializations and then select the most accurate factorization. While this increases the overall computational cost, this is offset by PMF’s computational

cost scaling linearly with input size and using lightweight low-rank approximations.

Stochastic gradient descent methods are a critical component of machine learning, and methods incorporating momentum and acceleration play an important role when used in conjunction with stochastic gradients (Assran and Rabbat, 2020). Momentum methods help accelerate stochastic gradient descent in the relevant direction and dampen oscillations as a minimum is approached by incorporating the momentum constant γ . The update step with respect to the parameters θ can be expressed as $v_t = \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta)$, $\theta = \theta - v_t$. However, simple momentum methods can be insufficient for complex surfaces. The Nesterov Accelerated Gradient (NAG) (Assran and Rabbat, 2020) improves on this method by “looking ahead” to where the parameters will be to calculate the gradient and is formalized as follows: $v_t = \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta - \gamma v_{t-1})$, $\theta = \theta - v_t$. Rather than computing the gradient at parameters θ , NAG looks ahead at a rough approximation of where the parameters will be, computing the gradient at $\theta - \gamma v_{t-1}$. This anticipatory update greatly increases optimization and performance of PMF as it approaches a minimum.

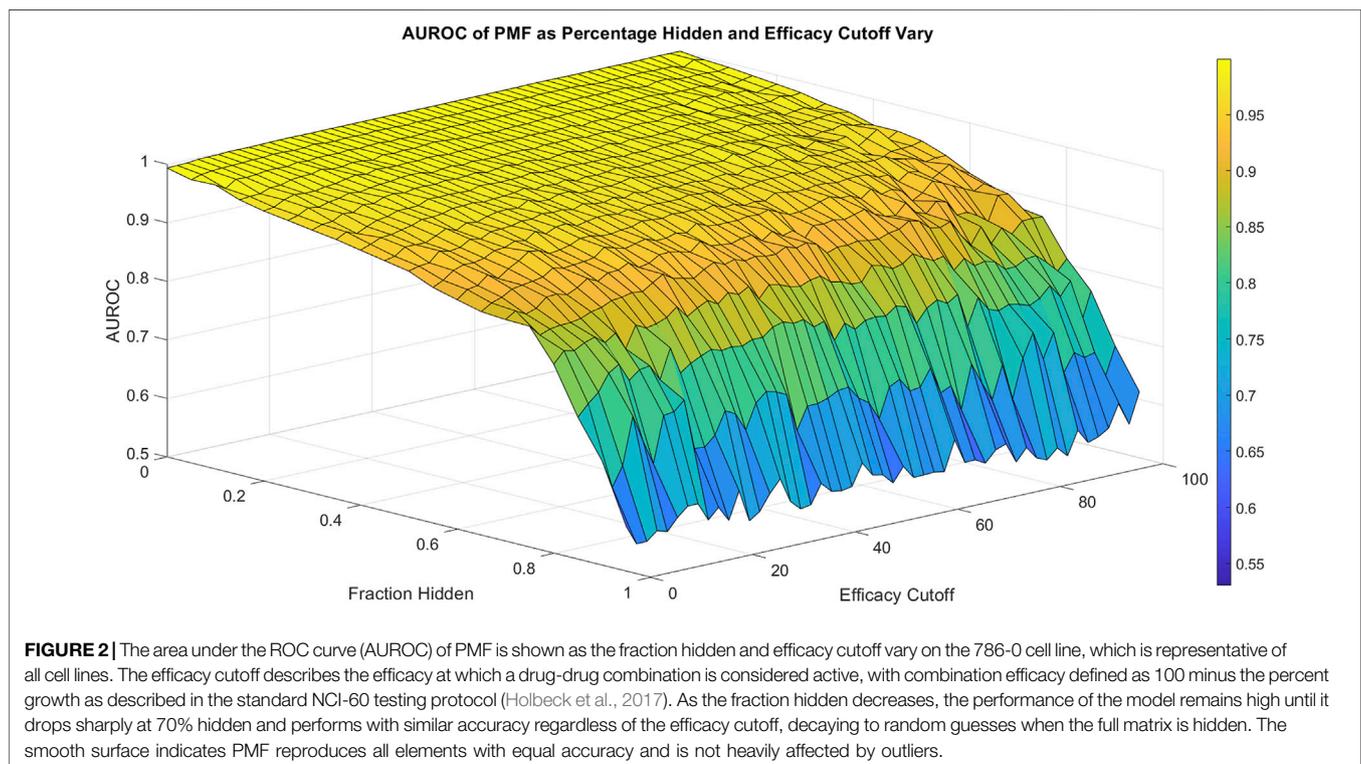
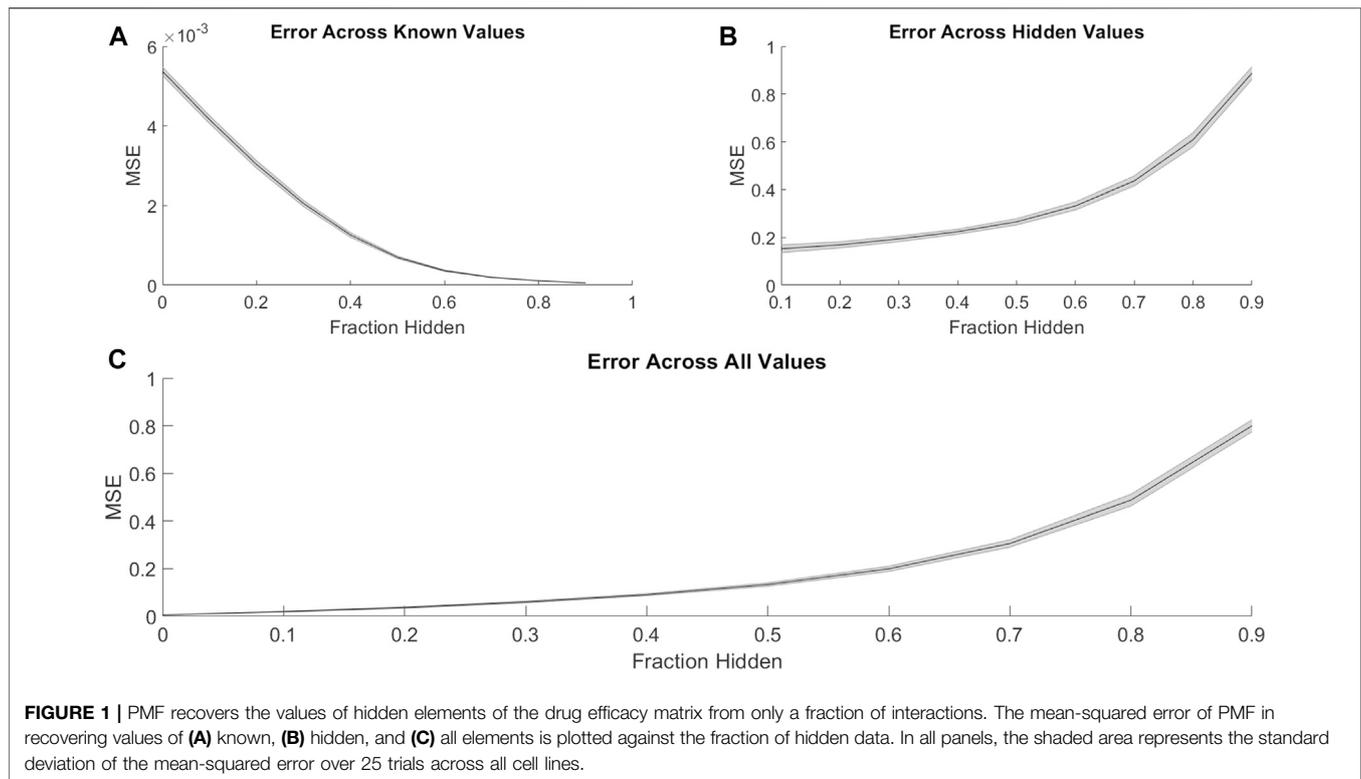
RESULTS

PMF Accurately Recovers Drug Synergies From Partial Data

We first investigated the ability of PMF to recover hidden elements in the drug combination efficacy matrix. For each cell line, we randomly hid a fraction of the combination efficacy matrix, creating non-overlapping “training” and “validation” sets. Then, we used PMF to predict the hidden values and complete the matrix. To guarantee a solution, we included only cases where all drugs were present in a single connected component; that is, where a path could be made from any drug to any other drug using common combination partners. PMF recovered training data to arbitrary precision (Figure 1A) and recovered test data well, provided a sufficiently large training set (i.e., small fraction of data hidden). Using empirically determined hyperparameters for the regularization rate (λ), learning rate (η), and momentum constant (γ), we found that knowing only 30–50% of the drug-drug interactions was sufficient to recover the remaining values in the matrix to within 10% (Figures 1B,C). When selecting combinations with efficacies above a given threshold, PMF performance did not vary strongly with the threshold value (Figure 2); that is, the method can predict whether a combination has an effect over 0.9 nearly as well as it can predict whether a combination has an effect over 0.2.

PMF Performance Is Largely Independent of Individual Drug Efficacies

Assuming compounds act independently (i.e., Bliss independence), the most efficacious compound combinations will be combinations of the independently most efficacious compounds. Reasoning that efficacious drugs are more likely



to influence pathologically relevant mechanisms, we next investigated whether PMF performed better when trained on combinations involving highly efficacious drugs. For each cell line, we rank-ordered the compounds by efficacy and then divided the drug efficacy matrix into halves and quarters to assess if PMF is more accurate when only provided singularly efficacious drugs.

On aggregate we find the differences in accuracy were small, and PMF performance was largely independent of the individual efficacies of the starting set outside of this edge case (**Supplementary Figures 1A,B**). More generally, we found that the most efficacious compounds neither led to the most efficacious combinations, nor were they the best at predicting the values of missing efficacies (**Supplementary Figure 1C**). In fact, individual drug identities did not greatly affect the accuracy of the prediction. We generated an occupancy matrix by randomly selecting 10% of the elements in the combination efficacy matrix. We then randomly shuffled the identities of the drugs while keeping the occupancy matrix static. Repeating this 1,000 times for 1,000 different occupancy matrices, we found PMF predicted the missing values of each matrix with a mean squared error of 0.938 ± 0.0145 , and thus performed equally well regardless of the individual drug identities for a given occupancy matrix.

Graph Topology's Influence on PMF Performance

Viewing the problem through a graph lens, the combination efficacy matrix describes an undirected graph in which the N drugs are nodes and known two-drug combinations represent weighted edges (for a primer on network science, see Barabási, 2016). The challenge of PMF is to reconstruct a fully connected graph from a seed network. By using different algorithms for selecting drug combinations for the training set, we investigated how seed network topology influences prediction accuracy. The method described above, where seed drug combinations are selected randomly and independently, is known as an Erdős-Rényi graph (Rényi and Erdős, 1959) that has a Poisson degree distribution (Barabási and Pósfai 2016).

An extension of the Erdős-Rényi graph is the Watts-Strogatz model (Watts and Strogatz, 1998). This method is motivated by the observation that often in real networks, almost any node can be reached by a short number of steps, known as the Small-World Property (Milgram, 1967). The Watts-Strogatz graph is generated by attaching each node to its nearest k neighbors, resulting in a regular lattice structure. Each edge is then randomly reassigned with probability β . When β is 0, no changes are accepted, and the method preserves the original lattice. As β increases, more links will be randomly assigned, and as β approaches 1, all links will be randomly reassigned, resulting in a completely random Erdős-Rényi network. Intermediate values of β result in small-world networks of low diameter (Barabási and Pósfai 2016).

When training data was arranged in a Watts-Strogatz model topology, the performance of PMF increased with β (**Figure 3A**). We attribute the poor performance near $\beta = 0$ to the difficulty of predicting combination effects of drugs that are separated by large

distances on the seed network. The adjacency matrix for a regular lattice is banded, with the unknown values comprising a contiguous block. Performance improves for values of β near $\frac{1}{2}$, where the small-world property emerges, and peaks at $\beta = 1$, the Erdős-Rényi network. Whereas the small-world Watts-Strogatz graph provides a short path between any pair of nodes, the Erdős-Rényi graph contains multiple paths, each carrying evidence for the value of the inferred combination efficacy.

Many real-world networks do not follow a binomial or Poisson degree distribution, and instead follow a power law or scale-free distribution. In a scale-free network, the probability that a node has k edges is proportional to $k^{-\gamma}$, where γ is a scaling exponent between 2 and 3. We explored whether a scale-free distribution in the input data influences the accuracy of the prediction. Using the hidden parameter model (Caldarelli et al., 2002; Söderberg, 2002; Boguñá and Pastor-Satorras, 2003), we generated scale-free seed networks for training PMF. The method performed equally well for scale-free distributions for all values of γ , and predicted unknown values with accuracy comparable to the Watts-Strogatz method (**Figure 3B**).

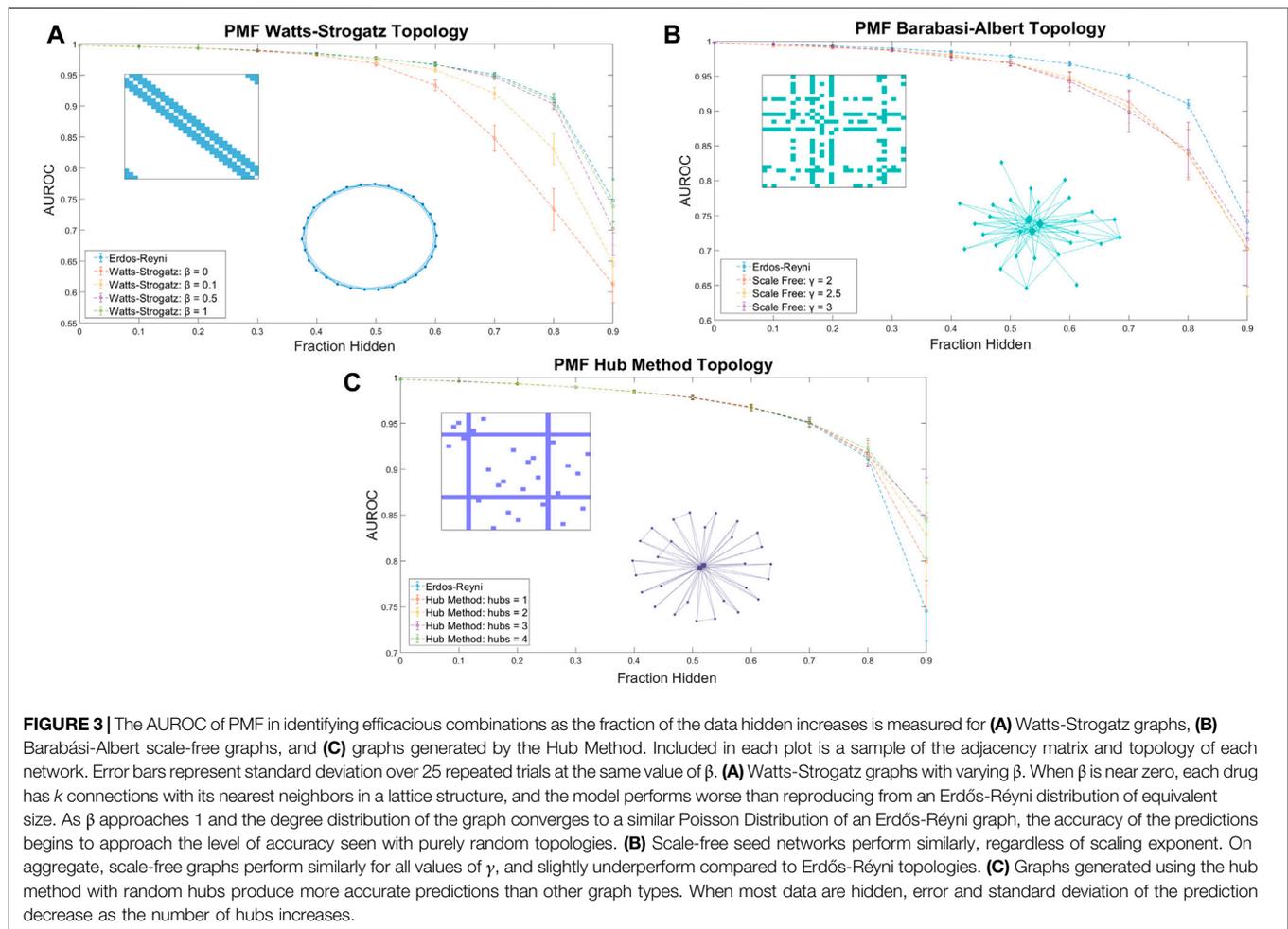
Designing a combination screen using the above-described graph topologies may not be experimentally convenient; instead, screeners are more likely to select a few well-known compounds and test them in combination with other compounds in a large library. The adjacency matrix of the seed graph in this approach has several rows/columns in which every value is known, while the large majority of have few or no known values (**Figure 3C**). The corresponding graph has several fully connected hubs, with the remaining nodes having very few connections.

We explored the accuracy of PMF by using a hub method construction defined as follows: First, we ensured that every node had exactly one connection. Then, we selected nodes at random to be hubs, and ensured that each hub was fully connected. Finally, the remaining edges were randomly assigned in an Erdős-Rényi random fashion. We found the PMF performed stronger on hub method topologies than random Erdős-Rényi topologies when more than 80% of the network was hidden and the graph was sparse. Moreover, when training data was arranged in a hub model topology, the performance of PMF increased as the number of hubs increased (**Figure 3C**).

Thus, we found that the specific seed topology of the training data did not greatly affect the accuracy of the prediction in identifying synergistic drugs if the topology was a random Erdős-Rényi graph or had a binomial degree distribution, such as the Watts-Strogatz for large β . However, PMF did perform worse when edges were evenly distributed following the Watts-Strogatz model for small values of β or when edges were distributed following a scale-free distribution. Moreover, we found that PMF was more accurate under hub topologies mirroring real drug combination assays when more than 80% of the network was hidden, which is exactly the region of interest if we want to test as few combinations as possible.

PMF Predicts Efficacy, But Not Synergy

The desired output of most phenotypic combination screens is an efficacious and non-toxic combination; however, *de novo*



development of combination therapeutics will benefit from identifying synergistic drug combinations, whether or not they are efficacious. For example, two drugs that individually have no efficacy may have a moderate effect in combination. Although such a combination may not be clinically useful, it carries structure and pathway information that may serve as the starting point for rational development of combination therapeutics.

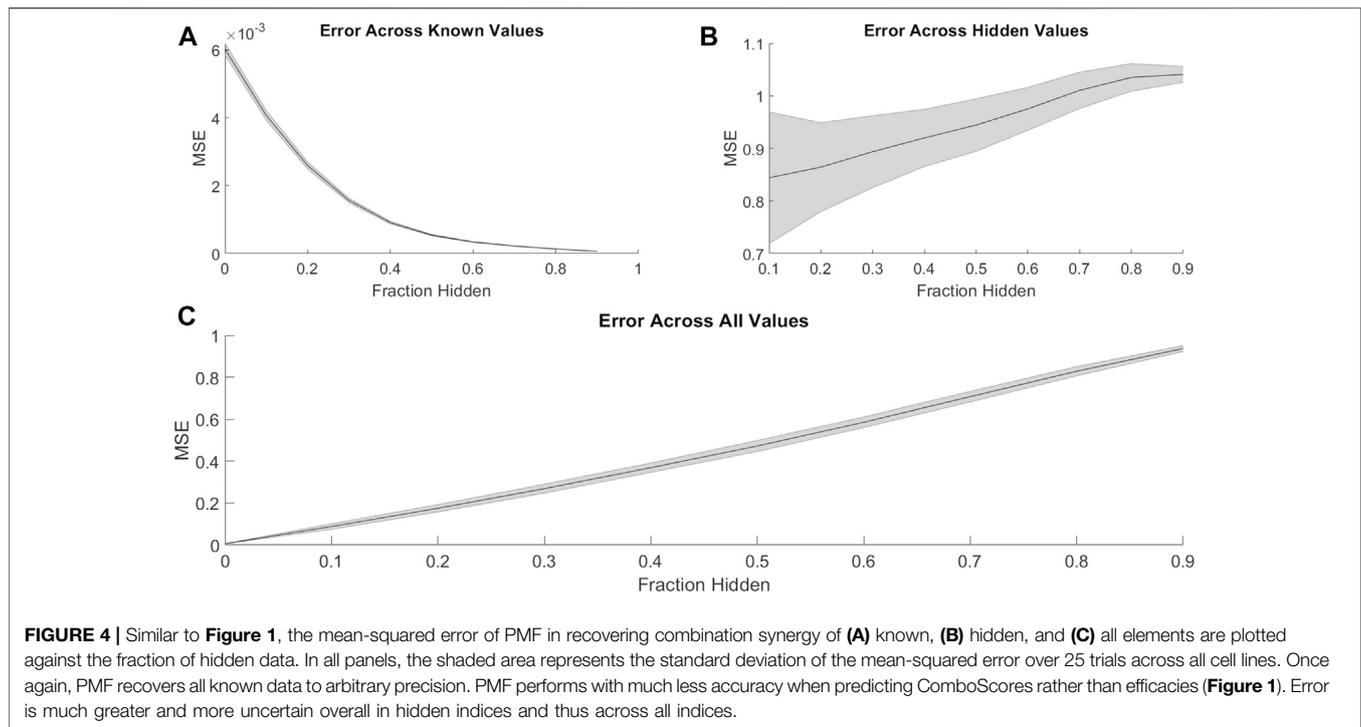
PMF was able to recover missing values much less accurately when predicting synergy rather than efficacy. Just as with efficacies (**Figure 1**), PMF recovered training data to arbitrary precision (**Figure 4A**), but it did not recover test data well, unless it had a sufficiently large training set (i.e., small fraction of data hidden) (**Figures 4B,C**). While the accuracy of PMF on predicting synergy was much weaker than PMF predicting efficacy, we still found that the model is robust and performed well in cases where 50–70% of the total matrix was known.

This stark decrease in accuracy and predictive power may result from the lopsided definition of synergy. The ComboScore of each drug combination represents the difference between the observed effect of the combination and the expected effect

assuming each drug acts independently. Because the upper bound on efficacy is the same for individual drugs and combinations, a combination of highly efficacious drugs cannot have a high ComboScore, even if it has optimal efficacy. Similarly, combinations with identical efficacies may have different ComboScores, depending on the efficacies of the individual drugs used in the combinations. Thus, a low ComboScore reveals nothing about the efficacy of the combination, but a high ComboScore indicates the combination's component drugs individually have low efficacy (**Supplementary Figure 2**). As ComboScores are calculated from individual and combination efficacies, one can still use PMF to predict combination efficacies, and use these to calculate ComboScores.

PMF as a Tool to Guide Combination Screening

In vitro phenotypic-based screens have several benefits for drug discovery, such as not needing to know the molecular target of a disease and being less restricted by hypotheses (Zheng et al., 2013). However, throughput can be low in such assays, and



increasing the number of compounds to be screened causes experimental effort and cost to rise exponentially. PMF may help combat this issue by guiding combination screens through iterative prediction and testing in an active learning scheme.

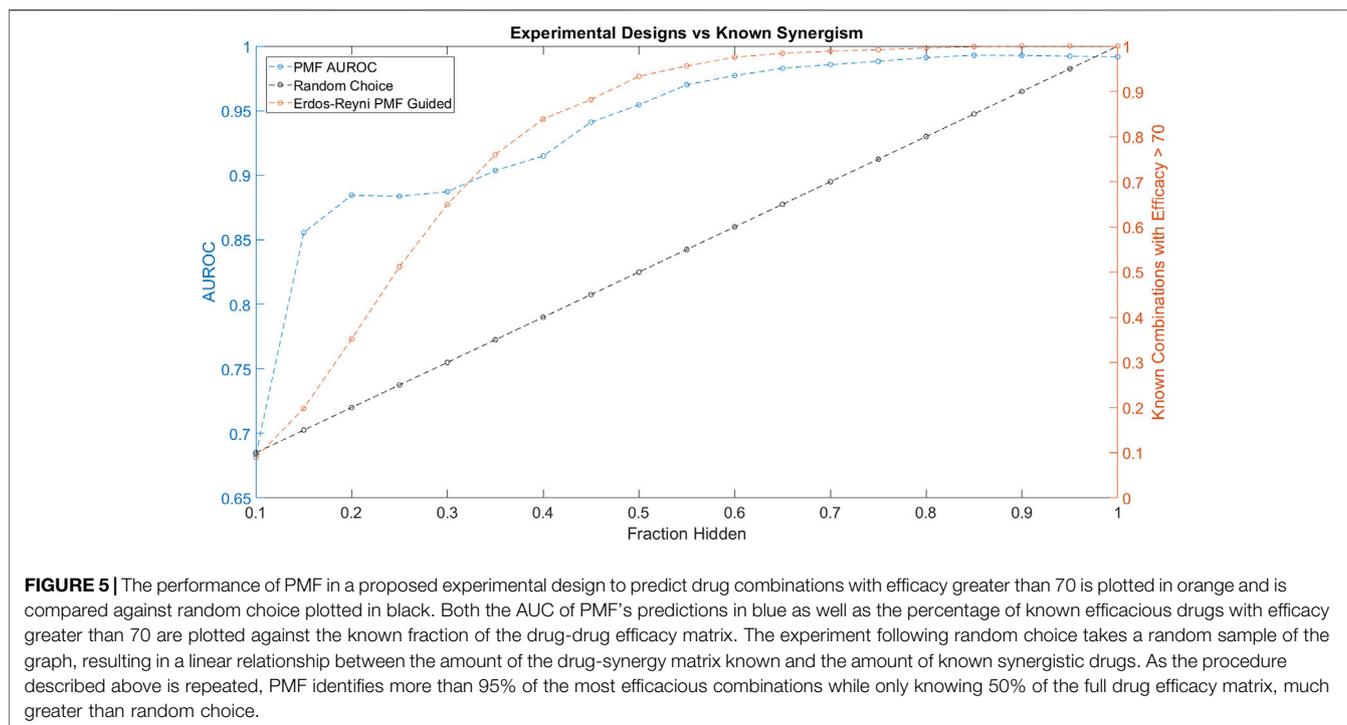
We simulated PMF being used in an active learning experimental design as follows. First, we created a random Erdős-Rényi graph topology with 10% of the total combinations known. Then, we used PMF to reproduce the entire combination efficacy matrix and identified the top 5% greatest efficacies as predicted by PMF. We then “tested” these identified efficacious combinations by adding the actual values of the efficacies to the list of known combinations, and then repeated the procedure to discover the next 5%, until the entire matrix is recovered.

PMF-guided screens identified efficacious combinations much more efficiently than naïve random tests (**Figure 5**). In our simulated experiment, PMF identified efficacious combinations at three times the rate of random choice and identified as much as 95% of all highly efficacious combinations while only testing 50% of all available combinations. This finding was consistent across cell lines and was not sensitive to the details of the starting point. Our results suggest that screeners may be able to test a small number of relevant combinations of direct interest and obtain the remaining synergistic combinations following a PMF-guided design. Future studies could fruitfully explore this issue further by optimizing PMF-guided screens as well as investigating its accuracy applied in a physical assay experiment.

Throughout the simulated experiment, we monitored the performance of PMF as measured by AUROC (**Figure 5**). The dip in AUROC observed around the fourth step of the simulated experiment may be due to bias introduced by the active learning. Efficacious combinations are not uniformly distributed across all drugs, and indeed a small subset of drugs is likely to contribute to many of the efficacious combinations. As the experiment progresses, PMF preferentially selects combinations from an efficacious minority of the nodes, mirroring the construction of a scale-free graph. PMF performs worse on scale-free graphs compared to Erdős-Rényi graphs (**Figure 3**), causing the accuracy to decrease as nodes are preferentially tested, and then increase as these nodes are saturated and the rest of the matrix is tested. Future studies might investigate ways to counteract this drop in error by using a more complex method than simply testing the top 5% most efficacious combinations as predicted by PMF.

The Method’s Performance Is Not Unique to Cancer

To test the performance of our method in diseases other than cancer, we applied it to data from a small combination screen for Huntington’s disease (HD), an autosomal dominant neurodegenerative disease caused by an abnormally long polyglutamine stretch in the huntingin protein (Zuccato et al., 2010). The clinical progression of HD starts with general loss of motor control around the third decade of life. This is followed by mood and personality changes, and eventual dementia and death. To date, there are no drug-like molecules that can prevent or slow



HD, and the pleiotropic nature of huntingtin makes it difficult to target directly.

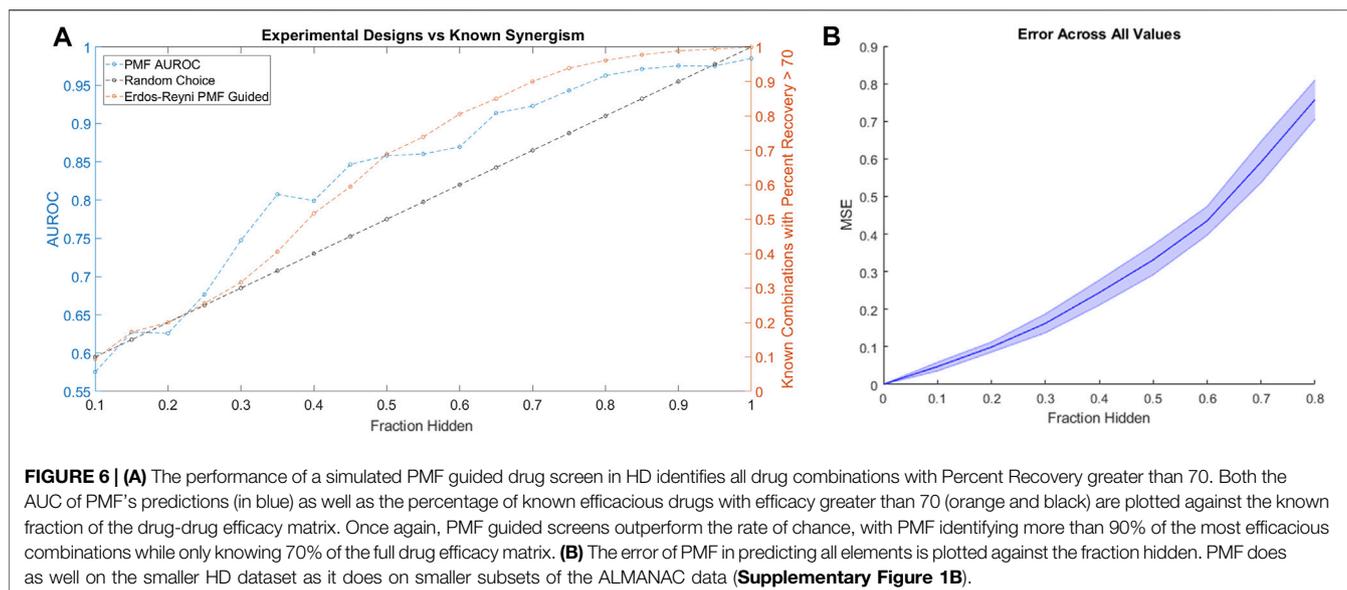
Pei and coworkers reported results from a combination screen in a murine cell-based model of HD (Pei et al., 2017). Briefly, their assay used serum deprivation to induce stress in neuronal progenitor cells derived from the *STHdh^{Q111}* murine cell line model of HD, which has an abnormally long glutamine stretch in its huntingtin protein. The phenotypic response of the cells was then compared to that in serum-deprived cells of their isogenic wild type *STHdh^{Q27}*; overall, serum deprivation killed about half of the HD cells and about 5% of the wild type cells. The group evaluated 268 two-compound combinations of 32 compounds for their ability to protect against cell death. As a metric, they used “Percent Recovery,” which captures the percent of HD cells that were rescued from cell death by a treatment (Pei et al., 2017).

Applying the same simulated experimental design to the HD data as we applied to the ALMANAC data, we found that PMF-guided screens identified combinations with the highest Percent Recovery much more efficiently than the rate of random chance: Over 90% of combinations with Percent Recovery over 70 were identified by testing only 70% of all available combinations (Figure 6A). Moreover, we found that PMF performed with nearly the same accuracy on the HD dataset as it did on similarly sized subsets of the ALMANAC data (Figure 6B, Supplementary Figure 1B). While the guided screen is more efficient than naïve guessing, the results are much weaker than on the much larger ALMANAC dataset. Starting with 10% of the 32 compound combinations known, PMF may only know a handful of combinations, and struggles until around 30% of all combinations are known.

DISCUSSION

Our results show that it is possible to use information on the effects of drug combinations to predict the effects of novel combinations. A strength of our approach is that does not require any outside knowledge of chemical structures, target profiles, or OMICS data. This lack of reliance on outside data contributes to the method's robust performance between data sets of similar sizes. Indeed, freedom from additional information augments our method's stability and flexibility: Rather than predicting the effects of combinations of drugs, it could be used for combinations of unknown substances, natural extracts, or even combinations of combinations. As such, our method may contribute to identifying mechanisms of action for novel compounds. Since PMF is lightweight and is not informed by structure or other data, it may provide a benchmark against which more complicated methods can be tested.

There are many constrained low-rank matrix approximation algorithms that could have been used other than PMF. Within this general framework, some powerful techniques include singular value decomposition (SVD) (Golub and Reinsch, 1971), principal component analysis (PCA) (Wold et al., 1987), non-negative matrix factorization (NMF) (Lee and Sebastian Seung, 1999), entropy maximization (Lezon et al., 2006; Lezon and Bahar, 2010), and deep matrix factorization (MF) (Pierre and Siebert, 2020). SVD's matrix factorization is unique and orthogonal unlike PMF. However, SVD's matrix reconstruction is a superposition of the orthogonal components with arbitrary signs, losing strong correlations that may exist and the



interpretability of the latent factor representation of PMF. Moreover, SVD has been shown to be less accurate than PMF (Salakhutdinov and Mnih, 2007), and is vulnerable to overfitting. PCA's orthogonal decomposition suffers from similar weaknesses. The parts-based decomposition of NMF has provided powerful, interpretable methods for matrix factorization in a variety of applications from text mining to gene expression. However, NMF is poorly suited to drug discovery due to its non-negativity constraint. Methods relaxing this constraint (Wang et al., 2015) may be able to provide similar results with a parts-based matrix representation as opposed to PMF's latent factor representation. Unlike PMF, which factors into only two matrices and captures a single layer of features, deep matrix factorization, inspired by the success of deep learning, aims to extract several features in a hierarchical way (Pierre and Siebert, 2020). Although many algorithms have been introduced for deep MF, it is still an emerging topic and questions of convergence, identifiability, and loss functions have not been fully explored. PMF on the contrary is well-explored algorithm with strong theoretical ground that has found success in a variety of matrix completion and collaborative filtering settings.

One limitation of PMF is that it suffers from the cold-start problem and is unable to predict the efficacy for compounds with no known values. Thus, any guided PMF assay must test at least one combination for every drug. It is also unknown how well PMF will scale to drug libraries larger than the ALMANAC. While our simulated experiment was more successful on the larger ALMANAC data than the smaller Huntington's disease data, PMF's performance on larger drug libraries remains to be seen. We additionally note that both the ALMANAC and HD data sets employed combinations of compounds that were individually effective as monotherapies. To fully evaluate our method's utility in

guiding combination screens, we will need to apply it to screens in which not all compounds are individually efficacious. That is, the proper test of the method is its application to a new screen, which is beyond the scope of this work.

CONCLUSION

We have shown that PMF can accurately impute missing values into the drug combination efficacy matrix for a screen, and that the performance of PMF does not depend on the efficacies of the drugs being tested. We further showed that PMF performs best when the input drug combination network has an Erdős-Rényi topology. Finally, we used simulated experiments to demonstrate that alternating PMF inference with experiments can efficiently identify the most efficacious two-drug combinations in a phenotypic screen.

There have been many other attempts at predicting the effects of drug combinations, and those that perform best include additional data, such as chemical structures, target profiles, or OMICS data (Wang et al., 2013; Huang et al., 2019; Menden et al., 2019). PMF has the advantage that its computation time scales linearly, and it can make accurate predictions for sparse and imbalanced data sets. Moreover, PMF is an unsupervised algorithm and by nature is easily interpretable as matrix factorizations easily provide a lens to determine relations, giving it several advantages over large deep learning networks. Our method is simpler by comparison, but it provides a baseline of performance against which more complicated prediction methods may be assessed. Indeed, not relying on additional information endows our method with flexibility: Instead of predicting the effects of combinations of drugs, it can be used to predict the effects of combinations of combinations, and we

have no reason to believe that it will perform worse on unannotated compounds. On the contrary, our method may contribute to identifying mechanisms of action for novel compounds. The very ability of PMF to predict efficacies of combinations points to hidden mechanistic similarities within the set of compounds. By interpreting the PMF in terms of underlying biochemistry, we may gain insight into the nature of disease.

DATA AVAILABILITY STATEMENT

The cancer datasets analyzed for this study can be found on the ALMANAC website (<https://wiki.nci.nih.gov/display/NCIDTPdata/NCI-ALMANAC/>). The HD data can be obtained from the authors. Source code is available online at <https://github.com/lezonlab/pmf-drug-combo/>.

REFERENCES

- Adam, G., Rampasek, L., Safikhani, Z., Smirnov, P., Haibe-Kains, B., and Goldenberg, A. (2020). 'Machine Learning Approaches to Drug Response Prediction: Challenges and Recent Progress. *NPJ Precis Oncol.* 4, 19. doi:10.1038/s41698-020-0122-1
- Aggarwal, C. C. (2016). *Recommender Systems*. London, UK: Springer International Publishing.
- Assran, M., and Rabbat, M. 2020. 'On the Convergence of Nesterov's Accelerated Gradient Method in Stochastic Settings', arXiv pre-print server.
- Barabási, A.-L. (2016). *Network Science*. Cambridge, England: Cambridge University Press.
- Bliss, C. I. (1939). The Toxicity of Poisons Applied Jointly. *Ann. Appl. Biol.* 26, 585–615. doi:10.1111/j.1744-7348.1939.tb06990.x
- Boguñá, M., and Pastor-Satorras, R. (2003). 'Class of Correlated Random Networks with Hidden Variables. *Phys. Rev. E* 68. doi:10.1103/physreve.68.036112
- Caldarelli, G., Capocci, A., De Los Rios, P., and Muñoz, M. A. (2002). 'Scale-Free Networks from Varying Vertex Intrinsic Fitness. *Phys. Rev. Lett.* 89. doi:10.1103/physrevlett.89.258702
- Calzolari, D., Bruschi, S., Coquin, L., Schofield, J., Feala, J. D., Reed, J. C., et al. (2008). 'Search Algorithms as a Framework for the Optimization of Drug Combinations. *Plos Comput. Biol.* 4, e1000249. doi:10.1371/journal.pcbi.1000249
- Cummings, J. L., Tong, G., and Ballard, C. (2019). Treatment Combinations for Alzheimer's Disease: Current and Future Pharmacotherapy Options. *Jad* 67, 779–794. doi:10.3233/jad-180766
- Gerlee, P., Schmidt, L., Monsefi, N., Kling, T., Jornsten, R., and Nelander, S. (2013). Searching for Synergies: Matrix Algebraic Approaches for Efficient Pair Screening. *PLoS One* 8, e68598. doi:10.1371/journal.pone.0068598
- Ghosh, J., Taiwo, B., Seedat, S., Aufran, B., and Katlama, C. (2018). Hiv. *The Lancet* 392, 685–697. doi:10.1016/s0140-6736(18)31311-4
- Golub, G. H., and Reinsch, C. (1971). *Singular Value Decomposition and Least Squares Solutions*. Springer Berlin Heidelberg.
- Gough, A., Soto-Gutierrez, A., Verneti, L., Ebrahimkhani, M. R., Stern, A. M., and Taylor, D. L. (2021). Human Biomimetic Liver Microphysiology Systems in Drug Development and Precision Medicine. *Nat. Rev. Gastroenterol. Hepatol.* 18, 252–268. doi:10.1038/s41575-020-00386-1
- Holbeck, S. L., Camalier, R., Crowell, J. A., Govindharajulu, J. P., Hollingshead, M., Anderson, L. W., et al. (2017). The National Cancer Institute ALMANAC: A Comprehensive Screening Resource for the Detection of Anticancer Drug Pairs with Enhanced Therapeutic Activity. *Cancer Res.* 77, 3564–3576. doi:10.1158/0008-5472.can-17-0489
- Huang, L., Brunell, D., Stephan, C., Mancuso, J., Yu, X., He, B., et al. (2019). Driver Network as a Biomarker: Systematic Integration and Network Modeling of Multi-Omics Data to Derive Driver Signaling Pathways for Drug Combination Prediction. *Bioinformatics* 35, 3709–3717. doi:10.1093/bioinformatics/btz109
- Ianevski, A., Giri, A. K., Gautam, P., Kononov, A., Potdar, S., Saarela, J., et al. (2019). Prediction of Drug Combination Effects with a Minimal Set of Experiments. *Nat. Mach. Intell.* 1, 568–577. doi:10.1038/s42256-019-0122-4
- Julkunen, H., Cichonska, A., Gautam, P., Szedmak, S., Douat, J., Pahikkala, T., et al. (2020). 'Leveraging Multi-Way Interactions for Systematic Prediction of Pre-clinical Drug Combination Effects. *Nat. Commun.* 11, 6136. doi:10.1038/s41467-020-19950-z
- Kashif, M., Andersson, C., Hassan, S., Karlsson, H., Senkowski, W., Fryknaas, M., et al. (2015). 'In Vitro Discovery of Promising Anti-cancer Drug Combinations Using Iterative Maximisation of a Therapeutic index. *Sci. Rep.* 5, 14118. doi:10.1038/srep14118
- Lee, D. D., and Seung, H. S. (1999). Learning the Parts of Objects by Non-negative Matrix Factorization. *Nature* 401, 788–791. doi:10.1038/44565
- Lezon, T. R., and Bahar, I. (2010). 'Using Entropy Maximization to Understand the Determinants of Structural Dynamics beyond Native Contact Topology. *Plos Comput. Biol.* 6, e1000816. doi:10.1371/journal.pcbi.1000816
- Lezon, T. R., Banavar, J. R., Cieplak, M., Maritan, A., and Fedoroff, N. V. (2006). Using the Principle of Entropy Maximization to Infer Genetic Interaction Networks from Gene Expression Patterns. *Proc. Natl. Acad. Sci.* 103, 19033–19038. doi:10.1073/pnas.0609152103
- Lou, Y.-R., and Leung, A. W. (2018). Next Generation Organoids for Biomedical Research and Applications. *Biotechnol. Adv.* 36, 132–149. doi:10.1016/j.biotechadv.2017.10.005
- Matlock, K., Berlow, N., Keller, C., and Pal, R. (2017). 'Combination Therapy Design for Maximizing Sensitivity and Minimizing Toxicity. *BMC Bioinformatics* 18, 116. doi:10.1186/s12859-017-1523-1
- Menden, M. P., Wang, D., Mason, M. J., Szalai, B., Bulusu, K. C., Guan, Y., et al. (2019). 'Community Assessment to advance Computational Prediction of Cancer Drug Combinations in a Pharmacogenomic Screen. *Nat. Commun.* 10, 2674. doi:10.1038/s41467-019-09799-2
- Milgram, S. (1967). SMALL-WORLD PROBLEM. *Psychol. Today* 1, 61–67. doi:10.1037/e400002009-005
- Mittal, R., Woo, F. W., Castro, C. S., Cohen, M. A., Karanxha, J., Mittal, J., et al. (2019). Organ-on-chip Models: Implications in Drug Discovery and Clinical Applications. *J. Cel Physiol* 234, 8352–8380. doi:10.1002/jcp.27729
- Mokhtari, R. B., Homayouni, T. S., Baluch, N., Morgatskaya, E., Kumar, S., Das, B., et al. (2017). Combination Therapy in Combating Cancer. *Oncotarget* 8, 38022–38043. doi:10.18632/oncotarget.16723
- Mulani, M. S., Kamble, E. E., Kumkar, S. N., Tawre, M. S., and Pardesi, K. R. (2019). 'Emerging Strategies to Combat ESKAPE Pathogens in the Era of Antimicrobial Resistance: A Review. *Front. Microbiol.* 10, 539. doi:10.3389/fmicb.2019.00539
- Ortiz-Orendain, J., Castiello-de Obeso, S., Colunga-Lozano, L. E., Hu, Y., Maayan, N., and Adams, C. E. (2017). Antipsychotic Combinations for Schizophrenia. *Cochrane Database Syst. Rev.* 6, CD009005. doi:10.1002/14651858.CD009005.pub2

AUTHOR CONTRIBUTIONS

RN and TL performed all research, analyzed all results, and wrote the manuscript.

FUNDING

This work was supported by the National Institutes of Health through grants UL1TR001857 and R56AG059612.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbinf.2021.708815/full#supplementary-material>

- Pei, F., Li, H., Henderson, M. J., Titus, S. A., Jadhav, A., Simeonov, A., et al. (2017). 'Connecting Neuronal Cell Protective Pathways and Drug Combinations in a Huntington's Disease Model through the Application of Quantitative Systems Pharmacology. *Sci. Rep.* 7, 17803. doi:10.1038/s41598-017-17378-y
- Pierre, N. G., and Siebert, X. 2020. 'Deep Matrix Factorizations', arXiv pre-print server.
- Rényi, P., and Erdős, A. (1959). On Random Graphs'. *Publicationes Mathematicae (Debrecen)* 6, 290–297.
- Salakhutdinov, Ruslan., and Mnih, Andriy. (2007). "Restricted Boltzmann Machines for Collaborative Filtering," in Proceedings of the 24th international conference on Machine learning - ICML '07 (Probabilistic Matrix Factorization). doi:10.1145/1273496.1273596
- Silva, A., Lee, B.-Y., Clemens, D. L., Kee, T., Ding, X., Ho, C.-M., et al. (2016). Output-driven Feedback System Control Platform Optimizes Combinatorial Therapy of Tuberculosis Using a Macrophage Cell Culture Model. *Proc. Natl. Acad. Sci. USA* 113, E2172–E2179. doi:10.1073/pnas.1600812113
- Singh, S., Osna, N. A., and Kharbanda, K. K. (2017). Treatment Options for Alcoholic and Non-alcoholic Fatty Liver Disease: A Review. *Wjg* 23, 6549–6570. doi:10.3748/wjg.v23.i36.6549
- Skardal, A., Murphy, S. V., Devarasetty, M., Mead, I., Kang, H. W., Seol, Y. J., et al. (2017). 'Multi-tissue Interactions in an Integrated Three-Tissue Organ-On-A-Chip Platform. *Sci. Rep.* 7, 8837. doi:10.1038/s41598-017-08879-x
- Söderberg, Bo. (2002). 'General Formalism for Inhomogeneous Random Graphs. *Phys. Rev. E* 66. doi:10.1103/physreve.66.066121
- Stern, A. M., Schurdak, M. E., Bahar, I., Berg, J. M., and Taylor, D. L. (2016). A Perspective on Implementing a Quantitative Systems Pharmacology Platform for Drug Discovery and the Advancement of Personalized Medicine. *J. Biomol. Screen.* 21, 521–534. doi:10.1177/1087057116635818
- Takahashi, T. (2019). Organoids for Drug Discovery and Personalized Medicine. *Annu. Rev. Pharmacol. Toxicol.* 59, 447–462. doi:10.1146/annurev-pharmtox-010818-021108
- Wang, Jue., Guan, Naiyang., Huang, Xuhui., and Luo, Zhigang. (2015). Constraint-Relaxation Approach for Nonnegative Matrix Factorization: A Case Study, IEEE International Conference on Systems, Man, and Cybernetics. Hong Kong, China: IEEE. doi:10.1109/smc.2015.383
- Wang, X., Chang, L., Wang, B., Wang, D., Li, P., Wang, L., et al. (2013). Comparative Proteomics of *Thellungiella Halophila* Leaves from Plants Subjected to Salinity Reveals the Importance of Chloroplastic Starch and Soluble Sugars in Halophyte Salt Tolerance. *Mol. Cell Proteomics* 12, 2174–2195. doi:10.1074/mcp.m112.022475
- Watts, D. J., and Strogatz, S. H. (1998). Collective Dynamics of 'small-World' Networks. *Nature* 393, 440–442. doi:10.1038/30918
- Weiss, A., Berndsen, R. H., Ding, X., Ho, C. M., Dyson, P. J., van den Bergh, H., et al. (2015). 'A Streamlined Search Technology for Identification of Synergistic Drug Combinations. *Sci. Rep.* 5, 14508. doi:10.1038/srep14508
- Wold, S., Esbensen, K., and Geladi, P. (1987). Principal Component Analysis. *Chemometrics Intell. Lab. Syst.* 2, 37–52. doi:10.1016/0169-7439(87)80084-9
- Zheng, W., Thorne, N., and McKew, J. C. (2013). Phenotypic Screens as a Renewed Approach for Drug Discovery. *Drug Discov. Today* 18, 1067–1073. doi:10.1016/j.drudis.2013.07.001
- Zuccato, C., Valenza, M., and Cattaneo, E. (2010). Molecular Mechanisms and Potential Therapeutic Targets in Huntington's Disease. *Physiol. Rev.* 90, 905–981. doi:10.1152/physrev.00041.2009
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Nafshi and Lezon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.