VB-MK-LMF: Fusion of drugs, targets and interactions using Variational Bayesian Multiple Kernel Logistic Matrix Factorization

Bence Bolgár[*] and Péter Antal

[*]Correspondence:
bolgar@mit.bme.hu
Department of Measurement and
Information Systems, Budapest
University of Technology and
Economics, Magyar tudósok krt.
2., 1117 Budapest, Hungary
Full list of author information is
available at the end of the article

Abstract

Background: Computational fusion approaches to drug-target interaction (DTI) prediction, capable of utilizing multiple sources of background knowledge, were reported to achieve superior predictive performance in multiple studies. Other studies showed that specificities of the DTI task, such as weighting the observations and focusing the side information are also vital for reaching top performance.

Method: We present Variational Bayesian Multiple Kernel Logistic Matrix Factorization (VB-MK-LMF), which unifies the advantages of (1) multiple kernel learning, (2) weighted observations, (3) graph Laplacian regularization, and (4) explicit modeling of probabilities of binary drug-target interactions.

Results: VB-MK-LMF achieves significantly better predictive performance in standard benchmarks compared to state-of-the-art methods, which can be traced back to multiple factors. The systematic evaluation of the effect of multiple kernels confirm their benefits, but also highlights the limitations of linear kernel combinations, already recognized in other fields. The analysis of the effect of prior kernels using varying sample sizes sheds light on the balance of data and knowledge in DTI tasks and on the rate at which the effect of priors vanishes. This also shows the existence of "small sample size" regions where using side information offers significant gains. Alongside favorable predictive performance, a notable property of MF methods is that they provide a unified space for drugs and targets using latent representations. Compared to earlier studies, the dimensionality of this space proved to be surprisingly low, which makes the latent representations constructed by VB-ML-LMF especially well-suited for visual analytics. The probabilistic nature of the predictions allows the calculation of the expected values of hits in functionally relevant sets, which we demonstrate by predicting drug promiscuity. The variational Bayesian approximation is also implemented for general purpose graphics processing units yielding significantly improved computational time.

Conclusion: In standard benchmarks, VB-MK-LMF shows significantly improved predictive performance in a wide range of settings. Beyond these benchmarks, another contribution of our work is highlighting and providing estimates for further pharmaceutically relevant quantities, such as promiscuity, druggability and total number of interactions.

Availability: Data and code are available at http://bioinformatics.mit.bme.hu.

Keywords: drug-target interaction prediction; matrix factorization; multiple kernel learning; variational Bayes; probabilistic graphical models

Background

Drug-target interactions (DTI) or compound-protein interactions (CPIs) have become a focal point in chemo- and bioinformatics. There are many factors behind this trend, such as the direct, quantitative nature of bioactivity data [1], its unprecedented amount, public availability [2, 3], and variety including also phenotypic and content-rich assays and screenings [4]. Further factors are the semantic, linked open nature of the data [5, 6], collaborative initiatives in the pharmaceutical policy [1] and the construction of DTI benchmarks [7, 8, 9, 10, 11, 12, 13].

An additional factor is the varying granularity and multiple facets of the DTI task: it was already attacked in the 90's in single target scenarios, e.g. by using neural networks of that time [14] and subsequently by kernel methods [15, 16]. A series of similarity-based methods were also developed for virtual screening [17, 18, 19]; in the early 2000's molecular docking became popular [20, 21]; from the late 2000's matrix factorization methods were developed [7, 22, 23]. As the importance of data and knowledge integration in drug discovery was further emphasized [24, 25, 26, 1], the incorporation of prior knowledge in DTI became mainstream and indeed improved predictive performance [23, 27, 28, 29].

Computational data and knowledge fusion approaches in the DTI problem seem to be especially relevant, as the growth of DTI datasets is limited by experimental and publication time and cost, while the cross-linked repertoire of side information expands at an enormous rate. This grand pool of information complementing the DTI data and the full scope of the DTI fusion challenge is best illustrated by the drug repositioning problem [30, 31]. In repositioning, i.e. in the finding of a novel indication for an already marketed drug, extra information sources could also be used, such as off-label drug usage patterns, patient-reported adverse-effects and official side-effects [32]. Notably, this information pool can be linked back to early stage compound discovery [33].

In this paper we investigate the multiple kernel-based fusion approach to the DTI task from a computational fusion perspective, by adopting widely used benchmark datasets, implementations and evaluation methodologies from Yamanishi et al. [7], Gönen [22], Pahikkala et al. [8] and Liu et al. [34]. Our contributions are as follows:

1. *VB-MK-LMF:* We present a Bayesian matrix factorization method with a novel variational Bayesian approximation, which unifies multiple kernel learning, importance weight for (positive) observations, network-based regularization and explicit modeling of probabilities of drug-target interactions.

2. *Performance in benchmarks:* We report the results of a comparison against three leading solutions using two benchmark datasets, in which VB-MK-LMF achieved significantly better performance in most settings.

3. *Effect of multiple kernels:* We systematically investigate factors behind its top performance, such as the type of the kernels, the role of neighborhood restriction and Bayesian averaging. Finally, we evaluate the effect of priors using varying sample sizes highlighting the regions where using side-information improves predictive performance.

4. *Posteriors for promiscuity and druggability:* We show that probabilistic predictions from VB-MK-LMF can be used to quantify the expected values for promiscuity or the number of hits in a DTI task.

5   *Dimensionality of the unified "pharmacological" space:* We investigate the learned unified latent representations of drugs and targets, and contrary to many studies we argue that drastically smaller dimensions are sufficient. We discuss the possibility that this low dimension, around 10, could be utilized in visual analytics and exploratory data analysis.

6   *Accessibility:* We report the adaptation of the developed variational Bayesian approximation to general purpose graphics processing units (GP-GPU). Evaluations show that $30\times$ speed-up can be achieved using a standard GP-GPU environment. To support the development of current DTI benchmarks towards "computational DTI fusion", we release the applied kernels, code and parameter settings for academic use.

Figure 2 shows the overview of Variational Bayesian Multiple Kernel Logistic Matrix Factorization (VB-MK-LMF).

Related works

To give an overview about related, earlier works [7, 35, 36, 37, 38, 39, 40, 27, 41, 28, 42, 43, 23, 44, 45, 29, 46, 47, 48, 49, 50, 51, 52, 53], we summarize the main properties of their applied datasets, side information, methods and evaluation methodologies in Additional file 2).

DTI data

Drug-target interaction data has become a fundamental resource in pharmaceutical research, which can be attributed to its public availability in and open, linked format, see e.g. [5, 54, 6, 55, 56, 1, 57]. The relative objectivity of interaction activities and the side information about drugs and targets renders a unique status to the comprehensive tabular DTI data, even compared to media and e-commerce data [58], despite the issues of quality [59, 60], duality of commercial and public repositories [61, 62, 63] and selection bias related to the lack of negative samples [12] and promiscuity [64]. However, at present the heterogeneous, real-valued activity data are usually treated as binary relations, even though the use of raw data together with information about the measurement context is expected in more realistic DTI prediction scenarios [8, 45, 51]. Another largely overlooked property of the binary drug-target interaction data is its possibly indirect nature, which influences the applicable target-target similarities, e.g. in the indirect case protein-protein networks may have relevance (for the explicit treatment of direct and indirect relations, see e.g. RBM [44]).

DTI prior knowledge

The molecular similarity property principle [65, 66], the drug-likeliness of a compound [67, 68] and druggability of proteins [69] are essential concepts in the broader drug discovery context, together with molecular docking [20, 21] and binding site, pocket predictors [70], if structure information is available. However, their use as priors in the computational DTI task is still largely unexplored. If the goal is the discovery of indirect drug-target interactions, possibly including multiple paths, which are especially relevant in polypharmacology [71], then the use of molecular interaction and regulatory networks alongside protein-protein similarities is another open issue.

Chemical similarity, the most widespread source of prior knowledge in DTI, was the basis of many "guilt-by-association" approaches in chemo- and bioinformatics. Earlier investigations helped to understand the use of multiple, heterogeneous representations, similarity measures and introduced the concept of fusion methods in ligand-based virtual screening [17, 18, 72, 73, 74]. Beyond chemical similarities, target-based similarities can also be used to exceed activity cliffs [32]; moreover, side-effect based and off-label usage based similarities can be constructed for compounds using FDA-approved drugs as canonical bases in a group-representation [33].

Target-target similarities are another diverse and voluminous source of prior information, which can be defined using sequence similarities, common motifs and domains, phylogenetic relations or shared binding sites and pockets [70]. In case of indirect drug-target interactions, a broader set of target-target similarities could be based on relatedness in pathways, protein-protein networks and functional annotations, e.g. from Gene Ontology [75].

We concentrate on predicting presumably direct activities in this paper, thus we demonstrate the capability of the developed method and the effect multiple information sources using multiple chemical similarities, although the method can incorporate symmetrically multiple target-target similarities. Furthermore, the method can also incorporate separate prior expectations about the success rates of drugs in a given DTI, which could be combined with drug-likeliness [76], promiscuity prediction [77] and decoy prediction in case of their use [78]. Symmetrically, it can also incorporate separate prior expectations about the success rates of targets in a given DTI, which could be combined with druggability predictions [69, 79, 80] and the presence of pockets [81]. For an overview of available resources relevant for the DTI task, see e.g. [82, 83].

DTI methods

The rapid growth, especially the public availability of tabular (dyadic) DTI data in the last decade caused a dramatic shift of the applied statistical methods. For an overview of classical single prediction oriented machine learning and data mining in drug discovery, especially in DTI and ADME predictions, see e.g. [84], for large-scale, comprehensive applications of DTI data, see e.g. [85]. The tabular nature of the DTI data called for new methods not only handling this type of data natively, but also capable of using side information. Transfer learning and multitask learning paradigms addressed this challenge [86, 87, 8], but in the DTI context, two groups of methods, the pairwise conditional methods and the matrix factorization based generative methods proved to be particularly successful.

Pairwise conditional approaches or pairwise kernel methods flatten the dyadic structure of the DTI data and use drug and target descriptors, optionally even explanatory descriptors about the drug-target relations to predict interaction properties of drug-target pairs (for the assumptions behind the conditional approach, see e.g. [88], for its early DTI application, see e.g. [89]). Classification and regression methods, such as MLPs, decision trees and SVMs remain directly applicable in this conditional approach (not modeling the distribution of the drug-target pairs), however, the high number of drug-target pairs is challenging for kernel based methods [90, 50], but recent developments in deep learning show promising results [91].

Using multiple representations for drugs and targets is directly possible in this pairwise approach, but the construction of an aggregate pair-pair (interaction-interaction) similarity or an efficient set of pair-pair similarities from drug-drug and target-target similarities is an open problem. In the case of single drug-drug and target-target similarities, the Kroneckerian combination was proposed in the work of van Laarhooven [90] with corresponding computational simplifications to maintain scalability. Additionally, kernel techniques were extended to use multiple kernels, which are potentially derived from heterogeneous representations and similarities [50]. Recent extensions include non-linear kernel fusion in the RLS-KF system [49] and using boosting to learn from unscreened controls [53].

Matrix factorization (MF) methods differ from pairwise approaches in multiple properties crucial in the DTI task. The central operation of these methods is the construction of a joint space with latent factors for drugs and targets and modeling their interactions based on the inner product of the respective vectorial representations. Contrary, pairwise approaches, such as kernel methods or deep learning cannot directly exploit the tabular prior constraint of the data. The MF approach also allows the direct incorporation of drug-drug similarities and target-target similarities. Additionally, the low dimensionality of the latent space supports data visualization, although its interpretation is still in its infancy. Finally, probabilistic MF methods construct a distribution over the latent representations of drugs and targets, which in fact means that they are full-fledged generative models.

Matrix factorization methods were adopted early in gene expression data analysis [92, 93]. They were used for dimensionality reduction and the construction of a unified space for ligands and receptors [94], applied in biomedical text-mining and [95] and chemogenomics [96]. Later in the 2000's media and e-commerce recommendation applications dominated the research of matrix factorization methods [97] and many developments were motivated and reported in these contexts, such as solutions for new items without interactions, selection bias, model regularization, automated parameter selection and incorporation of side information from multiple sources. An early work from Srebro et al. addressed the problems of using weights to represent importance or trust in the observations and the use of logistic regression as a non-linear transformation to predict probabilities of binary observations [98]. A special weighting of observations compared to unknowns were investigated in [99]. Salakhutdinov introduced Bayesian matrix factorization, which addressed regularization and automated parameter selection by Bayesian model averaging, also indicating the principled and flexible options for prior incorporation [100]. Severinski demonstrated the advantages of the full Bayesian approach versus a Maximum a Posteriori based alternative in this context [101]. Zhou introduced Gaussian process priors over the latent dimensions to enforce two kernels over row and column items [102]. Lobato et al. reported a variational Bayesian approach for logistic matrix factorization [103].

In the DTI context, an early kernel regression-based method (KRM) was reported in [7], and emphasized the advantages of a unified "pharmacological space". Gönen introduced a kernelized Bayesian matrix factorization (KBMF) [22], which applies kernel-based averaging over the latent vectorial representations of rows and columns. The paper also introduced an efficient variational Bayesian approximation and indicated the interpretability of the latent space. Zheng et al. proposed a

non-probabilistic multiple kernel learning approach, which achieved superior performance [23]. Multiple kernel learning was also realized in KBMF [27] and was also extended towards regression [104]. Special non-missing-at-random DTI data models were proposed in [51], which applied Gaussian priors to incorporate multiple kernels and used Gibbs sampling to approximate the posteriors. In an integrative work, Liu et al. proposed the combination of special neighborhood restricted kernels, network-based regularization, importance weights for the observations and logistic link functions in a non-Bayesian framework [47]. A recent extension applied a non-linear kernel diffusion technique to boost relevant, complementary information in similarity matrices [48].

DTI benchmarks

The most widely used DTI benchmark from Yamanishi et al. [7] defined DTI prediction as a binary prediction problem with a single source of drug-drug and a target-target similarity, which induced the development of variety of methods and datasets (see Additional file 2). These datasets are still in the range of 1000×1000 and contain $10k$ interactions, but they inherit the problem of the selection bias present in the DTI repositories [12, 64, 105, 106, 11, 82]. Pahikkala et al. stressed the importance of fully observed bioactivity values in benchmarks [8], such as from Davis [9], to avoid misleading results because of selection bias, indirect interactions and the binary nature of the interactions. Liu et al. [47] reported a comprehensive evaluation of methods and released a corresponding benchmark implementation, the *pyDTI* package. For real, experimental evaluation of DTI methods, see e.g. [107, 108].

Methods

Our work directly builds upon Gönen's work on kernel-based matrix factorization using twin kernels (KBMF-MKL), which applied variational Bayesian approximations [27]. Another direct predecessor of our work is Liu et al's neighborhood regularized logistic matrix factorization [47].

Materials

To maintain consistency with earlier works, we evaluated the methods on the data sets provided by Yamanishi *et al.* [7] and Pahikkala *et al.* [8]. While the latter comes with multiple similarity matrices based on various molecular fingerprints, the former is one-kernel and therefore needed to be extended to properly test the MKL performance. We used the RDKit package [109] to compute additional MACCS and Morgan fingerprints for the molecules and used these in conjunction with the Tanimoto and Gaussian RBF similarity measures. Target similarities were obtained from Nascimento *et al.* [50] which utilized sequential, GO- and PPI-based similarities.

Probabilistic model

Let $\mathbf{R} \in \{0,1\}^{I \times J}$ denote the matrix of the interactions, where $\mathbf{R}_{ij} = 1$ indicates a known interaction between the ith drug and jth target. In order to formulate a Bayesian model, we put a Bernoulli distribution on each $\mathbf{R}_{ij}$ with parameter $\sigma(\mathbf{u}_i^T \mathbf{v}_j)$ where σ is the logistic sigmoid function and $\mathbf{u}_i$, $\mathbf{v}_j$ are the ith and jth

columns of the respective factor matrices $\mathbf{U} \in \mathbb{R}^{L \times I}$ and $\mathbf{V} \in \mathbb{R}^{L \times J}$. One can think of $\mathbf{u}_i$ and $\mathbf{v}_j$ as L-dimensional latent representations of the ith drug and jth target, and the *a posteriori* probability of an interaction between them is modeled by $\sigma(\mathbf{u}_i^T \mathbf{v}_j)$.

Similarly to NRLMF, we utilize an augmented version of the Bernoulli distribution parameterized by $c \geq 1$ which assigns higher importance to observations (positive examples). NRLMF also uses a post-training weighted average to infer interactions corresponding to empty rows and columns in $\mathbf{R}$ (*i.e.* these would have to be estimated without using any corresponding observations). We account for them by introducing variables $\mathbf{m}^u, \mathbf{m}^v \in \{0,1\}$ indicating whether the row or column is empty. In these cases, only the side information will be used in the prediction. The conditional on the interactions can be written as

$$p(\mathbf{R} \mid \mathbf{U}, \mathbf{V}, c, \mathbf{m}^u, \mathbf{m}^v) \propto \prod_i \prod_j \left[\left(\sigma \left(\mathbf{u}_i^T \mathbf{v}_j \right) \right)^{c \mathbf{R}_{ij}} \left(1 - \sigma \left(\mathbf{u}_i^T \mathbf{v}_j \right) \right)^{1 - \mathbf{R}_{ij}} \right]^{\mathbf{m}_i^u \mathbf{m}_j^v}.$$
(1)

Specifying priors on $\mathbf{U}$ and $\mathbf{V}$ presents an opportunity to incorporate multiple sources of side information. In particular, we can use a Gaussian distribution with a weighted linear combination of kernel matrices $\mathbf{K}_n$, $n = 1, 2, \ldots$ in the precision matrix, which corresponds to a combined L_2-Laplacian regularization scheme [36]

$$p(\mathbf{U} \mid \alpha^u, \gamma^u, \mathbf{K}^u) \propto \prod_i \prod_k \exp \left\{ -\frac{1}{2} \sum_n \gamma_n^u \mathbf{K}_{n,ik}^u \left\| \mathbf{u}_i - \mathbf{u}_k \right\|^2 \right\} \cdot \prod_i \exp \left\{ -\frac{\alpha^u}{2} \left\| \mathbf{u}_i \right\|^2 \right\}.$$
(2)

The prior on $\mathbf{V}$ can be written similarly. To automate the learning of the optimal value of kernel weights γ_n^u, we introduce another level of uncertainty using Gamma priors:

$$p(\gamma_n^u \mid a, b) = \frac{b^a (\gamma_n^u)^{a-1} e^{-b \gamma_n^u}}{\Gamma(a)}.$$
(3)

Variational approximation

In the Bayesian approach, the combination of data $\mathbf{R}$ and prior knowledge through kernel matrices $\mathbf{K}_n$ and hyperparameters defines the posterior

$$p(\mathbf{U}, \mathbf{V}, \gamma^u, \gamma^v | \mathbf{R}, \mathbf{K}_n^u, a^u, b^u, \mathbf{R}, \mathbf{K}_n^v, a^v, b^v, \alpha^u, \alpha^v, c).$$

In the variational setting [110], we obtain an approximation

$$q(\mathbf{U}, \mathbf{V}, \gamma^u, \gamma^v)$$

by maximizing a lower bound on the expectation

$$p(\mathbf{R}) = \int p(\mathbf{R} \mid \mathbf{U}, \mathbf{V})p(\mathbf{U} \mid \gamma^u)p(\mathbf{V} \mid \gamma^v)p(\gamma^u)p(\gamma^v)d\mathbf{U}d\mathbf{V}d\gamma^u d\gamma^v,$$

with respect to $\mathbf{U}$, $\mathbf{V}$, γ^u, γ^v, where we suppressed the hyperparameters for notational simplicity. This is achieved by using a factorized variational distribution

$$q(\mathbf{U}, \mathbf{V}, \gamma^u, \gamma^v) = q(\mathbf{U})q(\mathbf{V})q(\gamma^u)q(\gamma^v)$$

and using the equality

$$\ln p(\mathbf{R}) = \mathcal{L}(q) + KL\left(q \mid\mid p\right),$$

where $KL(\cdot \mid\mid \cdot)$ is the Kullback–Leibler divergence and $\mathcal{L}(\cdot)$ is the expectation lower bound. In particular,

$$\mathcal{L}(q) = \int q(\mathbf{U})q(\mathbf{V})q(\gamma^u)q(\gamma^v) \ln \left\{ \frac{p(\mathbf{R}, \mathbf{U}, \mathbf{V}, \gamma^u, \gamma^v)}{q(\mathbf{U})q(\mathbf{V})q(\gamma^u)q(\gamma^v)} \right\} d\mathbf{U}d\mathbf{V}d\gamma^u d\gamma^v,$$

which is the quantity we aim to maximize with respect to q as it means an improved approximation to the posterior (note that the quantity $\ln p(\mathbf{R})$ on the left side is constant).

The optimal distribution satisfies

$$\ln q^*(\mathbf{U}) = E_{\mathbf{V}, \gamma^u, \gamma^v} \left[\ln \left\{ p(\mathbf{R} \mid \mathbf{U}, \mathbf{V})p(\mathbf{U} \mid \gamma^u)p(\mathbf{V} \mid \gamma^v)p(\gamma^u)p(\gamma^v) \right\} \right] + \text{const.}$$

which is non-conjugate due to the form of $p(\mathbf{R} \mid \mathbf{U}, \mathbf{V})$ and therefore the integral is intractable. However, by using Taylor approximation on the symmetrized logistic function (Jaakkola's bound [111, 103])

$$\sigma(z) \geq \tilde{\sigma}(z, \xi) = \sigma(\xi) \exp \left\{ \frac{z - \xi}{2} - \frac{1}{2\xi} \left(\sigma(\xi) - \frac{1}{2} \right) \left(z^2 - \xi^2 \right) \right\},$$

we can lower bound $p(\mathbf{R} \mid \mathbf{U}, \mathbf{V})$ at the cost of introducing local variational parameters ξ_{ij}, yielding a new bound $\tilde{\mathcal{L}}$ which contains at most quadratic terms. Collecting the terms containing $\mathbf{U}$ gives (see the proof in Additional file 1):

$$\ln q^*(\mathbf{U}) = -\frac{1}{2} \operatorname{tr}\left(\mathbf{U}^T \mathbf{Q}^u \mathbf{U}\right) + \sum_i \mathbf{u}_i^T \left(\sum_j \hat{\mathbf{R}}_{ij} \hat{\xi}_{ij} E\left[\mathbf{v}_j \mathbf{v}_j^T\right] \right) \mathbf{u}_i + \sum_i \mathbf{u}_i^T \left(\sum_j \mathbf{R'}_{ij} E\left[\mathbf{v}_j\right] \right)$$

where

$$\mathbf{Q}^u = \frac{E\left[\gamma_u\right]}{2} \left(\mathbf{K}^{uT}\mathbf{1} - \mathbf{K}^u \right) + \frac{\alpha_u}{2}\mathbf{I},$$

$$\hat{\xi}_{ij} = -\frac{1}{2\xi_{ij}} \left(\sigma(\xi_{ij}) - \frac{1}{2} \right),$$

$$\hat{\mathbf{R}}_{ij} = \mathbf{m}_i^u \mathbf{m}_j^v \left((c-1)\mathbf{R}_{ij} + 1 \right),$$

$$\mathbf{R'}_{ij} = \mathbf{m}_i^u \mathbf{m}_j^v c\mathbf{R}_{ij} + \frac{1}{2}\hat{\mathbf{R}}_{ij}.$$

Since this expression is quadratic in $\mathrm{vec}(\mathbf{U})$, we conclude that q^* is Gaussian and the parameters can be found by completing the square. In particular,

$$q^*(\mathrm{vec}(\mathbf{U})) = \mathcal{N}(\mathrm{vec}(\mathbf{U}) \mid \phi, \mathbf{\Lambda}^{-1})$$

$$\mathbf{\Lambda} = \mathbf{Q}^u \otimes \mathbf{I} - 2 \cdot \mathrm{blkdg}_i \left(\sum_j \hat{\mathbf{R}}_{ij} \hat{\xi}_{ij} E\left[\mathbf{v}_j \mathbf{v}_j^T\right] \right), \tag{4}$$

$$\phi = \mathbf{\Lambda}^{-1} \mathrm{vec}_i \left(\sum_j \mathbf{R}'_{ij} E\left[\mathbf{v}_j\right] \right), \tag{5}$$

where blkdg_i denotes the operator creating an $L \cdot I \times L \cdot I$ block-diagonal matrix from I $L \times L$-sized blocks. The variational update for $q(\mathbf{V})$ can be derived similarly. The most computationally intensive operation is computing

$$E\left[\mathbf{v}_j \mathbf{v}_j^T\right] = \mathrm{Cov}(\mathbf{v}_j) + E\left[\mathbf{v}_j\right] E\left[\mathbf{v}_j\right]^T \tag{6}$$

which requires the inversion of $\mathbf{\Lambda}$, performed using blocked Cholesky decomposition.

The optimal value of the local variational parameters ξ_{ij} can be computed by writing the expectation of the joint distribution in terms of ξ and setting its derivative to zero. In particular,

$$\tilde{\mathcal{L}}(\xi) = \sum_i \sum_j \hat{\mathbf{R}}_{ij} \left(\ln \sigma(\xi_{ij}) - \frac{\xi_{ij}}{2} - \frac{1}{2\xi_{ij}} \left(\sigma(\xi_{ij}) - \frac{1}{2} \right) \left(\xi_{ij}^2 - E\left[\left(\mathbf{u}_i^T \mathbf{v}_j\right)^2 \right] \right) \right),$$

from which [112, 103]

$$\xi_{ij}^2 = E\left[\left(\mathbf{u}_i^T \mathbf{v}_j\right)^2 \right]$$
$$= \left(E\left[\mathbf{u}_i\right]^T E\left[\mathbf{v}_j\right] \right)^2 + \sum_l E\left[\mathbf{U}_{li}\right]^2 V\left[\mathbf{V}_{lj}\right] + V\left[\mathbf{U}_{li}\right] E\left[\mathbf{V}_{lj}\right]^2 + V\left[\mathbf{U}_{li}\right] V\left[\mathbf{V}_{lj}\right]. \tag{7}$$

Since the model is conjugate with respect to the kernel weights, we can use the standard update formulas for the Gamma distribution

$$q^*(\gamma_n^u) = \mathcal{G}amma(\gamma_n^u \mid a', b')$$

$$a' = a + \frac{I^2}{2} \tag{8}$$

$$b' = b + \frac{1}{2} E_{\mathbf{U}} \left[\sum_i \sum_k \mathbf{K}_{n,ik}^u \|\mathbf{u}_i - \mathbf{u}_k\|^2 \right]$$

$$= b + \frac{1}{2} \sum_i \sum_k \mathbf{K}_{n,ik}^u \left(E\left[\mathbf{u}_i^T \mathbf{u}_i\right] - 2E\left[\mathbf{u}_i^T \mathbf{u}_k\right] + E\left[\mathbf{u}_k^T \mathbf{u}_k\right] \right), \tag{9}$$

which also requires the explicit inversion of $\mathbf{\Lambda}$. Figure 1 shows the pseudocode of the algorithm.

Results

We present the results of a systematic comparison with KBMF-MKL [27], NRLMF [47] and KronRLS-MKL [50]. Subsequently, our results show the effect of prior knowledge fading with increasing data size.

Experimental settings

Predictive performance was evaluated in a $5\times$ 10-fold cross-validation framework. To maintain consistency with the evaluations in earlier works, we utilized the CVS1-CVS2-CVS3 settings as presented in [47] and calculated the average AUROC and AUPRC values in each scenario. In particular, CVS1 corresponds to evaluating predictive performance after randomly blinding 10% of the interactions and using them as test entities. CVS2 corresponds to random drugs (entire rows blinded) and CVS3 corresponds to random targets. We used the same folds as the PyDTI tool to maximize comparability.

In the single-kernel setting, we compared the performance of the proposed method to KBMF, NRLMF and KronRLS. The optimal parameters for NRLMF were obtained from the original publication [47]. KBMF and KronRLS were parameterized using a grid search method. VB-MK-LMF was used with 3 neighbors in each kernel, $\alpha_u = \alpha_v = 0.1$, $a_u = a_v = 1$, $b_u = b_v = 10^3$ and $c = 10$. The number of latent factors was set to $L = 10$ in the Nuclear Receptor dataset and $L = 15$ in the others, and a more detailed investigation of this parameter was also conducted. The number of iterations was chosen manually as 20 since the variational parameters usually converged between $20 - 50$ iterations.

In the multiple-kernel setting, we compared the performance of the proposed method to KBMF-MKL and KronRLS-MKL using MACCS and Morgan fingerprints with RBF and Tanimoto similarities. Target kernels provided by KronRLS-MKL did not improve the results in either case, thus only the ones computed by Yamanishi et al. were utilized. We also investigated the weights assigned to the kernels and tested robustness by introducing kernels with random values.

Systematic evaluation

Single-kernel results are shown in Table 1. In most cases, VB-MK-LMF significantly outperforms NRLMF and one-kernel KBMF in terms of AUROC and AUPRC according to a pairwise t-test. Overall, the improvement is more modest on the Enzyme dataset, although still significant in some cases. This can be attributed to the fact that this dataset is by far the largest, which can mitigate the benefits of Bayesian model averaging and side information. On average, VB-MK-LMF yields 4.7% higher AUPRC values in the pairwise cross-validation setting than the second best method. In the drug and target settings, this is 2% and 7.6%, respectively. The lower AUROC and AUPRC values in these scenarios are explained by the lack of observations for the test drugs or targets in the training set, resulting in a harder task than in the pairwise scenario.

Following earlier investigations, we examined the number of latent factors, which has a crucial role from computational, statistical and interpretational aspects. Contrary to earlier works [23], which recommend $50 - 100$ as the number of latent factors, we found that these values do not yield better results; in fact, the AUPRC

values quickly become saturated. Conceptually, it is unclear what is to be gained going beyond the rank of the original matrix, which corresponds to perfect factorization with respect to the Frobenius norm when using SVD, and is also known to lead to serious overfitting in unregularized cases [98, 100]. Although overfitting is usually less of an issue with variational Bayesian approximations, a large number of latent factors significantly increases computational time. Figure 3 depicts the AUPRC values on the smaller datasets with varying number of latent factors. The Enzyme and Kinase datasets were not included in this experiment due to the rapidly increasing runtime.

Multi-kernel AUPRC values are shown in Table 2. Compared to the previous Table, it is clear that both VB-MK-LMF and KBMF benefits from using multiple kernels. Moreover, there is also an improvement in predictive performance when one combines instances of the same kernel but with different neighbor truncation values. However, advantages of using both of these combination schemes simultaneously are unclear as the results usually do not improve or even get worse (except for the Kinase dataset). This is a known property of linear kernel combinations, i.e. using large linear kernel combinations may not improve predictive performance beyond that of the best individual kernels in the combination [113].

Table 3 shows the normalized kernel weights in each of the datasets. For illustration purposes, we also included a unit-diagonal positive definite kernel matrix with random values. In the first four datasets, the algorithm assigned more or less uniform weights to the real kernels and a lower one to the random kernel. In the Kinase dataset, the random kernel is almost zeroed out. This underlines the validity of VB-MK-LMF's kernel combination scheme. Setting L to I (the rank of the kernels) yields an almost zero weight to the random kernel, i.e. allowing larger dimensions also allows sufficient separation of the latent representations, which makes spotting kernels with erroneous values easier for the algorithm. This property might also justify increasing the number of latent factors beyond the rank of the interaction matrix in the multi-kernel setting.

To understand the effect of priors behind the significantly improved performance, which is especially pronounced at smaller sample sizes, we investigated the difference in AUPRC and AUROC values while using and ignoring kernels, at varying training set sizes. The results suggest the existence of a "small sample size" region where using side information offer significant gains, and after which the effect of priors gradually vanishes. Figure 4 depicts the learning curves.

Discussion
VB-MK-LMF introduces a matrix factorization model incorporating multiple kernel learning, Laplacian regularization and the explicit modeling of interaction probabilities, for which a variational Bayesian inference method is proposed. The algorithm maps each drug and target into a joint vector space and interaction probabilities are derived from the inner products of the latent representations. Despite the suggested applicability of the unified "pharmacological space" [7], its semantics is still unexplored (for an early application in a ligand-receptor space, see [94], for a proof-of-concept illustration, see [22]). To facilitate a deeper understanding, we provide visual analytics tools alongside the factorization algorithm and allow arbitrary annotations to be mapped onto the latent representations.

We demonstrate this on the Ion Channel dataset. Using $L = 2$, the resulting latent representations can be visualized in a 2D Cartesian coordinate system as shown in Figure 5. Drugs are colored on the basis of their respective ATC classes, where only the classes with more than 5 members were used. Targets are colored according to their ion transporter activity as obtained from the Gene Ontology. Known interactions are represented as edges. Even in this low-dimensional case, drugs in the same class tend to cluster together. The only exception is the "Other antiepileptics" class, which is easily explained by its heterogeneity, also indicated by the name. Targets also cluster fairly nicely, albeit with somewhat more outliers. It can be also observed that the targets exhibiting potassium and sodium transporter activity are placed halfway between the sodium and potassium groups.

Similarly, Figure 6 depicts the joint space using a parallel coordinates visualization with $L = 10$, where ion transporter activity is denoted by different colors. Most of the dimensions tend to separate at least one class from the others and many of them seem to distinguish between more than two classes. This indicates that the algorithm manages to find biologically meaningful latent dimensions, possibly encoding pharmacophore properties and the properties of binding sites, but we leave it for further exploration.

From a more practical viewpoint, it is important to touch on the issue of drug promiscuity and polypharmacology. This refers to the observation that some drugs tend to act on multiple targets leading to distinct pharmacological effects, which is often considered an undesirable property [85], although partly unavoidable and potentially utilizable [114]. In either case, predicting the expected number of interactions in a restricted set of targets is a unique property of probabilistic DTI predictors, e.g. compared to ranking approaches. To illustrate this ability of VB-MK-LMF, we computed the expected value of the total number of interactions for every drug in all datasets, treating them independently, shown in Figure 7 together with the number known targets. Overall, the expected value of further hits approximates the number of interactions already discovered rather closely, although it tends to over-estimate, especially when only one or two interactions are known. We also conducted a $10\times$ cross-validation experiment for each drug in the GPCR dataset and performed the same comparison with similar results (Figure 8). It is worth to mention that the number of currently unobserved positive interactions in large-scale settings and in comprehensive DTI repositories is vital for the pharmaceutical industry and an open scientific question, as indicated by research on drug-likeliness and druggability. Assuming total independence, the expected value provides a raw estimate for this. However, as the relative frequency of positive interactions among the unobserved cases should influence the selection of weight for the observed cases (c), and the value of c influences the expected value, resolving this circular situation and tuning c requires further investigations.

We also performed a case-based evaluation by obtaining the top 5 novel predictions in the incomplete datasets and examining whether they are present in the current version of the DrugBank database. Most interactions were confirmed and some of the unconfirmed hits are known to bind to other members of that particular protein family. This shows the ability of VB-MK-LMF to predict novel interactions. The predicted lists are similar to those of the NRLMF method. Table 4 illustrates these

results and also contains the rank of the predicted interactions among the NRLMF predictions.

Finally, we discuss computational issues. Due to the explicit computation of inverse matrices, the variational approximation is highly compute-intensive, however, it is straightforward to parallelize and many steps can be written as BLAS operations. GPUs are particularly well-suited for this task. All computations presented in this work can be performed on a mid-range graphics card. Figure 9 shows the runtime of GPU and CPU implementations in terms of latent factors 200×200 matrix factorization task, which showed a $30\times$ speedup using an NVIDA Titan X graphics card. However, in larger dimensions or with many latent factors, one can quickly run out of GPU memory, i.e. scaling remains an open question. Although GPUs provide excellent performance with single precision, double precision performance typically lags far behind, especially with modern consumer-level graphics cards. This raises the issue of numerical stability. To cope with the memory footprint of the algorithm, we provide a sparse implementation beside the standard dense solver. To address the issue of numerical stability, we also provide a QR factorization-based implementation which is more stable but significantly slower than the default Cholesky-based method. The computation in VB-MK-LMF is dominated by the inversion in Eq. 6, which gives $\mathcal{O}(DL^3 \max(I^3, J^3))$ for the total time complexity (D is the number of iterations). Comparison with the time complexity of NRLMF, $\mathcal{O}(DLIJ)$, clearly shows the burden of Bayesian computation in the current implementation and calls for the usage of approximative inversion techniques, which we consider as a future work.

Conclusion

We presented Variational Bayesian Multiple Kernel Logistic Matrix Factorization (VB-MK-LMF), integrating multiple kernel learning, weighted observations, graph Laplacian regularization, and explicit modeling of probabilities of binary drug-target interactions. Compared to other state-of-the-art methods, VB-MK-LMF achieved significantly better predictive performance in standard benchmarks.

Admittedly, benchmarking the pure predictive performance on a given dataset gives a very focused view about the real-world applicability of the methods, but helps comparability. On the other hand, the release of new and updated datasets as shown in Additional file 2 in fact quickly create an impractical fragmentary situation. In general, the definition of a standard background knowledge pool for a benchmarking is even more complicated, as earlier attempts show in computational fusion methods for gene prioritization [115, 116].

Additionally, currently the possible utilizations of a DTI prediction method in real-world applications are at least as diverse as the methodological repertoire. For example, DTI prediction methods could be applied in data quality control phase for anomaly detection, especially in the case of merging different bioactivity values from public and private sources. Screening design, hit triage and prioritization for further validation [117], possibly in an active learning framework [16, 118], are standard usages. Finally, DTI prediction methods may also provide essential data to support visualization and visual data analytics, as we demonstrated in a new range of dimensionality $(10 - 20)$, which proved to be sufficient with VB-MK-LMF.

Another key property of VB-MK-LMF is the explicit modeling of probabilities, which allows the prediction of interaction probabilities and their credibility. We demonstrated the use of probabilistic predictions by proposing DTI dataset specific versions of promiscuity and druggability, through the expected number of hits in a dataset for a drug or a target respectively. In general, the predicted posteriors for the interactions can be seen as a probabilistic "data-analytic" knowledge base, which allows new functionalities in post-processing, beyond enrichment methods available for ranking methods [37, 33]. To utilize the Bayesian predictions of VB-MK-LMF, we also plan to investigate their decision theoretic usage, when certainty for expected gains and losses of prioritization of interactions is expected, e.g. in functional validations.

Further interesting research directions are the regression version of VB-MK-LMF directly approximating the continuous activity data [8, 51] and the use of multiple instances of VB-MK-LMF for overlapping DTI matrices, which are linked to each other by weighted common observations. The latter could improve the scalability of the method using parallel implementations for mid-sized DTI tasks with 10^5 drugs and 10^4 targets, going beyond the current benchmarks.

Competing interests
The authors declare that they have no competing interests.

Author's contributions
Designed the experiments: BB AP. Developed the software: BB. Performed the experiments: BB. Analyzed the data: BB AP. Wrote the paper: BB AP. All authors read and approved the final manuscript.

References
1. Williams, A.J., Ekins, S., Tkachenko, V.: Towards a gold standard: Regarding quality in public domain chemistry databases and approaches to improving the situation. Drug Discovery Today **17**(13-14), 685–701 (2012). doi:10.1016/j.drudis.2012.02.013
2. Goldmann, D., Montanari, F., Richter, L., Zdrazil, B., Ecker, G.F.: Exploiting open data: a new era in pharmacoinformatics. Future medicinal chemistry **6**(5), 503–514 (2014). doi:10.4155/fmc.14.13
3. Chen, X., Yan, C.C., Zhang, X., Zhang, X., Dai, F., Yin, J., Zhang, Y.: Drug-target interaction prediction: Databases, web servers and computational models. Briefings in Bioinformatics **17**(4), 696–712 (2016). doi:10.1093/bib/bbv066
4. Zheng, W., Thorne, N., McKew, J.C.: Phenotypic screens as a renewed approach for drug discovery. Drug discovery today **18**(21), 1067–1073 (2013)
5. Orchard, S., Al-Lazikani, B., Bryant, S., Clark, D., Calder, E., Dix, I., Engkvist, O., Forster, M., Gaulton, A., Gilson, M., Glen, R., Grigorov, M., Hammond-Kosack, K., Harland, L., Hopkins, A., Larminie, C., Lynch, N., Mann, R.K., Murray-Rust, P., Lo Piparo, E., Southan, C., Steinbeck, C., Wishart, D., Hermjakob, H., Overington, J., Thornton, J.: Minimum information about a bioactive entity (MIABE). Nature reviews. Drug discovery **10**(9), 661–669 (2011). doi:10.1038/nrd3503
6. Samwald, M., Jentzsch, A., Bouton, C., Kallesøe, C.S., Willighagen, E., Hajagos, J., Scott Marshall, M., Prud'hommeaux, E., Hassanzadeh, O., Pichler, E., Stephens, S.: Linked Open drug data for pharmaceutical research and development. Journal of Cheminformatics **3**(5), 19 (2011). doi:10.1186/1758-2946-3-19
7. Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., Kanehisa, M.: Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics **24**(13), 232–240 (2008). doi:10.1093/bioinformatics/btn162
8. Pahikkala, T., Airola, A., Pietilä, S., Shakyawar, S., Szwajda, A., Tang, J., Aittokallio, T.: Toward more realistic drug-target interaction predictions. Briefings in Bioinformatics **16**(2), 325–337 (2015). doi:10.1093/bib/bbu010
9. Davis, M.I., Hunt, J.P., Herrgard, S., Ciceri, P., Wodicka, L.M., Pallares, G., Hocker, M., Treiber, D.K., Zarrinkar, P.P.: Comprehensive analysis of kinase inhibitor selectivity. Nature Biotechnology **29**(11), 1046–1051 (2011). doi:10.1038/nbt.1990. 0402594v3
10. Schomburg, I., Chang, A., Placzek, S., Söhngen, C., Rother, M., Lang, M., Munaretto, C., Ulas, S., Stelzer, M., Grote, A., Scheer, M., Schomburg, D.: BRENDA in 2013: Integrated reactions, kinetic data, enzyme function data, improved disease classification: New options and contents in BRENDA. Nucleic Acids Research **41**(D1), 1–9 (2013). doi:10.1093/nar/gks1049
11. Lindh, M., Svensson, F., Schaal, W., Zhang, J., Sköld, C., Brandt, P., Karlén, A.: Toward a Benchmarking Data Set Able to Evaluate Ligand- and Structure-based Virtual Screening Using Public HTS Data. Journal of Chemical Information and Modeling, 150128112200004 (2015). doi:10.1021/ci5005465
12. Mervin, L.H., Afzal, A.M., Drakakis, G., Lewis, R., Engkvist, O., Bender, A.: Target prediction utilising negative bioactivity data covering large chemical space. Journal of Cheminformatics **7**(1), 1–16 (2015). doi:10.1186/s13321-015-0098-y

13. Liu, C., Su, J., Yang, F., Wei, K., Ma, J., Zhou, X.: Compound signature detection on LINCS L1000 big data. Mol. BioSyst. **11**(3), 714–722 (2015). doi:10.1039/C4MB00677A

14. Kövesdi, I., Dominguez-Rodriguez, M.F., Ôrfi, L., Náray-Szabó, G., Varró, A., Papp, J.G., Matyus, P.: Application of neural networks in structure–activity relationships. Medicinal research reviews **19**(3), 249–269 (1999)

15. Burbidge, R., Trotter, M., Buxton, B., Holden, S.: Drug design by machine learning: support vector machines for pharmaceutical data analysis. Computers & chemistry **26**(1), 5–14 (2001)

16. Warmuth, M.K., Liao, J., Rätsch, G., Mathieson, M., Putta, S., Lemmen, C.: Active learning with support vector machines in the drug discovery process. Journal of chemical information and computer sciences **43**(2), 667–673 (2003)

17. Willett, P., Barnard, J.M., Downs, G.M.: Chemical similarity searching. Journal of chemical information and computer sciences **38**(6), 983–996 (1998)

18. Ginn, C.M., Willett, P., Bradshaw, J.: Combination of molecular similarity measures using data fusion. In: Virtual Screening: An Alternative or Complement to High Throughput Screening?, pp. 1–16. Springer, ??? (2000)

19. Ding, H., Takigawa, I., Mamitsuka, H., Zhu, S.: Similarity-based machine learning methods for predicting drug-target interactions: a brief review. Briefings in bioinformatics, 056 (2013). doi:10.1093/bib/bbt056

20. Kitchen, D.B., Decornez, H., Furr, J.R., Bajorath, J.: Docking and scoring in virtual screening for drug discovery: methods and applications. Nature reviews Drug discovery **3**(11), 935–949 (2004)

21. Sousa, S.F., Fernandes, P.A., Ramos, M.J.: Protein–ligand docking: current status and future challenges. Proteins: Structure, Function, and Bioinformatics **65**(1), 15–26 (2006)

22. Gönen, M.: Predicting drug–target interactions from chemical and genomic kernels using bayesian matrix factorization. Bioinformatics **28**(18), 2304–2310 (2012)

23. Zheng, X.: Collaborative Matrix Factorization with Multiple Similarities for Predicting Drug-Target Interactions. Proceeding KDD '13 Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, 1025–1033 (2013). doi:10.1145/2487575.2487670

24. Waller, C.L., Shah, A., Nolte, M.: Strategies to support drug discovery through integration of systems and data. Drug discovery today **12**(15), 634–639 (2007)

25. Muresan, S., Petrov, P., Southan, C., Kjellberg, M.J., Kogej, T., Tyrchan, C., Varkonyi, P., Xie, P.H.: Making every SAR point count: The development of Chemistry Connect for the large-scale integration of structure and bioactivity data. Drug Discovery Today **16**(23-24), 1019–1030 (2011). doi:10.1016/j.drudis.2011.10.005

26. Agrafiotis, D.K., Alex, S., Dai, H., Derkinderen, A., Farnum, M., Gates, P., Izrailev, S., Jaeger, E.P., Konstant, P., Leung, A., Lobanov, V.S., Marichal, P., Martin, D., Rassokhin, D.N., Shemanarev, M., Skalkin, A., Stong, J., Tabruyn, T., Vermeiren, M., Wan, J., Xu, X.Y., Yao, X.: Advanced Biological and Chemical Discovery (ABCD): Centralizing discovery knowledge in an inherently decentralized world. Journal of Chemical Information and Modeling **47**(6), 1999–2014 (2007). doi:10.1021/ci700267w

27. Gönen, M., Khan, S., Kaski, S.: Kernelized bayesian matrix factorization. In: International Conference on Machine Learning, pp. 864–872 (2013)

28. Cheng, F., Liu, C., Jiang, J., Lu, W., Li, W., Liu, G., Zhou, W., Huang, J., Tang, Y.: Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS Computational Biology **8**(5) (2012). doi:10.1371/journal.pcbi.1002503

29. Fu, G., Ding, Y., Seal, A., Chen, B., Sun, Y., Bolton, E.: Predicting drug target interactions using meta-path-based semantic network analysis. BMC bioinformatics **17**(1), 160 (2016)

30. Ashburn, T.T., Thor, K.B.: Drug repositioning: identifying and developing new uses for existing drugs. Nature reviews Drug discovery **3**(8), 673–683 (2004)

31. Li, J., Zheng, S., Chen, B., Butte, A.J., Swamidass, S.J., Lu, Z.: A survey of current trends in computational drug repositioning. Briefings in bioinformatics **17**(1), 2–12 (2016)

32. Arany, A., Bolgár, B., Balogh, B., Antal, P., Mátyus, P.: Multi-aspect candidates for repositioning: data fusion methods using heterogeneous information sources. Current medicinal chemistry **20**(1), 95–107 (2013)

33. Temesi, G., Bolgár, B., Arany, Á., Szalai, C., Antal, P., Mátyus, P.: Early repositioning through compound set enrichment analysis: a knowledge-recycling strategy. Future medicinal chemistry **6**(5), 563–575 (2014)

34. Liu, Z., Guo, F., Gu, J., Wang, Y., Li, Y., Wang, D., Lu, L., Li, D., He, F.: Similarity-based prediction for anatomical therapeutic chemical classification of drugs by integrating multiple data sources. Bioinformatics **31**(11), 1788–1795 (2015)

35. Bleakley, K., Yamanishi, Y.: Supervised prediction of drug-target interactions using bipartite local models. Bioinformatics **25**(18), 2397–2403 (2009). doi:10.1093/bioinformatics/btp433

36. Xia, Z., Wu, L.-Y., Zhou, X., Wong, S.T.C.: Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. BMC systems biology **4**(S6), 6 (2010). doi:10.1186/1752-0509-4-S2-S6

37. Agarwal, S., Dugar, D., Sengupta, S.: Ranking chemical structures for drug discovery: A new machine learning approach. Journal of Chemical Information and Modeling **50**(5), 716–731 (2010). doi:10.1021/ci9003865

38. van Laarhoven, T., Nabuurs, S.B., Marchiori, E.: Gaussian interaction profile kernels for predicting drug-target interaction. Bioinformatics **27**(21), 3036–3043 (2011). doi:10.1093/bioinformatics/btr500

39. Perlman, L., Gottlieb, A., Atias, N., Ruppin, E., Sharan, R.: Combining Drug and Gene Similarity Measures for Drug-Target Elucidation. Computational Biology **18**(2), 133–145 (2011). doi:10.1089/cmb.2010.0213

40. Chen, B., Ding, Y., Wild, D.J.: Improving integrative searching of systems chemical biology data using semantic annotation. Journal of cheminformatics **4**(1), 6 (2012). doi:10.1186/1758-2946-4-6

41. Yu, H., Chen, J., Xu, X., Li, Y., Zhao, H., Fang, Y., Li, X., Zhou, W., Wang, W., Wang, Y.: A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data. PLoS ONE **7**(5) (2012). doi:10.1371/journal.pone.0037608

42. Mei, J.P., Kwoh, C.K., Yang, P., Li, X.L., Zheng, J.: Drug-target interaction prediction by learning from local information and neighbors. Bioinformatics **29**(2), 238–245 (2013). doi:10.1093/bioinformatics/bts670

43. van Laarhoven, T., Marchiori, E.: Predicting Drug-Target Interactions for New Drug Compounds Using a

Weighted Nearest Neighbor Profile. PLoS ONE **8**(6), 1–6 (2013). doi:10.1371/journal.pone.0066952

44. Wang, Y., Zeng, J.: Predicting drug-target interactions using restricted Boltzmann machines. Bioinformatics **29**(13), 126–134 (2013). doi:10.1093/bioinformatics/btt234

45. Simm, J., Arany, A., Zakeri, P., Haber, T., Wegner, J.K., Chupakhin, V., Ceulemans, H., Moreau, Y.: Macau: Scalable Bayesian Multi-relational Factorization with Side Information using MCMC. In: Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (2017). Roppongi, Tokyo, Japan

46. Yuan, Q., Gao, J., Wu, D., Zhang, S., Mamitsuka, H., Zhu, S.: DrugE-Rank: Improving drug-target interaction prediction of new candidate drugs or targets by ensemble learning to rank. Bioinformatics **32**(12), 18–27 (2016). doi:10.1093/bioinformatics/btw244

47. Liu, Y., Wu, M., Miao, C., Zhao, P., Li, X.L.: Neighborhood Regularized Logistic Matrix Factorization for Drug-Target Interaction Prediction. PLoS Computational Biology **12**(2), 1–26 (2016). doi:10.1371/journal.pcbi.1004760

48. Hao, M., Bryant, S.H., Wang, Y., Iorio, F., Rittman, T., Ge, H., Menden, M., Saez-Rodriguez, J., Bartlett, J.B., Dredge, K., Dalgleish, A.G., Steinbach, G., Koehl, G.E., Schlitt, H.J., Geissler, E.K., Cappelli, C., Gu, S., Keiser, M.J., Wang, L., Haupt, V.J., Schroeder, M., Ma, D.L., Chan, D.S., Leung, C.H., Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., Kanehisa, M., Bleakley, K., Yamanishi, Y., van Laarhoven, T., Nabuurs, S.B., Marchiori, E., Mei, J.-P., Kwoh, C.-K., Yang, P., Li, X.-L., Zheng, J., Hao, M., Wang, Y., Bryant, S.H., Wang, B., Liu, Y., Wu, M., Miao, C., Zhao, P., Li, X.L., Kanehisa, M., Schomburg, I., Günther, S., Wishart, D.S., Kuang, Q., Smith, T.F., Waterman, M.S., Hattori, M., Okuno, Y., Goto, S., Kanehisa, M., Ma, H., King, I., Lyu, M.R., Duchi, J., Hazan, E., Singer, Y., Gonen, M., Kaski, S., Cao, Y., Charisi, A., Cheng, L.-C., Jiang, T., Girke, T., Guha, R., Sievers, F., Leslie, C., Eskin, E., Noble, W.S., Langham, J.J., Cleves, A.E., Spitzer, R., Kirshner, D., Jain, A.N., Collins, I., von Coburg, Y., Kottke, T., Weizel, L., Ligneau, X., Stark, H., Wishart, D., Alaimo, S., Sui, J.: Predicting drug-target interactions by dual-network integrated logistic matrix factorization. Scientific Reports **7**(January), 40376 (2017). doi:10.1038/srep40376

49. Hao, M., Wang, Y., Bryant, S.H.: Improved prediction of drug-target interactions using regularized least squares integrating with kernel fusion technique. Analytica Chimica Acta **909**, 41–50 (2016). doi:10.1016/j.aca.2016.01.014. 15334406

50. Nascimento, A.C.a., Prudêncio, R.B.C., Costa, I.G.: A multiple kernel learning algorithm for drug-target interaction prediction. BMC bioinformatics **17**(1), 46 (2016). doi:10.1186/s12859-016-0890-3

51. Bolgár, B., Antal, P.: Bayesian matrix factorization with non-random missing data using informative Gaussian process priors and soft evidences. In: Antonucci, A., Corani, G., Campos, C.P. (eds.) Proceedings of the Eighth International Conference on Probabilistic Graphical Models, pp. 25–36 (2016)

52. Wu, Z., Cheng, F., Li, J., Li, W., Liu, G., Tang, Y.: SDTNBI: an integrated network and chemoinformatics tool for systematic prediction of drug–target interactions and drug repositioning. Briefings in Bioinformatics (October 2015), 012 (2016). doi:10.1093/bib/bbw012

53. Keum, J., Nam, H.: Self-blm: Prediction of drug-target interactions via self-training svm. PloS one **12**(2), 0171839 (2017)

54. Visser, U., Abeyruwan, S., Vempati, U., Smith, R.P., Lemmon, V., Schürer, S.C.: BioAssay Ontology (BAO): a semantic description of bioassays and high-throughput screening results. BMC bioinformatics **12**(1), 257 (2011). doi:10.1186/1471-2105-12-257

55. Chen, B., Dong, X., Jiao, D., Wang, H., Zhu, Q., Ding, Y., Wild, D.J.: Chem2Bio2RDF: a semantic framework for linking and data mining chemogenomic and systems chemical biology data. BMC bioinformatics **11**, 255 (2010). doi:10.1186/1471-2105-11-255

56. Gaulton, A., Hersey, A., Nowotka, M., Bento, A.P., Chambers, J., Mendez, D., Mutowo, P., Atkinson, F., Bellis, L.J., Cibrián-Uhalte, E., *et al.*: The chembl database in 2017. Nucleic acids research **45**(D1), 945–954 (2016)

57. Mathias, S.L., Hines-Kay, J., Yang, J.J., Zahoransky-Kohalmi, G., Bologa, C.G., Ursu, O., Oprea, T.I.: The CARLSBAD database: A confederated database of chemical bioactivities. Database **2013**, 1–8 (2013). doi:10.1093/database/bat044

58. Said, A., Bellogín, A.: Comparative recommender system evaluation: benchmarking recommendation frameworks. In: Proceedings of the 8th ACM Conference on Recommender Systems, pp. 129–136 (2014). ACM

59. Tiikkainen, P., Bellis, L., Light, Y., Franke, L.: Estimating error rates in bioactivity databases. Journal of Chemical Information and Modeling **53**(10), 2499–2505 (2013). doi:10.1021/ci400099q

60. Hersey, A., Chambers, J., Bellis, L., Patrícia Bento, a., Gaulton, A., Overington, J.P.: Chemical databases: curation or integration by user-defined equivalence? Drug Discovery Today: Technologies **xxx**(xx) (2015). doi:10.1016/j.ddtec.2015.01.005

61. Lipinski, C.A., Litterman, N.K., Southan, C., Williams, A.J., Clark, A.M., Ekins, S.: Parallel worlds of public and commercial bioactive chemistry data: Miniperspective. Journal of medicinal chemistry **58**(5), 2068 (2015)

62. Southan, C., Vrkonyi, P., Muresan, S.: Quantitative assessment of the expanding complementarity between public and commercial databases of bioactive compounds. Journal of Cheminformatics **1**(1), 1–17 (2009). doi:10.1186/1758-2946-1-10

63. Tiikkainen, P., Franke, L.: Analysis of commercial and public bioactivity databases. Journal of Chemical Information and Modeling **52**(2), 319–326 (2012). doi:10.1021/ci2003126

64. Hu, Y., Bajorath, J.: Growth of ligand-target interaction data in ChEMBL is associated with increasing and activity measurement-dependent compound promiscuity. Journal of Chemical Information and Modeling **52**(10), 2550–2558 (2012). doi:10.1021/ci3003304

65. Johnson, M.A., Maggiora, G.M.: Concepts and Applications of Molecular Similarity. Wiley New York, ??? (1990)

66. Maggiora, G., Vogt, M., Stumpfe, D., Bajorath, J.: Molecular similarity in medicinal chemistry: miniperspective. Journal of medicinal chemistry **57**(8), 3186–3204 (2013)

67. Lipinski, C.A.: Lead-and drug-like compounds: the rule-of-five revolution. Drug Discovery Today: Technologies

1(4), 337–341 (2004)

68. Tian, S., Wang, J., Li, Y., Li, D., Xu, L., Hou, T.: The application of in silico drug-likeness predictions in pharmaceutical research. Advanced drug delivery reviews **86**, 2–10 (2015)

69. Rask-Andersen, M., Masuram, S., Schiöth, H.B.: The druggable genome: evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. Annual review of pharmacology and toxicology **54**, 9–26 (2014)

70. Gao, M., Skolnick, J.: A comprehensive survey of small-molecule binding pockets in proteins. PLoS Comput Biol **9**(10), 1003302 (2013)

71. Hopkins, A.L.: Network pharmacology: the next paradigm in drug discovery. Nature chemical biology **4**(11), 682–690 (2008)

72. Kubinyi, H.: Similarity and dissimilarity: a medicinal chemist's view. Perspectives in Drug Discovery and Design **9**, 225–252 (1998)

73. Eckert, H., Bajorath, J.: Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. Drug discovery today **12**(5), 225–233 (2007)

74. Ding, H., Takigawa, I., Mamitsuka, H., Zhu, S.: Similarity-based machine learning methods for predicting drug–target interactions: a brief review. Briefings in Bioinformatics **15**(5), 734–747 (2013)

75. Gönen, M.: Predicting drug-target interactions from chemical and genomic kernels using Bayesian matrix factorization. Bioinformatics **28**(18), 2304–2310 (2012). doi:10.1093/bioinformatics/bts360

76. Daina, A., Michielin, O., Zoete, V.: Swissadme: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific Reports **7**, 42717 (2017)

77. Hopkins, A.L.: Drug discovery: predicting promiscuity. Nature **462**(7270), 167–168 (2009)

78. Cereto-Massagué, A., Guasch, L., Valls, C., Mulero, M., Pujadas, G., Garcia-Vallvé, S.: Decoyfinder: an easy-to-use python gui application for building target-specific decoy sets. Bioinformatics **28**(12), 1661–1662 (2012)

79. Hussein, H.A., Geneix, C., Petitjean, M., Borrel, A., Flatters, D., Camproux, A.-C.: Global vision of druggability issues: applications and perspectives. Drug discovery today **22**(2), 404–415 (2017)

80. Jamali, A.A., Ferdousi, R., Razzaghi, S., Li, J., Safdari, R., Ebrahimie, E.: Drugminer: comparative analysis of machine learning algorithms for prediction of potential druggable proteins. Drug discovery today **21**(5), 718–724 (2016)

81. Hussein, H.A., Borrel, A., Geneix, C., Petitjean, M., Regad, L., Camproux, A.-C.: Pockdrug-server: a new web server for predicting pocket druggability on holo and apo proteins. Nucleic acids research **43**(W1), 436–442 (2015)

82. Chen, X., Yan, C.C., Zhang, X., Zhang, X., Dai, F., Yin, J., Zhang, Y.: Drug–target interaction prediction: databases, web servers and computational models. Briefings in bioinformatics **17**(4), 696–712 (2015)

83. Cheng, T., Hao, M., Takeda, T., Bryant, S.H., Wang, Y.: Large-scale prediction of drug-target interaction: a data-centric review. The AAPS Journal, 1–12 (2017)

84. Lavecchia, A.: Machine-learning approaches in drug discovery: methods and applications. Drug Discovery Today **20**(3), 318–331 (2014). doi:10.1016/j.drudis.2014.10.012

85. Lounkine, E., Keiser, M.J., Whitebread, S., Mikhailov, D., Hamon, J., Jenkins, J.L., Lavan, P., Weber, E., Doak, A.K., Côté, S., et al.: Large-scale prediction and testing of drug activity on side-effect targets. Nature **486**(7403), 361–367 (2012)

86. Jacob, L., Vert, J.-P.: Protein-ligand interaction prediction: an improved chemogenomics approach. Bioinformatics **24**(19), 2149–2156 (2008)

87. Xu, Q., Yang, Q.: A survey of transfer and multitask learning in bioinformatics. Journal of Computing Science and Engineering **5**(3), 257–268 (2011)

88. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: Bayesian Data Analysis vol. 2. Chapman & Hall/CRC Boca Raton, FL, USA, ??? (2014)

89. Nagamine, N., Sakakibara, Y.: Statistical prediction of protein–chemical interactions based on chemical structure and mass spectrometry data. Bioinformatics **23**(15), 2004–2012 (2007)

90. van Laarhoven, T., Nabuurs, S.B., Marchiori, E.: Gaussian interaction profile kernels for predicting drug-target interaction. Bioinformatics **27**(21), 3036–3043 (2011). doi:10.1093/bioinformatics/btr500

91. Wen, M., Zhang, Z., Niu, S., Sha, H., Yang, R., Yun, Y., Lu, H.: Deep-learning-based drug–target interaction prediction. Journal of Proteome Research **16**(4), 1401–1409 (2017)

92. Srebro, N., Jaakkola, T.: Sparse matrix factorization of gene expression data. Internal report, MIT Artificial Intelligence Laboratory (2001).
www.ai.mit.edu/~research/abstracts/abstracts2001/genomics/01srebro.pdf

93. Dueck, D., Morris, Q.D., Frey, B.J.: Multi-way clustering of microarray data using probabilistic sparse matrix factorization. Bioinformatics **21**(suppl 1), 144–151 (2005)

94. Bock, J.R., Gough, D.A.: A new method to estimate ligand-receptor energetics. Molecular & Cellular Proteomics **1**(11), 904–910 (2002)

95. Agarwal, P., Searls, D.B.: Literature mining in support of drug discovery. Briefings in bioinformatics **9**(6), 479–492 (2008)

96. Parsons, A.B., Lopez, A., Givoni, I.E., Williams, D.E., Gray, C.A., Porter, J., Chua, G., Sopko, R., Brost, R.L., Ho, C.-H., et al.: Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. Cell **126**(3), 611–625 (2006)

97. Takács, G., Pilászy, I., Németh, B., Tikk, D.: Matrix factorization and neighbor based algorithms for the netflix prize problem. In: Proceedings of the 2008 ACM Conference on Recommender Systems, pp. 267–274 (2008). ACM

98. Srebro, N., Jaakkola, T., et al.: Weighted low-rank approximations. In: Icml, vol. 3, pp. 720–727 (2003)

99. Pan, R., Zhou, Y., Cao, B., Liu, N.N., Lukose, R., Scholz, M., Yang, Q.: One-class collaborative filtering. In: Data Mining, 2008. ICDM'08. Eighth IEEE International Conference On, pp. 502–511 (2008). IEEE

100. Salakhutdinov, R., Mnih, a.: Bayesian probabilistic matrix factorization using Markov chain Monte Carlo.,

880–887 (2008). doi:10.1145/1390156.1390267

101. Severinski, C., Salakhutdinov, R.: Bayesian Probabilistic Matrix Factorization: A User Frequency Analysis. ArXiv e-prints (2014). https://arxiv.org/abs/1407.7840

102. Zhou, T., Shan, H., Banerjee, A., Sapiro, G.: Kernelized probabilistic matrix factorization: Exploiting graphs and side information. In: SDM, pp. 403–414. SIAM / Omnipress, ??? (2012)

103. Hernandez-Lobato, J.M., Houlsby, N., Ghahramani, Z.: Stochastic inference for scalable probabilistic modeling of binary matrices. In: Proceedings of the 31st International Conference on Machine Learning (ICML-14), pp. 379–387 (2014)

104. Gönen, M., Kaski, S.: Kernelized bayesian matrix factorization. IEEE Transactions on Pattern Analysis and Machine Intelligence **36**(10), 2047–2060 (2014)

105. Koutsoukas, A., Lowe, R., KalantarMotamedi, Y., Mussa, H.Y., Klaffke, W., Mitchell, J.B., Glen, R.C., Bender, A.: In silico target predictions: defining a benchmarking data set and comparison of performance of the multiclass naïve bayes and parzen-rosenblatt window. Journal of chemical information and modeling **53**(8), 1957–1966 (2013)

106. Schomburg, K.T., Rarey, M.: Benchmark data sets for structure-based computational target prediction. Journal of chemical information and modeling **54**(8), 2261–74 (2014). doi:10.1021/ci500131x

107. Wale, N., Karypis, G.: Target fishing for chemical compounds using target-ligand activity data and ranking based methods. Journal of Chemical Information and Modeling **49**(10), 2190–2201 (2009). doi:10.1021/ci9000376. NIHMS150003

108. Peón, A., Dang, C.C., Ballester, P.J.: How Reliable Are Ligand-Centric Methods for Target Fishing? Frontiers in chemistry **4**(April), 15 (2016). doi:10.3389/fchem.2016.00015

109. Landrum, G.: Rdkit: Open-source cheminformatics. Online). http://www. rdkit. org. Accessed **3**(04), 2012 (2006)

110. Jordan, M.I., Ghahramani, Z., Jaakkola, T.S., Saul, L.K.: An introduction to variational methods for graphical models. Machine learning **37**(2), 183–233 (1999)

111. Jaakkola, T.S., Jordan, M.I.: Bayesian parameter estimation via variational methods. Statistics and Computing **10**(1), 25–37 (2000). doi:10.1023/A:1008932416310

112. Bishop, C.M.: Pattern recognition. Machine Learning **128**, 1–58 (2006)

113. Cortes, C., Mohri, M., Rostamizadeh, A.: Learning non-linear combinations of kernels. In: Proceedings of the 22Nd International Conference on Neural Information Processing Systems. NIPS'09, pp. 396–404. Curran Associates Inc., USA (2009). http://dl.acm.org/citation.cfm?id=2984093.2984138

114. Maggiora, G., Gokhale, V.: Non-specificity of drug-target interactions–consequences for drug discovery. In: Frontiers in Molecular Design and Chemical Information Science-Herman Skolnik Award Symposium 2015: Jürgen Bajorath, pp. 91–142 (2016). ACS Publications

115. Börnigen, D., Tranchevent, L.-C., Bonachela-Capdevila, F., Devriendt, K., De Moor, B., De Causmaecker, P., Moreau, Y.: An unbiased evaluation of gene prioritization tools. Bioinformatics **28**(23), 3081–3088 (2012)

116. Moreau, Y., Tranchevent, L.-C.: Computational tools for prioritizing candidate genes: boosting disease gene discovery. Nature Reviews Genetics **13**(8), 523–536 (2012)

117. Paricharak, S., Méndez-Lucio, O., Chavan Ravindranath, A., Bender, A., IJzerman, A.P., van Westen, G.J.: Data-driven approaches used for compound library design, hit triage and bioactivity modeling in high-throughput screening. Briefings in bioinformatics, 105 (2016). Epub ahead of print

118. Cobanoglu, M.C., Liu, C., Hu, F., Oltvai, Z.N., Bahar, I.: Predicting drug–target interactions using probabilistic matrix factorization. Journal of chemical information and modeling **53**(12), 3399–3409 (2013)

Figures

Figure 1 Pseudocode of the VB-MK-LMF algorithm.

Figure 2 Overview of the VB-MK-LMF workflow. A priori information (left) are combined with DTI data through a Bayesian model (middle). Learning is carried out using a Variational Bayesian method which approximates the latent factors and optimal kernel weights. The model provides quantitative predictions of interaction probabilities and estimates of drug promiscuity (right). Finally, VB-MK-LMF supports the visualization and exploration of the unified "pharmacological" space. Gray indicates functionalities which may also be utilized in the VB-MK-LMF model but not explored in this paper.

Figure 3 AUPRC values on the three smallest datasets with varying number of latent factors. The results become saturated around 10 dimensions.

Tables

Additional Files

Additional file 1 — Derivation of the lower bound using Jaakkola's bound on the logistic sigmoid

Additional file 2 — The properties of DTI methods related to the development or evaluation of VB-MK-LMF

Figure 4 The effect of priors on predictive performance with varying sample sizes. The difference between the values using and not using kernels gradually vanishes as the training size increases. 95% confidence intervals are indicated by gray ribbons.

Figure 5 Latent representations of drugs and targets in the Ion Channel dataset using 2 latent dimensions. Drugs are colored on the basis of their respective ATC classes and targets are colored according to their ion transporter activity as obtained from the Gene Ontology. Known interactions are represented as edges.

Figure 6 Parallel coordinates visualization of 10 latent dimensions in the Ion Channel dataset. Each curve corresponds to a latent representation of a drug or a target. Targets are colored on the basis of their ion transporter activity.

Figure 7 Drug promiscuity vs. the expected number of interactions. The number of targets of each drug in the datasets are depicted on the horizontal axis. The expected number of interactions as predicted by VB-MK-LMF are depicted on the vertical axis.

Figure 8 Expected number of interactions as predicted by VB-MK-LMF for each drug in the GPCR dataset. The number of targets are depicted on the horizontal axis. A $10\times$ cross-validation setting was used.

Figure 9 Runtime of the GPU and CPU implementations in terms of the number of latent factors. This benchmark was conducted on a 200×200 matrix factorization. The GPU implementation brings a $30\times$ speedup on an NVIDIA GTX Titan X graphics card.

Table 1 Single-kernel results on gold standard data sets.

AUROC (CV1)			
	VB-MK-LMF	NRLMF	KBMF
Nuclear Receptor	$\mathbf{0.957} \pm 0.010$	0.949 ± 0.011	0.860 ± 0.024
GPCR	$\mathbf{0.976} \pm 0.003$	0.960 ± 0.004	0.911 ± 0.004
Ion Channel	$\mathbf{0.989} \pm 0.001$	0.984 ± 0.002	0.941 ± 0.003
Enzyme	$\mathbf{0.987} \pm 0.001$	0.976 ± 0.002	0.887 ± 0.003
Kinase	$\mathbf{0.921} \pm 0.002$	0.919 ± 0.001	0.916 ± 0.001

AUPRC (CV1)			
	VB-MK-LMF	NRLMF	KBMF
Nuclear Receptor	$\mathbf{0.773} \pm 0.030$	0.723 ± 0.042	0.533 ± 0.047
GPCR	$\mathbf{0.777} \pm 0.016$	0.703 ± 0.023	0.541 ± 0.012
Ion Channel	$\mathbf{0.916} \pm 0.007$	0.863 ± 0.012	0.763 ± 0.009
Enzyme	$\mathbf{0.890} \pm 0.006$	0.876 ± 0.007	0.656 ± 0.008
Kinase	$\mathbf{0.850} \pm 0.003$	0.845 ± 0.003	0.844 ± 0.003

AUROC (CV2)			
	VB-MK-LMF	NRLMF	KBMF
Nuclear Receptor	$\mathbf{0.939} \pm 0.021$	0.896 ± 0.023	0.845 ± 0.023
GPCR	0.878 ± 0.014	$\mathbf{0.883} \pm 0.012$	0.847 ± 0.018
Ion Channel	$\mathbf{0.812} \pm 0.026$	0.800 ± 0.026	0.785 ± 0.021
Enzyme	$\mathbf{0.851} \pm 0.021$	0.811 ± 0.024	0.718 ± 0.028
Kinase	$\mathbf{0.894} \pm 0.004$	0.891 ± 0.004	0.838 ± 0.004

AUPRC (CV2)			
	VB-MK-LMF	NRLMF	KBMF
Nuclear Receptor	$\mathbf{0.593} \pm 0.058$	0.547 ± 0.053	0.447 ± 0.048
GPCR	$\mathbf{0.368} \pm 0.023$	0.363 ± 0.023	0.365 ± 0.024
Ion Channel	$\mathbf{0.345} \pm 0.035$	0.343 ± 0.033	0.287 ± 0.035
Enzyme	0.349 ± 0.042	$\mathbf{0.360} \pm 0.041$	0.269 ± 0.037
Kinase	$\mathbf{0.803} \pm 0.009$	0.797 ± 0.010	0.735 ± 0.009

AUROC (CV3)			
	VB-MK-LMF	NRLMF	KBMF
Nuclear Receptor	$\mathbf{0.917} \pm 0.026$	0.847 ± 0.029	0.735 ± 0.050
GPCR	$\mathbf{0.941} \pm 0.009$	0.920 ± 0.014	0.839 ± 0.020
Ion Channel	$\mathbf{0.966} \pm 0.007$	0.958 ± 0.008	0.911 ± 0.012
Enzyme	$\mathbf{0.962} \pm 0.005$	0.947 ± 0.006	0.859 ± 0.012
Kinase	$\mathbf{0.767} \pm 0.018$	0.763 ± 0.018	0.740 ± 0.022

AUPRC (CV3)			
	VB-MK-LMF	NRLMF	KBMF
Nuclear Receptor	$\mathbf{0.601} \pm 0.081$	0.456 ± 0.079	0.352 ± 0.070
GPCR	$\mathbf{0.596} \pm 0.040$	0.553 ± 0.040	0.437 ± 0.047
Ion Channel	$\mathbf{0.826} \pm 0.021$	0.788 ± 0.028	0.695 ± 0.024
Enzyme	0.794 ± 0.017	$\mathbf{0.808} \pm 0.018$	0.573 ± 0.028
Kinase	$\mathbf{0.608} \pm 0.039$	0.597 ± 0.038	0.594 ± 0.039

CV indicates the cross-validation setting (pairwise, drug and target, respectively). AUROC and AUPRC values were averaged over 5×10 runs and 95% confidence intervals were computed. In most cases, VB-MK-LMF significantly outperforms the other methods using t-test.

Table 2 AUPRC values on gold standard data sets in the pairwise cross-validation setting.

Nuclear Receptor (KBMF-MKL: 0.566, KronRLS-MKL: 0.522)

Neighbors	MrgRbf	MrgTan	McsRbf	McsTan	Orig	All
2	0.749	0.758	0.742	0.735	0.754	**0.779**
3	0.744	0.771	0.761	0.734	0.773	0.775
5	0.732	0.757	0.739	0.724	0.755	0.756
2+3	0.750	0.765	0.754	0.736	0.757	0.758
2+3+5	0.760	0.765	0.740	0.738	0.764	0.760

GPCR (KBMF-MKL: 0.622, KronRLS-MKL: 0.696)

Neighbors	MrgRbf	MrgTan	McsRbf	McsTan	Orig	All
2	0.743	0.759	0.754	0.762	0.764	0.793
3	0.755	0.774	0.772	0.780	0.777	**0.802**
5	0.762	0.787	0.782	0.783	0.787	0.796
2+3	0.763	0.782	0.781	0.786	0.785	**0.802**
2+3+5	0.777	0.798	0.793	0.789	0.796	0.800

Ion Channel (KBMF-MKL: 0.826, KronRLS-MKL: 0.885)

Neighbors	MrgRbf	MrgTan	McsRbf	McsTan	Orig	All
2	0.909	0.911	0.910	0.911	0.910	0.909
3	0.911	0.914	0.915	0.914	0.912	0.916
5	0.915	0.914	0.913	0.916	0.916	**0.917**
2+3	0.912	0.914	0.916	0.914	0.913	0.909
2+3+5	0.912	0.915	0.915	0.915	0.916	0.906

Enzyme (KBMF-MKL: 0.704, KronRLS-MKL: 0.893)

Neighbors	MrgRbf	MrgTan	McsRbf	McsTan	Orig	All
2	0.885	0.887	0.879	0.883	0.888	0.884
3	0.885	0.890	0.885	0.882	0.890	**0.895**
5	0.883	0.886	0.880	0.881	0.884	0.883
2+3	0.888	0.889	0.880	0.881	0.888	0.881
2+3+5	0.887	0.889	0.881	0.878	0.888	0.875

Kinase (KBMF-MKL: 0.846, KronRLS-MKL: 0.561)

Neighbors	-	2D	3D	ECFP	All
2		0.850	0.849	0.849	0.850
3		0.850	0.848	0.850	0.851
5	-	0.850	0.849	0.850	0.851
2+3		0.850	0.850	0.850	0.853
2+3+5		0.851	0.851	0.850	**0.854**

Rows correspond to the cut-off value of the number of closest neighbors and the combinations of the resulting truncated kernels. Columns correspond to individual kernels. The last column was obtained by combining all kernels.

Table 3 Normalized kernel weights with an extra positive definite, unit-diagonal, random valued kernel matrix.

	MrgRbf	MrgTan	McsRbf	McsTan	Yam	Random
Nuclear Receptor	0.175	0.176	0.175	0.175	0.175	0.123
GPCR	0.173	0.173	0.172	0.172	0.172	0.138
Ion Channel	0.176	0.176	0.176	0.176	0.176	0.120
Enzyme	0.176	0.176	0.176	0.176	0.176	0.119
	-	2D	3D	ECFP	Random	
Kinase	-	0.300	0.283	0.398	0.019	

The number of latent factors was not altered in this experiment. Setting the number of latent factors to I (the rank of the kernel matrix) zeroes out the weight of the random kernel.

Table 4 Top 5 predicted interactions which are not present in the datasets.

			Nuclear Receptor			
Probability	Drug	Target	Drug name	Target name	DrugBank	NRLMF
0.943	D00316	hsa6096	Etretinate	RARB	yes	1
0.671	D01132	hsa6097	Tazarotene	RARC	*RARB	6
0.662	D01132	hsa190	Tazarotene	NR0B1		18
0.529	D00898	hsa2100	Dienestrol	ESR2	yes	7
0.445	D00094	hsa6095	Tretinoin	RARA	yes	26

			GPCR			
Probability	Drug	Target	Drug name	Target name	DrugBank	NRLMF
0.966	D00283	hsa1814	Clozapine	DRD3	yes	1
0.956	D00110	hsa1813	Cocaine	DRD2		188
0.938	D02358	hsa154	Metoprolol	ADRB2	yes	2
0.937	D02614	hsa154	Denopamine	ADRB2	yes	4
0.937	D04625	hsa154	Isoetharine	ADRB2	yes	3

			Ion Channel			
Probability	Drug	Target	Drug name	Target name	DrugBank	NRLMF
0.990	D00538	hsa6331	Zonisamide	SCN5A	yes	9
0.986	D00294	hsa3767	Diazoxide	KCNJ11	yes	244
0.985	D00552	hsa6331	Tetracaine	SCN5A	yes	5
0.983	D00438	hsa779	Nimodipine	CACNA1S	yes	2
0.983	D00649	hsa8911	Amiloride	CACNA1I		83

			Enzyme			
Probability	Drug	Target	Drug name	Target name	DrugBank	NRLMF
0.999	D00542	hsa1571	Halothane	CYP2E1	yes	1
0.995	D00097	hsa5743	Salicylic acid	PTGS2	yes	4
0.995	D00437	hsa1559	Nifedipine	CYP2C9	yes	5
0.987	D00501	hsa50940	Pentoxifylline	PDE11A	*PDE5A	2
0.986	D00501	hsa5150	Pentoxifylline	PDE7A	*PDE5A	3

Many of the hits were confirmed by the current version of DrugBank. The * symbol indicates a known interaction with another member of the protein family. The last column denotes the rank of the interaction among the NRLMF predictions.