



# SPVec: A Word2vec-Inspired Feature Representation Method for Drug-Target Interaction Prediction

Yu-Fang Zhang<sup>1</sup>, Xiangeng Wang<sup>1</sup>, Aman Chandra Kaushik<sup>1,2</sup>, Yanyi Chu<sup>1</sup>, Xiaoqi Shan<sup>1</sup>, Ming-Zhu Zhao<sup>3</sup>, Qin Xu<sup>1\*</sup> and Dong-Qing Wei<sup>1,4\*</sup>

<sup>1</sup> State Key Laboratory of Microbial Metabolism, and SJTU-Yale Joint Center for Biostatistics and Data Science, School of Life Sciences and Biotechnology, and Joint Laboratory of International Cooperation in Metabolic and Developmental Sciences, Ministry of Education, Shanghai Jiao Tong University, Shanghai, China, <sup>2</sup> Wuxi School of Medicine, Jiangnan University, Wuxi, China, <sup>3</sup> Instrumental Analysis Center, Shanghai Jiao Tong University, Shanghai, China, <sup>4</sup> Peng Cheng Laboratory, Shenzhen, China

### **OPEN ACCESS**

#### Ed ted by

Zunnan Huang, Guangdong Medical University, China

#### ev ewed by

Francesco Ortuso, University of Catanzaro, Italy Ling Wang, South China University of Technology, China

### Correspondence

Qin Xu xuqin523@sjtu.edu.cn Dong-Qing Wei dqwei@sjtu.edu.cn

#### pec a ty sect on

This article was submitted to Medicinal and Pharmaceutical Chemistry, a section of the journal Frontiers in Chemistry

ece ved 11 October 2019 Accepted 12 December 2019 Pub shed 10 January 2020

#### C tat on

Zhang Y-F, Wang X, Kaushik AC, Chu Y, Shan X, Zhao M-Z, Xu Q and Wei D-Q (2020) SPVec: A Word2vec-Inspired Feature Representation Method for Drug-Target Interaction Prediction. Front. Chem. 7:895. doi: 10.3389/fchem.2019.00895

Drug discovery is an academical and commercial process of global importance. Accurate identification of drug-target interactions (DTIs) can significantly facilitate the drug discovery process. Compared to the costly, labor-intensive and time-consuming experimental methods, machine learning (ML) plays an ever-increasingly important role in effective, efficient and high-throughput identification of DTIs. However, upstream feature extraction methods require tremendous human resources and expert insights, which limits the application of ML approaches. Inspired by the unsupervised representation learning methods like Word2vec, we here proposed SPVec, a novel way to automatically represent raw data such as SMILES strings and protein sequences into continuous, information-rich and lower-dimensional vectors, so as to avoid the sparseness and bit collisions from the cumbersomely manually extracted features. Visualization of SPVec nicely illustrated that the similar compounds or proteins occupy similar vector space, which indicated that SPVec not only encodes compound substructures or protein sequences efficiently, but also implicitly reveals some important biophysical and biochemical patterns. Compared with manually-designed features like MACCS fingerprints and amino acid composition (AAC), SPVec showed better performance with several state-of-art machine learning classifiers such as Gradient Boosting Decision Tree, Random Forest and Deep Neural Network on BindingDB. The performance and robustness of SPVec were also confirmed on independent test sets obtained from DrugBank database. Also, based on the whole DrugBank dataset, we predicted the possibilities of all unlabeled DTIs, where two of the top five predicted novel DTIs were supported by external evidences. These results indicated that SPVec can provide an effective and efficient way to discover reliable DTIs, which would be beneficial for drug reprofiling.

Keywords: drug-target interaction, representation learning, Word2vec, machine learning, feature embedding

# INTRODUCTION

is old 🗸 po **₽**Ø0 D 🖉 is iss n 0 **,** 🛛 s n .sp 🖕 'n ions 1 no . ín , ions D. Is 5 – A ni ion o D. Id n st⊠n ( is of po n 11 . Τĭ n \$ 0 'n n D. Us ò 0 ⊠.s ş po ín 📕 🕹 🚽 po n.s n 0 3 i ions o D. Isο 🖉 🚬 1 15 3 1 X n po 'nop **∦**⊠s n spon na 🔊 n Min Moi n . Ś ð p , • .s no 0 ions ips 3 *in vivo* n s.sp n р 0 Q nd Yuk 'n jn⊠ł , ons in vitro M. n ડે ્રા હેડે ູກູ່ຫ 0⊠ ≴ o in silico 0 3 n 0 nʻq \$ p . ions on **,** 🛛 🤇 ni **,**⊠, .sso 1 po 'n nʻp⊠ os is on 'n , pp o ín 💊 n′n∎ ó.s X 10 .5 op 'n s. 'n os .,.s 'n .ssi n ö 0 ppo 'n Со pni 1 D 0 n 3 in Boos in D ision D BD. n n so on- 🔳 "p " o n **"**0 nʻn⊠ 0 .s ίn ş on íon -51 р **á**⊠n o p po sky⊠ n ion is o 🖉 🖕 ns o 'n ns on n 🕹 ppo n′µ⊠ íon (n 'n 3 **N**o p opos ss. 0 Ś ò 5 n n ∎ no n⊠ p in s o ⊠ n \$.J şp, ₩,, í s. ņ 'n⊠ n 0 0.5 3 **N** on⊠ íno n A 'n position AAC ö n 図 n p sio р ор С 0 po ins so ot 0 s, no n n .sp ņ 0 ро ýn s g ۹'n n 0 0 'n 0 **₽**⊠s D. dор n 0 nní **₽**Øŝ ₽ p⊠ p′ns n .s . b s u ons on \* 1 n n op o o is n s. s i ion on 33 . ion o X n so oni.s .s psn р on 🚬 on 'nΜ Ø 'n po.s p⊠ p 'n s 0 3 n n ion 🖉 o ps o . 'nρ o AAC o po ún n ps 0 'n ş Ø р i.s-C 'n р 0 'n **\*** 🛛 **,**⊠.s jn š n ., ро n 0 0 n . , o⊠n o \* íon n 🔥 no n 0 S. 'n 24 **₽**Ø0 1 n 🔥 0 n n ö ņ 'n "...p⊠nosn", ion o D. Ison is 0 0 , poi **√**⊠o s'n⊠ 5 0 5 ín o 'on o ⊠ n ₩ íon n ≴–B .s \_ on i ion **"**p n o \$ . S. 0 4 4

s o p nín 🛛 🗸 🖾 o 🛝 i jons o Dis p o🕅 Ø 0 0 q × nos ion 0 .s npo Ð. . 🛛 ⊠n st o .so n o p , po⊠ ions o nos io s po <u>,</u>\$\$ S , n⊠ʻn 'n⊠ on n ín o nd n o 0 \* n′µ⊠ no oli s isn. pp on o 'n n on n**Ø**'n рņ

n′n⊠ B n⊠ʻo P 'n íon is is \$ 'n 🙀 n **"**Ø., oín А p ío no n on 'n o s ō n ín o ion n , o.s nsion n **√**⊠o .s . 'n 、 s<sup>s</sup>−An p.s 0 is o 0 ş "pii 'n 33 ions-0 p 0 \* 'n⊠ sson nín 'n n .so⊠ýp⊠ ро in o Sistion p ٦p n ion nØ p o 🛛 śś'n⊠ n n ⊠∜n ions o \$ n P 14.16 Ĩ1 ns n'n 'n 0 0 p 🔥 0 0X 'nΜ on 🚬 o os pop 0.5 y.d1 0 nspi íŚ nk on n.s. s n′n⊠s o n in. N ₩ s. 0 onisn 0 • p ∎ís o 🖌 🛛 🐼 👖 st S. s in 'n o 0 n o s 0 5 .⊠, .p.s.s n🕅 on s Co 0 n s on n′n⊠o 0 0 0 \$ on on 0 0 s o M n′nµ⊠ íon 0 o⊠í 'n⊠p ʻn⊠slo n′n⊠ 'n 0 ions on p s'n⊠ Mo. \$ n 0 'n⊠ "pi 'n n b. c 0 jn⊠ is ns-.s p isin⊠ n ş n p 0 s ín " n ionsns ,00, is 'n 0 0 0 n ાંડે 📓 i ion o p.s. 00, os Ó , juž , o isi o so iso 🏹 11 s. 0 3 o non pa a s s ún.s.g ₩ n i N N рØ

o ⊠s ín o po n പൂ⊠ .ട്.ജ്. ņ 3 \* -Con 🛛 on n 👡 ýn sp n 💊 .sp n ро 0 .s on o s. \$3 🖌 🛛 р pp n ş pi ions o Со n 0 0 3 ní s 'n op nʻq p ş 0 ŝì A₫Ø õ 0 ş ʻp o .s– 0 n 0 s. opoínsa ņş 0 s,n \* .ss i iono po in P 🙀 ion o ş n 0 iso ро ún.sn n Zņ⊠ 'n⊠ 'n po n ions n o n Ð 0 o⊾⊠n p⊠\_p 'n.s– y.s Ó .s \$ 0 p in possi s. 0 'n ₽Ø., .sp .st in s. рØ **"**p pp o p opos . 0 0 🕅 l.s × • 0 ns 🔊 on 'n o s n



FIGURE 1 | Flowchart of the whole pipeline for DTI prediction in this article (left) in comparison to the traditional pipeline (right), with the procedures of feature representations squared in dashed lines.

## METHOD

### Datasets

s. 0 3 .s stat s íons o 🖕 🛛 🖕 11 0 S o ■ o Bin maDB Net o o \_ftpo in a sin  $\begin{array}{c} m & 1 & 1 & 0 \\ m & 1 & 1 & 0 \\ m & 0 & 1 & 0 \\ m & 0 & 0 & 0 \\ m & 0$  $\begin{array}{c} \mathbf{M} \\ \mathbf{M} \\ \mathbf{M} \\ \mathbf{S} \\ \mathbf{N} \\ \mathbf{N} \\ \mathbf{S} \\ \mathbf{N} \\ \mathbf$ > \$ 8 > n s s s m **,⊠,**p∴s /\_,\$  $\begin{array}{c} \mathbf{S} & \mathbf{n} \\ \mathbf{n} & \mathbf{0} \\ \mathbf{n} & \mathbf{0} \\ \mathbf{0} \\ \mathbf{D} \\ \mathbf{\Sigma} \\ \mathbf{B} & \mathbf{n} \\ \mathbf{n} \\ \mathbf{0} \\ \mathbf{D} \\ \mathbf{M} \\ \mathbf{B} & \mathbf{n} \\ \mathbf{n} \\ \mathbf{0} \\ \mathbf$ >> n 🛛 🕷 .s .o o **,⊠s** , so o o **, ⊠s** n o D. Is "Ap ( , 1, s s s o Api Api5

n as -1 1 1 1 .s. , on o p⊠ p in s 4 s on p n n  $\mathbb{A}$  os  $-\hat{A}$  s on s on s  $\mathbb{A}$  of p n  $\mathbb{A}$   $\mathbb$ on p , , n .ક .કં po.s oop p m on s po m \$ 8 Vino isg ng station spate to past ," " po'n s– 🕷 "po "ío B n po o ist n po in ⊠s in o MB n po o ist n psion sp in o ns'on sp in o D**´₂⊠**B´n´ 、 n n s ion 0

# Machine Learning Models

jn⊠s– on ș po ist po⊠. Mo is n ns ssi s o o – BD 🔳 is .ssi ion po po po s p i ion o in o o p i ion o s pi ision s- $n'n \boxtimes \boxtimes O$   $n \circ O$  non  $n \circ f$  n  $s^{d}$  o p = O f n  $f \circ f$  n n  $n \circ n^{d}$  n  $s^{d} \otimes \mathbb{N}n_{s}^{d}$ únp n sd o i ņ

# **RESULTS AND DISCUSSION**

## **Biochemical Implications of SPVec** Features

6 po io i pi ions o o X8 in D XB n p 3 o B sp son in Figure 2A poin p s ns so n in Figure 2A poin p s ns so n in Figure 2A poin p s ns so n in Figure 2A poin p s ns so n in Figure 2A poin p s ns so n in Figure 2A poin p s ns so n in Figure 2A poin p s ns so n in Figure 2A poin p s ns so n in Figure 2A poin p s ns so n in Figure 2A poin p s ns so n in Figure 2A poin p s ns so n in Figure 2A poin p s ns so n in Figure 2A p s s so n in Figure 2A 0.5 

⊠**\_**¶o \_, s n n ion s iiiso poind d do s n o p o poins site o A ou D DB 1 - j n n ips o 1 - 2B s o s o 1 - 2B s o s op n 'n D 🏹 0 po n 0 1 ່ຫຼິ່ຫ⊠ີ່ກ s p pions-Figure 2B s o s o o s op 1 still s as o p n i in s s s A op n in  $\mathbb{M}$  o o o on  $\mathbb{M}$  s D po n D A op in po - 11 in D DB M n A op n in  $\mathbb{M}$  o o n A op n 'n , , o o on , , , , D DB ( n A op n n  $\mathbb{M}$  n o o o o o n  $\mathbb{M}$ D  $(A op ) n po n An <math>\mathbb{M}$  n po - ppost( ppoB nppoB o OB o  $\mathbb{M}$ 

FIGURE 2 | Biochemical implications from SMILES2Vec features. (A) Visualizations of the SMILES2Vec vector space of drugs in DrugBank using t-SNE. (B) The top 10 drugs most similar to Acetophenazine (DrugBank ID: DB01063) according to their SMILES2Vec vectors. Red values show their cosine distances with Acetophenazine. The smaller the value, the more similar in the chemical structures. FIGURE 3 | Normalized distributions of biochemical and biophysical properties in a 2D space projected by t-SNE from the 100-dimensional ProtVec protein-space. In these plots, each point represents a protein, and the colors indicate the scale for each property.

BD. 🛍 y.s n D .331 A\_Cs sin⊠ Ø 0 ACC 手了. n AAC ₩, n on 0 A 2 0 n n ACC 0 3 0 ņ \_ ACC Ø 0 ياتي . n n ACC AACí n D. d p ions sin⊠ ş n ions o .so p P , son .ssi 'on p is .so n 0 **"**р メン **,** ⊠k ∫in ∭ati .s .s n \$ no o on р 0 3 s o po n р ор оро ind s ions n po in р .ssi ions-Table 2 .so 1 'n ş ş "n P ssd-0 on ŝ ~sd 'n⊠ĎB BD. 🔳 D n 0 .,s 4 Ş .s. s o 'n no .ssi íon po n 0 ,Ø D. Is n Cso ision 5 0 p ¥. 1 n 4 , i fond D. 🛋 p s. .s .so n š. s'n⊠ pop .ss i in Table 3s. 0 .s .s 33 ņ ₽ \* ns'ons δΩoί 3 0 'n nsion ×sd С .s n 0 ۵. ņ 0 ~ n .sp ín so ş "р 0 'n s. 0 0 in s ín ş pos. n ро 🕅 n 🔊 🕺 o 0 n – Ł. 0 o no 3 .s ., Yo D**⊿**B n n s. \* ູກ 📓 , s'ons o⊠– n 🔥 s. on n 4 ∽<sup>s</sup>ł **,**0 ₩ **,** p í 11 0 ions sp ,⊠. s'on -**⊠**B n ō p D 2 4 1 'n on OM O 33

 TABLE 1 | Number of entries of the five different datasets obtained from

 DrugBank dataset.

Datasets	Dataset_1	Dataset_2	Dataset_3	Dataset_4	Dataset_5
Drug	6,068	6,068	537	6,068	537
Target	3,839	3,839	3,839	160	160
Interactions	15,434	3,348	1,735	264	37

. fond Øł 'n , n n po in in " p **.**.s 'n .s. s. o í'n íon n 0 0 n l n <u>p</u>⊠ ., s'n⊠ •₽•, ° Table S1 \$0 0 no BD. 🔳 n X sions o 1 ņ s. 0 D 3 3 on 'n 'n şр **5**0 n • 0 1 р 0 0 р s in Table 3 is s in 🛛 n **\*\*** 

# **Evaluation of the Robustness of SPVec**

n o 0 o snjso in D. Is p 🚬 i ions ş sp ji 'n ζ'n X n o n ions .s .,s ∎B 'n.s., s. ons D n ion D ín ín 🛛 🕏 0 .s n ons n n 0 ş р ś on .s .,s n ion p n 1 🖬 1 odd 🕅 ion n ş on n s. 1 × 0 р 0 yd i on ín р'n ş s. Ś ņ n 3 - 1 .s′n⊠ BD. 🖿 D in Table 4 n \$ С 📓 🛃 o n in Figure 4 – As in Table 4 · \* **≰**⊠o ş pp o íφp q -A A\_C on 'n m 🙀 🗴 id 0 s on Bín ín⊠DB−Á o**,**⊠ D s o р., 0



**TABLE 2** | Results of classification performance of four feature combinations using three classifiers on BindingDB via 10  $\times$  5-fold cross-validation, with the highest scores highlighted in the bold font.

Feature combinations	Model	AUC	Accuracy	Precision	Recall	F1-score
SPVec (SMILES2Vec-ProtVec)	GBDT	0.9923	0.9680	0.9695	0.9667	0.9681
	RF	0.9927	0.9675	0.9808	0.9540	0.9672
	DNN	0.9617	0.9332	0.9287	0.9248	0.9197
SMILES2Vec-AAC	GBDT	0.9037	0.8272	0.8563	0.7873	0.8204
	RF	0.8770	0.7974	0.8657	0.7050	0.7772
	DNN	0.8708	0.8124	0.7993	0.7879	0.7126
MACCS-ProtVec	GBDT	0.9479	0.8810	0.8908	0.8690	0.8798
	RF	0.9302	0.8542	0.8712	0.8322	0.8512
	DNN	0.9136	0.8034	0.8025	0.8097	0.8074
MACCS-AAC	GBDT	0.8588	0.7811	0.8077	0.7392	0.7719
	RF	0.8360	0.7468	0.8366	0.6150	0.7089
	DNN	0.8451	0.7832	0.7884	0.7726	0.7724

TABLE 3 | AUCs of SPVec and other models on DTI predictions using DrugBank.

**, ⊠**, s no s n po n D D ds 0 nc ,n 'n 5 no ju íŚ .so ons n õ s ion in s po íon o ni post 🗑 Τí po p 📓 n 🔊 👢 ,s−A 1 s. 5 ions o 3 4 ∍⊠ D. Is 'n **,**0 o .sp n 5 ín íp 🛛 🕏 Ś 0 n ss 0 n 0 4 ¥gn BD, 0 ş onØ p ... n 155 s in D p ons **8**<sup>d</sup> Cś s. on o**, 🛛** BD. n А BD. s o 5 on 0 3 "s ş 0 3 'n on p BD. × **Ø** "n ion p 0 ø íon ø ۵, ş n o n \* , 🛛 í'n no ions in c íon 0 4 BD. p i ion .s. 🔯 👌 o jis 3 p 'n n , p ,i ion o no⊠, D.Isin possi ø, 0 D position in A

# Prediction and Validation on Unidentified DTIs

0, ⊠ ⊠\_\_ o.s 1 on ín o 🕏 📓 \$ \$ 0 p , 🔞 • 'n , p , o n , sto D.™stp., i ions ovinniss in pit iono D. Adis on on , post , in post 🛛 🙀 in 🗸 ions 3.3 1 os post , n no n o ions 'n n 0 3 3 3 34 鮾 ions in D n 0 р o n 'n n 5 o**, ⊠** n 🛛 🗑 s., TΪ 'n possi i ions on D. Is ₩ р., 0 11 4 D. 🛋 n D 🖉 B n 0 n s. ø 33 \* BD.™ n s'n⊠ , op ⊠, n ín 🔪 ʻons 4

Drug features	Drug dim.	Protein features	Protein dim.	ML method	AUC	References
Drug structure information	2,216	AAC, DC <sup>a</sup> and TC <sup>b</sup>	11,943	DNN	0.81	You et al., 2019 <sup>c</sup>
Constitutional, topological and molecular descriptors, 2D autocorrelations, topological charge indices, eigenvalue-based indices	1,664	AAC; DC <sup>a</sup> ; autocorrelation; Composition, Transition, Distribution descriptors; Quasi-sequence-order	1,080	RF	0.8950	Yu et al., 2012 <sup>c</sup>
Constitutional, topological and geometrical descriptors	193	AAC; DC <sup>a</sup> ; autocorrelation; composition, transition and distribution; quasi-sequence-order; amphiphilic pseudo-amino acid composition and total amino acid properties	1,260	DT RF	0.760 0.855	Ezzat et al., 2016
PubChem fingerprints indicating presence or absence of 881 known chemical substructures	881	Fingerprints of 876 different protein domains that are obtained from the Pfam database	876	EnsemDT	0.882	Ezzat et al., 2017
SMILES2Vec	100	ProtVec	100	GBDT RF DNN	0.9467 0.9469 0.8637	This work

<sup>a</sup>DC, dipeptide composition; <sup>b</sup>TC, tripeptide composition; <sup>c</sup>These models are trained on different versions of DrugBank, whose AUCs are only as references.

**TABLE 4** | Results of classification performance using three classifiers on datasets obtained from DrugBank, with the highest scores highlighted in the bold font.

Dataset	Model	AUC	Accuracy	Precision	Recall	F1-score			
Training set	10 × 5-fold cr	10 x 5-fold cross-validation							
Dataset_1	GBDT	0.9506	0.9323	0.9456	0.9367	0.9343			
	RF	0.9557	0.9234	0.9378	0.9369	0.9337			
	DNN	0.8952	0.8732	0.8345	0.8437	0.8654			
Test sets	Independent	Independent validation							
	GBDT	0.8945	0.8628	0.8747	0.8696	0.8637			
Dataset_2	RF	0.8930	0.8753	0.8645	0.8467	0.8555			
	DNN	0.8201	0.8026	0.8138	0.8199	0.8144			
	GBDT	0.7502	0.7389	0.7340	0.7245	0.7333			
Dataset_3	RF	0.7448	0.7299	0.7198	0.7243	0.7230			
	DNN	0.6999	0.6922	0.6825	0.6798	0.6832			
	GBDT	0.7356	0.7223	0.7167	0.7177	0.7201			
Dataset_4	RF	0.7235	0.7034	0.7108	0.7078	0.71			
	DNN	0.7173	0.6899	0.6884	0.6896	0.6866			
	GBDT	0.68	0.6703	0.6679	0.6664	0.6688			
Dataset_5	RF	0.5689	0.5605	0.5398	0.5321	0.5411			
	DNN	0.6267	0.6098	0.607	0.6122	0.6114			

FIGURE 4 | ROC curves with different models on the test sets obtained from DrugBank.



#### TABLE 5 | Top five novel DTIs predicted by SPVec-GBDT.

Drug ID	Target ID	Drug name	Target name	Validation source
DB11805	P07947	Saracatinib	The tyrosine-protein kinase Yes	Patel et al., 2013
DB09282	P42262	Molsidomine	Glutamate receptor 2	None
DB05524	Q99640	Pelitinib	Membrane-associated tyrosine and threonine-specific cdc2-inhibitory kinase	https://pubchem.ncbi. nlm.nih.gov/compound/ 6445562
DB03017	Q16620	Lauric acid	BDNF/NT-3 growth factors receptor	None
DB13165	P11362	Ripasudil	Fibroblast growth factor receptor 1	None

(n(D, D, D, D, D, N)) (n(1, 0, 1))• • • • os'm n on'n sp i in i to in s  $\mathbb{Z}$ D i i s on C  $\mathbb{Z}$ n i  $\mathbb{Z}$  is on C  $\mathbb{Z}$ i in i b i s po n o o  $\mathbb{Z}$  i s  $\mathbb{Z}$ n i  $\mathbb{Z}$  of in to o pi  $\mathbb{Z}$  o po i in  $\mathbb{Z}$  or i o n i n  $\mathbb{Z}$  n i i o in s is in s o in o n i sp i in to in s is in s o in o n i -  $\mathbb{Z}$  s s ons is instant in n in n if n is a standard s

# CONCLUSION

Co in in 🛛 🔪 ínýp⊠ no o ns súp⊠so ⊠o ponsin poýnsig nistíno víoní no í piston ⊠os o í – ns 👡 'n o 

 ino
 ion i
 n
 o
 in sion
 is os
 o
 i

 is
 ion o
 is
 o s ni
 is
 s
 o
 in
 is
 o

 is
 ion o
 is
 is
 s
 o
 o
 in
 is
 o
 in
 is
 is
 o
 in
 is
 is</t

# REFERENCES

- As M = n = 0, - 1, -Con (n o s + s + 1),  $p \leq n$  for o (o M),  $s \in q$ ,  $n \leq 0$ ,  $p \neq 0$ ,  $0 \neq 1$ , M,  $n \neq 0$ ,  $0 \neq 0$ .  $1 = 1^{-1} 1^{-1} = 0$ ,  $1 = 1^{-1} -$
- $C \uparrow Y_{-} \uparrow X_{-} X_{-} X_{-} n C \circ = ppo \blacksquare \circ n s \circ p \uparrow ion \circ p \circ n s \circ o ion in \circ p \circ n 2 q sisq p \circ J. Cell. Biochem. <math>A_{-} \uparrow f + f \circ h + - f$



# DATA AVAILABILITY STATEMENT

n on in p<u>s</u> 

# **AUTHOR CONTRIBUTIONS**

YZ, X, nD, -YZ, nX, o, n, on pion n, Man, ZA, n, YCp, o, s, ist, n, sist-X, n, YZ o, n, s, ip -X, on i, op o, s, o n, s, ip -, ZZ, on i, op o, n, s, ip -A o, o, o, i, o, o, n, pp oM d, i, M, dion, n, pp oM d, i, M, YZ, X, n D s -Y Z n X st N s ⊠ ston-

# FUNDING

™is o sisppo , Manison ion in on iono Cin Con nos-jii ii , 1/ n  $V_{j}$   $V_{j}$ n  $V_{j}$ o  $V_{j}$ inis o  $v_{n}$ , n  $V_{n}$ o  $V_{j}$ inis o  $v_{n}$ , n  $V_{n}$ n so  $V_{n}$ n  $V_{n}$ n  $V_{j}$ n  $v_{n}$ 

## SUPPLEMENTARY MATERIAL

- $Co_{y_1} = -n_{y_2} p_{y_3} - \mathbf{I}_{y_3} v_{y_3}$

- Co  $\leq C_n$  pni -1 sppo  $\square$  o n, o  $\leq$ -Mach. Learn. -1
- sʻp⊠ –I. Mach. Learn. Res. \_

- oit -i \_\_n  $\mathbb{M}$   $\mathbb{M}$ i." o \_\_-B\_- o s \_\_ \_  $\mathbb{M}$   $\mathbb{Q}$ , C\_- n  $\mathbb{R}$  n  $\mathbb{A}$   $\mathbb{C}$   $\mathbb{N}$   $\mathbb{R}$  n in p  $\mathbb{M}$  n p pos p n n o o p  $\mathbb{N}$   $\mathbb{M}$   $\mathbb{N}$ p op  $\mathbb{N}$  arXiv [Preprint]. All  $\mathbb{N}$  on  $\mathbb{N}$  n  $\mathbb{R}$  -1:  $\mathbb{R}$   $\mathbb{Q}$   $\mathbb{Q}$   $\mathbb{R}$   $\mathbb{N}$   $\mathbb{Q}$   $\mathbb{Q}$   $\mathbb{R}$   $\mathbb{Q}$   $\mathbb{R}$   $\mathbb{Q}$   $\mathbb{R}$   $\mathbb{Q}$   $\mathbb{R}$   $\mathbb{R}$   $\mathbb{Q}$   $\mathbb{R}$   $\mathbb{Q}$   $\mathbb{R}$   $\mathbb{$ 1.\*\_\_\_
- اللہ ہے۔ میں جے opdi -C-B ہے -A-n ہے۔ -- اور ۱۹۹۰ ہے، میں میں معامی معامی معامی میں میں معامی میں معامی معامی میں معامی معامی میں معامی میں معامی معام عام چند چند ہے ہے کام اللہ جے اممار کے تعامی معامی معامی معامی کی معامی کی معامی معامی معامی معامی کی معامی مع