The best practices of multilearning for (Big) data analysis in chemoinformatics

Dr. Igor V. Tetko

Institute of Structural Biology, HelmholtzZentrum München

Strasbourg, June 26, 2018

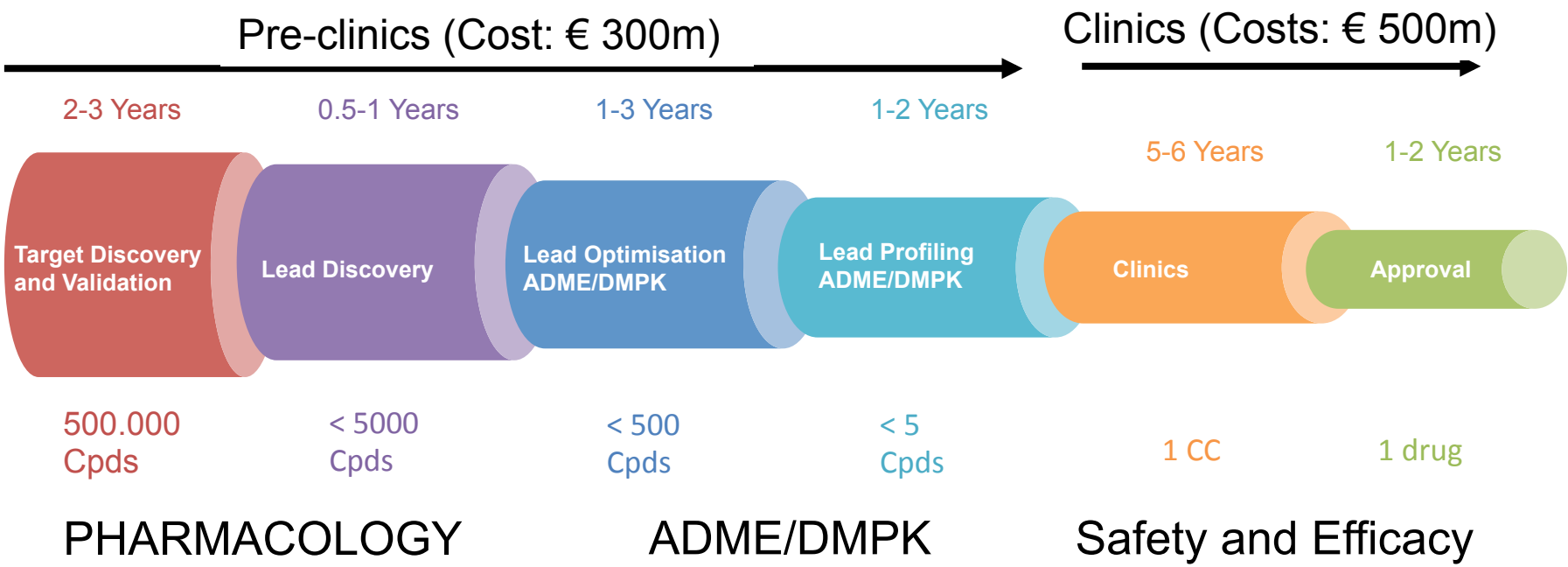
Institute of Structural Biology

HelmholtzZentrum münchen
German Research Center for Environmental Health

Outline

- Overview and motivation
- Multi-learning using neural networks
- Other methods
- Overview of available tools

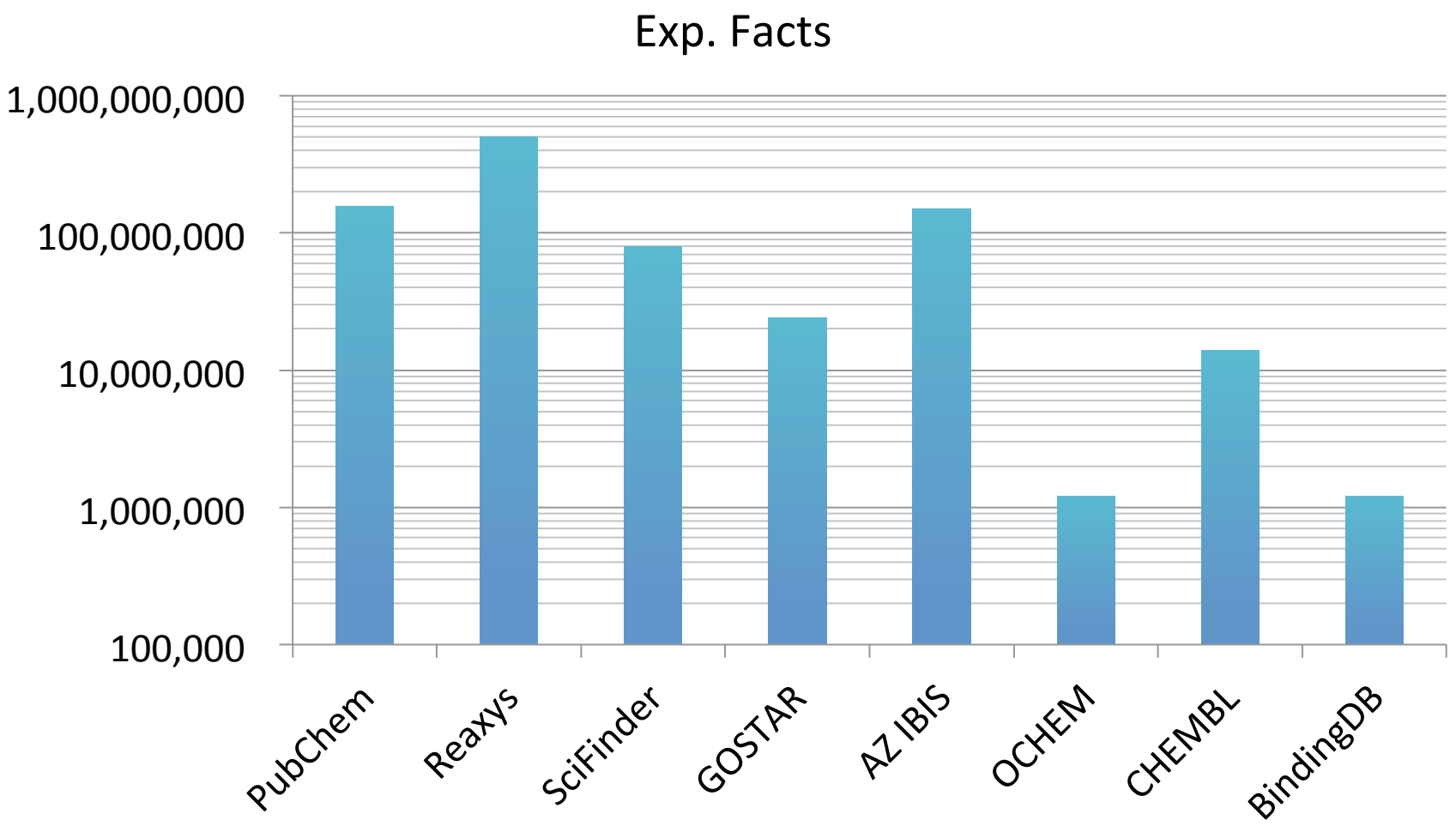
Big Data Sources - Process of Drug Discovery



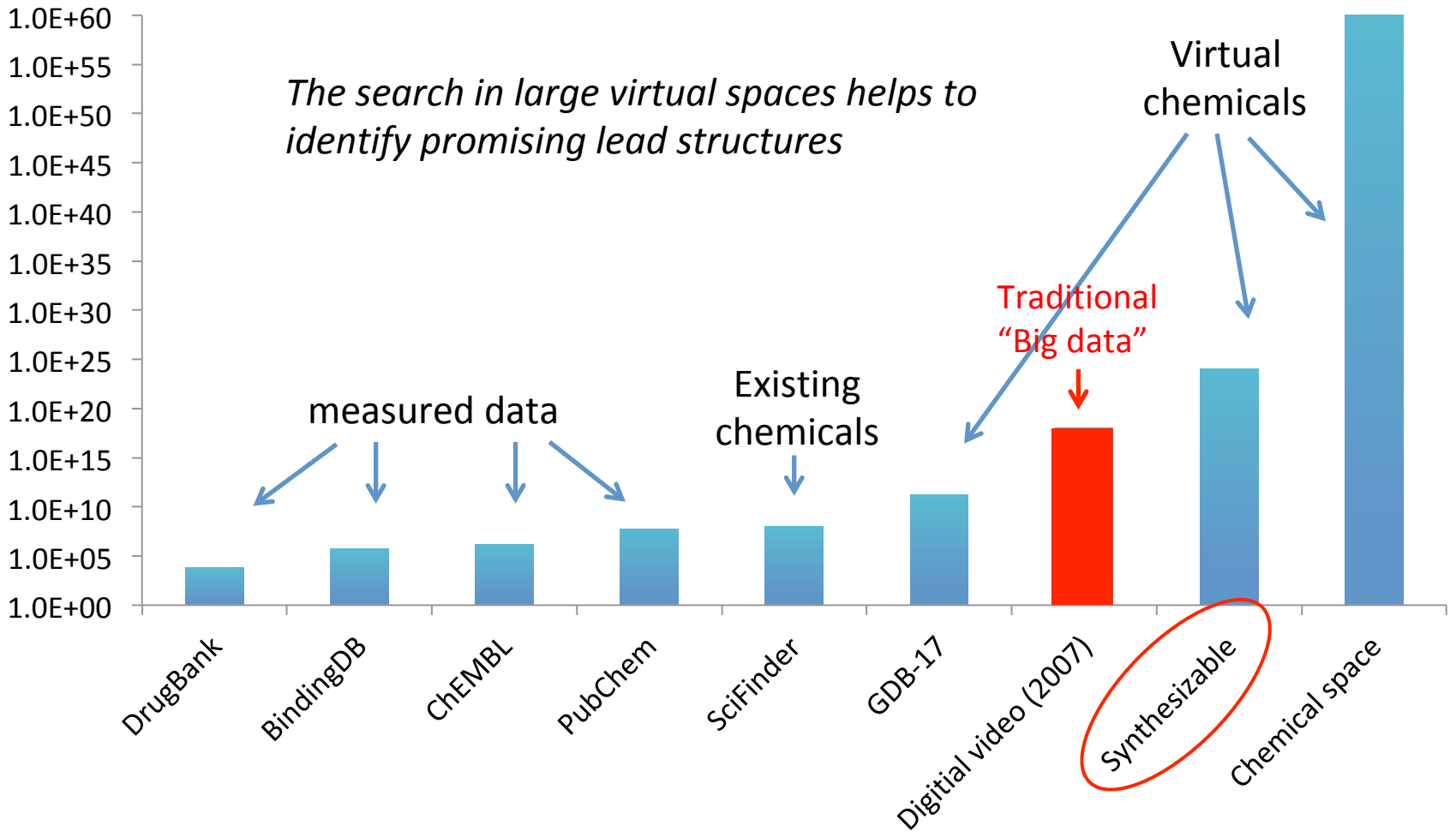
Profiling and screening in the virtual space helps to identify most promising candidates

Slide courtesy of Dr. C. Höfer, Vitilis

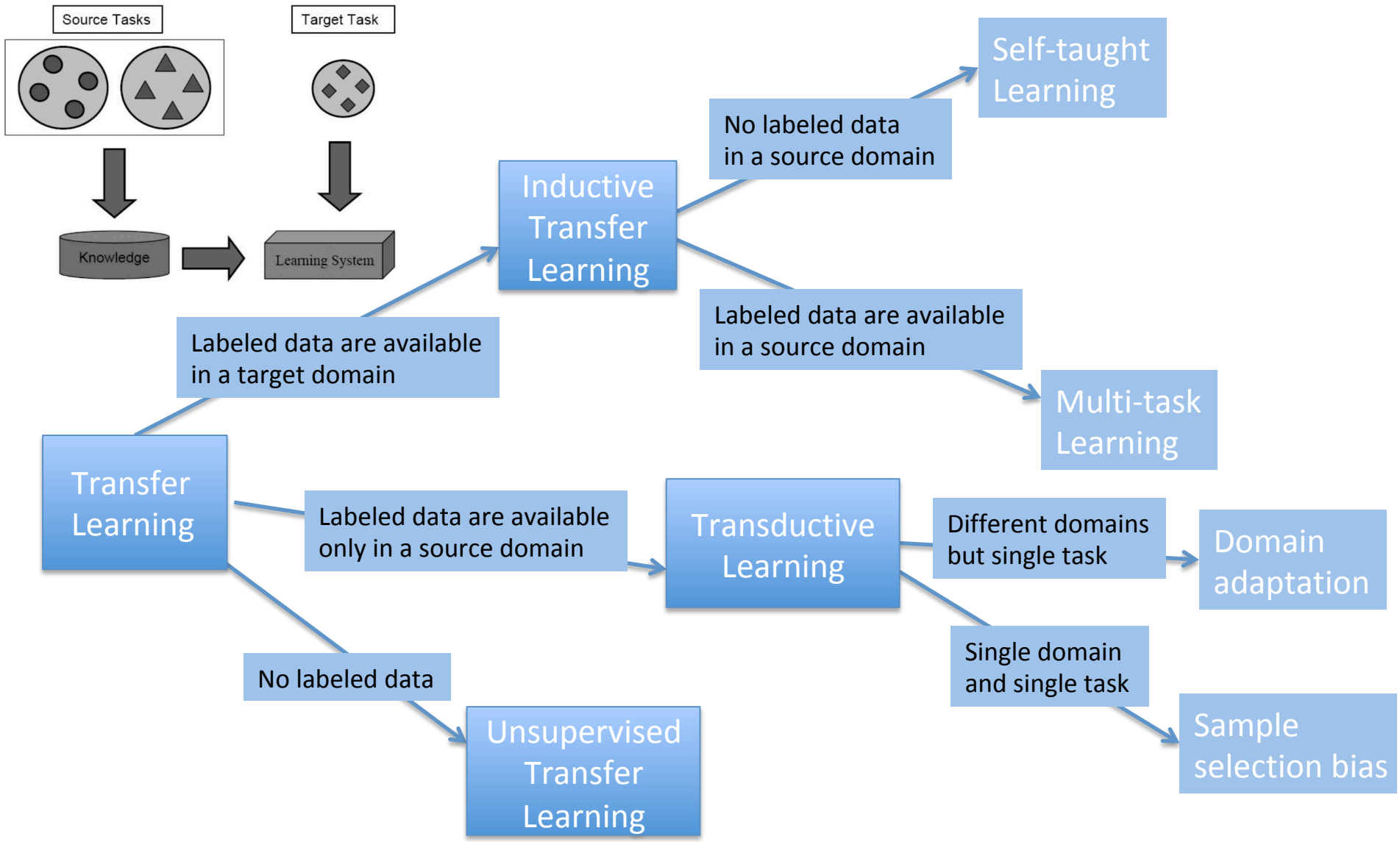
Big Data Sources - Large Chemical Database



Large virtual chemical spaces - Search of compounds in virtual chemical spaces: Big Data challenges

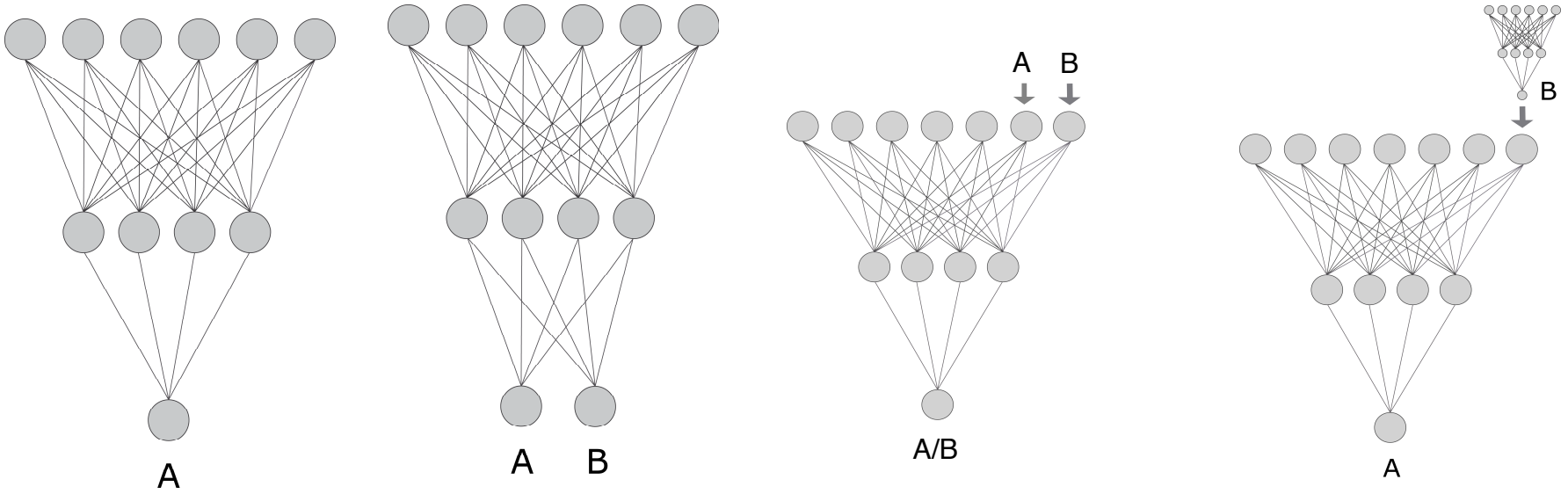


Tetko, I.V.; Engkvist, O.; Koch, U.; Reymond, J.L.; Chen, H. Bigchem: Challenges and opportunities for big data analysis in chemistry. *Mol. Inform.* **2016**, *35*, 615-621.

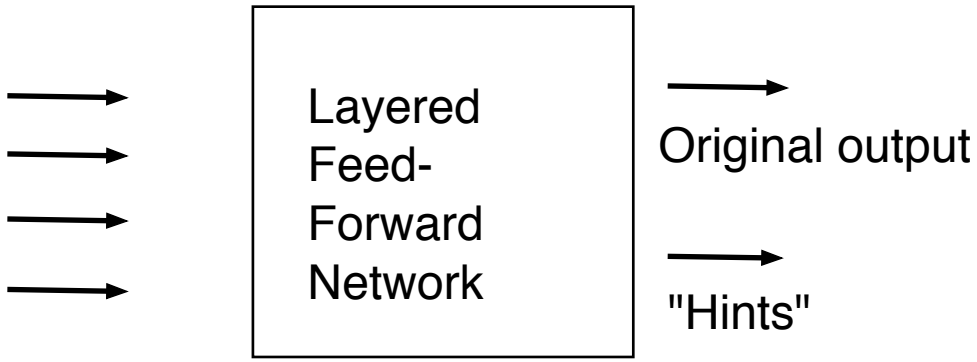


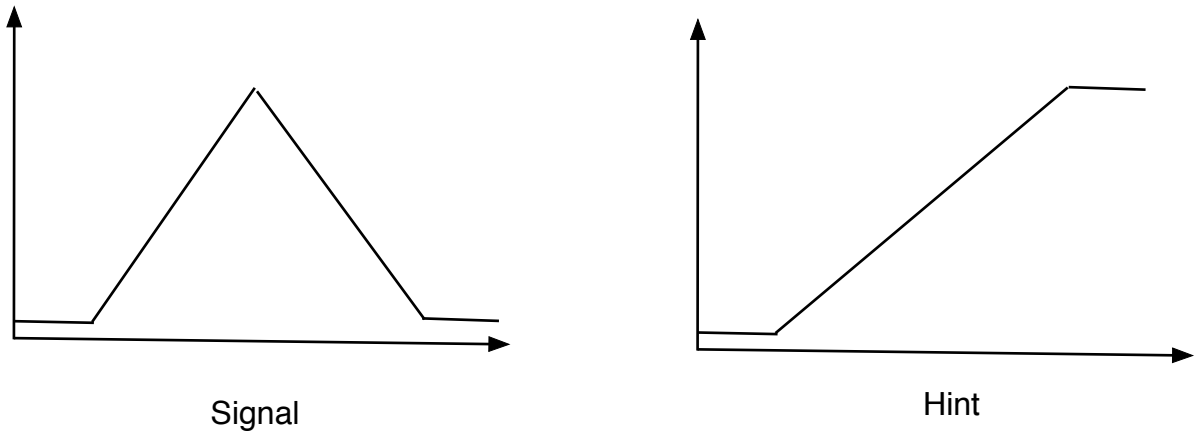
Adapted from: Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* **2010**, *22*, 1345-1359.

Multi-task learning



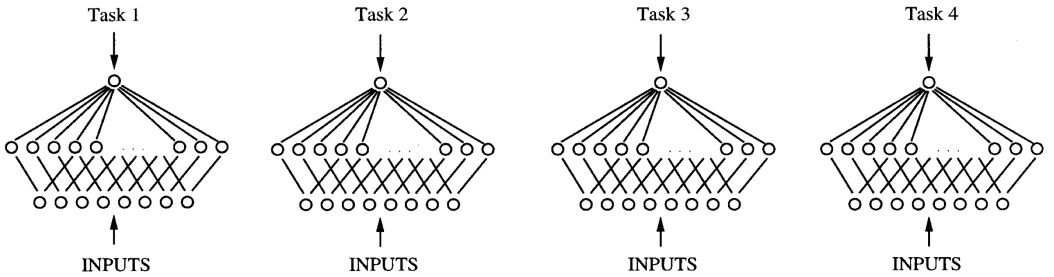
Neural networks with hints



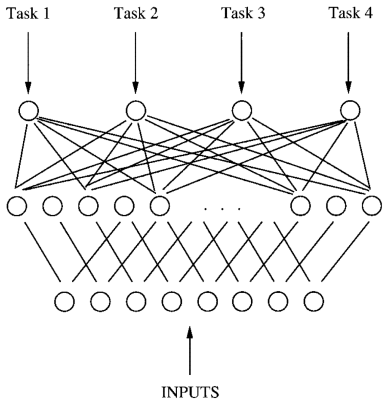


Suddarth, S.C.; Kergosien, Y.L. In *Rule-injection hints as a means of improving network performance and learning time*, Neural Netw., Berlin, Heidelberg, 1990, pp 120-129.

Caruana "multi-learning"



http://cs.cornell.edu/~caruana



Caruana, R. Multitask learning. *Machine Learn.* 1997, *28*, 41-75.

Why does multi-learning (can) work better?

- Amplification of statistical data (noise reductions)
- Attention focusing (finding better signal in a noisy data)
- Eavesdropping (learning “hints” from simpler tasks)
- Representation bias/feature selection (selection of common features)
- Regularisation (less overfitting)

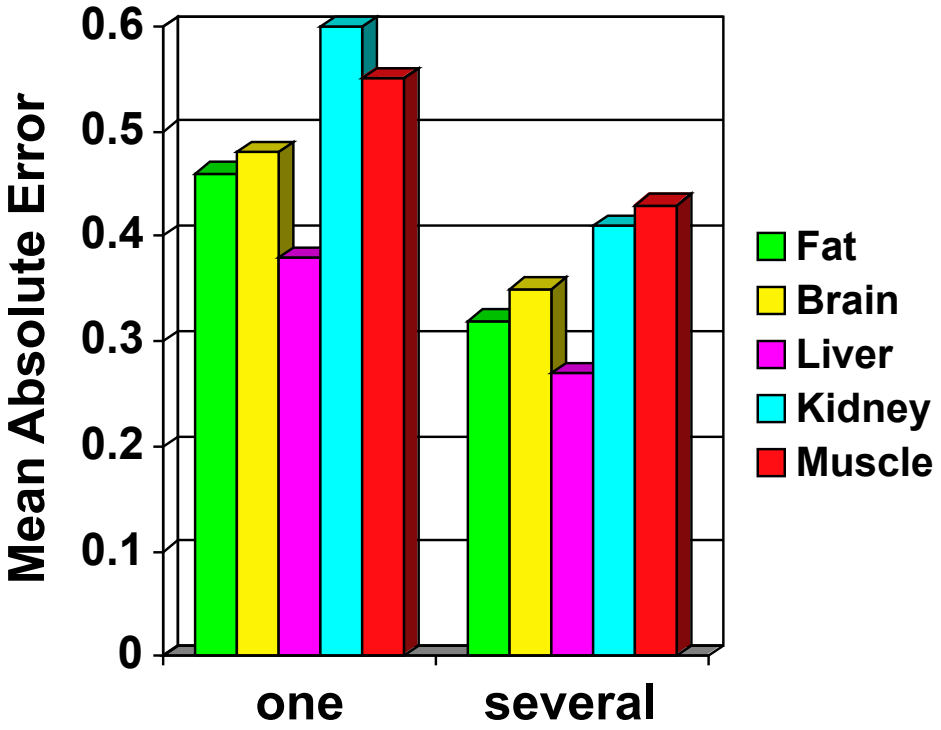
Multi-task learning

Problem:

• prediction of tissue-air partition coefficients
• small datasets 30-100 molecules (human & rat data)

Results:

simultaneous prediction of several properties increased the accuracy of models



Varnek, A. et al J. Chem. Inf. Model. 2009, 49, 133-44.

Prediction of toxicity of chemical compounds:
REGISTRY OF TOXIC EFFECTS OF CHEMICAL SUBSTANCES (RTECS®)

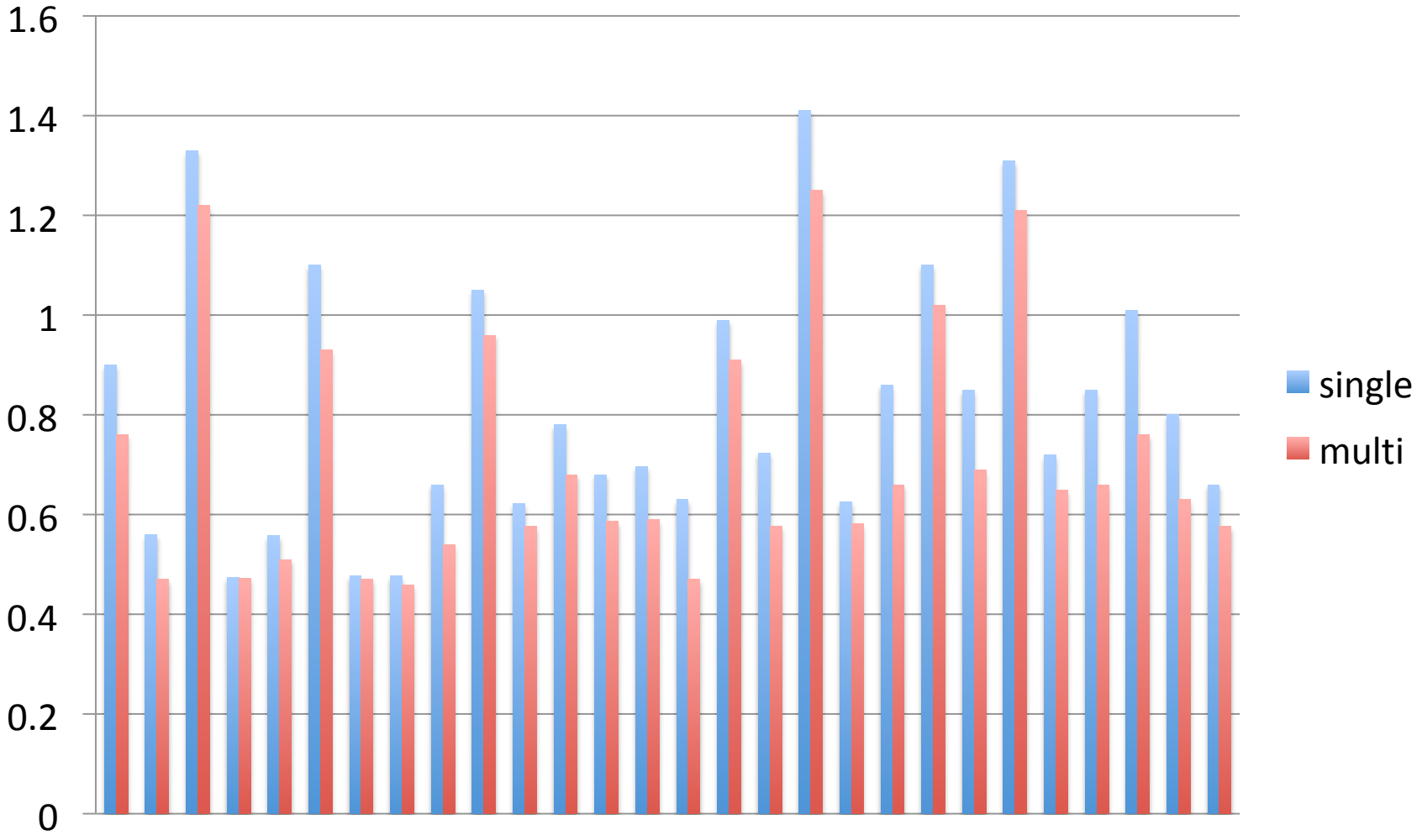
Different species
- Rat
- Mouse
- Rabbit
- …
- Human

~ 129k records
~ 87k compounds
29 properties

- Different toxicities
  - LD50
  - TDL
  - NOEL
  - LDLo
- Administartion
  - Oral
  - IPR (intraperitoneal)
  - IVR (intravenous)

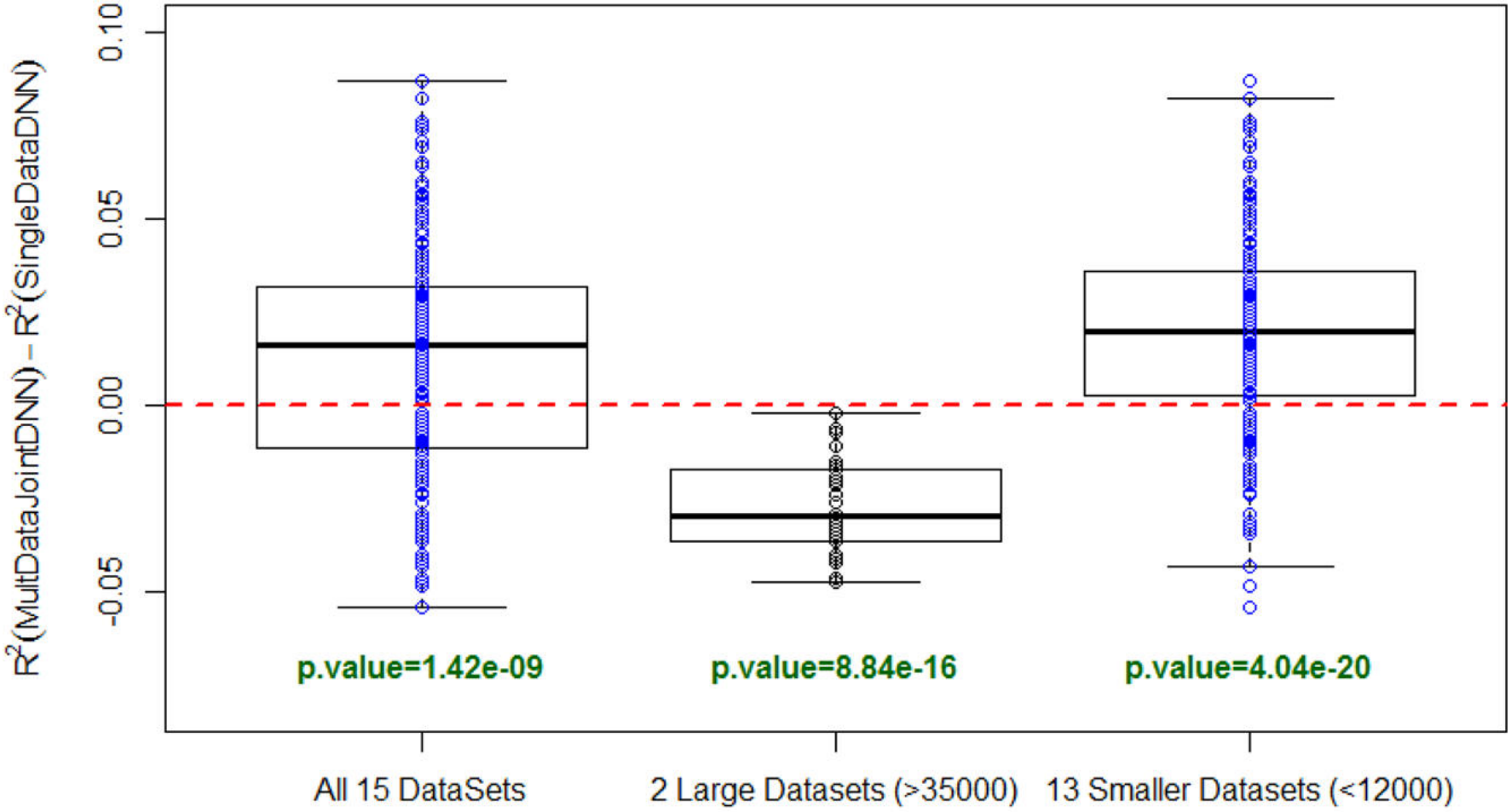
Sosnin, S.; Karlov, D.; Tetko, I.V.; Fedorov, M.V. A comparative study of prediction of multi-target toxicity for a broad chemical space. *Chem. Res. Toxicol.* 2018, *in prep*.

RMSE for different toxicities using CDK descriptors



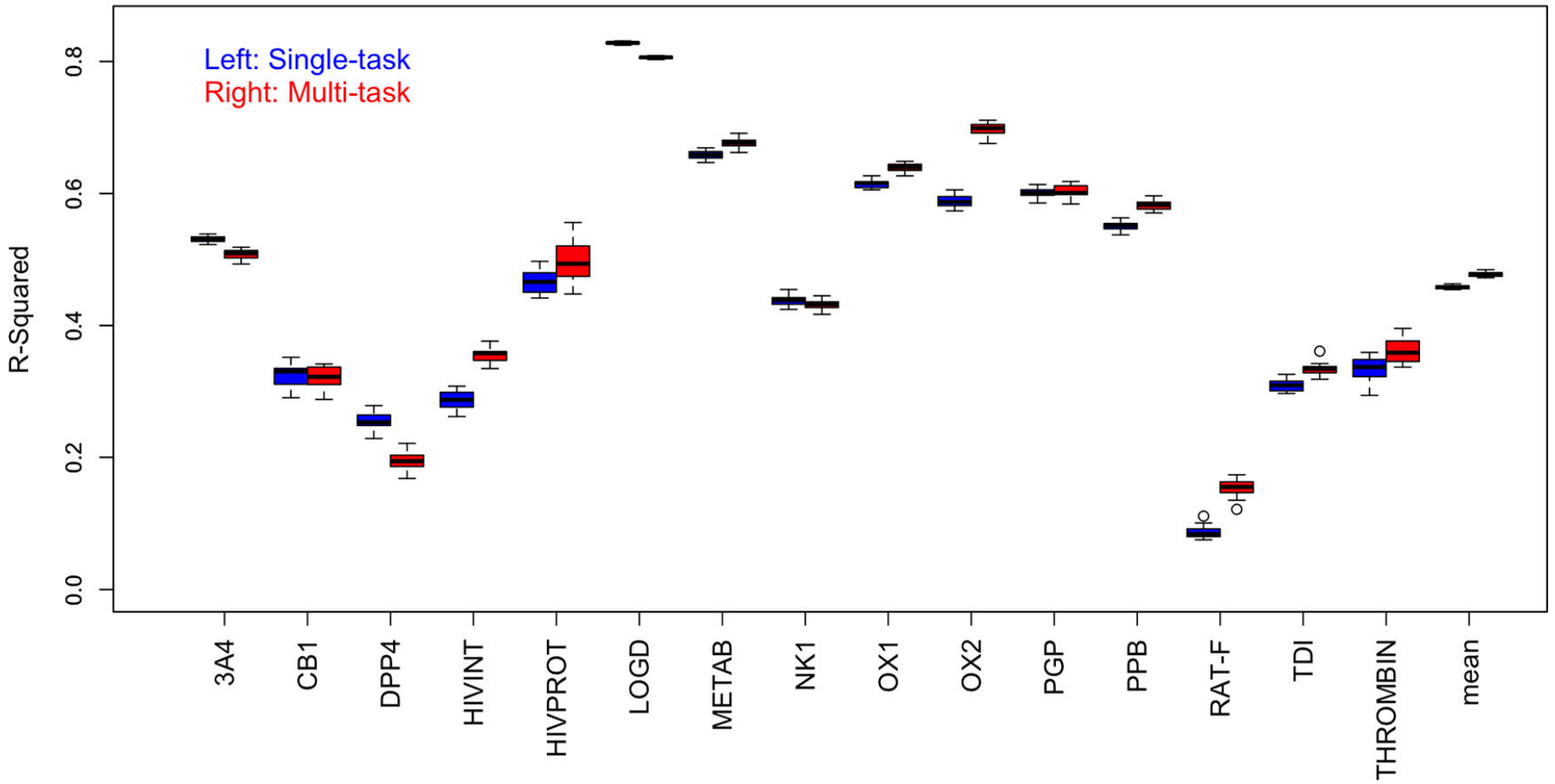
Sosnin, S. et al. A comparative study of prediction of multi-target toxicity for a broad chemical space. *Chem. Res. Toxicol.* 2018, *in prep*.

Multi-learning vs. single-task learning using DNN



Ma, J.; Sheridan, R.P.; Liaw, A.; Dahl, G.E.; Svetnik, V. Deep neural nets as a method for quantitative structure-activity relationships. *J. Chem. Inf. Model.* 2015, *55*, 263-274.

Comparison of MTL and STL



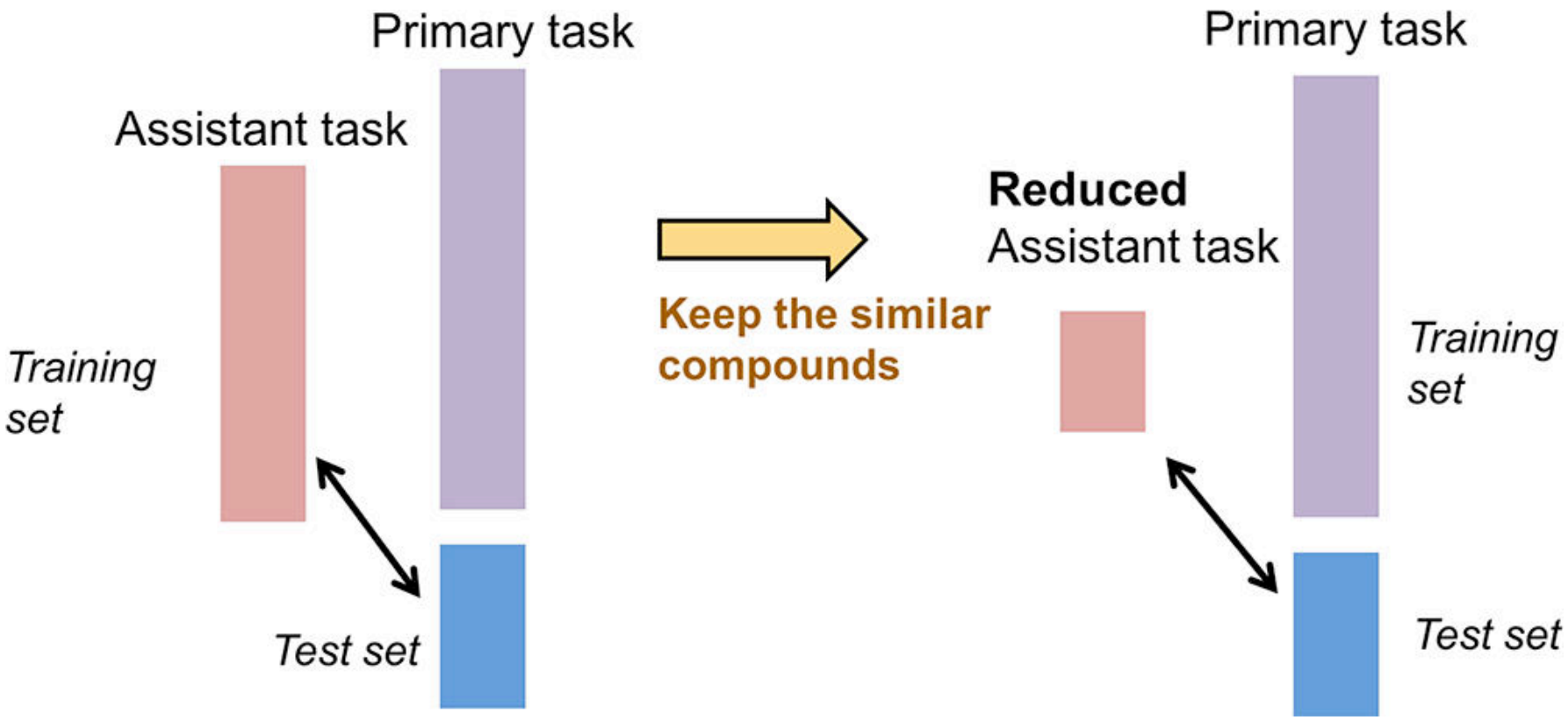
Xu, Y.; Ma, J.; Liaw, A.; Sheridan, R.P.; Svetnik, V. Demystifying multitask deep neural networks for quantitative structure-activity relationships. *J. Chem. Inf. Model.* 2017, *57*, 2490-2504.

When MTL can over-perform STL?

Similar molecules modelling correlated properties will boost the predictive performance of the DNN, and likewise uncorrelated properties will degrade performance.

Structurally dissimilar molecules have no influence on the predictive performance of the MTL DNN, regardless of whether or not tasks are correlated.

Proposed selection of an "assistant task"



Xu, Y.; Ma, J.; Liaw, A.; Sheridan, R.P.; Svetnik, V. Demystifying multitask deep neural networks for quantitative structure-activity relationships. *J. Chem. Inf. Model.* 2017, *57*, 2490-2504.

When can MTL over-perform STL?

Molecular structure	Molecular activity	results
primary test set molecules are more similar to assistant training set molecules	primary data set and assistant data set have correlated activities (positive or negative)	improved prediction R^2 for the primary test set
	uncorrelated biological activities	decreased prediction R^2 for the primary test set
primary test set molecules are very different from assistant training set molecules	correlated or not	no significant change of prediction for the primary test set

Xu, Y.; Ma, J.; Liaw, A.; Sheridan, R.P.; Svetnik, V. Demystifying multitask deep neural networks for quantitative structure-activity relationships. *J. Chem. Inf. Model.* 2017, *57*, 2490-2504.

Feature selection/transformation methods

Partial Least Squares

Singular Value Decomposition

Matrix factorization

Common idea: decrease dimensionality of the descriptor space by finding (transformation of) features correlated with the analyzed activities

Partial Least Squares (PLS)

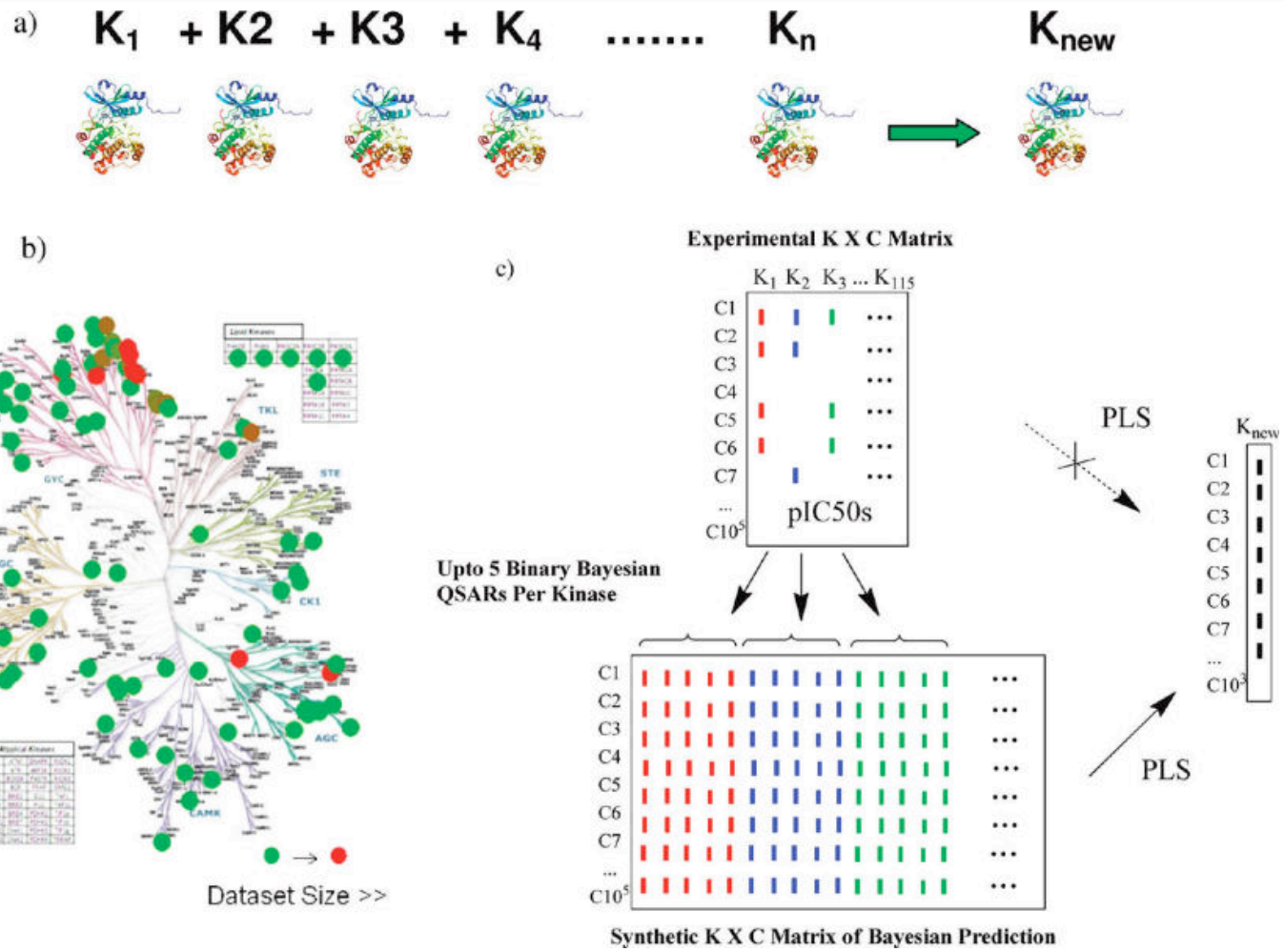
$$Y = XB + F$$

Y – output target property
X – descriptors
B – regression coefficients
F – residuals

$$B = X^T U (T^T X X^T U)^{-1} T^T Y$$

T - scores and **U -** latent variables

Profile-like QSAR



Martin, E.; Mukherjee, P.; Sullivan, D.; Jansen, J. Profile-QSAR: A novel meta-QSAR method that combines activities across the kinase family to accurately predict affinity, selectivity, and cellular activity. *J. Chem. Inf. Model.* **2011**, *51*, 1942-1956.

Macau

$$X \approx UV^T$$

$$X = \begin{bmatrix} 1 & 2 & \dots & M \\ 2 & & & \\ \dots & & & \\ N & & & N,M \end{bmatrix}$$

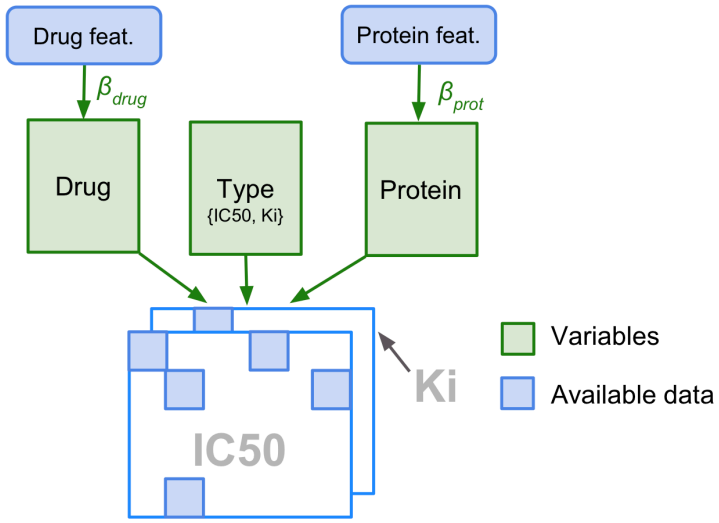
N,M > k

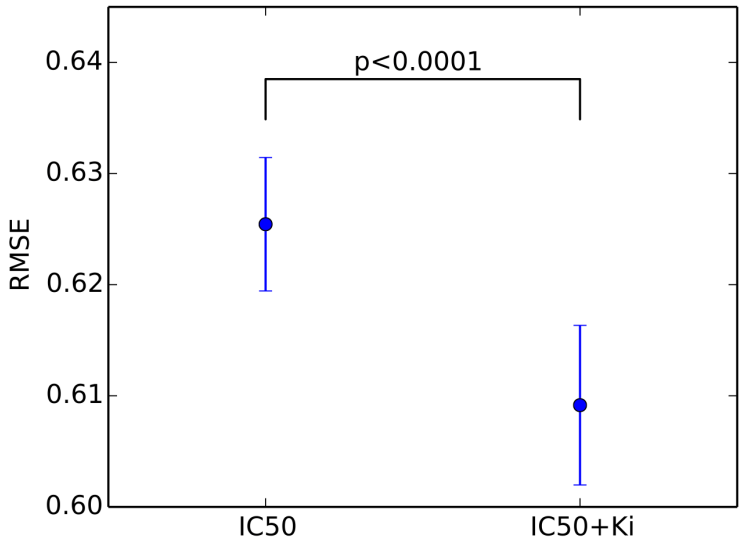
$$\min_{u,v} \sum_{(i,j)\in I_x} \left(X_{ij} - u_i v_j^T \right)^2 + \lambda_u \left\| u \right\|^2 + \lambda_v \left\| u \right\|^2$$

$$U = \begin{bmatrix} 1,1 & 1,k \\ \dots & \\ \dots & \\ N & N,k \end{bmatrix} \qquad V = \begin{bmatrix} 1,1 & 1,k \\ \dots & \\ \dots & \\ M & M,k \end{bmatrix}$$

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. *ArXiv e-prints* **2015**, 1509.04610.

Macau





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. *ArXiv e-prints* **2015**, 1509.04610.

Learning similarity of tasks: k Nearest Neighbors

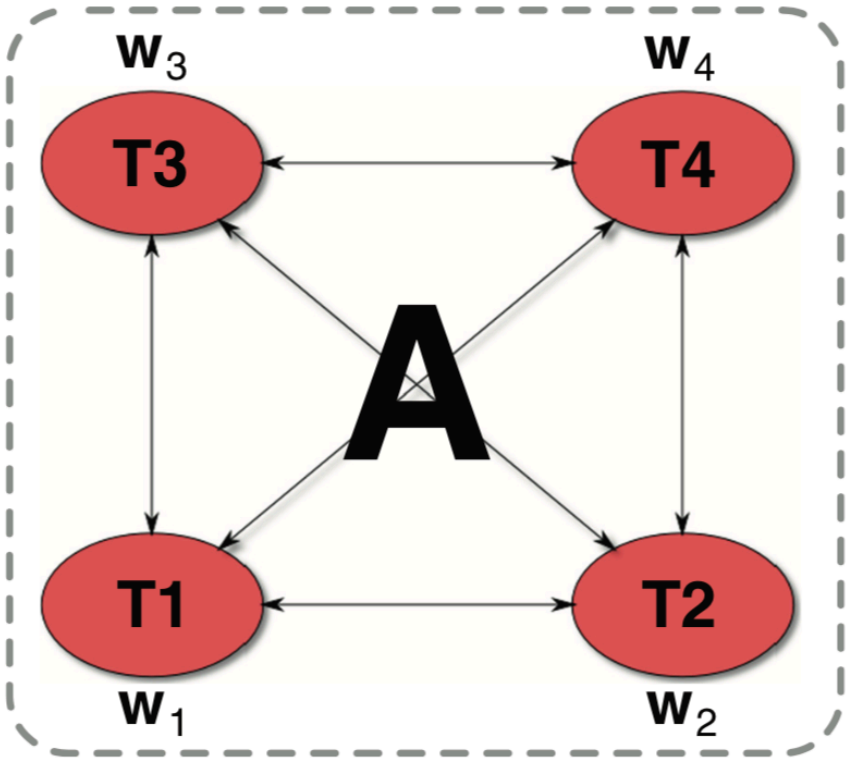
$$d_M(x_i, x_j) = \sqrt{(x_i - x_j)^T M (x_i - x_j)}$$

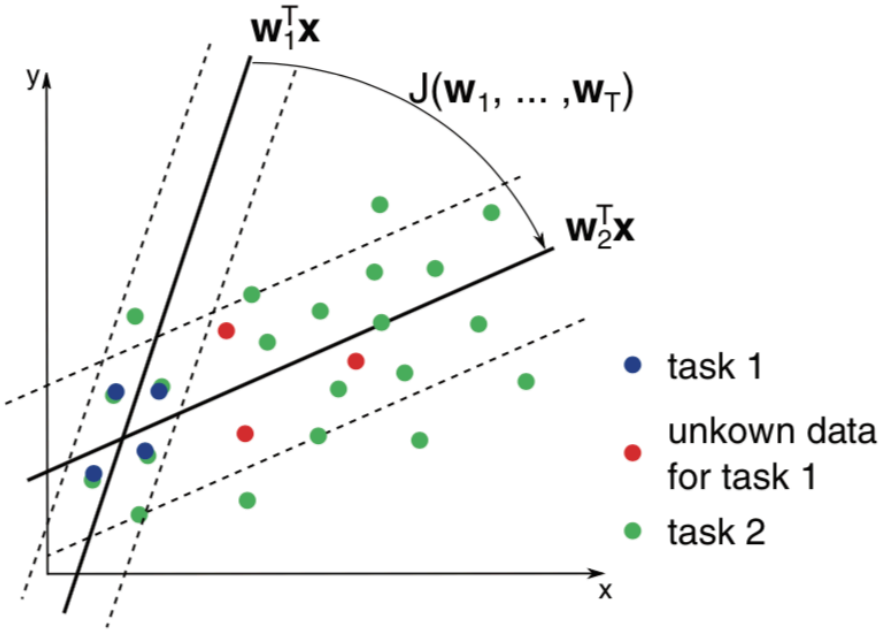
$$d_t(x_i, x_j) = \sqrt{(x_i - x_j)^T (M_0 + M_t)(x_i - x_j)}$$

Multi-learning: M_0 captures main variation + Mt task specific

Parameswaran, S.; Weinberger, K.Q. Large margin multi-task metric learning. In *Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 2*, Curran Associates Inc.: Vancouver, British Columbia, Canada, 2010; pp 1867-1875.

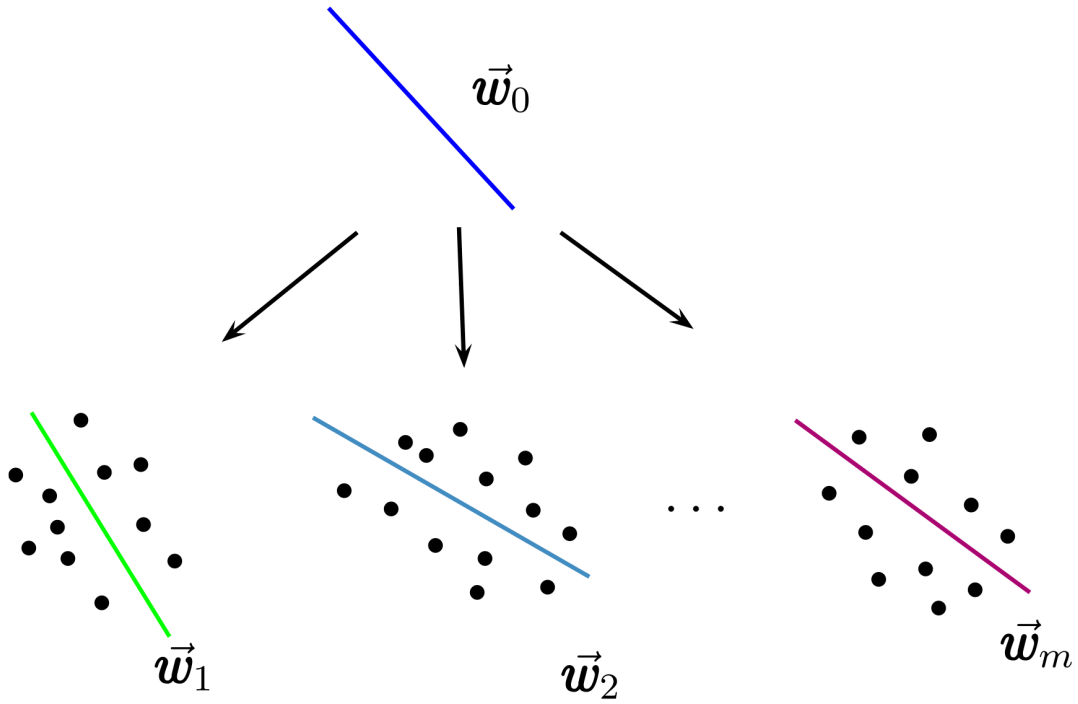
Graph-regularized multi-task Support Vector Regression





Rosenbaum, L.; Dorr, A.; Bauer, M.R.; Boeckler, F.M.; Zell, A. Inferring multi-target QSAR models with taxonomy-based multi-task learning. *J. Cheminform.* **2013**, *5*, 33.

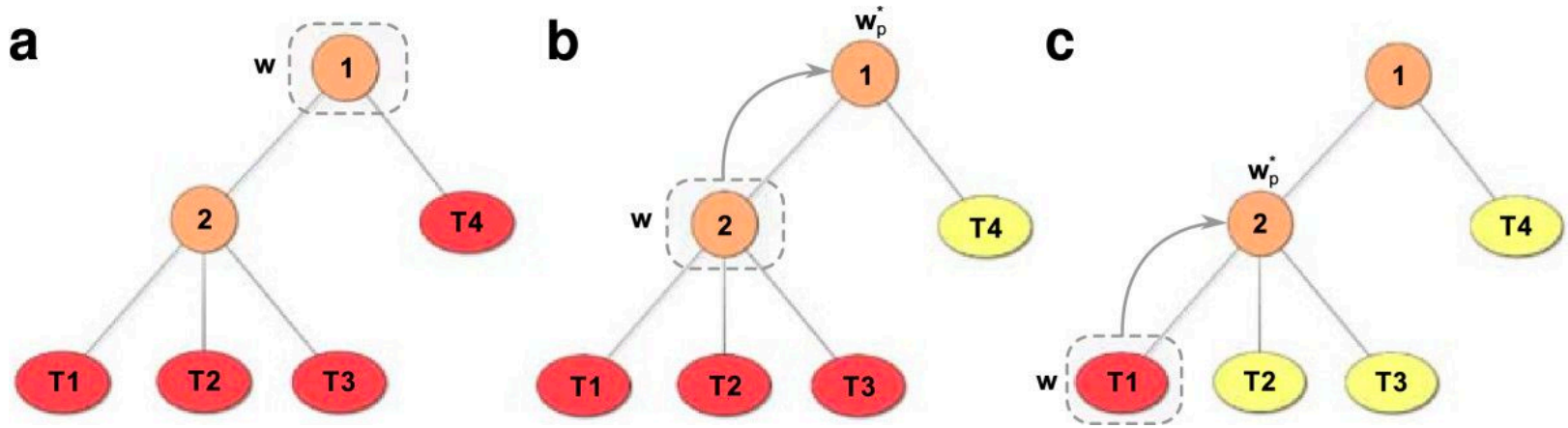
Non-neural network approaches to multi-learning: Least Squares Support Vector Regression (LSSVM)



Suykens, J.A.K.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* 1999, *9*, 293-300.

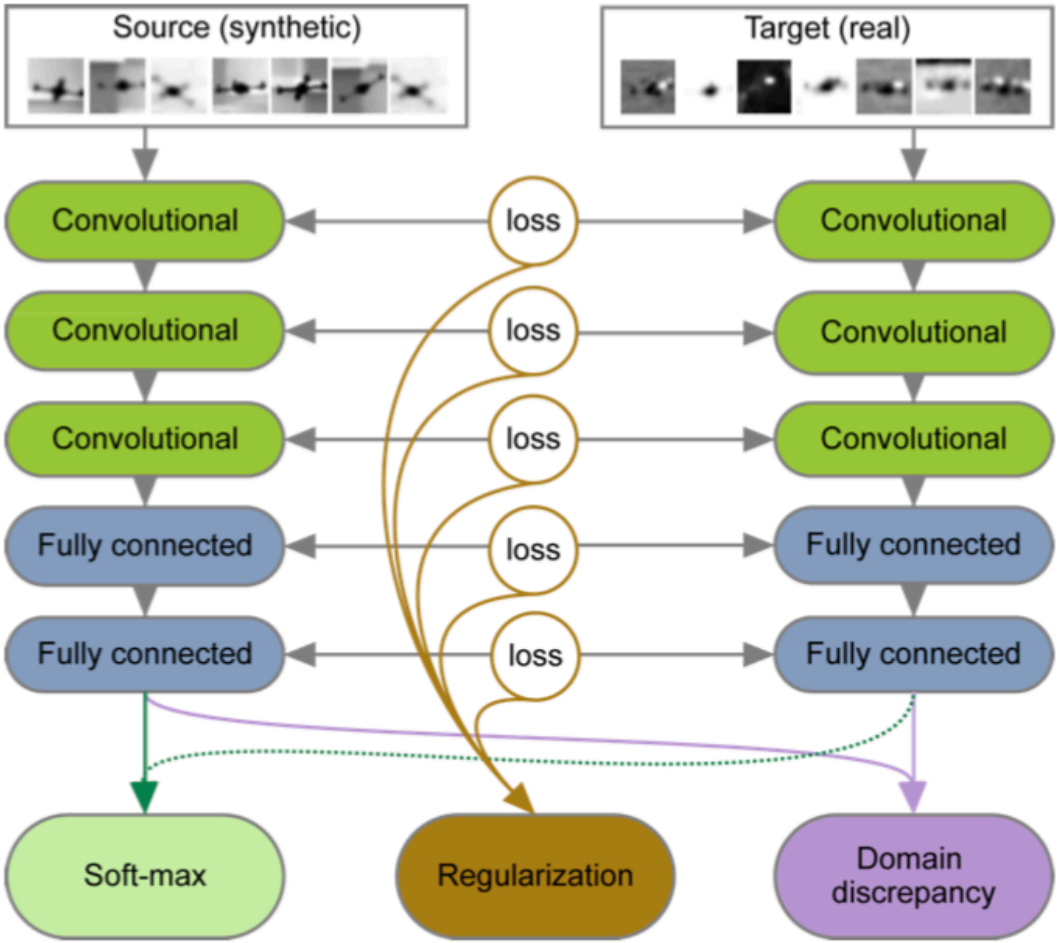
Xu, S.; An, X.; Qiao, X.; Zhu, L.; Li, L. Multi-output least-squares support vector regression machines. Pattern Recognition Letters 2013, 34, 1078-1084.

Hierarchical Classification



Rosenbaum, L.; Dorr, A.; Bauer, M.R.; Boeckler, F.M.; Zell, A. Inferring multi-target QSAR models with taxonomy-based multi-task learning. *J. Cheminform.* **2013**, *5*, 33.

Learning features and tasks similarity simultaneously



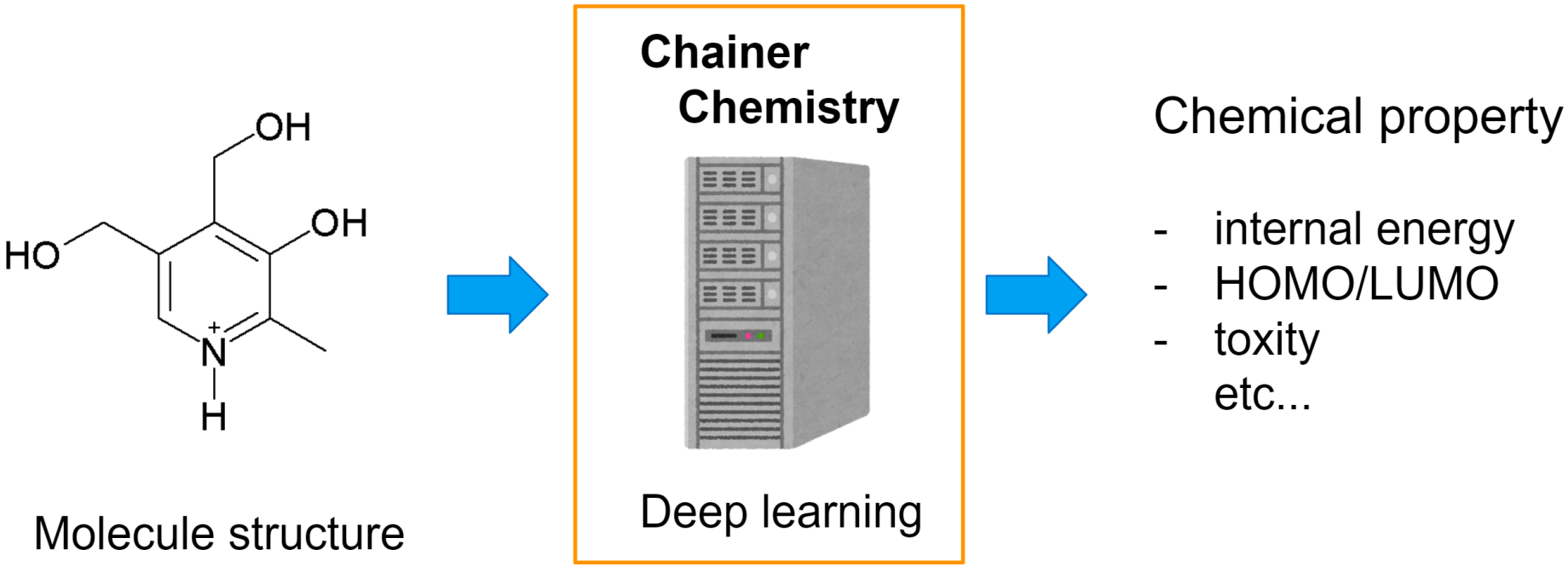
Rozantsev, A.; Salzmann, M.; Fua, P. Beyond sharing weights for deep domain adaptation. *eprint arXiv:1603.06432* **2016**, arXiv:1603.06432.

"Ready" available tools

- GitHub: "the treasure house" of available methods
- Specialized tools
  - Chainer Chemistry
  - DEEPCHEM
  - OCHEM

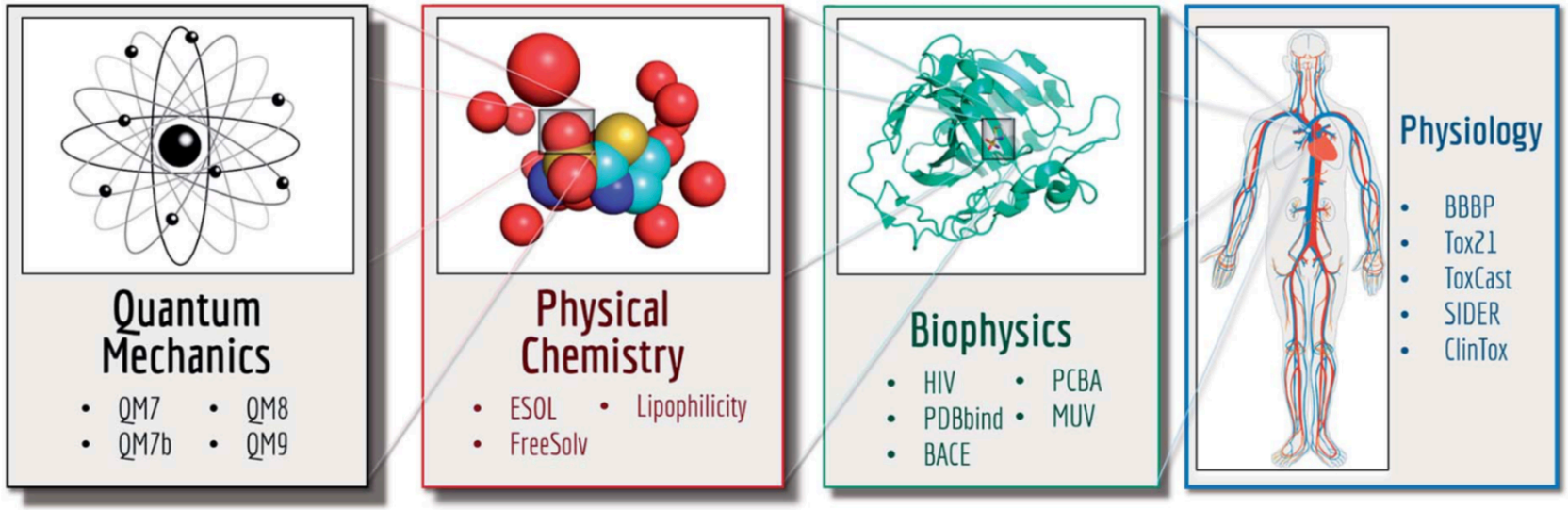
Chainer Chemistry ("ChemChainer")

- Chainer – one of popular frameworks for Deep Learning
- Algorithms provided by Chainer developers
- Can be installed using Python tools
- https://github.com/pfnet-research/chainer-chemistry



DEEPCHEM

- Based on TensorFlow (google)
- Available as part of Python (Anaconda) or as a Docker
- Supports multiple MTL and STL approaches
- https://github.com/deepchem/deepchem



Wu, Z.; Ramsundar, B.; Feinberg, E.N.; Gomes, J.; Geniesse, C.; Pappu, A.S.; Leswing, K.; Pande, V. Moleculenet: A benchmark for molecular machine learning. *Chem Sci* **2018**, *9*, 513-530.

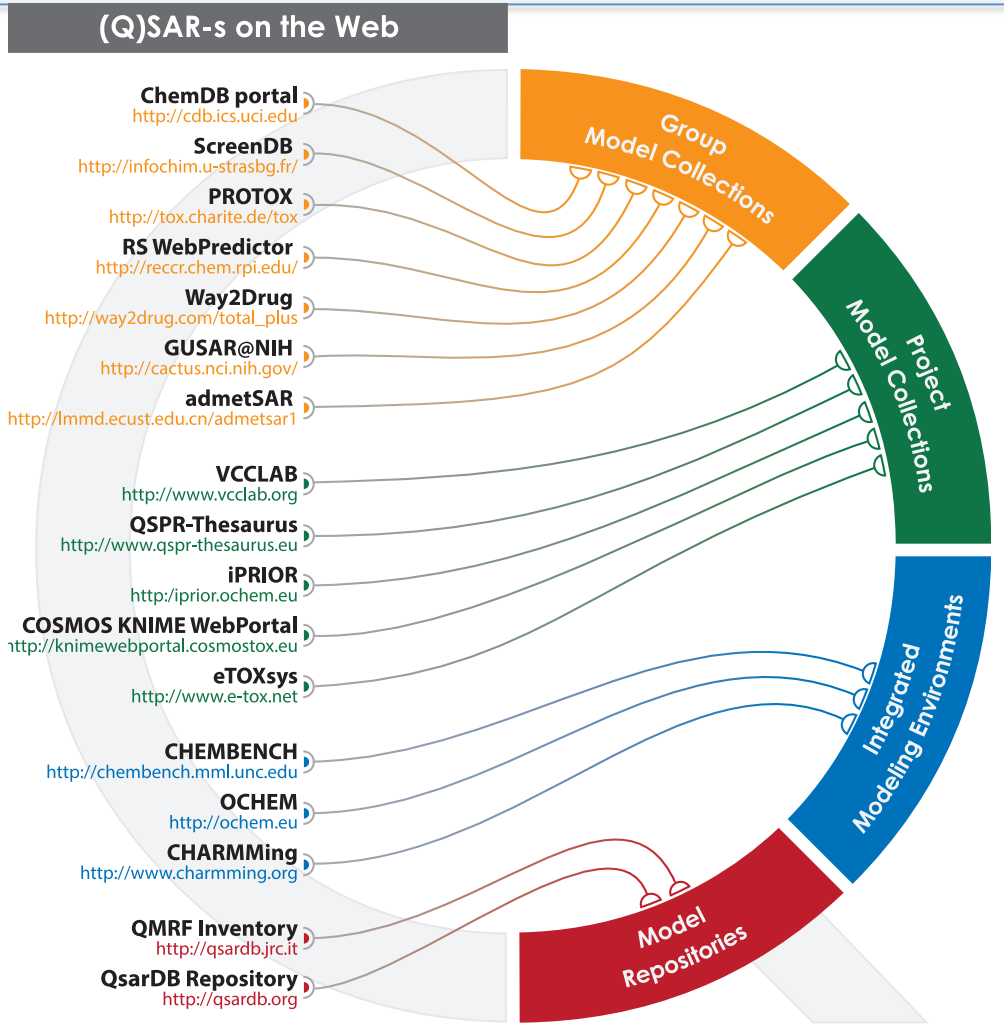
OCHEM

Supports algorithms from ChemChainer, DEEPCHEM

Provides a number of original algorithms

Models can be published and be freely shared on web

http://ochem.eu



Tetko, I.V.; Maran, U.; Tropsha, A. Public (Q)SAR services, integrated modeling environments, and model repositories on the web: State of the art and perspectives for future development. *Mol. Inform.* **2017**, *36*.

Summary of "readily" available methods

Package	Examples of supported algorithms
Chainer Chemistry	NFP, GGNN, RSGCN, WeaveNet, SchNet
DeepChem	DAG, NNF, MPNN, TEXTCNN, WEAVE, IRV
OCHEM	Above methods + DNN, LSSVM, Macau, feature learning as well as use of tasks classes as descriptors

NFP/NNF - Neural Fingreprint; GGNN - Gated Graph Neural Network; MPNN - Message Passing Neural Networks; SchNet - continuous-filter convolutional neural network for modeling quantum interactions; DAG - Directed Acyclic Graphs; IRV - Influence Relevance Voters ; LSSVM − Least Squares Support Vector Machines

Comparison of different models to predict toxicity (RMSE)

single multi single

s RMSE - Root Mean Square Error for Training set Validation: Cross-Validation (63 models)

	DNN	DNN(2)	XGBOOST
CDK2 (constitutional, topological, geometrical, electronic, ...	0.9 0.56 1.33 0.474 0.56 1.1 0.478 0.477 0.66 1.05 0.623 0.78 0.68 0.7 0.63 0.99 0.724 1.41 0.63 0.86 1.1 0.85 1.31 0.72 0.85 1.01 0.8 0.66 1.27 (0.834)	0.76 0.47 1.22 0.472 0.51 0.93 0.471 0.459 0.54 0.96 0.576 0.68 0.59 0.591 0.47 0.91 0.577 1.25 0.581 0.66 1.02 0.69 1.21 0.65 0.66 0.76 0.63 0.58 1.14 (0.725)	0.8 0.47 1.29 0.454 0.5 1.02 (0.439 0.56 1.04 0.584 0.75 0.65 0.59 0.95 0.66 1.33 0.5 0.75 1.08 0.764 1.3 0.67 0.81 0.76 0.63 1.2 (0.779)
Dragon6 (blocks: 1-29)	0.89 0.58 1.3 0.458 0.56 1.06 0.481 0.472 0.6 1.06 0.63 0.74 0.66 0.686 0.63 0.97 0.69 1.32 0.622 0.82 1.09 0.83 1.33 0.76 0.83 0.98 0.8 0.7 1.24 (0.82)	0.78 0.44 1.31 0.445 0.474 0.96 0.461 0.446 0.52 1 0.555 0.68 0.55 0.581 0.47 0.95 0.57 1.31 0.574 0.65 1.08 0.68 1.2 0.68 0.67 0.74 0.64 0.59 1.22 (0.732)	0.8 0.49 1.3 0.454 0.523 1.01 0.439 0.59 1.02 0.588 0.73 (0.66 0.602 0.94 0.67 1.33 0. 0.76 1.09 0.77 1.38 0.68 0.82 0.74 0.63 1.24 (0.786)
ALogPS, OEstate	0.91 0.61 1.32 0.461 0.54 1.1 0.478 0.469 0.6 1.1 0.617 0.75 0.7 0.652 0.64 1 0.69 1.36 0.617 0.84 1.11 0.87 1.43 0.76 0.85 0.95 0.8 0.71 1.2 (0.832)	0.79 0.44 1.23 0.447 0.49 0.94 0.467 0.444 0.53 0.99 0.554 0.66 0.55 0.59 0.49 0.9 0.58 1.21 0.571 0.65 1.05 0.69 1.22 0.65 0.7 0.74 0.64 0.6 1.17 (0.724)	0.84 0.5 1.42 0.456 0.519 1 (0.44 0.56 1.03 0.58 0.73 0.5 0.65 0.61 0.95 0.64 1.34 0.59 1.11 0.79 1.33 0.69 0.8 0.81 0.63 1.21 (0.786)
Fragmentor (Length 2 - 4)	0.96 0.61 1.43 0.463 0.542 1.14 0.491 0.484 0.62 1.1 0.647 0.81 0.71 0.71 0.64 1.04 0.74 1.38 0.643 0.79 1.14 0.86 1.33 0.82 0.86 0.94 0.84 0.66 1.22 (0.849)	0.73 0.45 1.25 0.44 0.48 0.95 0.465 0.448 0.502 0.99 0.554 0.65 0.55 0.56 0.46 0.92 0.575 1.28 0.564 0.63 1.07 0.69 1.24 0.7 0.66 0.73 0.63 0.62 1.2 (0.724)	0.78 0.45 1.38 0.447 0.52 1 0.476 0.436 0.58 1.09 0.592 0.61 0.67 0.59 0.94 0.67 1.3 (0.77 1.14 0.79 1.43 0.69 0.83 0.77 0.64 1.29 (0.797)

Sosnin, S. et al. A comparative study of prediction of multi-target toxicity for a broad chemical space. *Chem. Res. Toxicol.* 2018, *in prep*.

Comparison of MTL and STL

Multiple models overview

Predicted property: Cblood/Cair(Human)
Training set: tissue/air set

Metrics RMSE - Root Mean Square Error for Training set Validation: Cross-Validation (16 models)

	ASNN	MTL	DNN	ASNN(2)	STL	DNN(2)
CDK2 (constitutional, topological, geometrical, electronic, ...	0.45 0.28 0.21 0.29 0.39 0.33 0.28 0.32 0.4 0.33 0.4 (0.335)		0.54 0.33 0.38 0.35 0.4 0.45 0.321 0.43 0.44 0.49 0.52 (0.423)	0.41 0.41 0.45 0.42 0.44 0.56 0.279 0.5 0.39 0.37 0.44 (0.424)		0.549 0.45 0.54 0.48 0.71 0.66 0.35 0.6 0.46 0.44 0.71 (0.541)
OEstate	0.44 0.35 0.31 0.33 0.4 0.44 0.32 0.33 0.33 0.31 0.36 (0.356)		0.42 0.29 0.31 0.32 0.38 0.41 0.31 0.33 0.41 0.37 0.4 (0.359)	0.41 0.47 0.44 0.51 0.66 0.6 0.37 0.57 0.5 0.39 0.48 (0.491)		0.44 0.35 0.46 0.41 0.4 0.46 0.38 0.48 0.47 0.41 0.57 (0.439)

	DAG	GRAPH_CONV	TEXTCNN	WEAVE
MTL	0.75 0.55 0.6 0.35 0.94 0.67 0.44 0.64 0.58 0.57 0.92 (0.637)	0.93 0.64 0.8 0.58 1 1 0.6 0.79 0.85 0.89 0.8 (0.807)	0.53 0.4 0.43 0.33 0.48 0.53 0.35 0.53 0.47 0.48 0.5 (0.457)	0.7 0.69 0.8 0.61 0.9 0.64 0.41 0.74 0.57 0.61 0.7 (0.67)
STL	0.63 0.52 0.9 0.47 1.1 1 0.38 0.8 0.62 0.62 1 (0.731)	0.8 0.61 0.9 0.7 0.9 0.78 0.65 0.8 0.86 0.92 0.9 (0.802)	0.58 0.54 0.57 0.51 0.7 0.63 0.39 0.66 0.51 0.62 0.48 (0.563)	0.62 0.52 0.7 0.59 0.8 1.1 0.48 0.71 0.72 0.72 0.8 (0.705)

Multi-learning

Can be useful method for analysis of correlated properties

Can provide better accuracy of predictions (Xu et al, 2017)

Provides smaller models

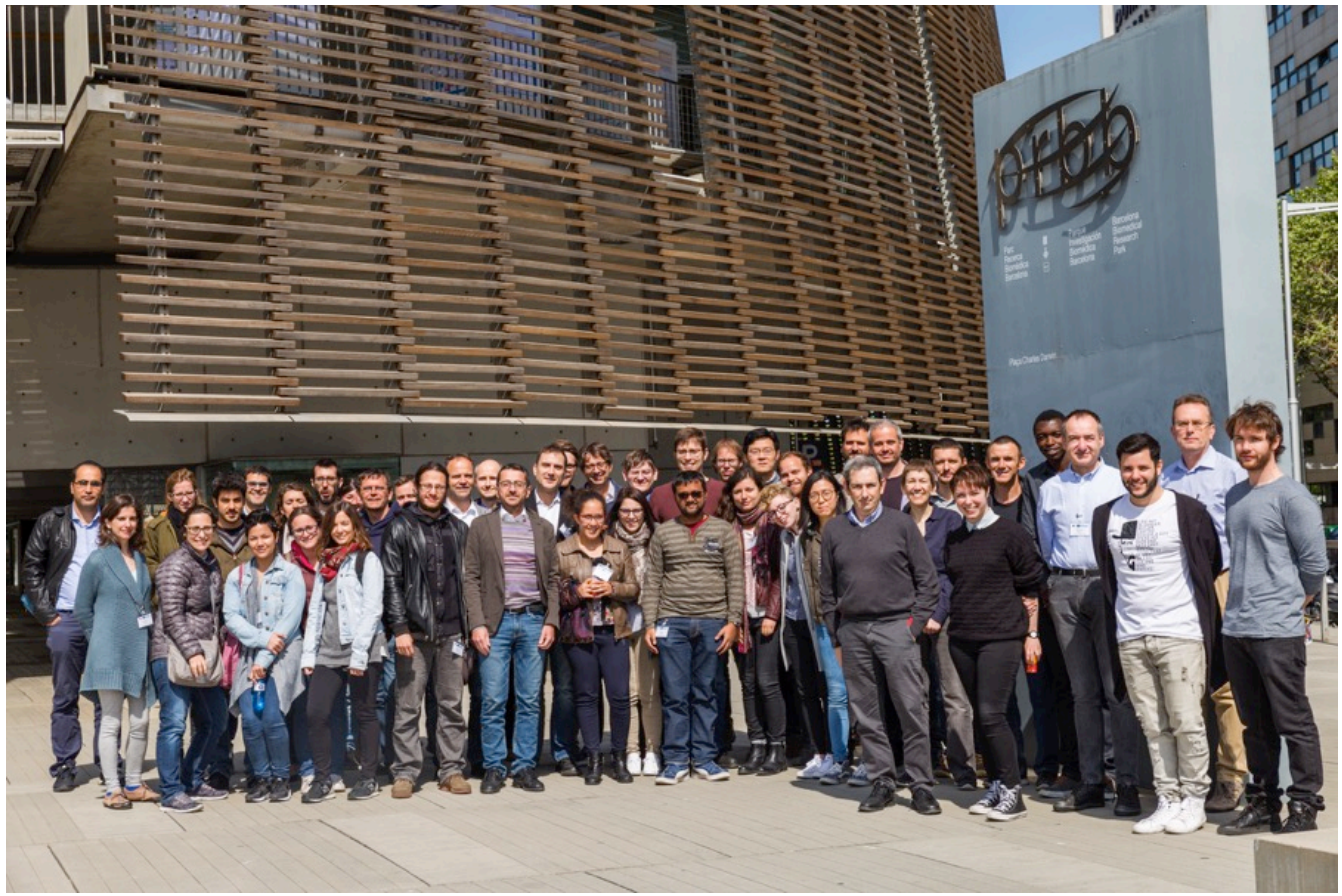
Faster training

Covers large applicability domain

Different problems may require different approaches

Multiples tools exist and continuously appear

Acknowledgements



Pavel Karpov
Dipan Ghosh
Michael Withnall
Maria Vashurina
Zhonghua Xia
Barbara Gasset

Sergey Sosnin (Skoltech)
Maxim Fedorov (Skoltech)

Yura Sushko
Sergey Novotarskyi
Robert Körner

M. Sattler (HMGU)
G. Wess (HMGU)

MARIE CURIE ACTIONS

BigChem

ERA CVD