We use molecular features and metric learning to predict gene expression

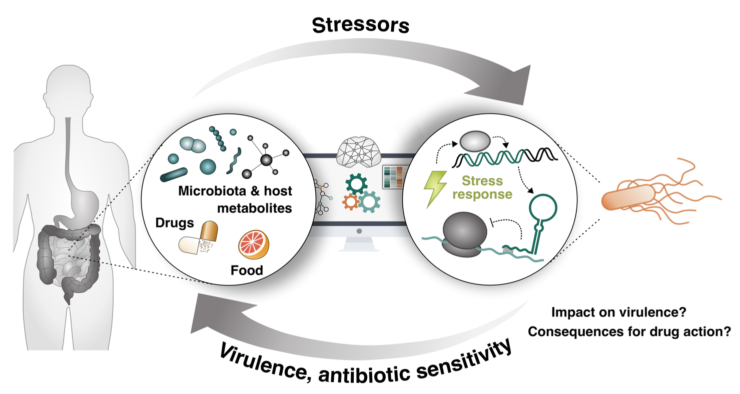
Cheminformatics deciphers stress response and virulence pathways in infection

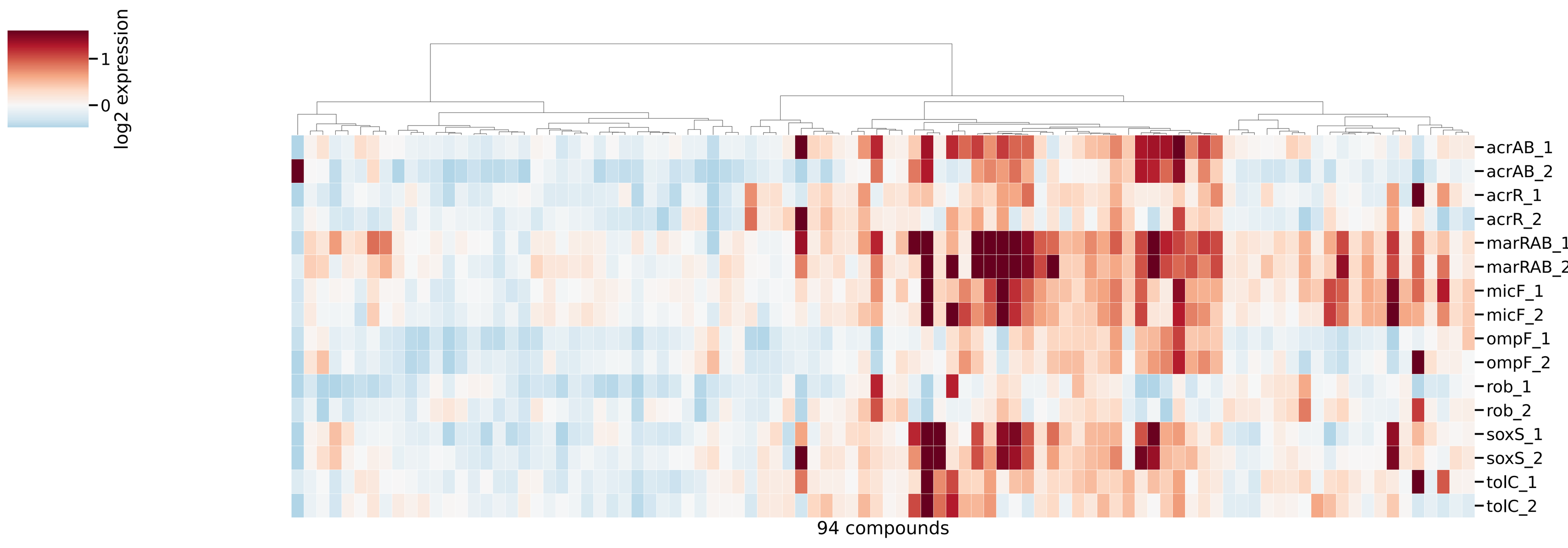
Roberto Olayo Alarcon[1], Christoph Binsfeld[2], Ana Rita Brochado[2], Christian L. Müller[1]

[1]Ludwig-Maximilians-Universität München, Germany; [2]Department of Microbiology, University of Würzburg, Biocenter, Germany

Small compounds and their effect on microbes

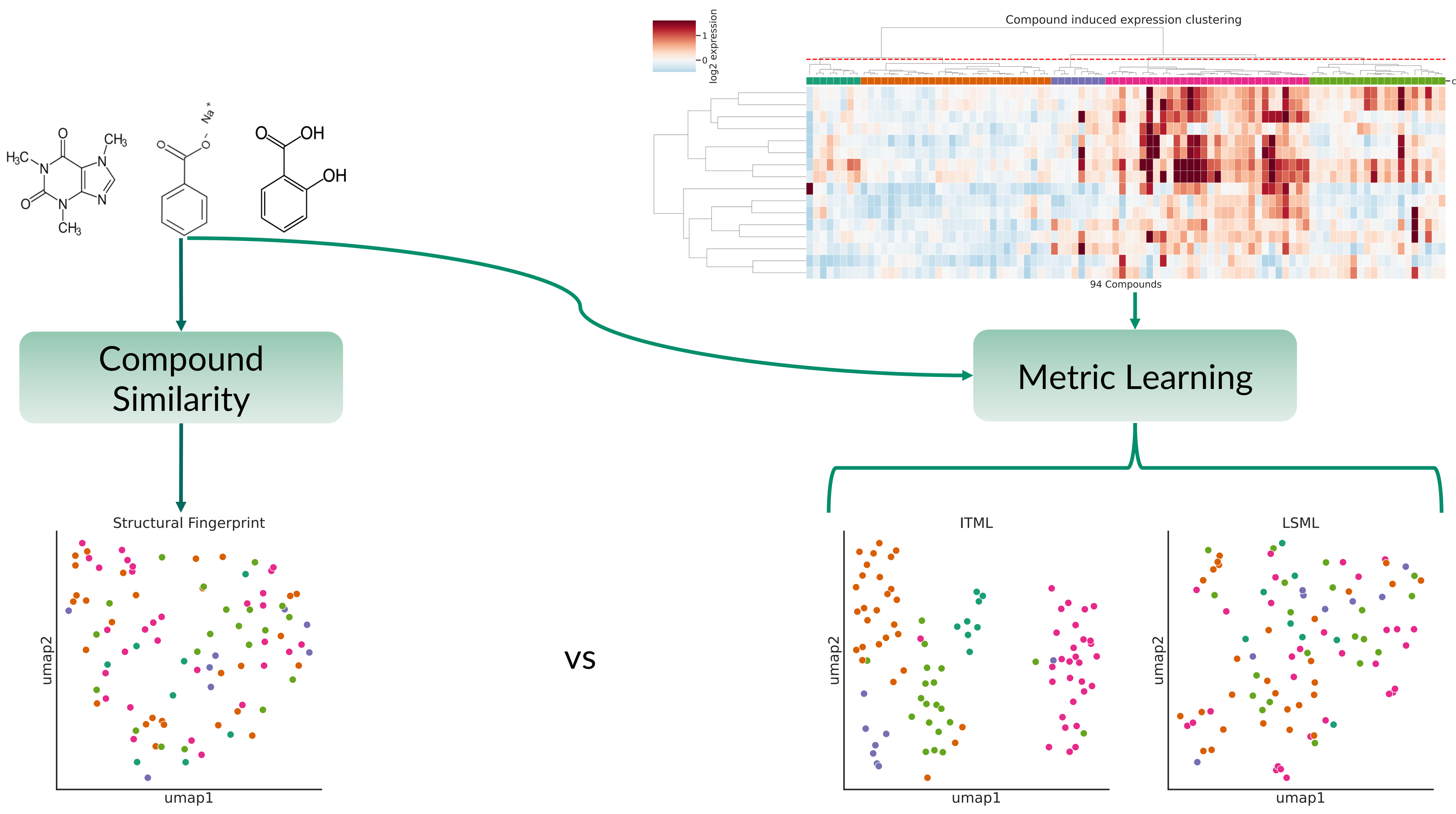
Much is unknown about which chemical signals trigger different stress responses in bacterial pathogens. Reporter assay experiments are time-consuming and expensive. **Compounds must be prioritized for further study.**





Structural fingerprints and metric learning

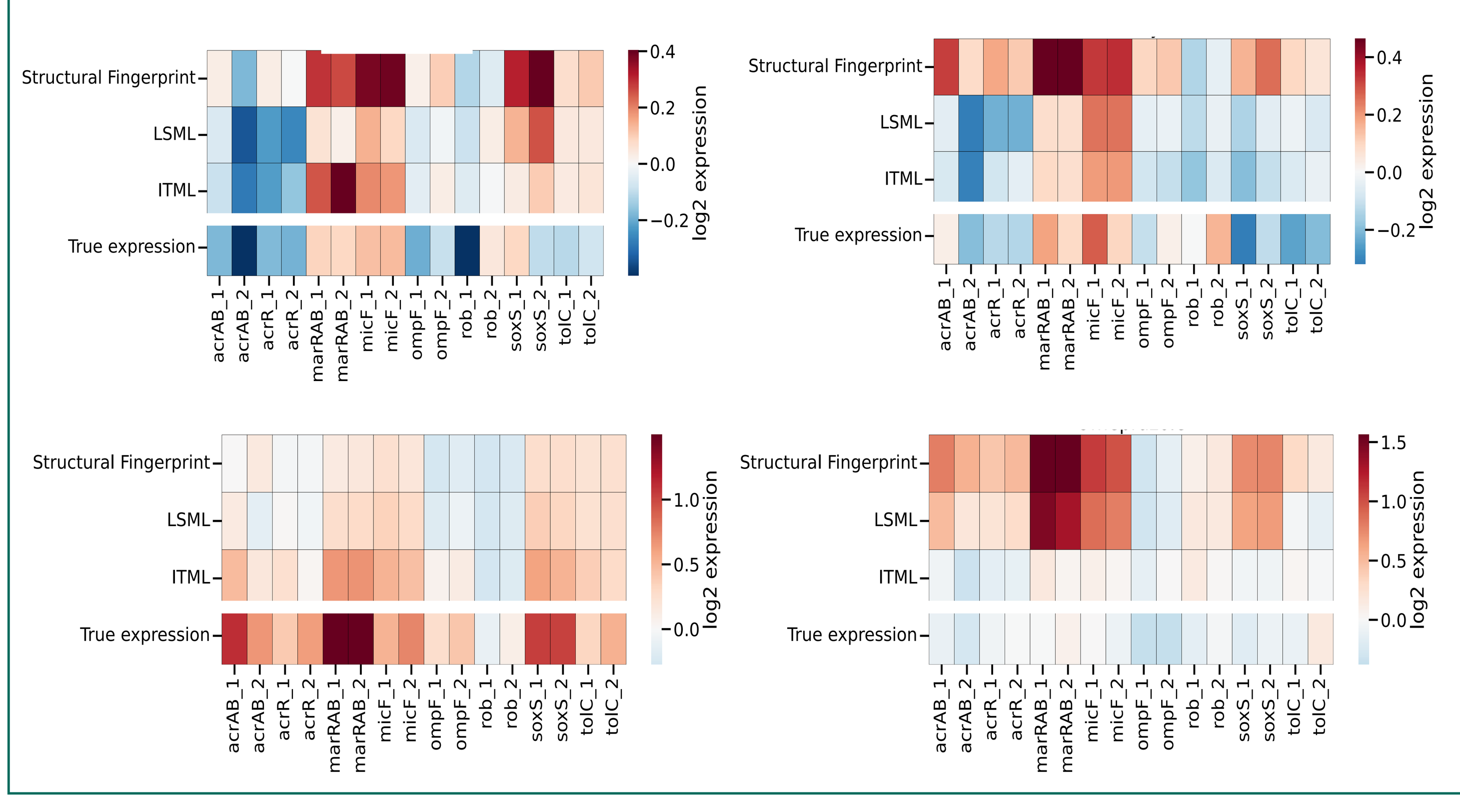
Structural fingerprints alone are partially predictive of the expression pattern they induce. We use metric learning to **learn a molecular representation that is more predictive of gene expression.**



Predicting the effect of new compounds

Using the learned molecular representation, we **predict the expression patterns induced by *unseen* compounds.**

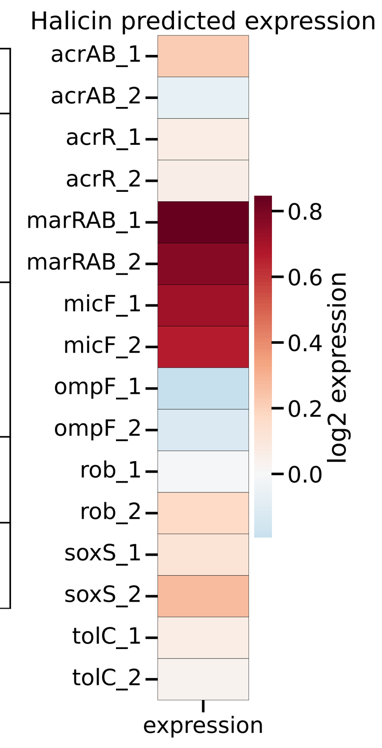
This can **enable the characterization of novel antibiotics and adjuvants**, such as Halicin.



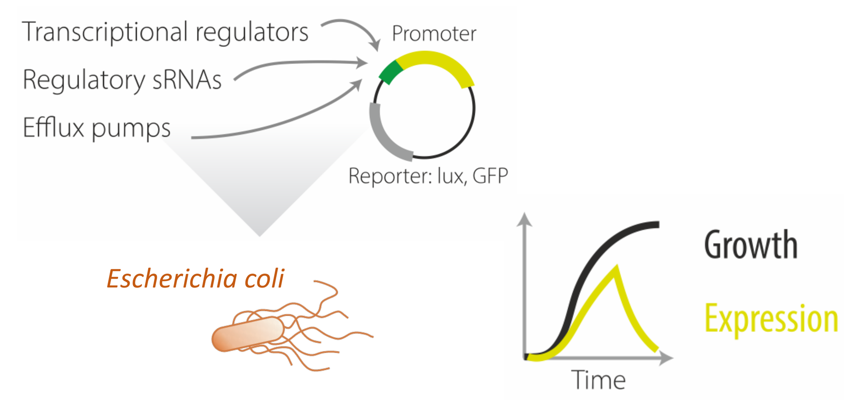
Halicin Nearest Neighbours

Name	Class
4-aminosalicylic acid	Antibiotic
Sodium benzoate	Antifungal
Metronidazole	Antibiotic
Trimethoprim	Antibiotic

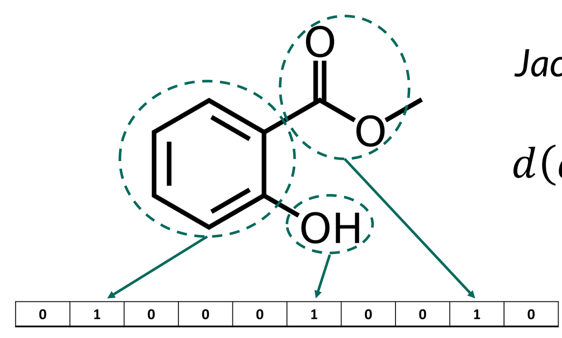
Halicin predicted expression



Reporter assay experiments

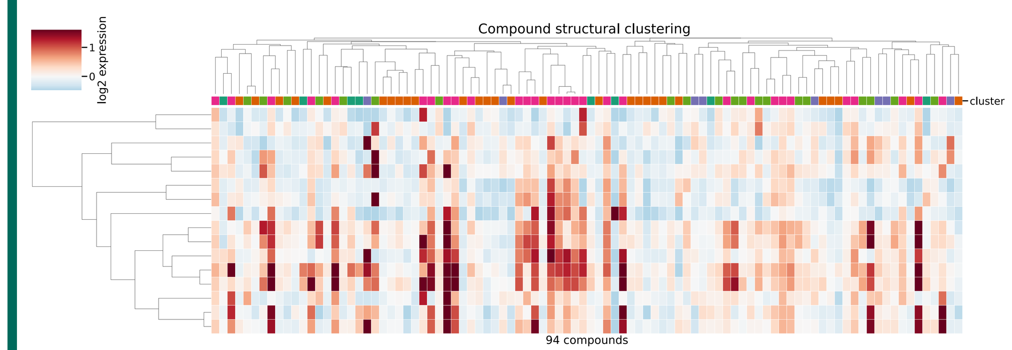


Structural fingerprints



Jaccard distance

$$d(a,b) = \frac{|a \cup b| - |a \cap b|}{|a \cup b|}$$



Metric Learning

Information Theoretic Metric Learning:

$$\min_{A} KL(p(\boldsymbol{x};A_0)||p(\boldsymbol{x};A))$$

$$s.t. \quad d_A(\boldsymbol{x}_i,\boldsymbol{x}_j) \le u \quad (i,j) \in S$$

$$d_A(\boldsymbol{x}_i,\boldsymbol{x}_j) \ge l \quad (i,j) \in D$$

Model selection

Makine a new prediction

$$u(\boldsymbol{x}) = \frac{\sum_{i=1}^{N} w_i(\boldsymbol{x}) u_i}{\sum_{i=1}^{N} w_i(\boldsymbol{x})}$$

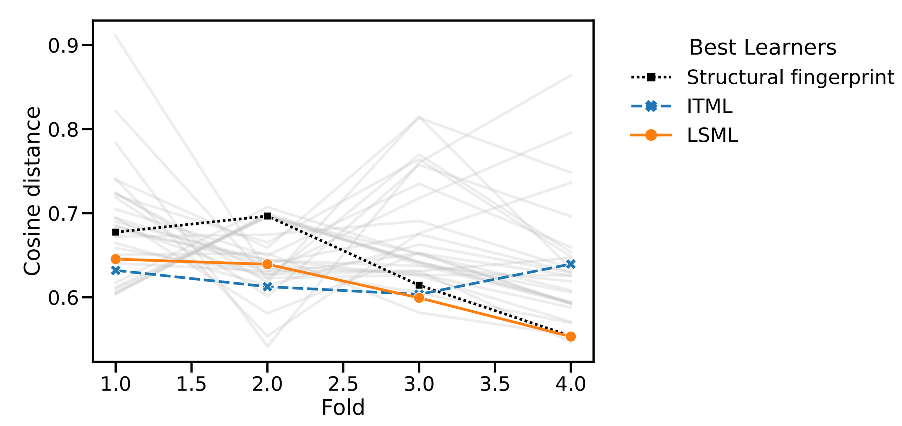
Inverse distance weighting

$$w_i(\boldsymbol{x}) = \frac{1}{d(\boldsymbol{x}, xi)}$$

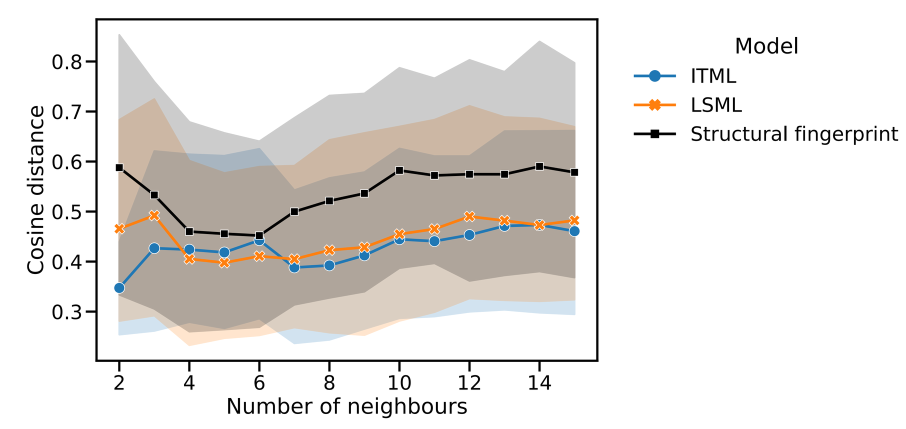
Evaluating a new prediction

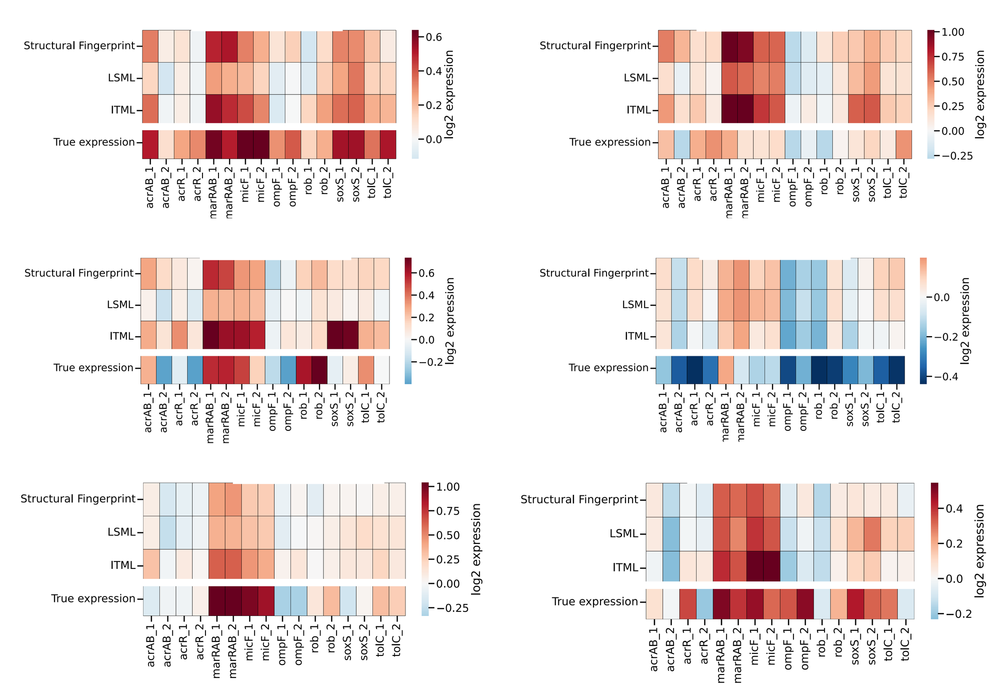
$$d(u,v) = 1 - \frac{u \cdot v}{||u||_2 ||v||_2}$$

Cosine distance



Out of bag predictions





Literature:

1. Davis, J., Kulis, B., Jain, P., Sra, S., & Dhillon, I. (2007). Information-theoretic metric learning. *Proceedings Of The 24Th International Conference On Machine Learning - ICML '07*. doi: 10.1145/1273496.1273523
2. Mediati, D., Wu, S., Wu, W., & Tree, J. (2021). Networks of Resistance: Small RNA Control of Antibiotic Resistance. *Trends In Genetics*, *37*(1), 35-45. doi: 10.1016/j.tig.2020.08.016
3. Stokes, J., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., & Donghia, N. et al. (2020). A Deep Learning Approach to Antibiotic Discovery. *Cell*, *180*(4), 688-702.e13. doi: 10.1016/j.cell.2020.01.021

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Julius-Maximilians-UNIVERSITÄT WÜRZBURG

bayresq.net Neue Strategien gegen multiresistente Krankheitserreger mittels digitaler Vernetzung