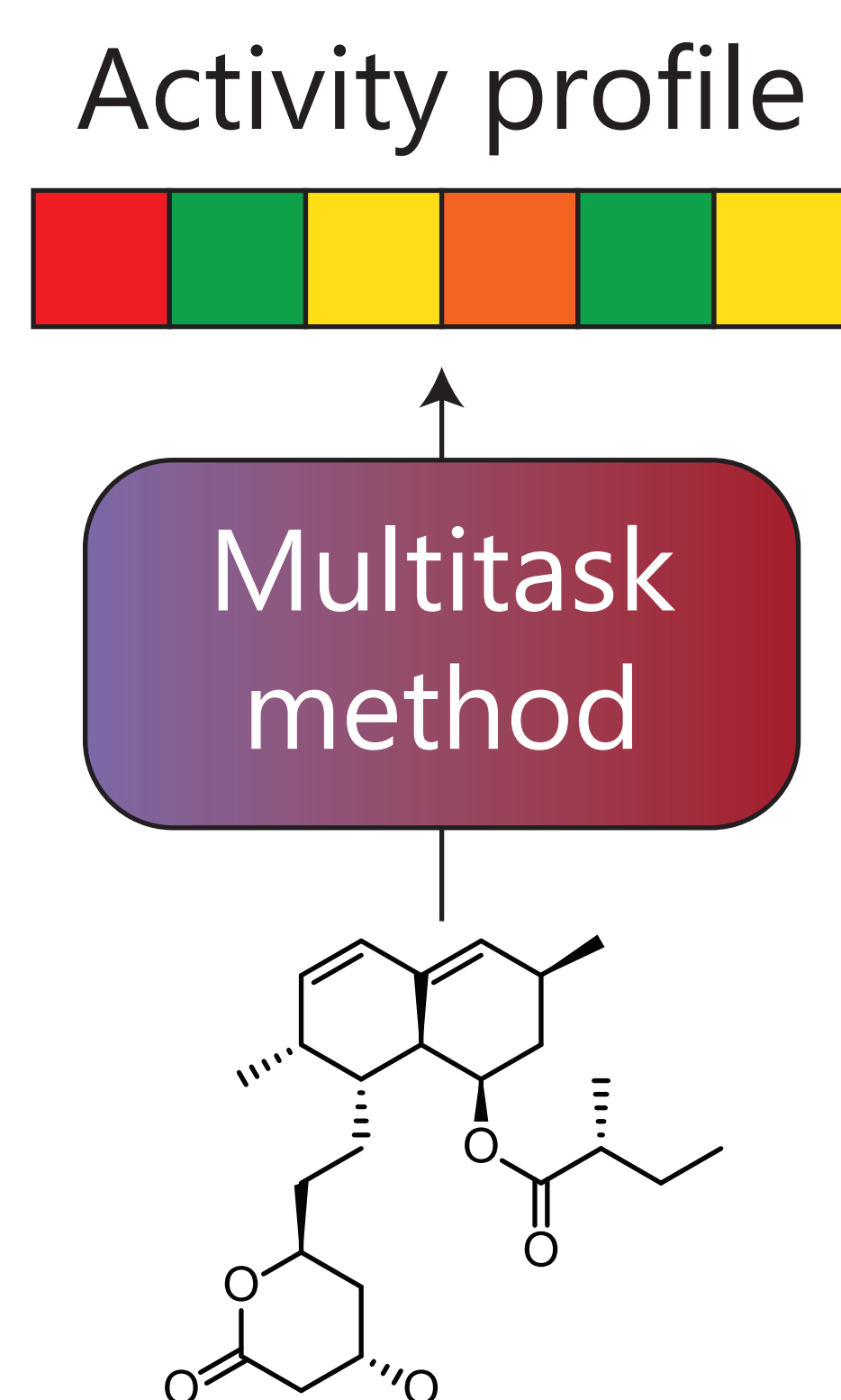


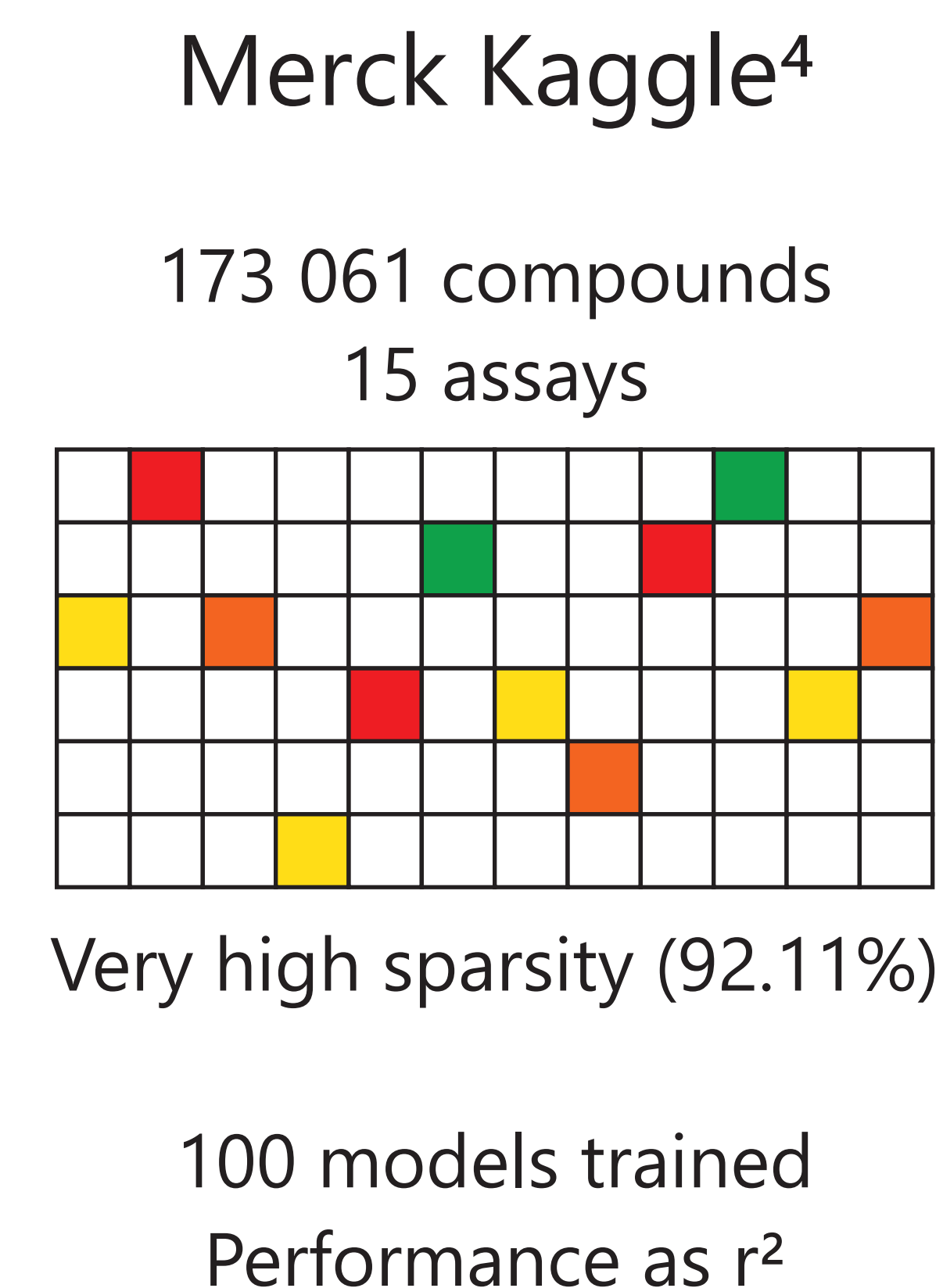
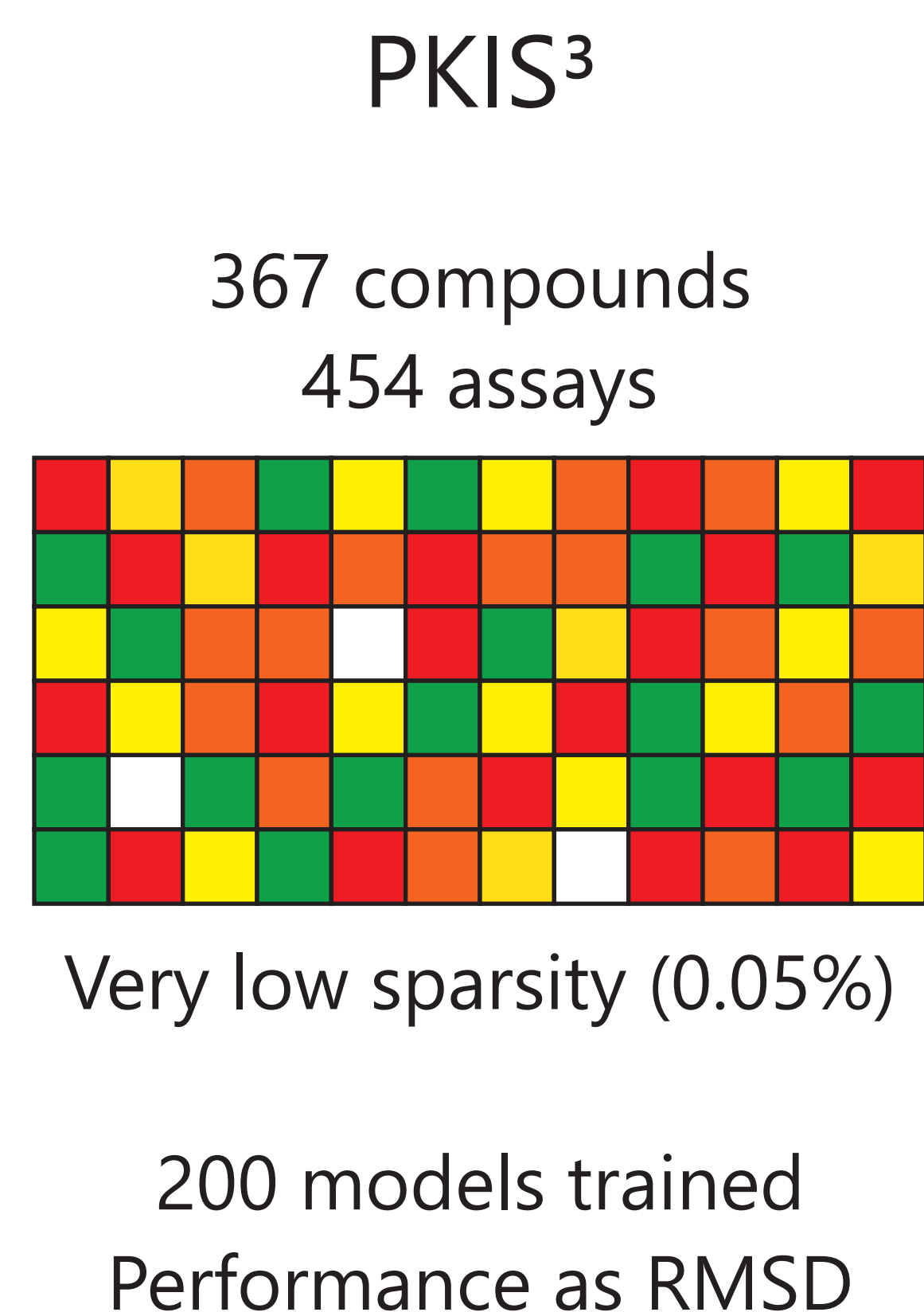


Introduction

Multitask machine learning techniques are able to predict several outputs with a single model. These techniques have become popular in the cheminformatics field, thanks to deep neural networks (DNNs).¹ However, DNNs are not the only multitask method available. A novel method based on matrix factorization called Macau has also been applied to cheminformatics problems.² In this work, we test these two methods on different multitarget datasets.

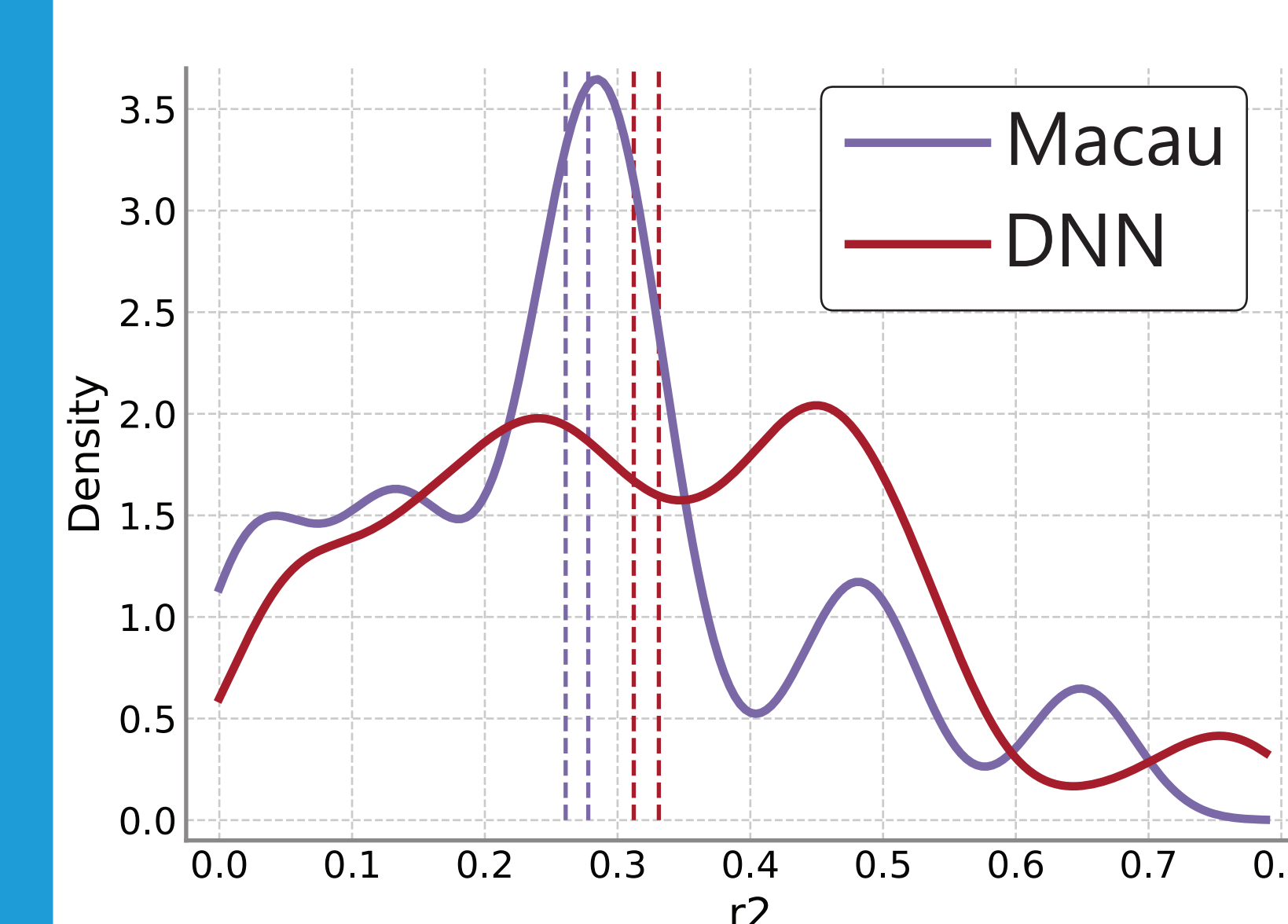


Datasets



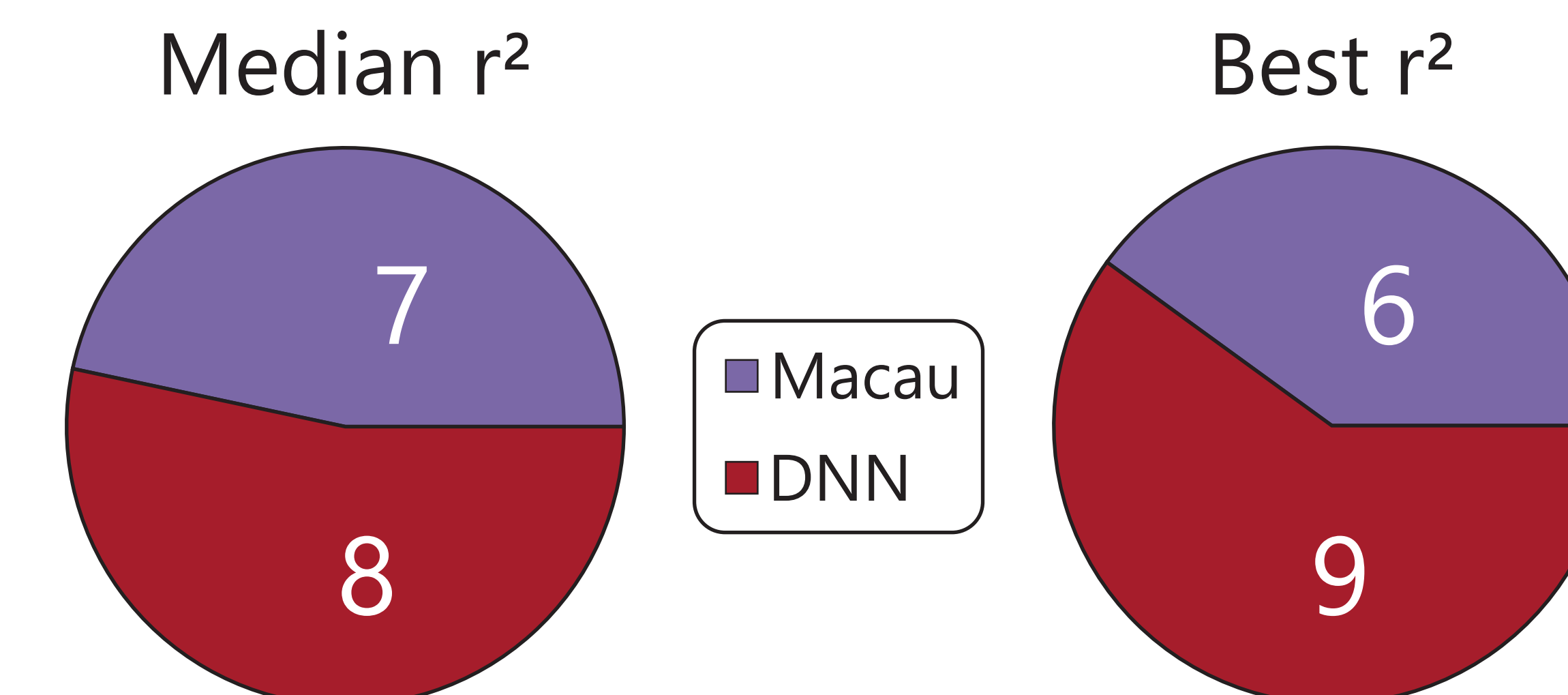
Merck Kaggle Results

Overall results



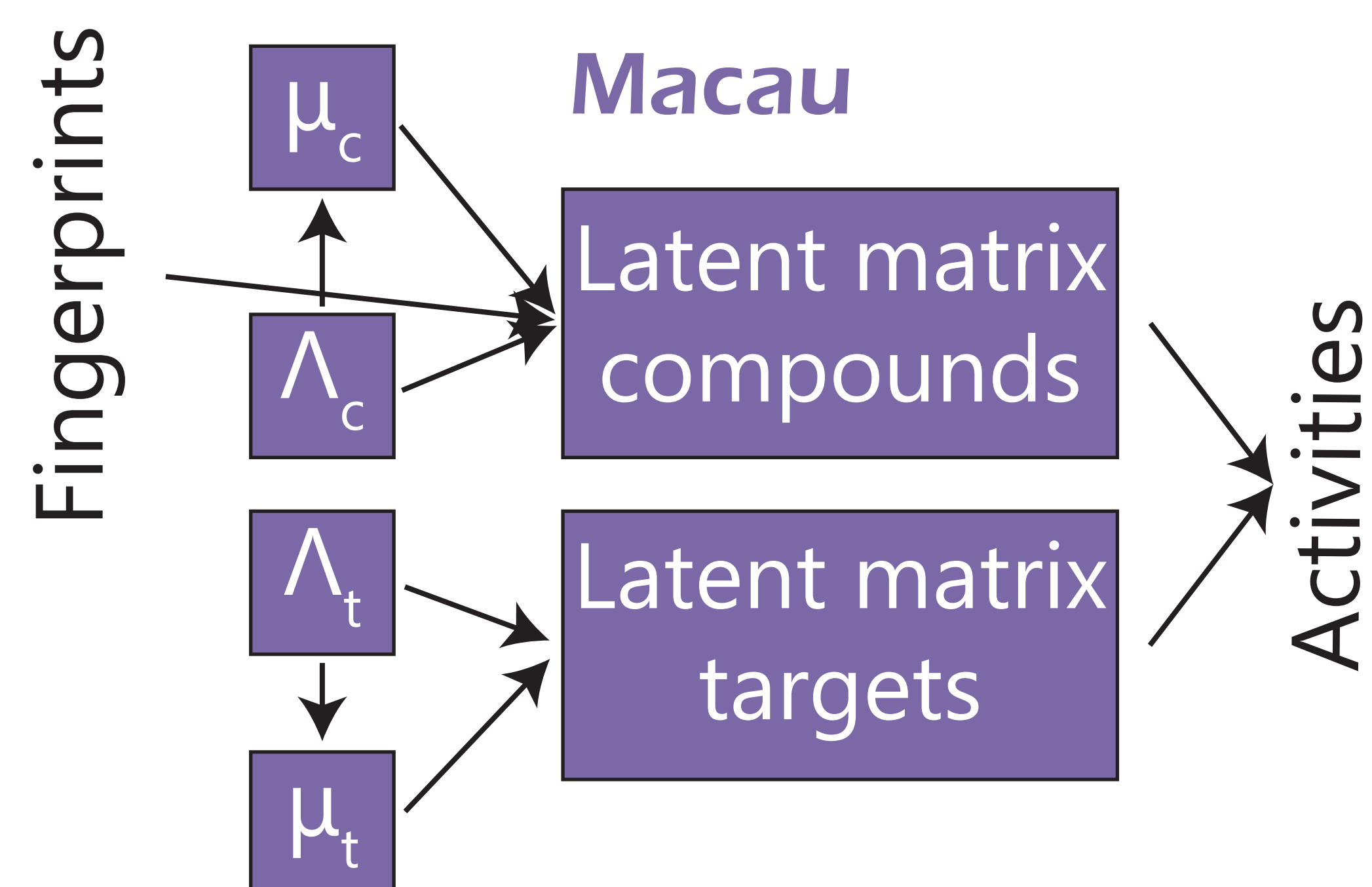
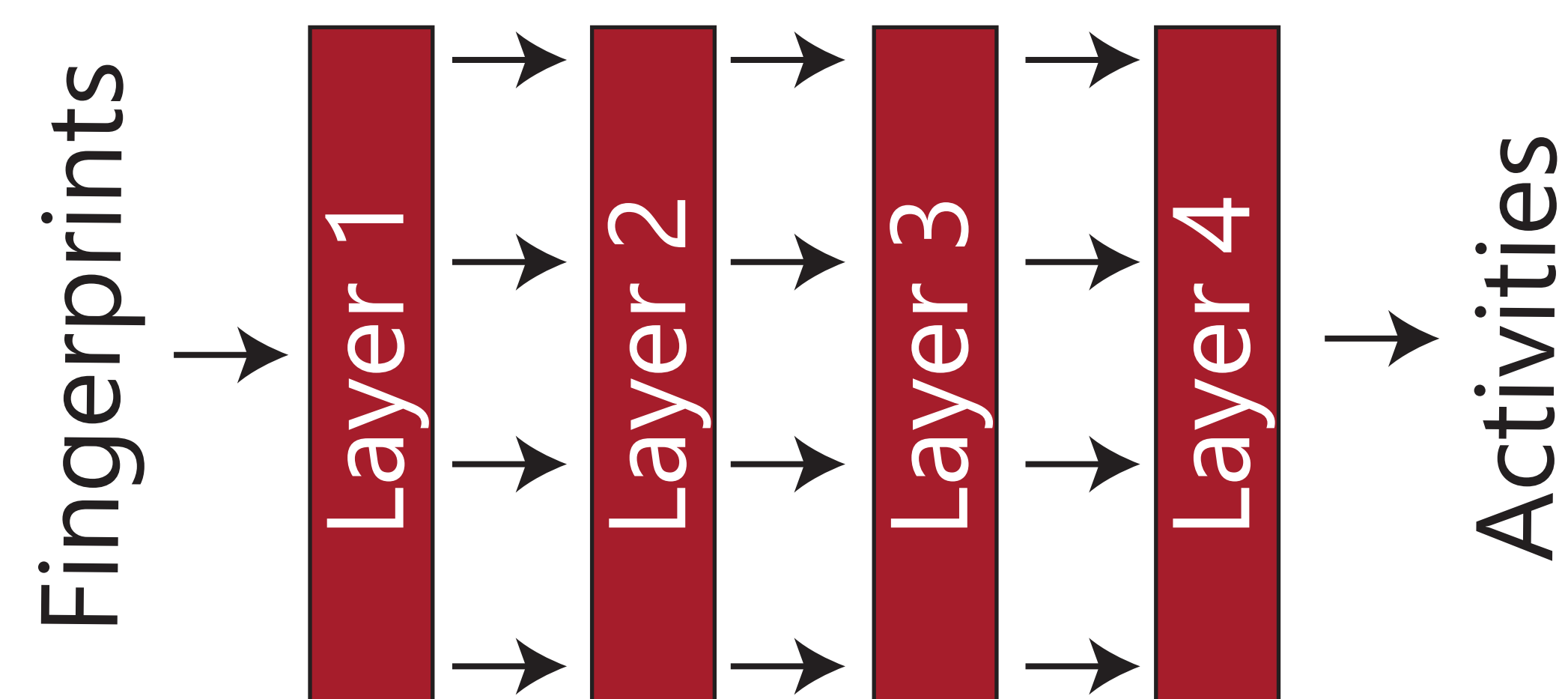
Left, 15x100 r² values for both techniques. Dotted lines represent bootstrapped mean confidence intervals. Below, we compare which method has higher median and best (maximum) r² for each target.

Target based results



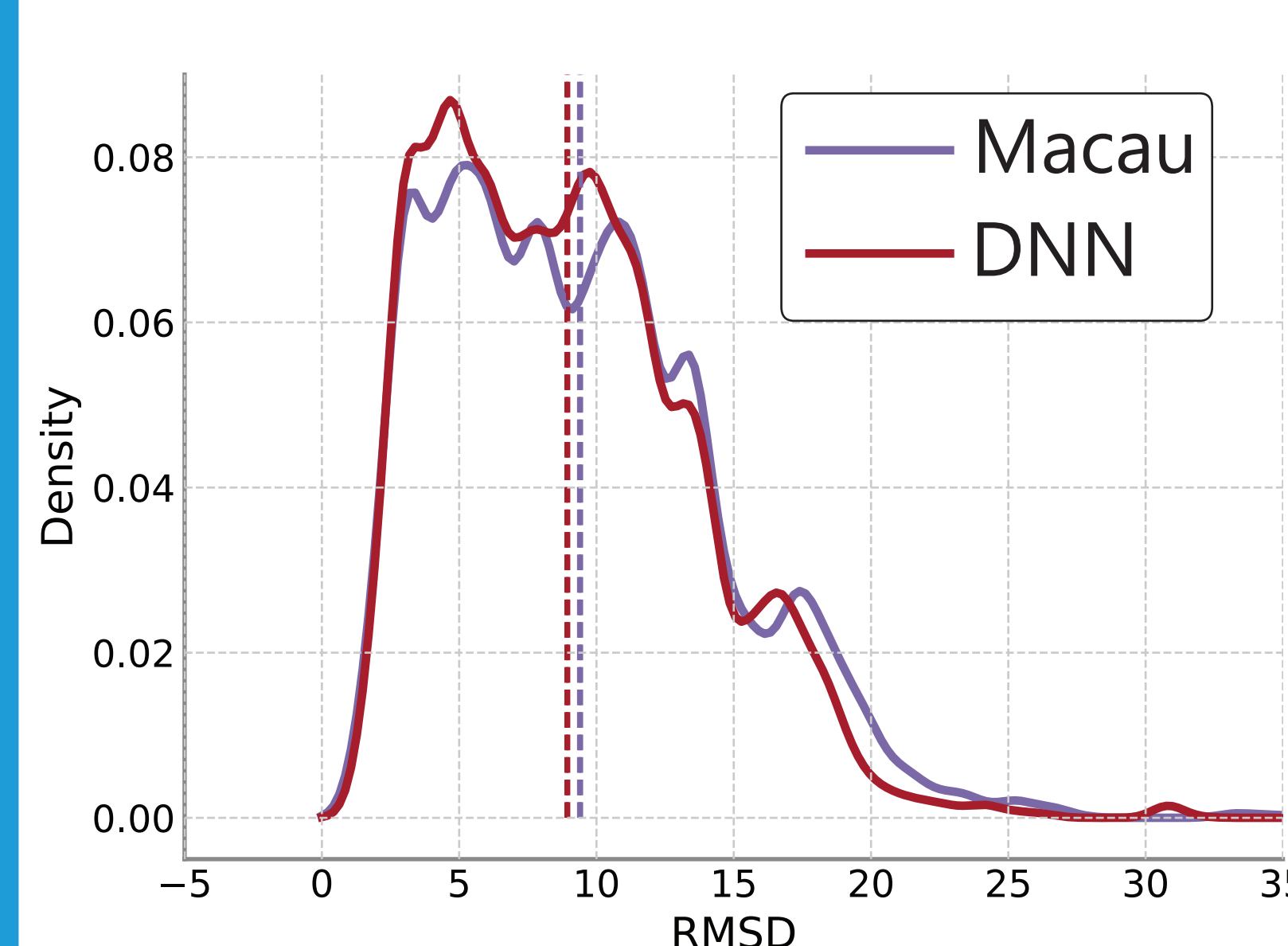
Methods

Deep neural networks



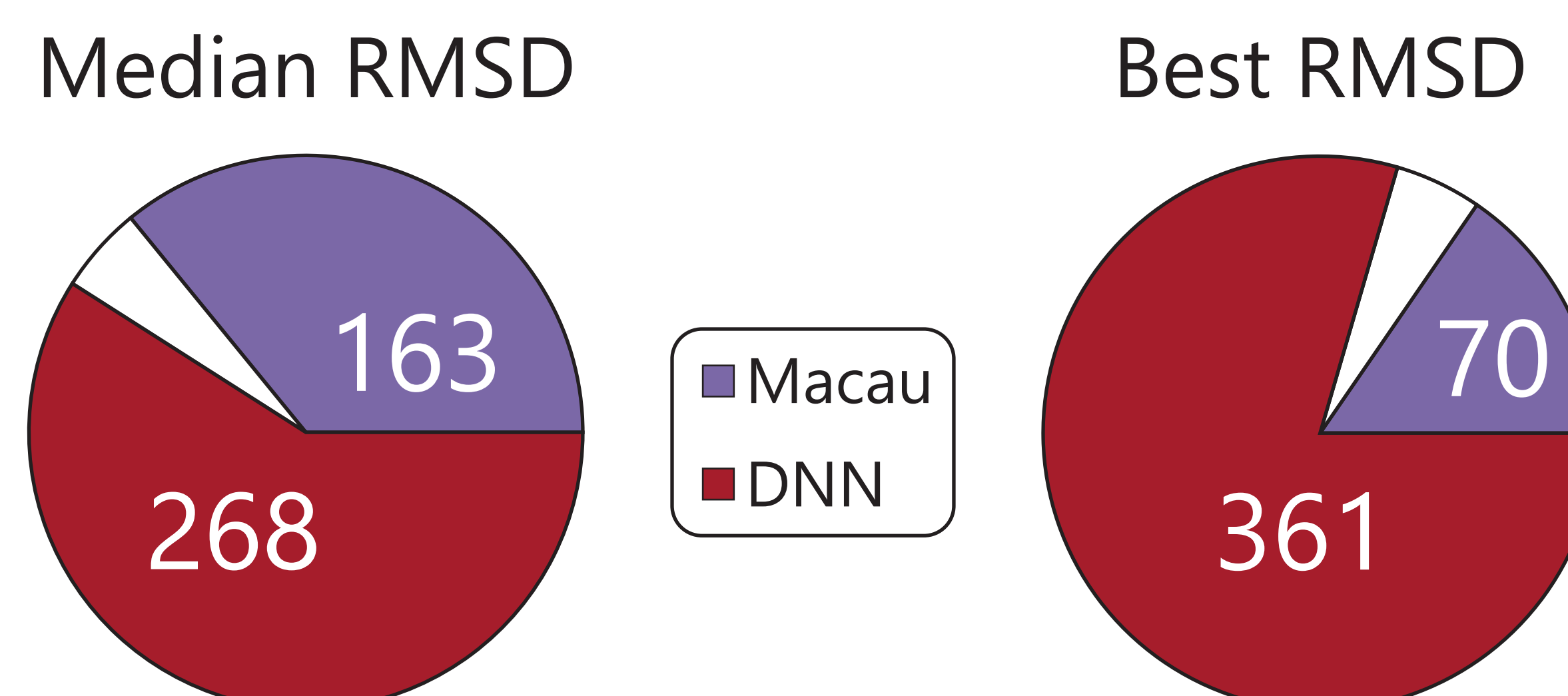
PKIS Results

Overall results



Left, 454x200 RMSD values for both techniques. Dotted lines represent bootstrapped mean confidence intervals. Below, we compare which method has lower median and best (minimum) RMSD for each target. White represents targets where there was no statistical difference.

Target based results



Conclusions

Multitask prediction methods are very promising in cheminformatics. Although the spotlight has been focused on deep neural networks, other methods such as Macau provide comparable performance. In addition, Macau does not require GPU resources to train models in adequate time.

References

- 1) LeCun et al. Nature 521, 436-444, 2015
- 2) Simm et al. arXiv 1509.04610, 2015
- 3) <https://www.ebi.ac.uk/chembl/db/extra/PKIS>
- 4) Ma et al. JCI 55, 263-274, 2015

Funding

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n°612347.

