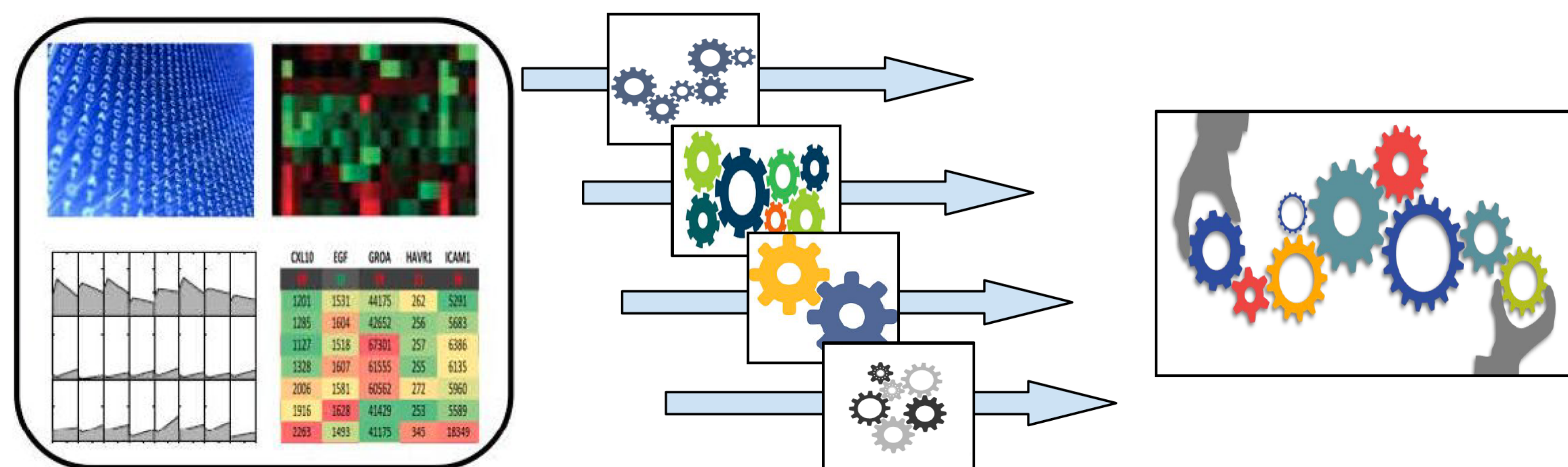


# Crowdsourced ensembles for the DREAM Respiratory Viral Challenge

Ana Stanescu<sup>1,2</sup>, Mehmet Eren Ahsen<sup>2</sup>, Robert Vogel<sup>3</sup>, Gustavo Stolovitzky<sup>2,3</sup>, Solveig K. Sieberts<sup>4</sup>, Gaurav Pandey<sup>2</sup>.  
<sup>1</sup>University of West Georgia, GA, <sup>2</sup>Department of Genetics and Genomic Sciences and Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, NY, <sup>3</sup>IBM Research, NY, <sup>4</sup>Sage Bionetworks, WA

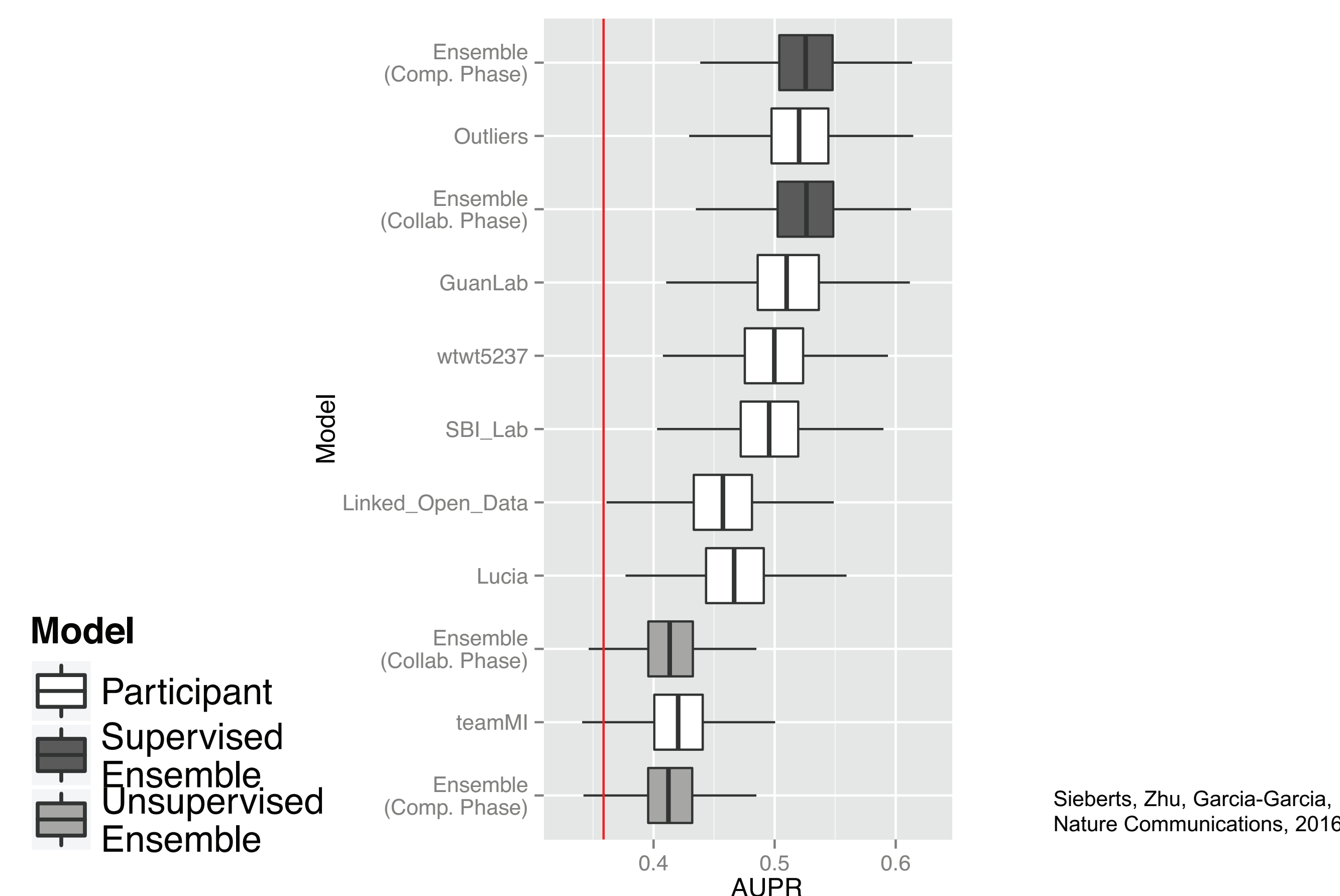
## Heterogeneous ensembles for DREAM Challenges



DREAM Challenges are very good at identifying the best individual (base) model(s).

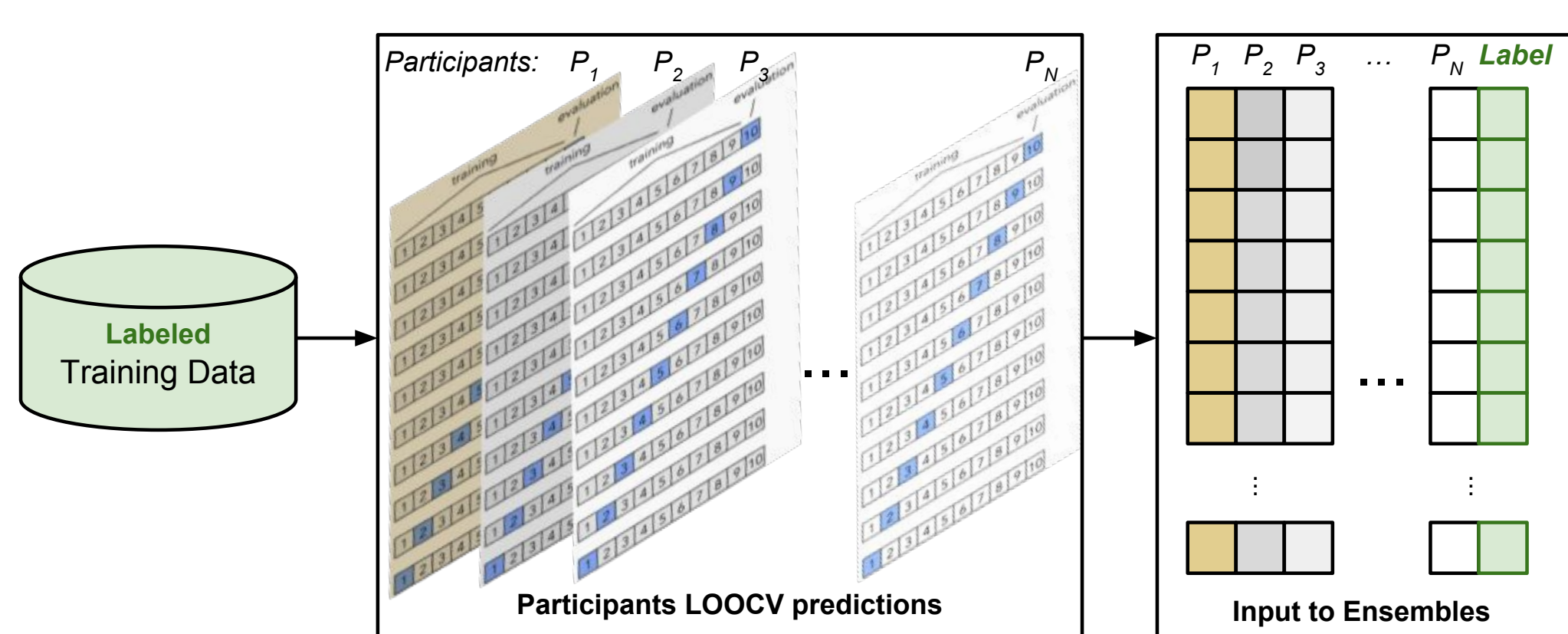
- By "ensembling" the individual models, we can synthesize (accumulate) their inherent knowledge
- Performance boost, reduction in variance
- Participants are free to train their own models and generate predictions
- Different setting from traditional methods, such as Boosting
- Heterogeneous ensembles

## Prior results from the DREAM Rheumatoid Arthritis Challenge



## Heterogeneous Ensemble Methods for the Respiratory Viral DREAM Challenge

### RV DREAM Predictions

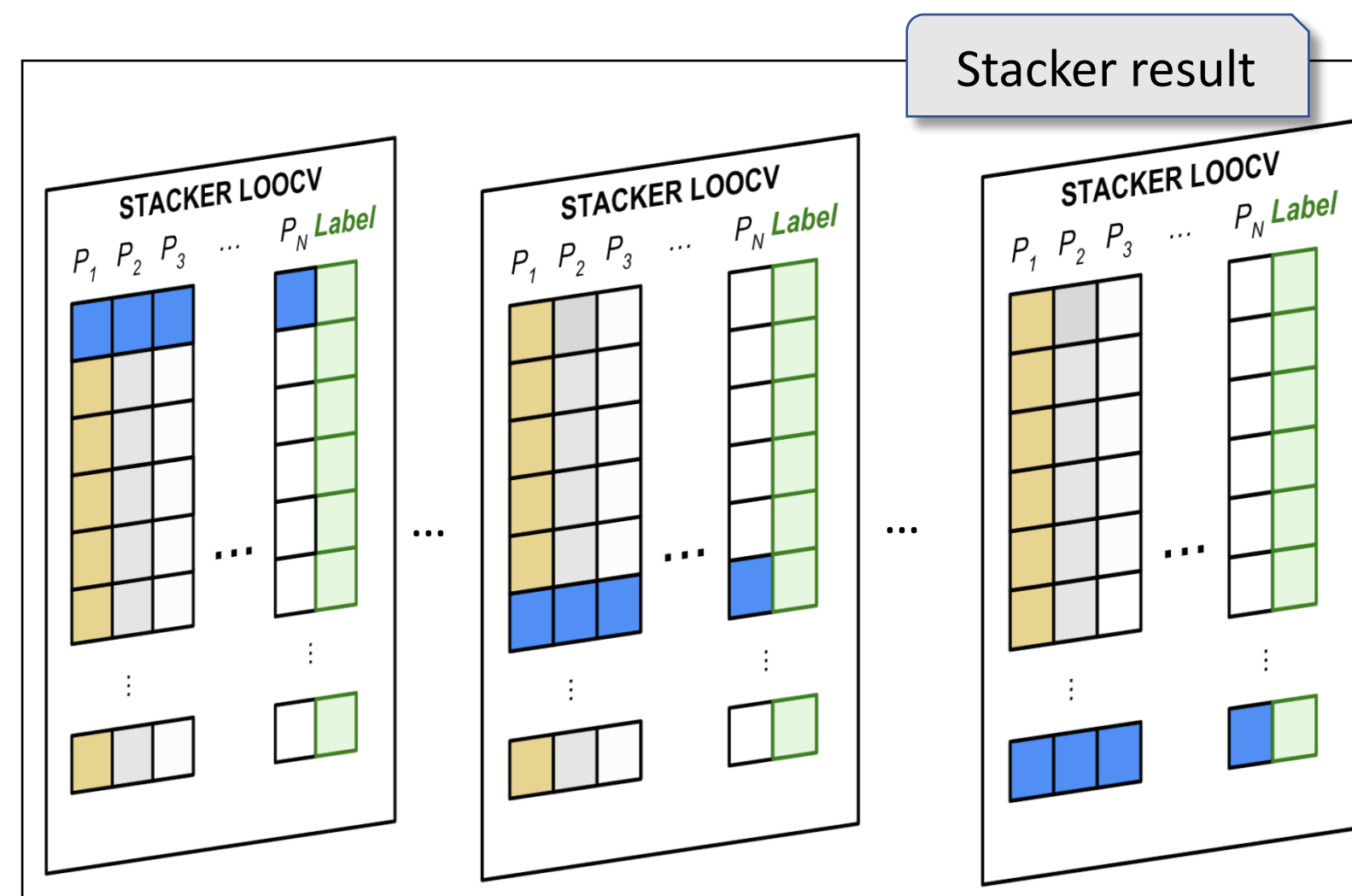


We cannot use the training set for learning base predictors as well as ensembles: **OVERFITTING**. Hence, to build heterogeneous ensembles, use LOOCV (deterministic)

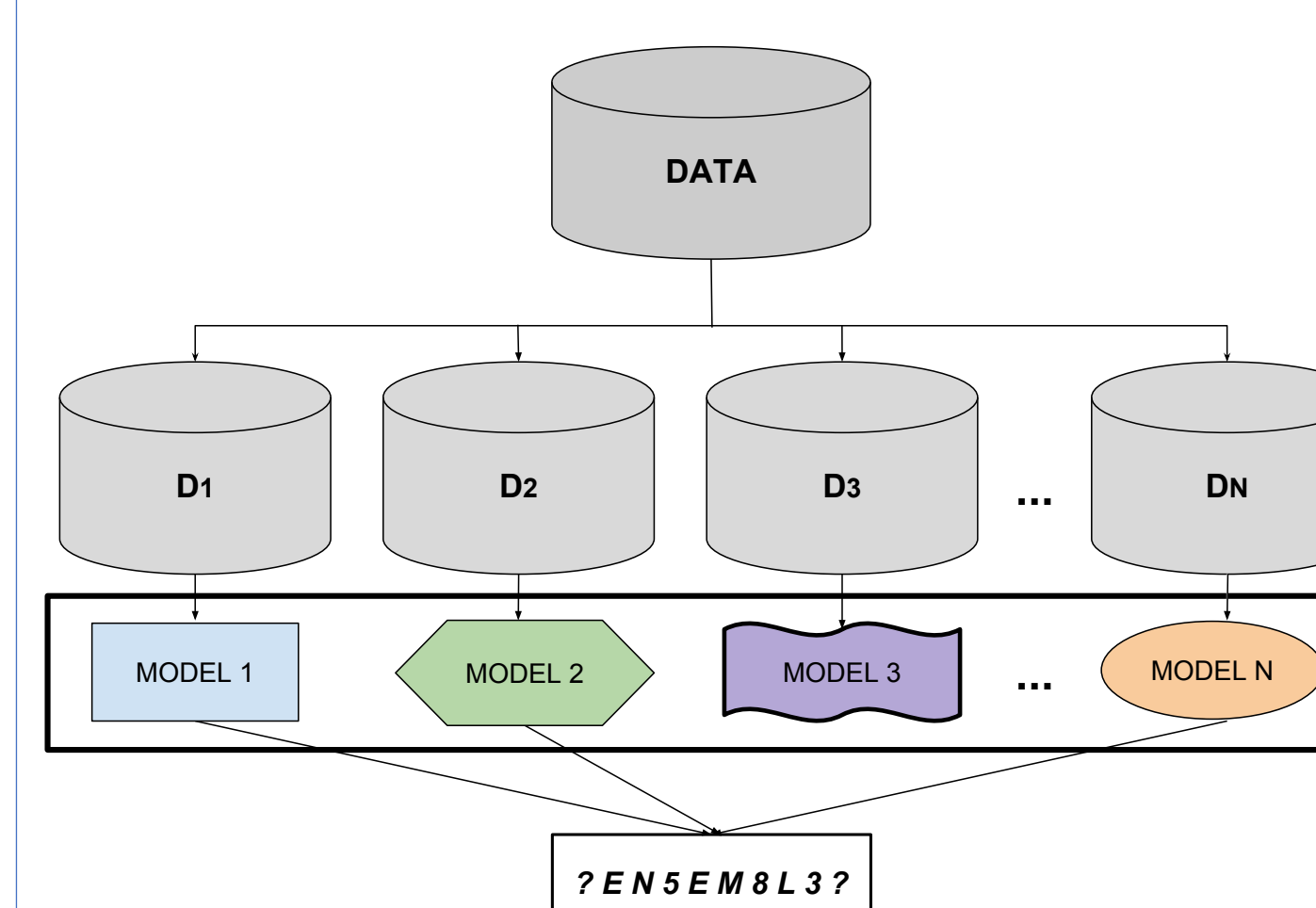
- Use N-1 LOOCV predictions from the participants for training the ensemble
- Use the learned ensemble to classify the N<sup>th</sup> instance

Supervised Ensembles  
Stacking: plain, intra-/inter-cluster

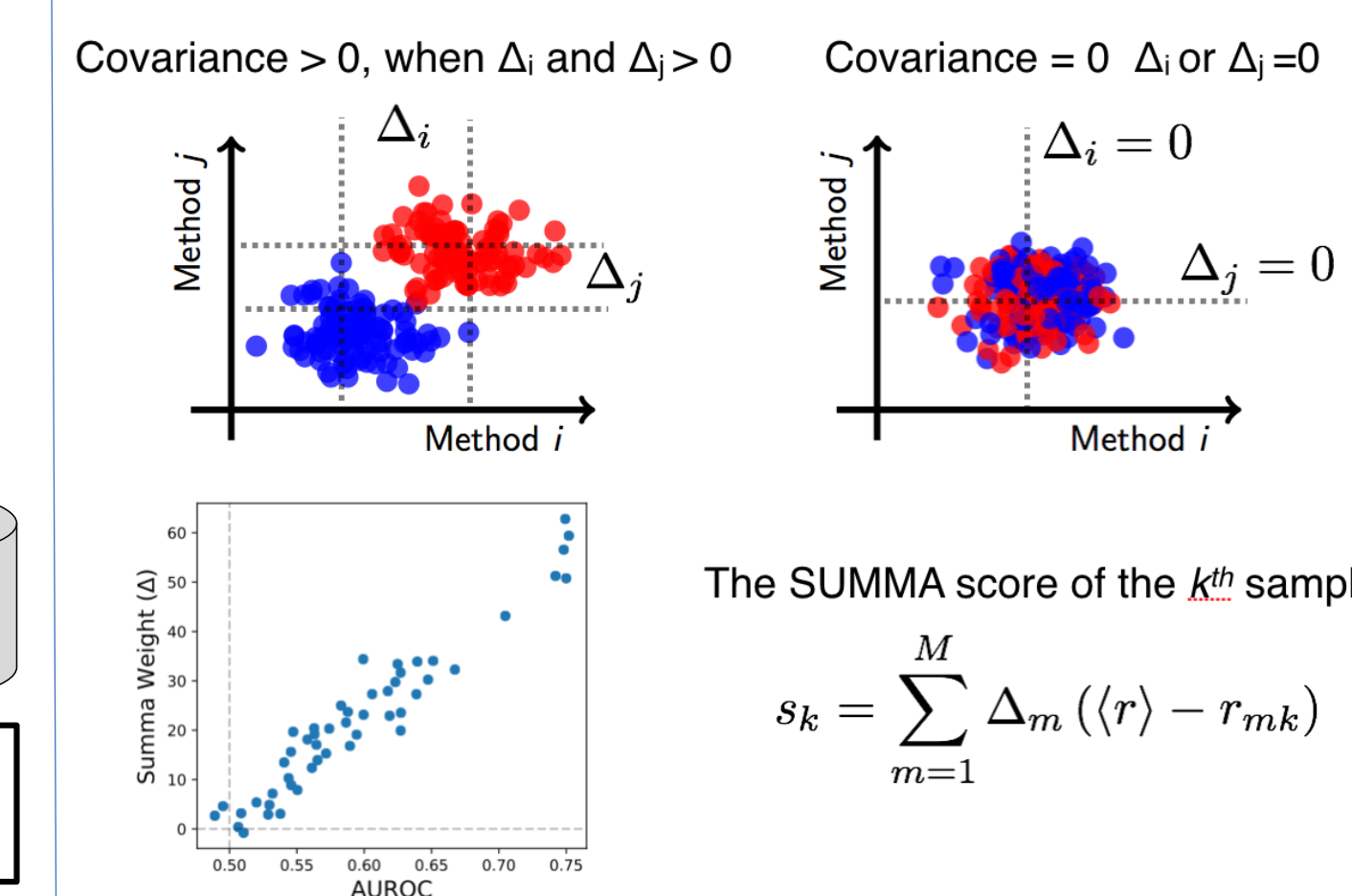
The stackers were trained in a LOOCV procedure



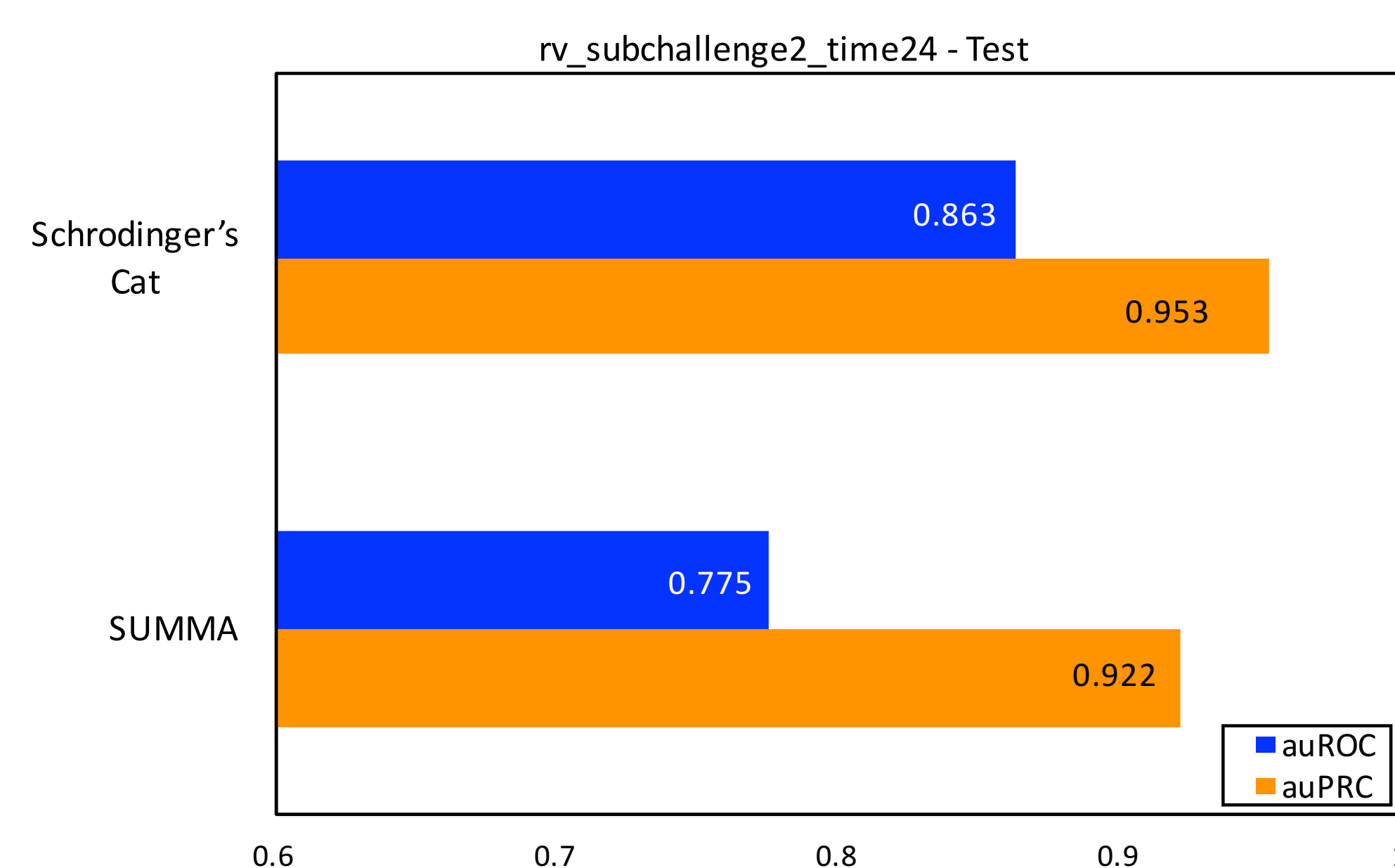
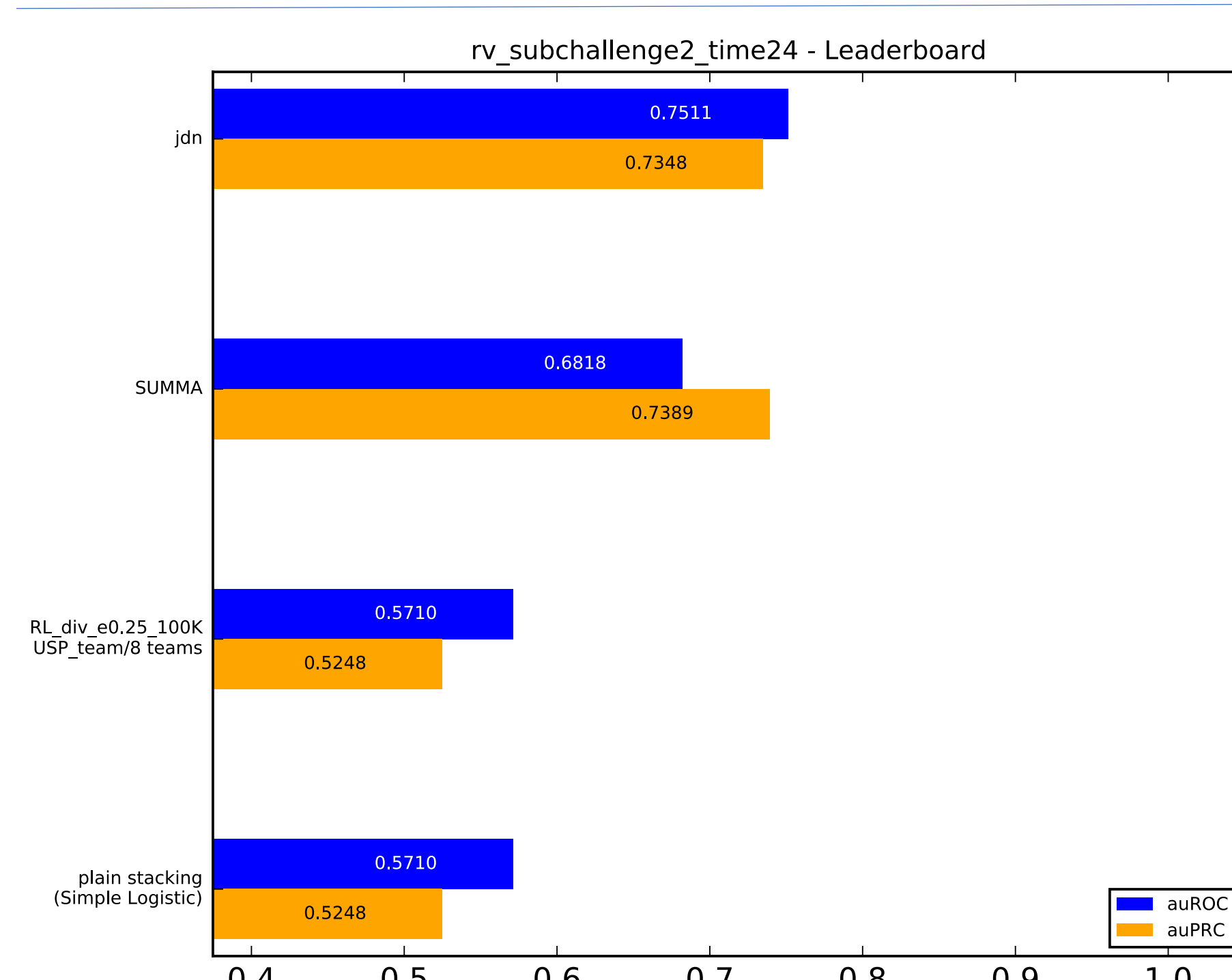
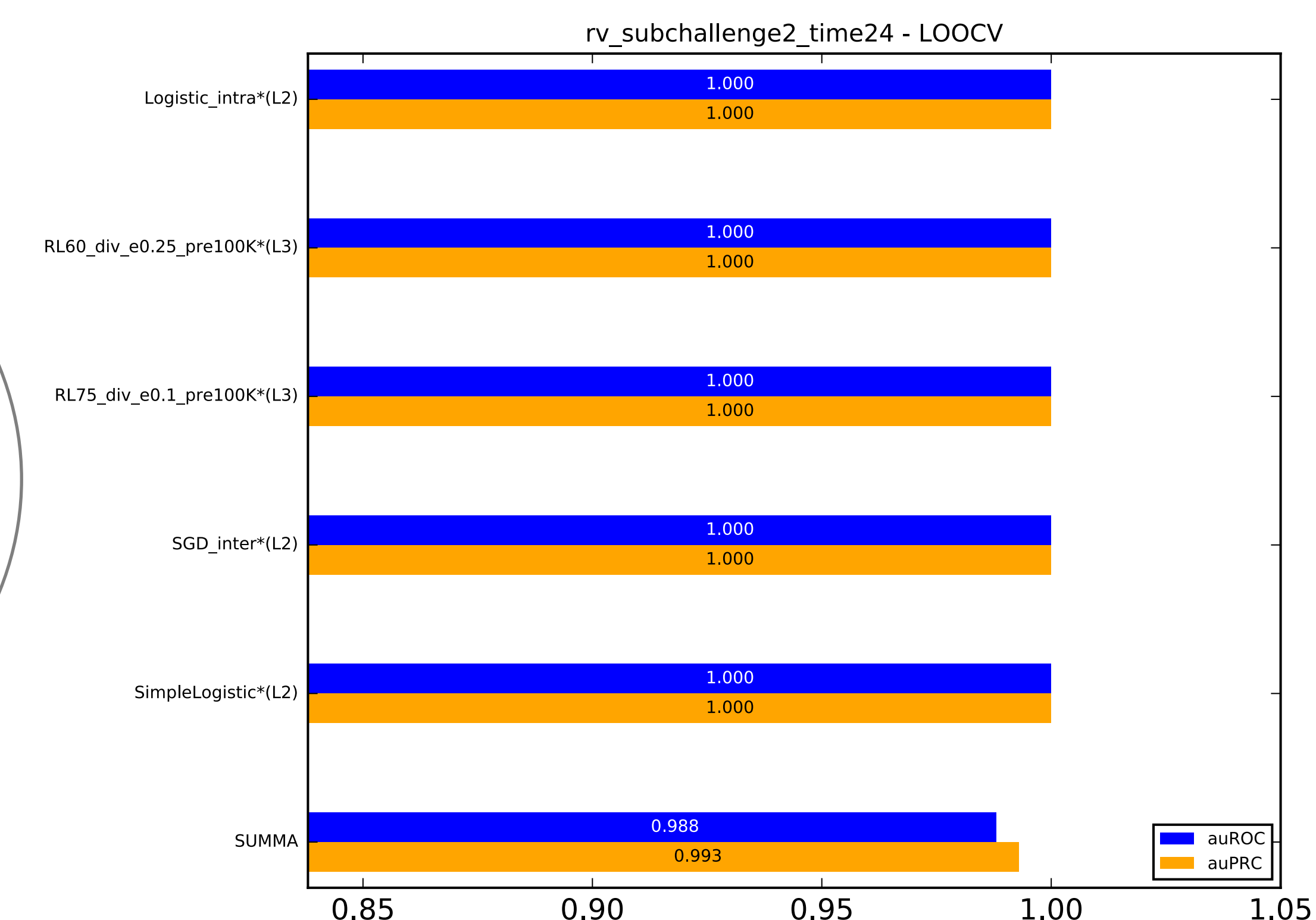
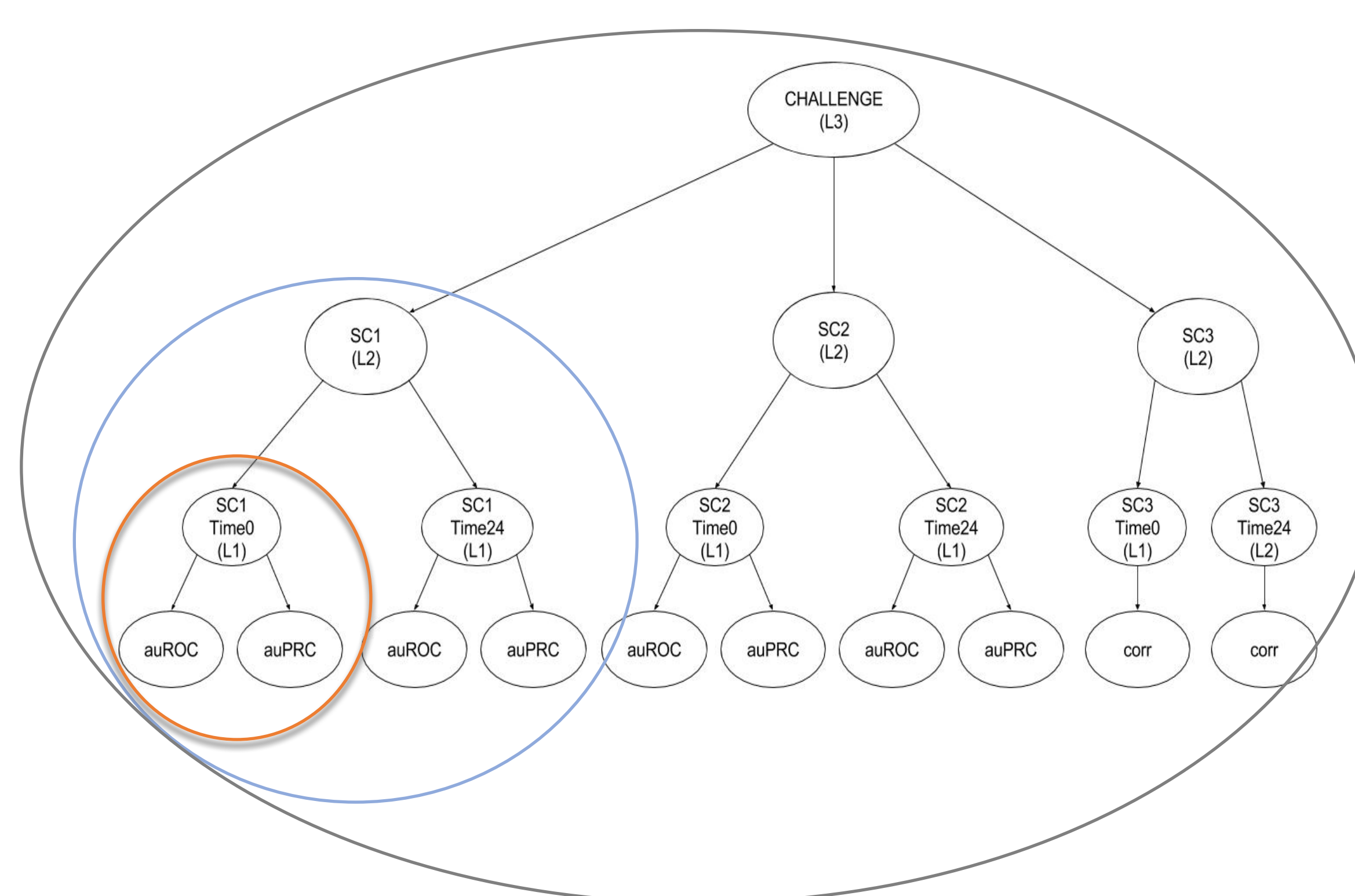
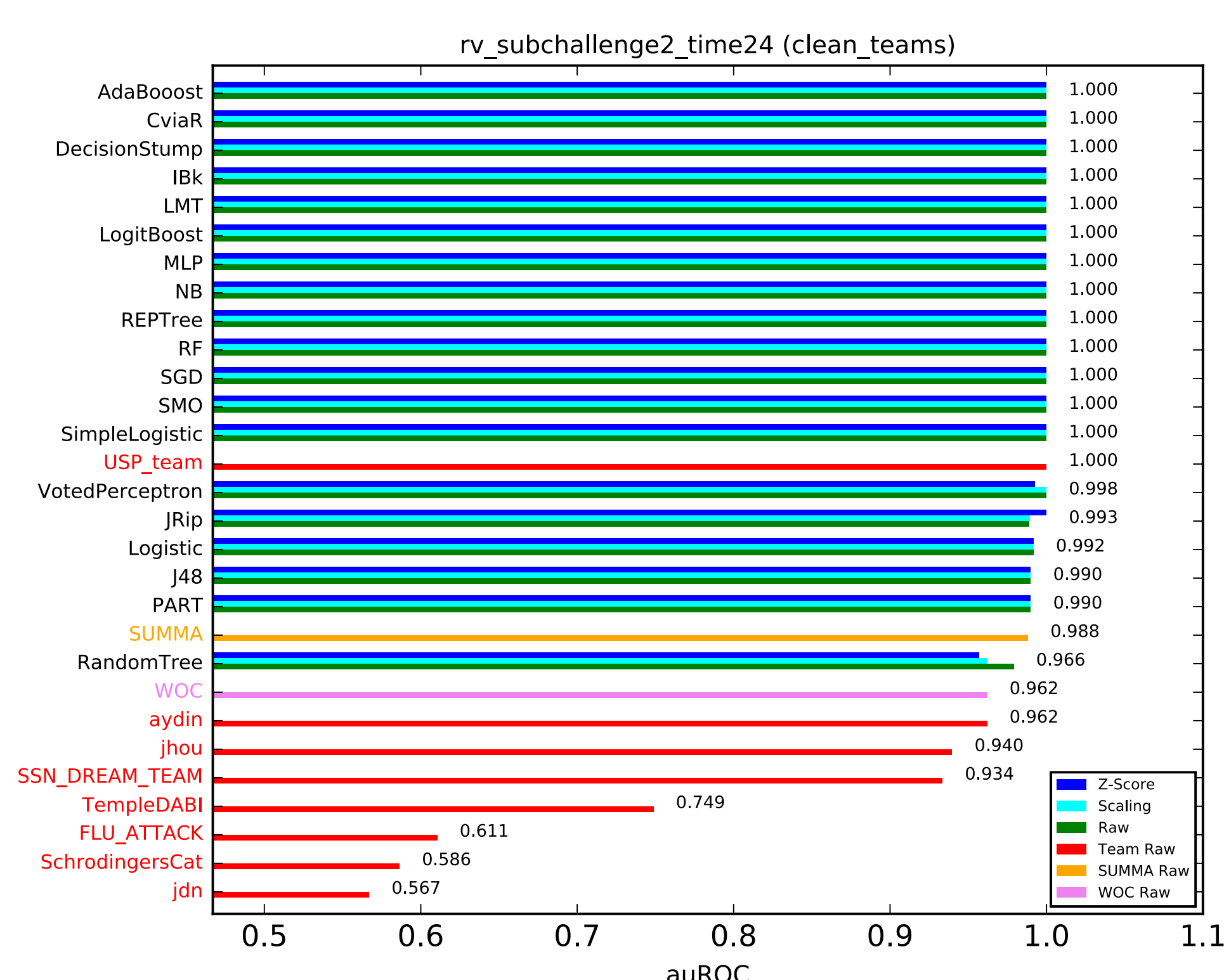
Supervised Ensembles  
Ensemble selection:  
Reinforcement Learning (two variants)



Unsupervised Ensembles  
SUMMA



## Results



## References

- Stanescu A. and Pandey G. *Learning parsimonious ensembles for unbalanced computational genomics problems*. Proceedings of Pacific Symposium on Biocomputing, 22: 288-299, 2017.
- Whalen S., Pandey O.P., and Pandey G., *Predicting protein function and other biomedical characteristics*. Methods 93(15): 92-102, 2016.
- Sieberts S.K., Zhu F., Garcia-Garcia J., Stahl E., Pratap A., Pandey G., Pappas D., Aguilar D., Anton Bernat, et al. *Crowdsourced assessment of common genetic contribution to predicting anti-TNF treatment response in rheumatoid arthritis*. Nature Communications 7:12460, 2016.