A COMPREHENSIVE COMPARISON OF METHODS FOR KINASE ACTIVITY INFERENCE

Sophia Müller-Dott[1], Aurelien Dugourd[1], Martin Garrido-Rodriguez[1,2], Julio Saez-Rodriguez[1]

[1]Heidelberg University, Faculty of Medicine, and Heidelberg University Hospital, Institute for Computational Biomedicine, Bioquant, Heidelberg, Germany
[2]Genome Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany
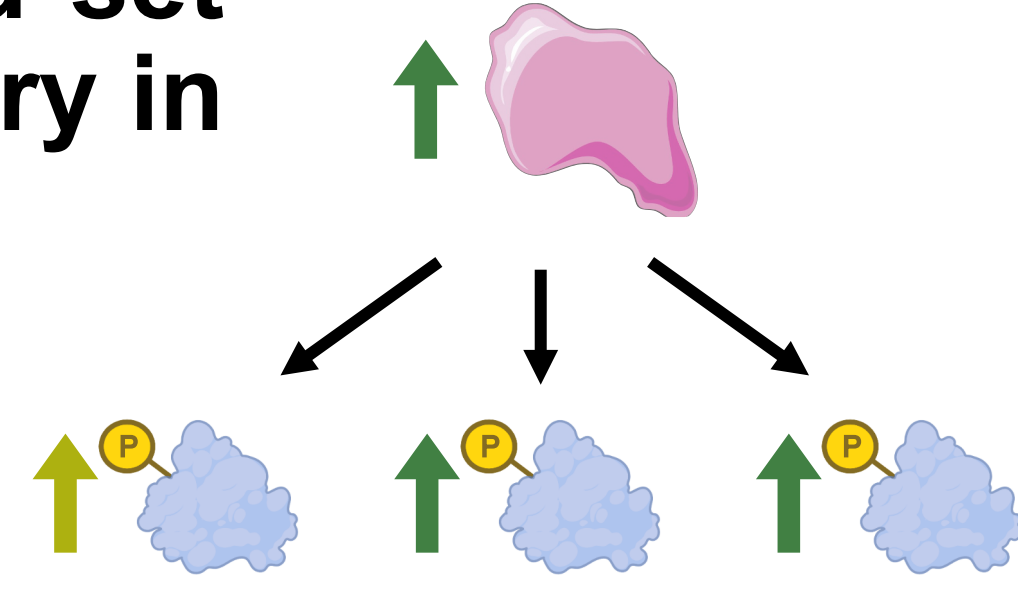
@saezlab
@sophia_muellerd

INTRODUCTION

In recent years, several methods have been developed to **infer the activity of kinases from phosphoproteomic data.**

These methods usually **rely on a fixed set of prior knowledge** interactions and **vary in the complexity** of their algorithms.
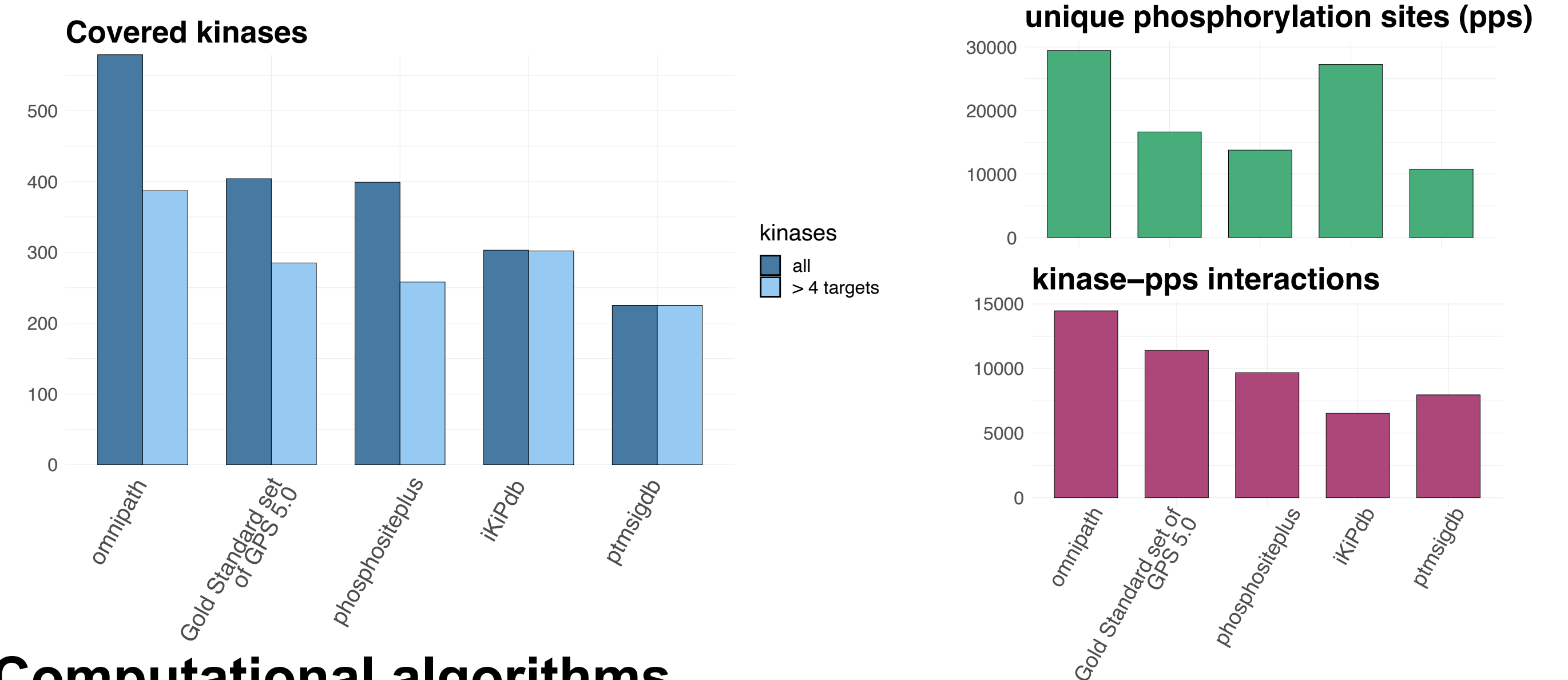
So far a **systematic comparison and evaluation is still missing.**

KINASE PERTURBATION DATASETS

We **estimated kinase activities for 103 manually curated perturbation experiments** (Hernandez-Armenta, 2017).

In these experiments **30 different kinases were perturbed,** supposedly leading to an increase or decrease in activity.

up-regulation: 105, down-regulation: 79 cases

METHOD OVERVIEW

The methods can be divided into two components: A **prior knowledge resource** and a **computational algorithm**, which can be flexibly combined.

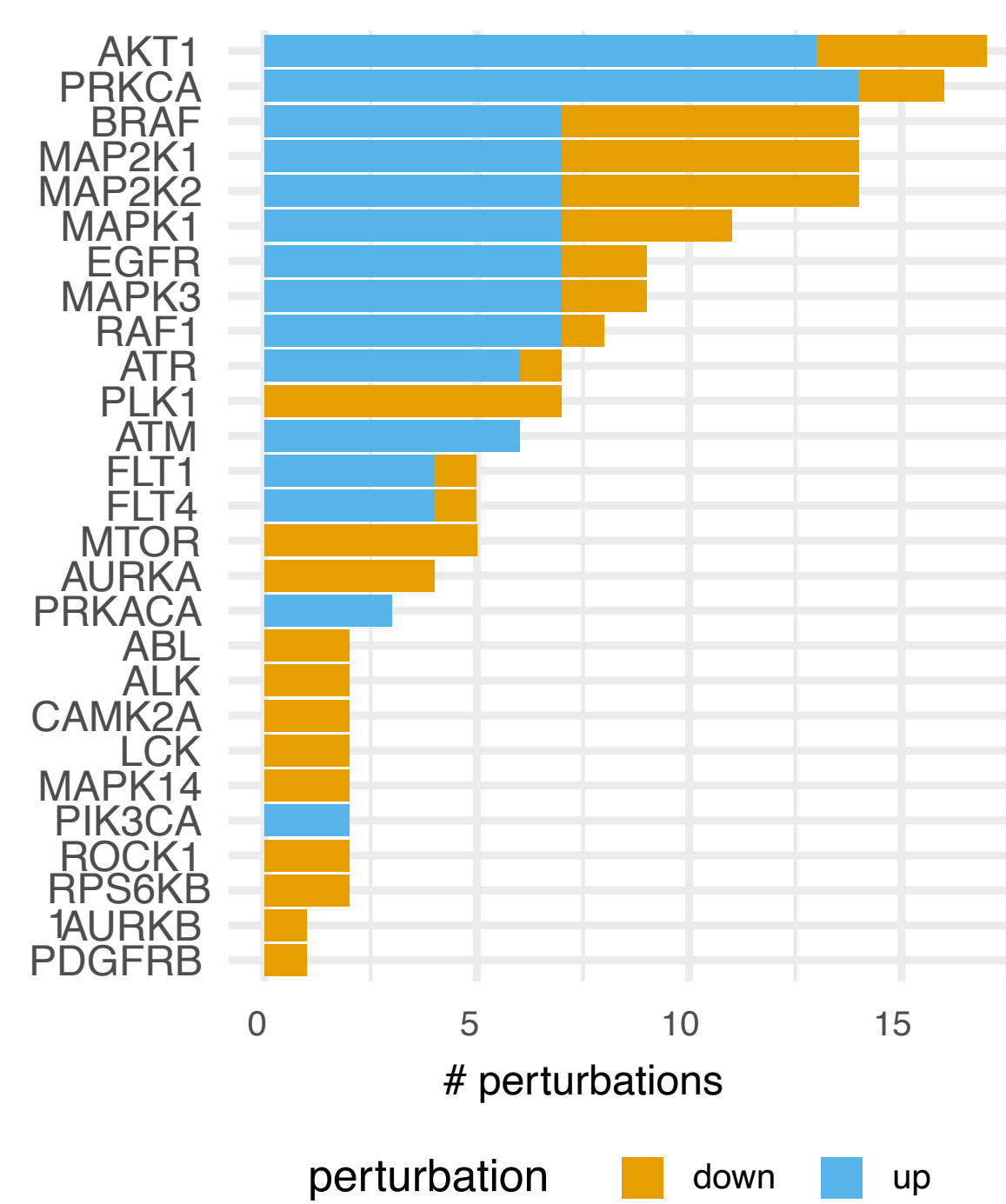### Prior knowledge resources

Computational algorithms

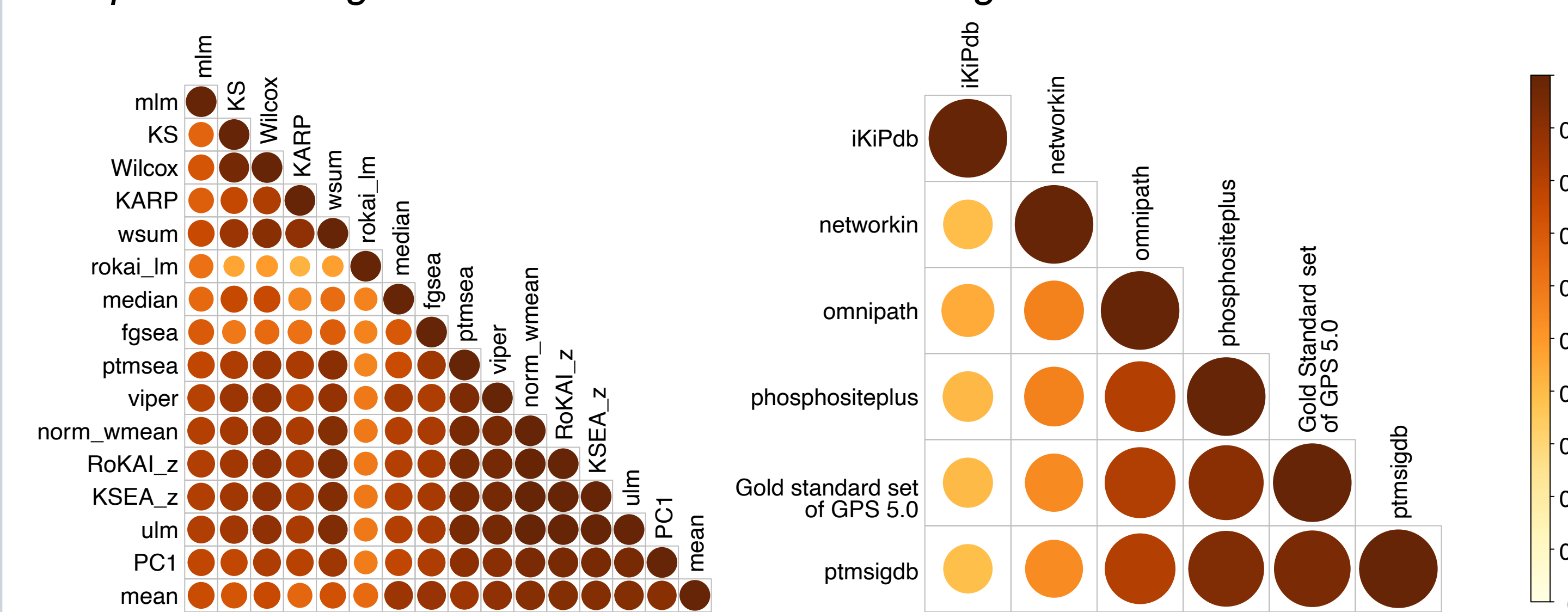| Method	Description	Link
fgsea	Fast gene set enrichment	(Sergushichev, 2016, Badia-i-Mompel et al. 2022)
KARP	K-score as implemented in KARP	(Wilkes et al. 2017)
KS	Kologomorov-Smirnov test comparing targets to non-targets	
lm (RoKAI)	Linear model as implemented in RoKAI	(Yilmaz et al. 2021)
mean	Mean of target sites	
Norm mean (decoupler)	Mean of target sites normalized by random permutations	(Badia-i-Mompel et al. 2022)
median	Median of target sites	
mlm	Multivariate linear model	(Badia-i-Mompel et al. 2022)
PC1	PC1 loading for target site set	
PTM-SEA	PTM-Signature Enrichment Analysis	(Krug et al. 2019)
sum	Sum of target sites	
ulm (decoupler)	Univariate linear model as implemented in decoupler	(Badia-i-Mompel et al. 2022)
viper	Virtual Inference of Protein-activity by Enriched Regulon analysis	(Alvarez et al., 2016, Badia-i-Mompel et al. 2022)
Wilcox	Wilcoxon rank sum test comparing targets to non-targets	
z score	Normalized mean of target sites as implemented in KSEA or RoKAI	(Wiredja et al. 2017, Yilmaz et al. 2021)

SYSTEMATIC COMPARISON & EVALUATION

We **compared the inferred activities by calculating the mean Pearson** correlation between method and prior knowledge resource (left) and then **evaluated the methods in terms of identifying the perturbed kinase** in each experiment (right).
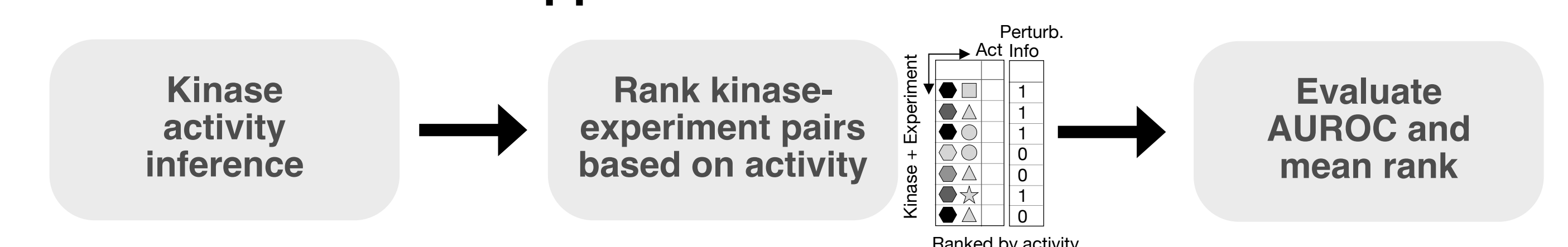
Lower mean Pearson correlation between different prior knowledge resources than between methods
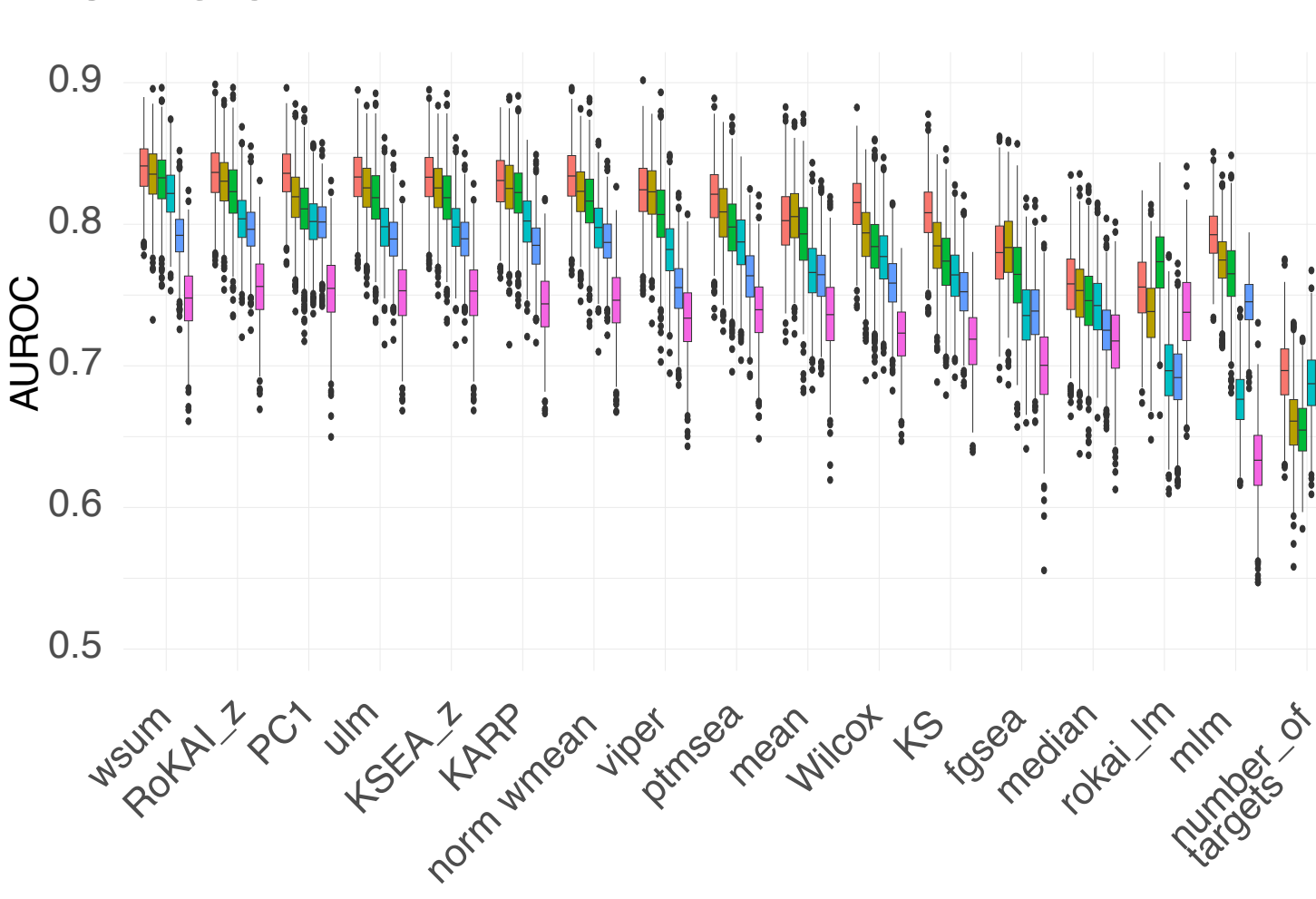
Computational algorithms: Prior knowledge resources:

Overview evaluation approach

Kinase activity inference → Rank kinase-experiment pairs based on activity → Evaluate AUROC and mean rank

Simpler approaches such as z-score or ulm already show a good performance in identifying perturbed kinases

AUROC:

Mean rank:
(lower is better)

⚠ **Remember**: The coverage of the resources must also be taken into consideration

CONCLUSION & OUTLOOK

Overall we observe that the prior knowledge resource has a greater impact on the inferred kinase activities, compared to the computational algorithms. For the future, we aim to incorporate the following points in assessing kinase activity inference methods:

- **Inclusion of further computational methods/prior knowledge resources**
- **Combination of prior knowledge resources**
- **Extending the benchmark metric to validate the methods for more kinases**

ACKNOWLEDGMENT

Krug, Karsten, Philipp Mertins, Bin Zhang, Peter Hornbeck, Rajesh Raju, Rushdy Ahmad, et al. 2019. "A Curated Resource for Phosphosite-Specific Signature Analysis." Molecular & Cellular Proteomics: MCP 18 (3): 576–93.
Alvarez, Mariano J., Yao Shen, Federico M. Giorgi, Alexander Lachmann, B. Belinda Ding, B. Hilda Ye, and Andrea Califano. 2016. "Functional Characterization of Somatic Mutations in Cancer Using Network-Based Inference of Protein Activity." Nature Genetics 48 (8): 838–47.
Badia-i-Mompel, Pau, Jesús Vélez Santiago, Jana Braunger, Celina Geiss, Daniel Dimitrov, Sophia Müller-Dott, Petr Taus, et al. 2022. "decoupleR: Ensemble of Computational Methods to Infer Biological Activities from Omics Data." Bioinformatics Advances 2 (1): vbac016.
Hernandez-Armenta, Claudia, David Ochoa, Emanuel Gonçalves, Julio Saez-Rodriguez, and Pedro Beltrao. 2017. "Benchmarking Substrate-Based Kinase Activity Inference Using Phosphoproteomic Data." Bioinformatics 33 (12): 1845–51.
Hornbeck, Peter V., Jon M. Kornhauser, Sasha Tkachev, Bin Zhang, Elzbieta Skrzypek, Beth Murray, Vaughan Latham, and Michael Sullivan. 2012. "PhosphoSitePlus: A Comprehensive Resource for Investigating the Structure and Function of Experimentally Determined Post-Translational Modifications in Man and Mouse." Nucleic Acids Research 40 (Database issue): D261–70.
Korotkevich, Gennady, Vladimir Sukhov, Nikolay Budin, Boris Shpak, Maxim N. Artyomov, and Alexey Sergushichev. 2016. "Fast Gene Set Enrichment Analysis." bioRxiv. bioRxiv. https://doi.org/10.1101/060012.
Linding, Rune, Lars Juhl Jensen, Adrian Pasculescu, Marina Olhovsky, Karen Colwill, Peer Bork, Michael B. Yaffe, and Tony Pawson. 2008. "NetworKIN: A Resource for Exploring Cellular Phosphorylation Networks." Nucleic Acids Research 36 (Database issue): D695–99.
Mari, Tommaso, Kirstin Mösbauer, Emanuel Wyler, Markus Landthaler, Christian Drosten, and Matthias Selbach. 2022. "In Vitro Kinase-to-Phosphosite Database (iKiP-DB) Predicts Kinase Activity in Phosphoproteomic Datasets." Journal of Proteome Research 21 (6): 1575–87.