

Benchmarking regulatory network inference algorithms for single-cell RNA-seq datasets

Matthew Stone^{1,4}, Sunnie Grace McCalla^{2,4}, Viswesh Periyasamy^{3,4}, Alireza Fotuhi Siahpirani^{3,4}, Sushmita Roy^{1,3,4}

Departments of ¹Biostatistics and Medical Informatics, ²Genetics, and ³Computer Science – University of Wisconsin, Madison

⁴Wisconsin Institute for Discovery

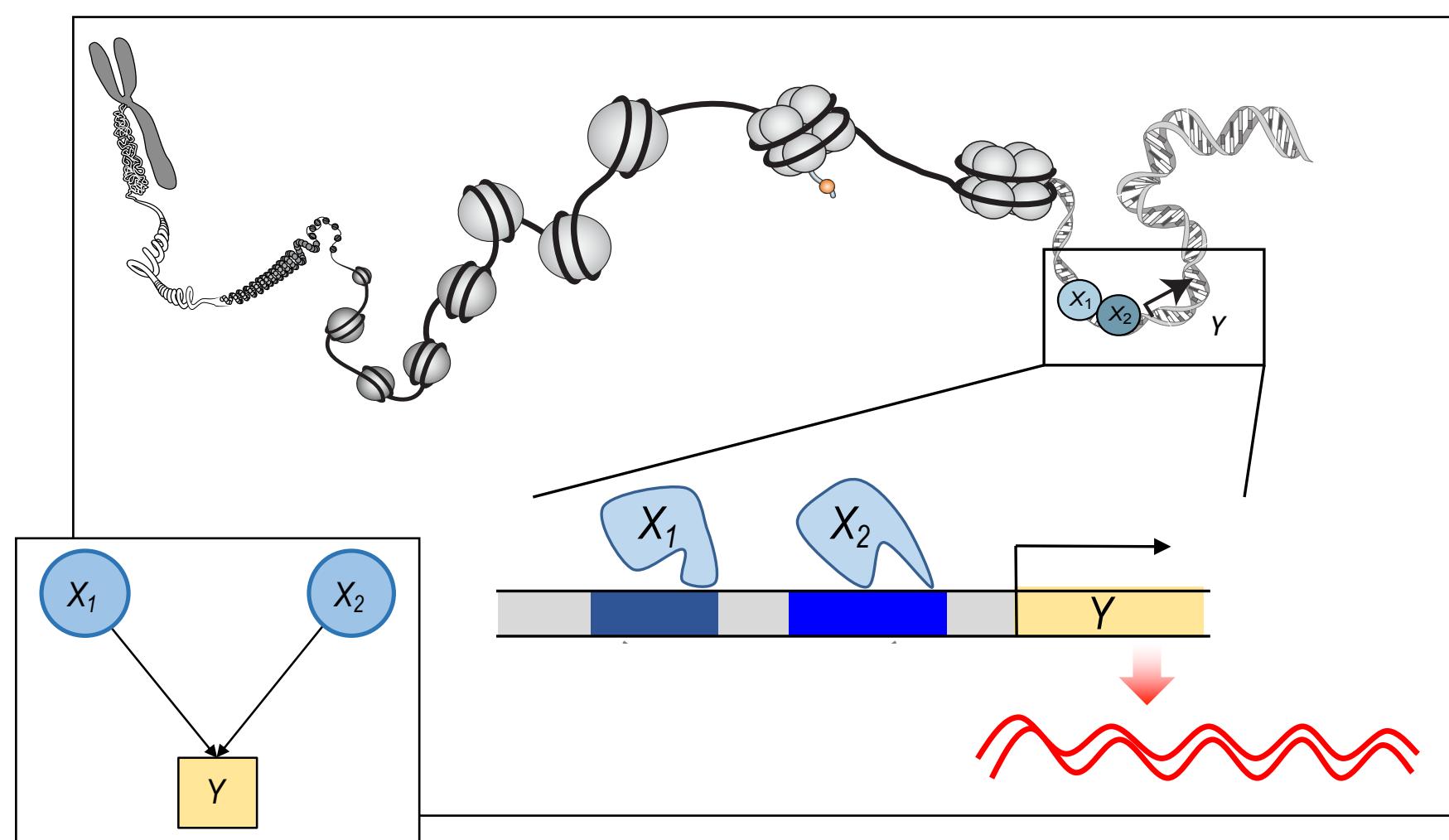


WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

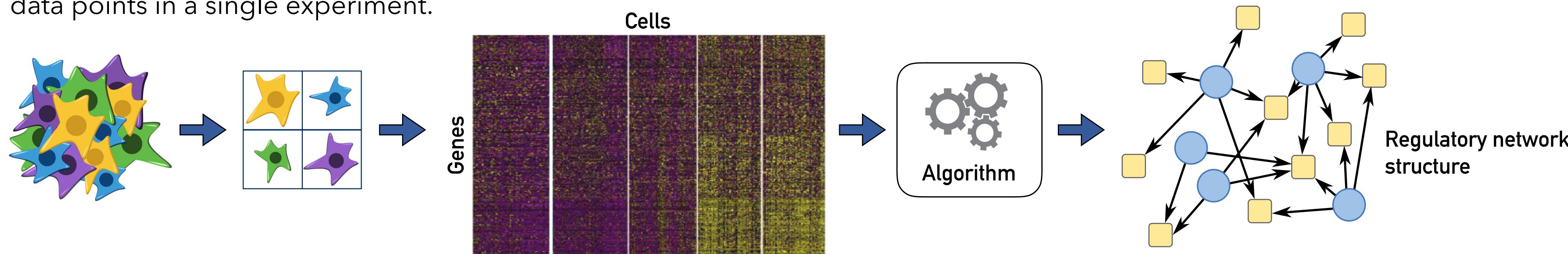
WISCONSIN
INSTITUTE FOR DISCOVERY
AT THE UNIVERSITY OF WISCONSIN-MADISON

Motivation

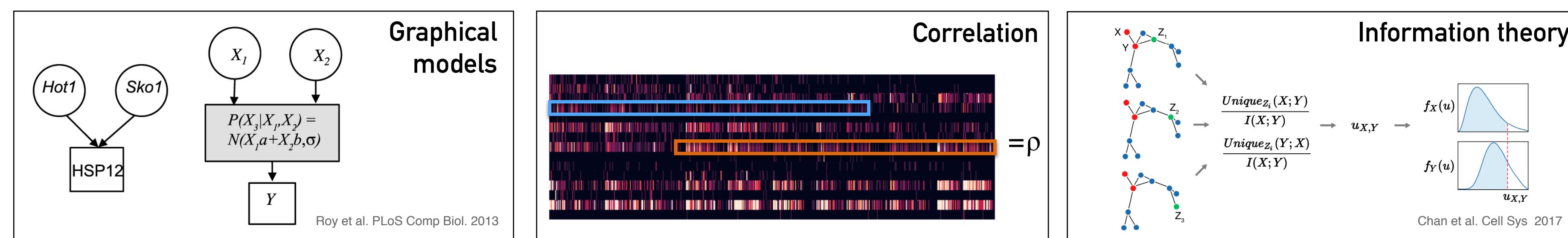
A **transcriptional regulatory network** specifies the regulators (transcription factors; TFs) that control which genes must be expressed when and where. Regulatory networks govern context-specific gene expression and are important in many different biological processes such as cell-fate specification, environmental stress response, and species evolution. Therefore, determining regulatory network structure is essential for improving our understanding of both normal and disease processes.



Network inference problem. Experimentally determining a genome-wide regulatory network is expensive, but networks can be inferred algorithmically from gene expression. Efforts from bulk transcriptomics required the curation of multiple experiments to obtain sufficient sample sizes, but single-cell RNA-seq (scRNA-seq) can now generate thousands of data points in a single experiment.



Algorithmic approaches. Many algorithms have been developed to infer regulatory networks from scRNA-seq data, and it is not yet clear which approach yields the most accurate networks.



Experimental design. We evaluated 13 network inference algorithms on seven published scRNA-seq datasets from human, mouse, and yeast cells.

Algorithm	Citation	Methodology
HurdleNormal	McDavid et al., arXiv 1610.05857	Graphical models and dependency networks
Inferelator	Jackson et al., bioRxiv 2019	
MERLIN**	Roy et al., PLoS Comp Biol 2013	
SCENIC	Aibar et al., Nature Methods 2017	
SCHiRM	Intosalmi et al., bioRxiv 2018	
SILGGM	Zhang et al., PLoS Comp Biol 2018	
kNN-DREMI	van Dijk et al., Cell 2018	Information theoretic
PIDC	Chan et al., Cell Systems 2017	
Scribe*	Qiu et al., bioRxiv 2018	
BTR	Lim et al., BMC Bioinformatics 2016	
SCODE*	Matsumoto et al., Bioinformatics 2017	
LEAP*	Specht and Li, Bioinformatics 2017	
Pearson	n/a	Correlation

Dataset	Cell type	# cells	# genes
Gasch et al., PLoS Biol 2017	Yeast	163	3,847
Jackson et al., bioRxiv 2019		17,396	5,736
Shalek et al., Nature 2014	mDC**	1,211	9,411
Tran et al., Cell Reports 2019*		2,369	6,618
Tran et al., Cell Reports 2019*	mESC	3,324	6,621
Zhao et al., Cell Stem Cell 2018		36,199	8,442
Han et al., Genome Biol 2018	hESC	5,520	7,465
Gasch et al., PLoS Biol 2017		1,211	9,411

*Tran et al. report two experiments

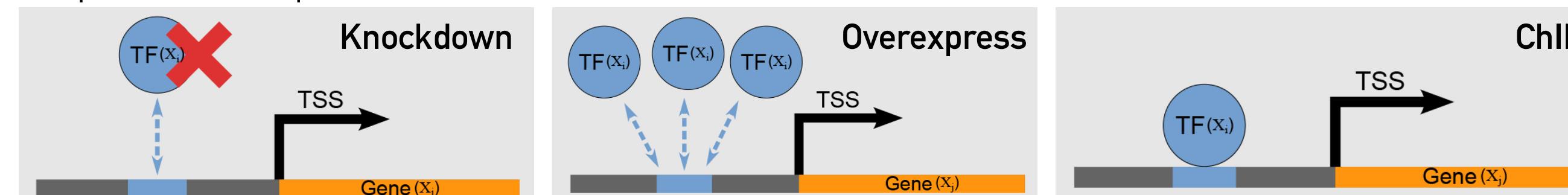
**mouse dendritic cells

We learned cellular trajectories and pseudotimes using Monocle, and imputed missing counts with MAGIC.

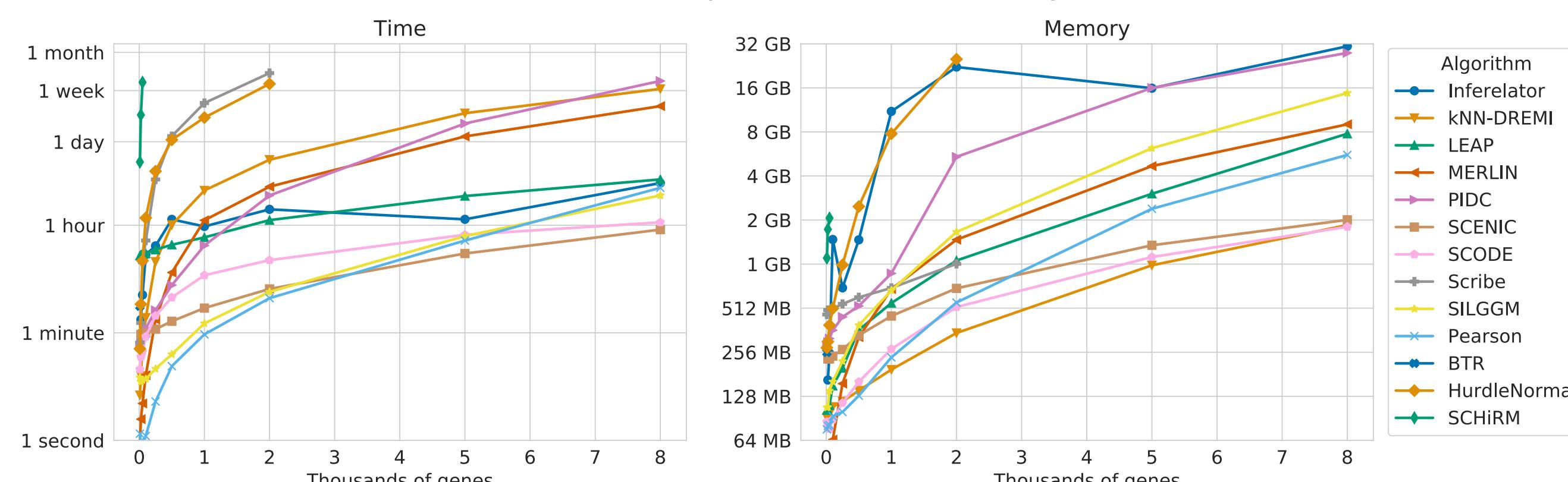
Qiu et al., Nat Methods 2017
Dijk et al., Cell 2018

*Incorporates pseudotime/trajecory
**Designed for bulk transcriptomic data

Evaluating network accuracy. We compared the inferred networks to networks constructed from experimental evidence for physical TF binding or expression changes in response to TF perturbation.

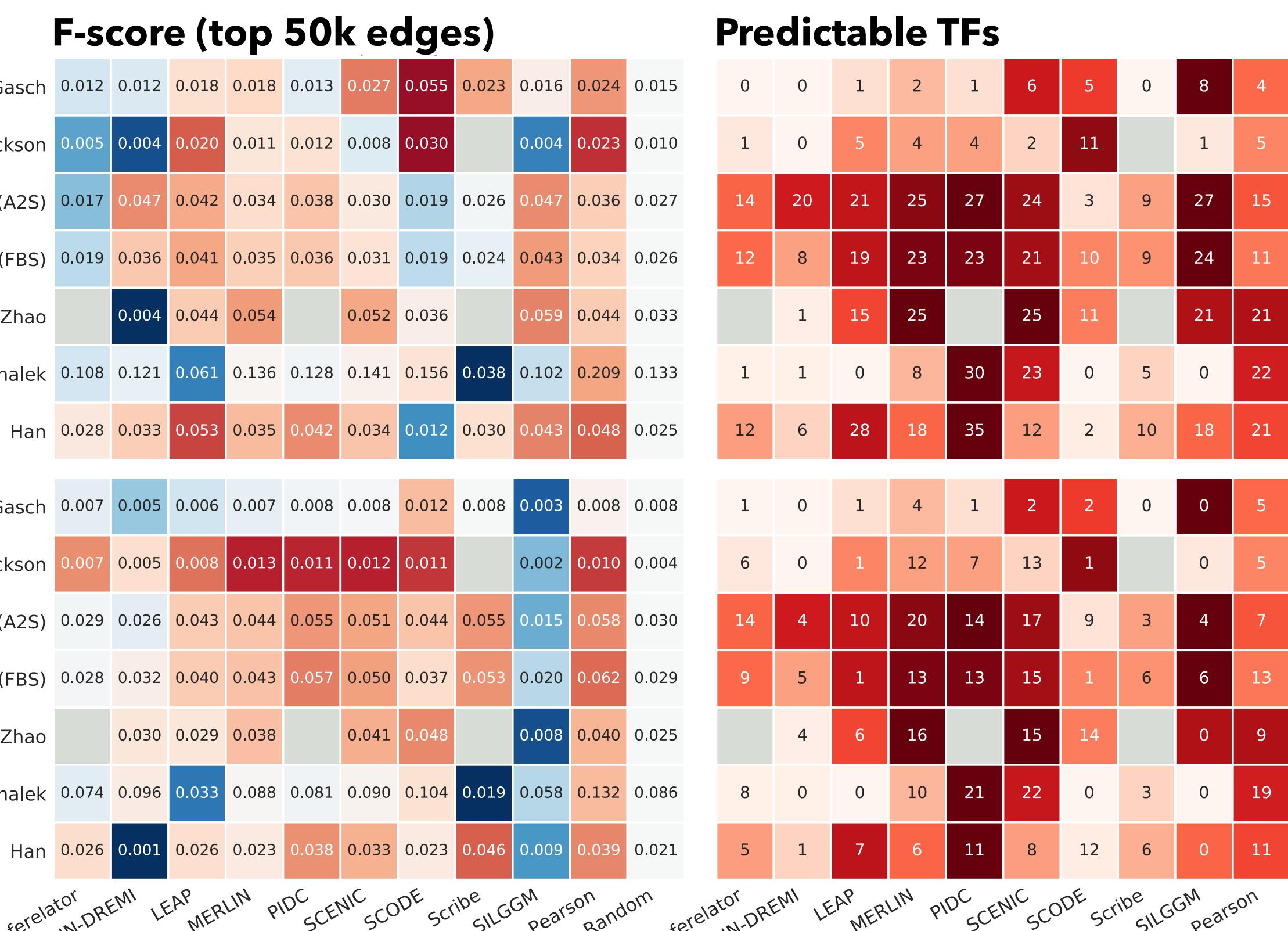


Benchmarking. We benchmarked each algorithm's runtime and memory consumption with respect to the number of genes in an expression matrix (n=5,520 cells). We excluded BTR, SCHiRM, and HurdleNormal as they did not scale to a genome-wide network.

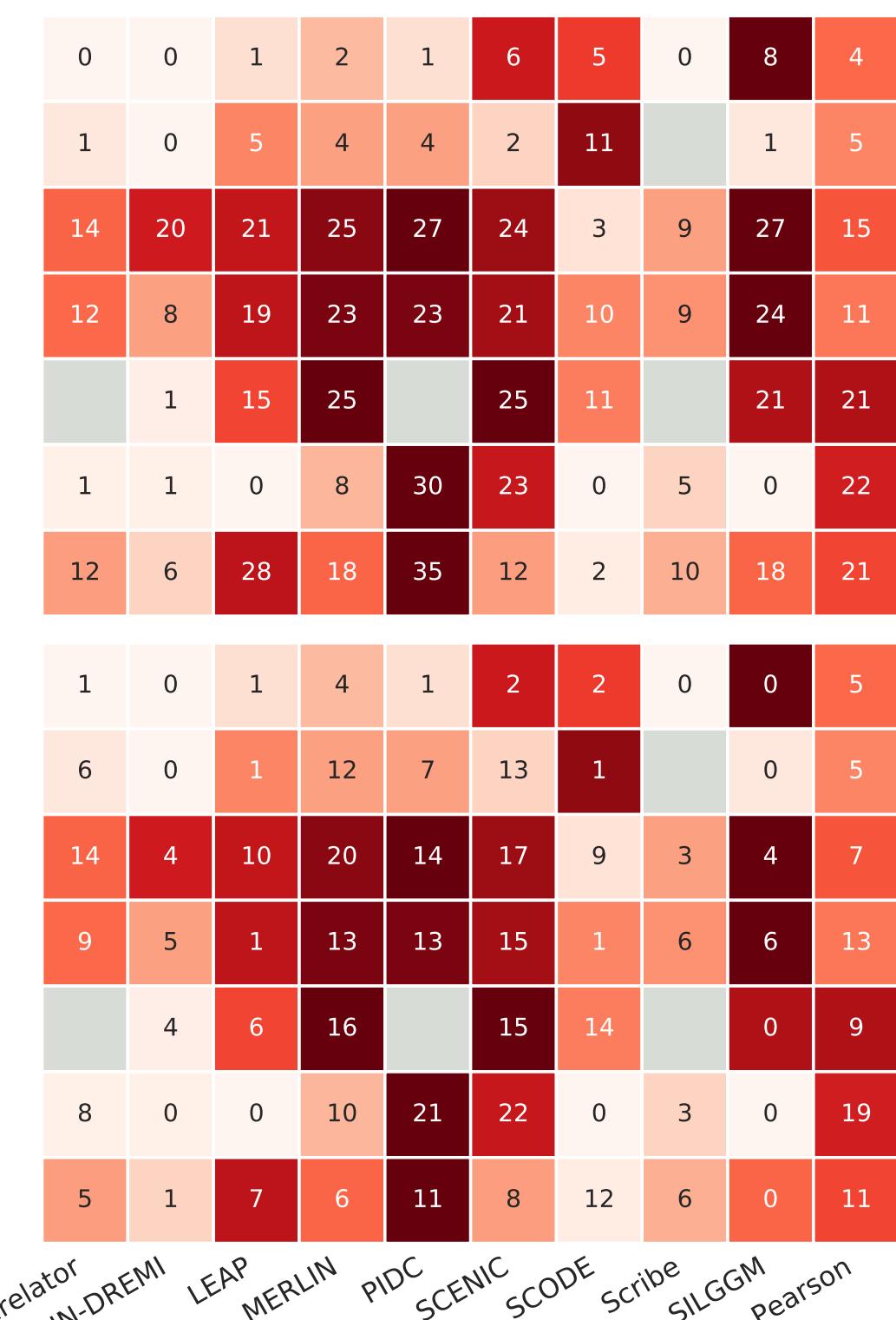


Approach

Perturbation



Predictable TFs



Ranked performance



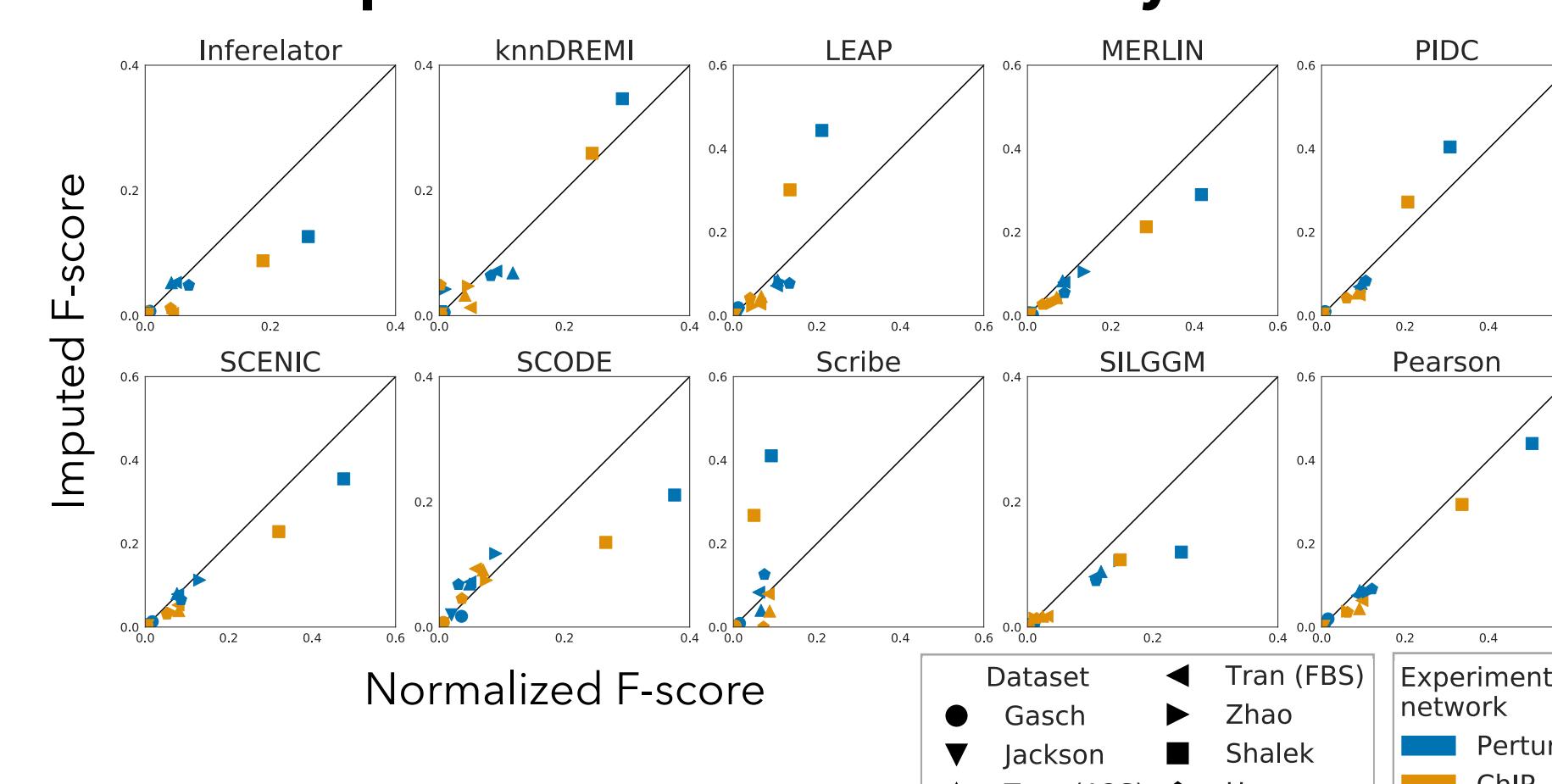
(Above) Numbers are median ranks of performance across datasets and evaluation metrics. Brighter yellow indicates better relative performance; darker blue, worse.

(Right) Scatter plots show the F-score of top 50k edges before and after imputation. Each point represents a dataset/algorithim pair from the heat maps at left. Points above the line indicate networks whose accuracy improved after imputation.

Results

ChIP

Effect of imputation on network accuracy



Summary

- We present a comprehensive study of network inference in yeast and mammals.
- Algorithms vary substantially in time and memory requirements.
- Imputation with MAGIC was generally not beneficial.
- Top performing methods are Pearson, SCENIC, PIDC, SILGGM, although results varied across datasets and reference networks.
- Different measures of network accuracy offer complementary information about algorithm performance.

Acknowledgments

We acknowledge the support from NLM Training Grant 5T15LM007359, the UW Data Science Initiative, and the ISCB Travel Fellowship.

Numbers are F-scores; colors are proportional to relative performance to a random network (last column). Darker red indicates better relative performance; darker blue, worse.

Numbers are counts of TFs whose predicted targets are significantly enriched for experimental targets. Darker red indicates more TFs and better performance.