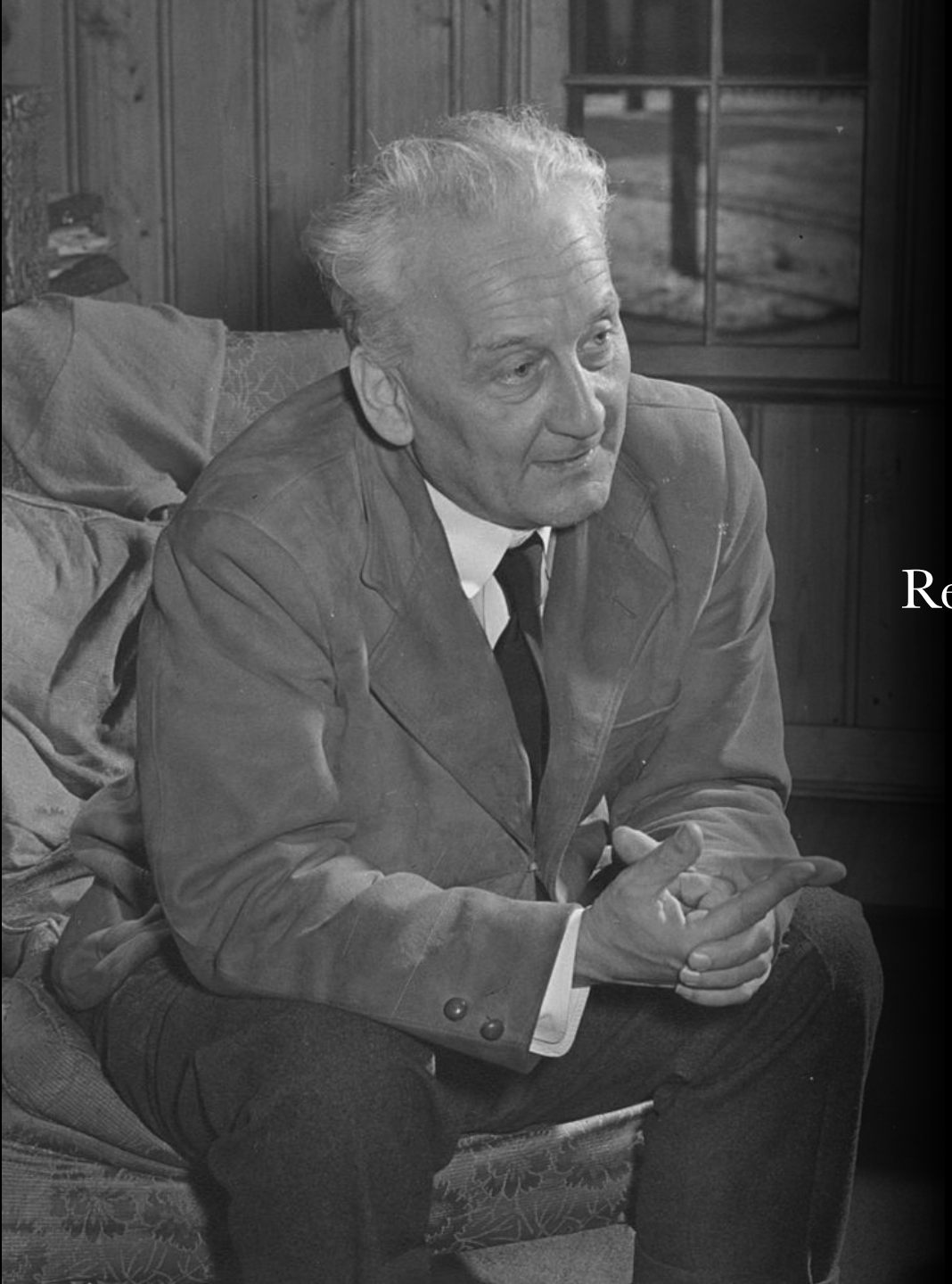
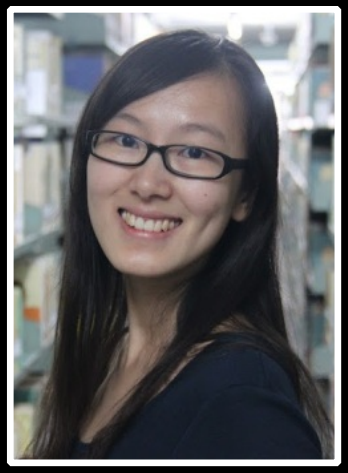


Research is to see what everybody else has seen and to think what nobody else has thought.

- Albert Szent-Györgyi

Image by J.W. McGuire/NIH



impossible question

Image from You Don't Know Jack. Vol 3.

Perelman School of Medicine
University of Pennsylvania

Data-driven pathway analysis with ADAGE

Casey Greene

@GreeneScientist

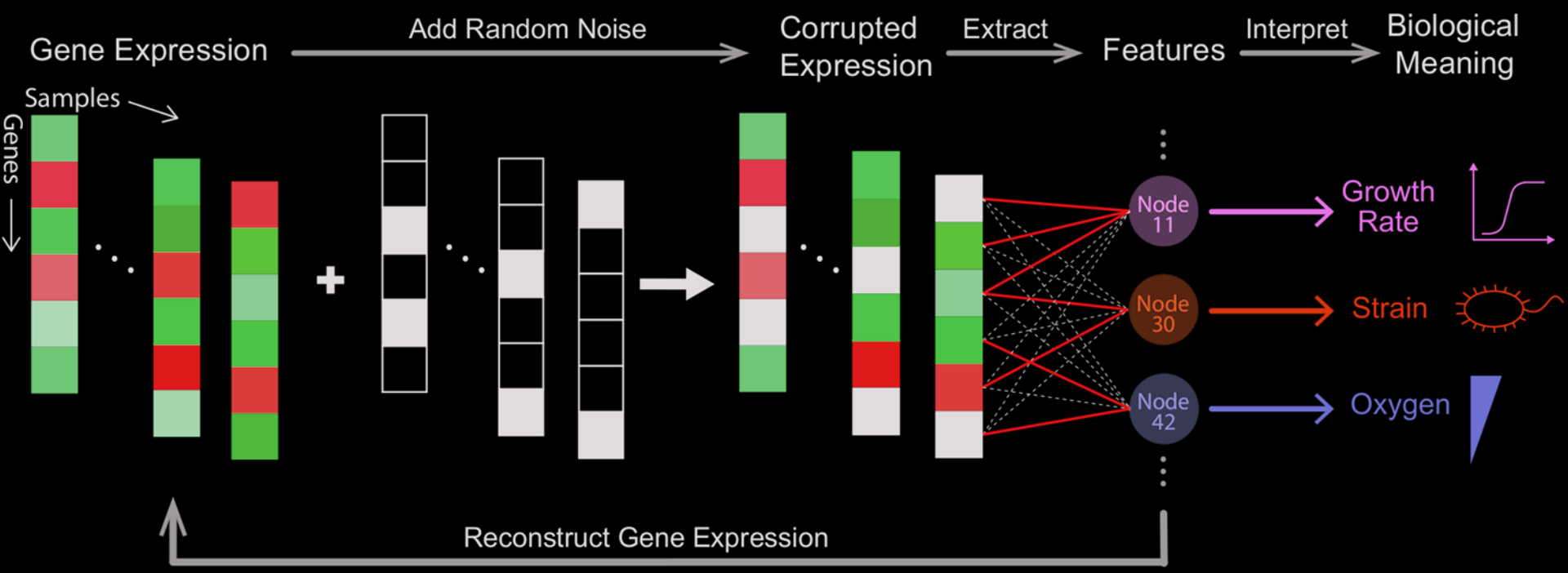
Greene Lab

Calvin and Hobbes. Bill Watterson

Le et al. 2012

If you showed 16,000 computers 10 million images from youtube, what would they see?

Analysis with Denoising Autoencoders of Gene Expression (ADAGE)

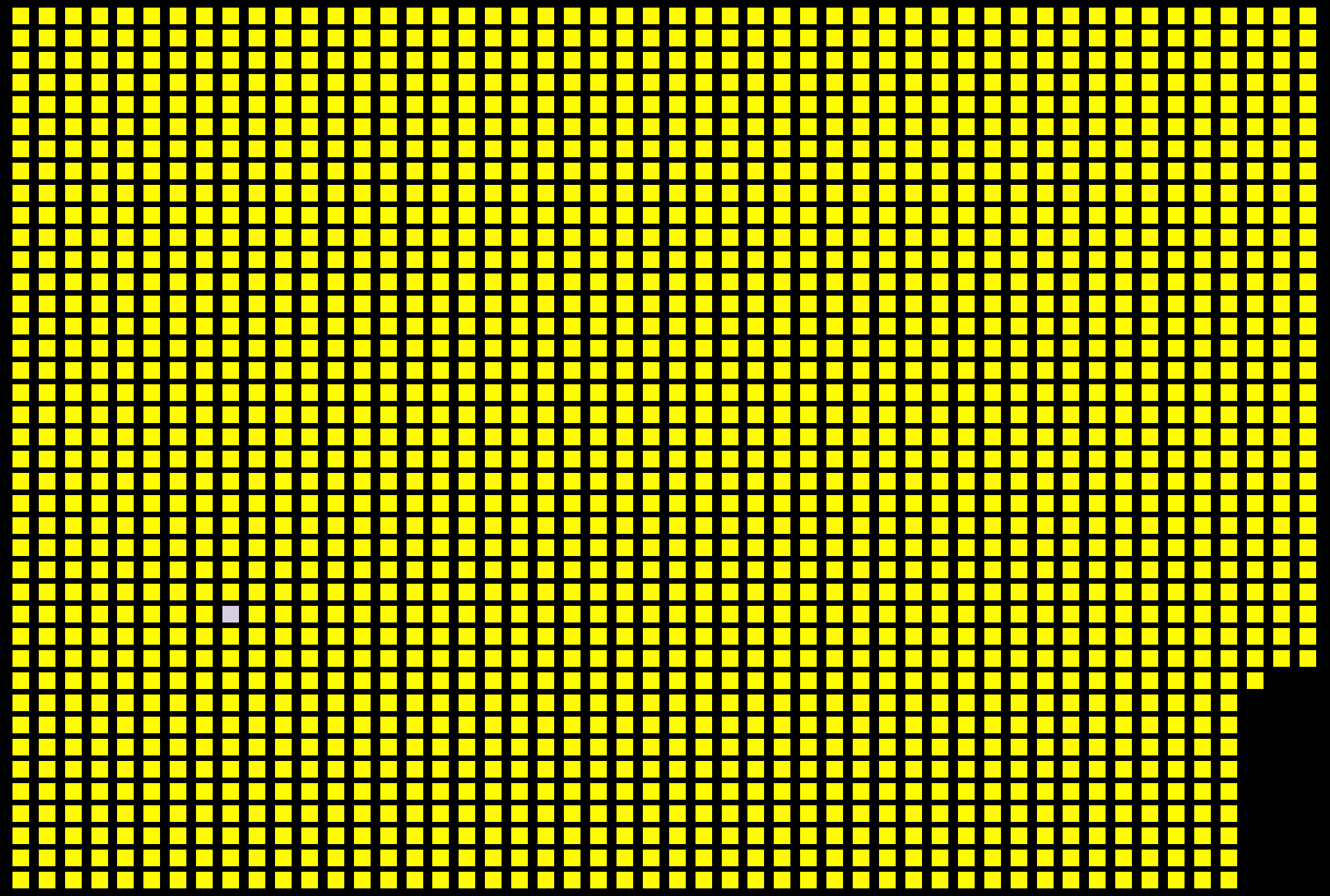


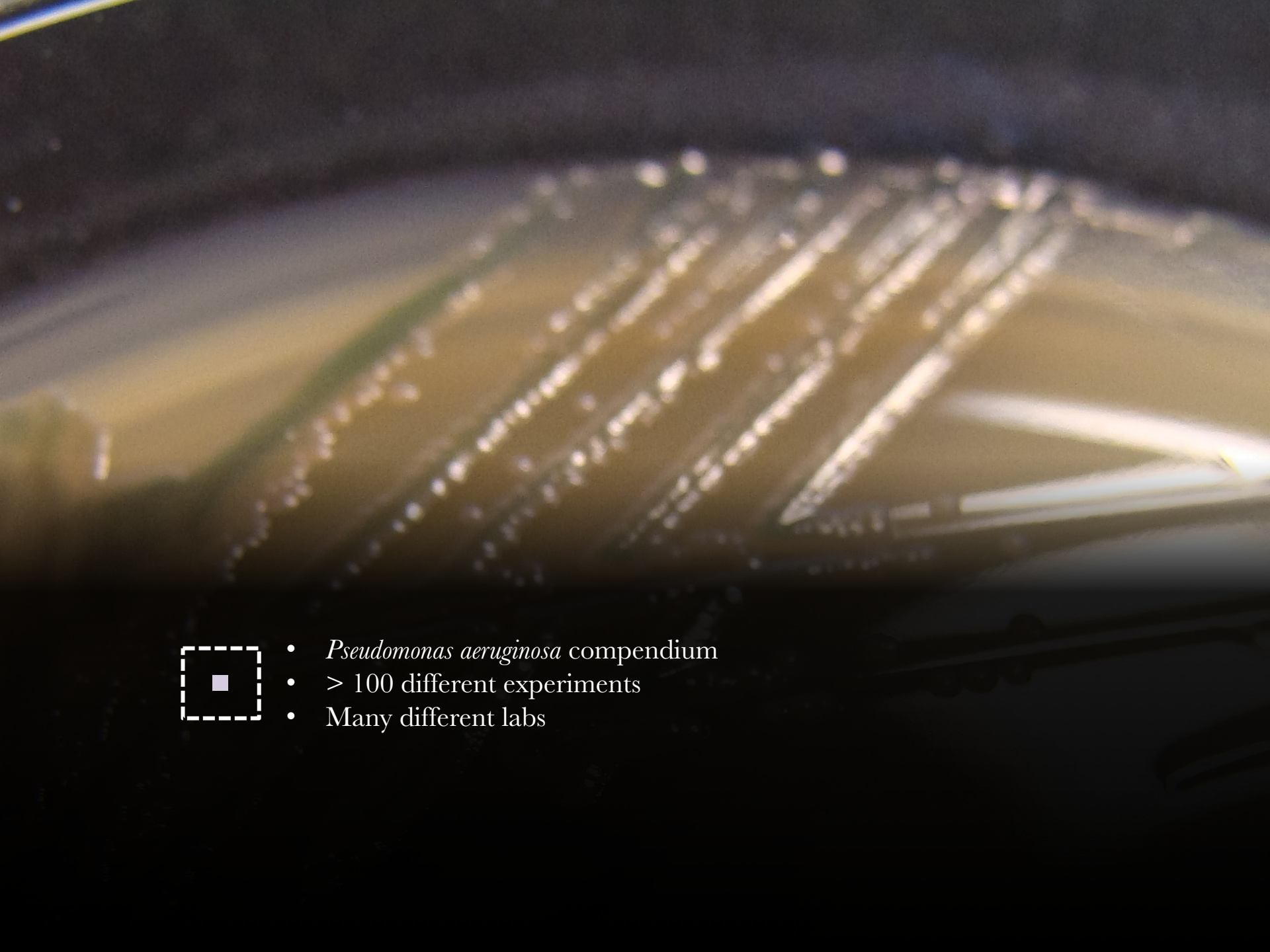
Tan et al. *Pac Sym Bio* 2015; Tan et al. *mSystems* 2016.

The future of deep learning

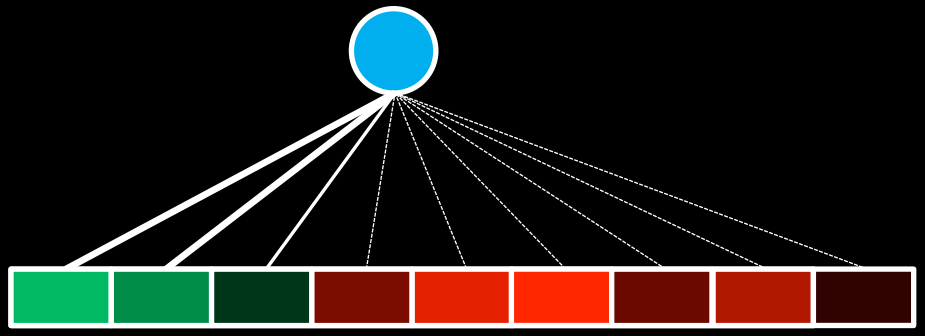
"Although we have not focused on it in this Review, we expect unsupervised learning to become far more important in the longer term."

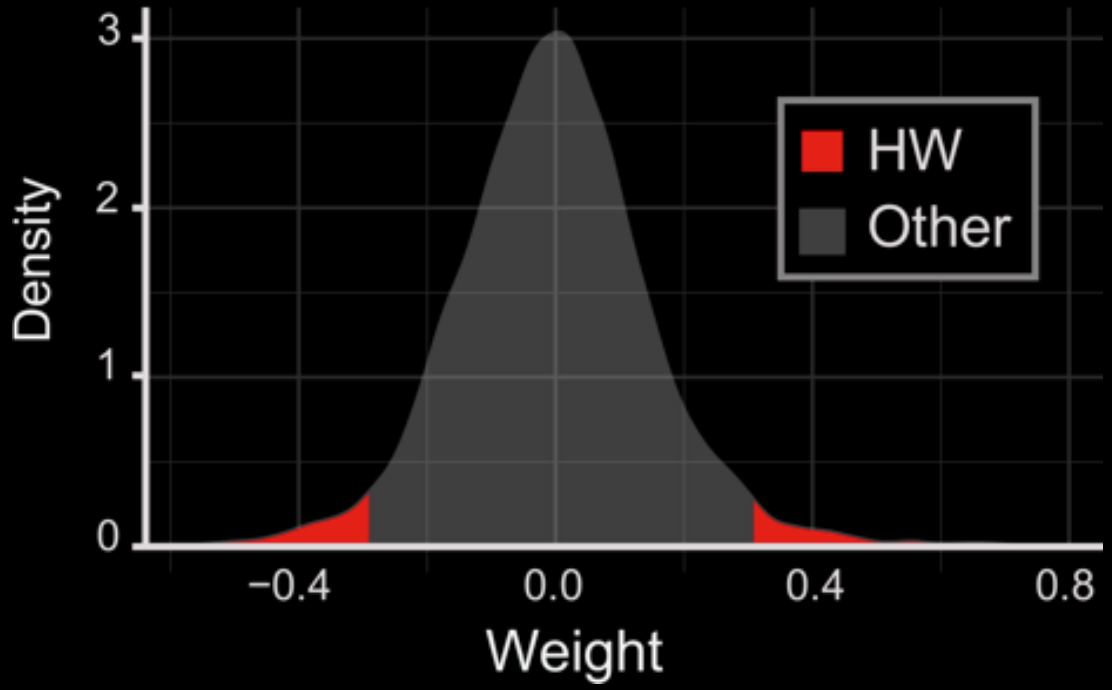
LeCun, Bengio, and Hinton. *Nature* 2015.



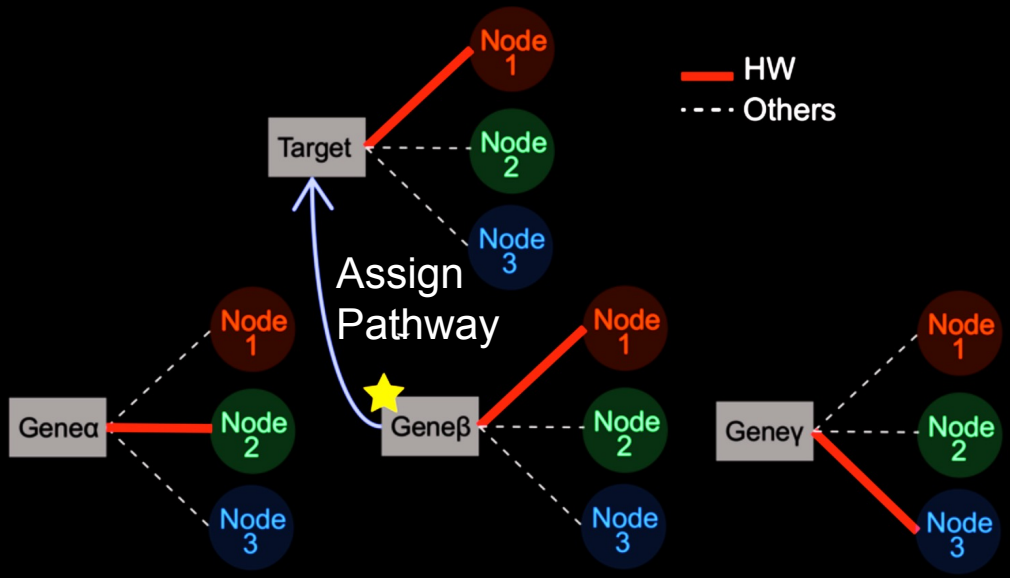

• *Pseudomonas aeruginosa* compendium
• > 100 different experiments
• Many different labs

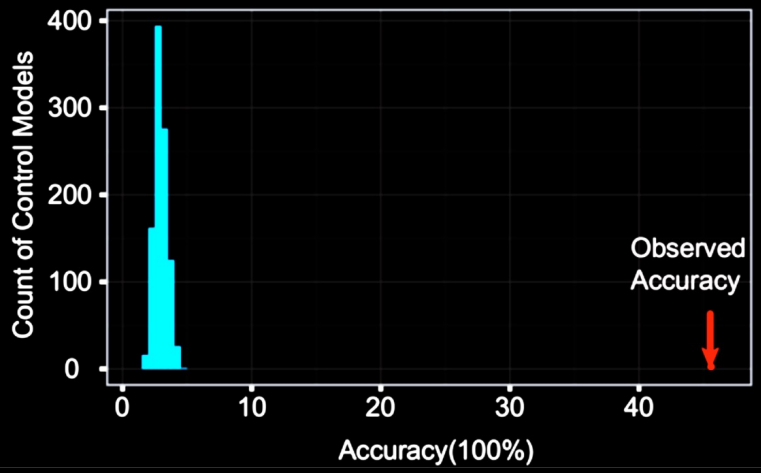
High-weight genes





HW relationships capture genes' pathways



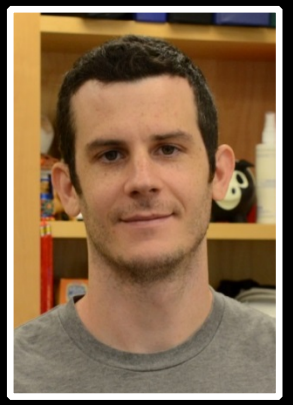


Tan et al. *mSystems.* 2016.

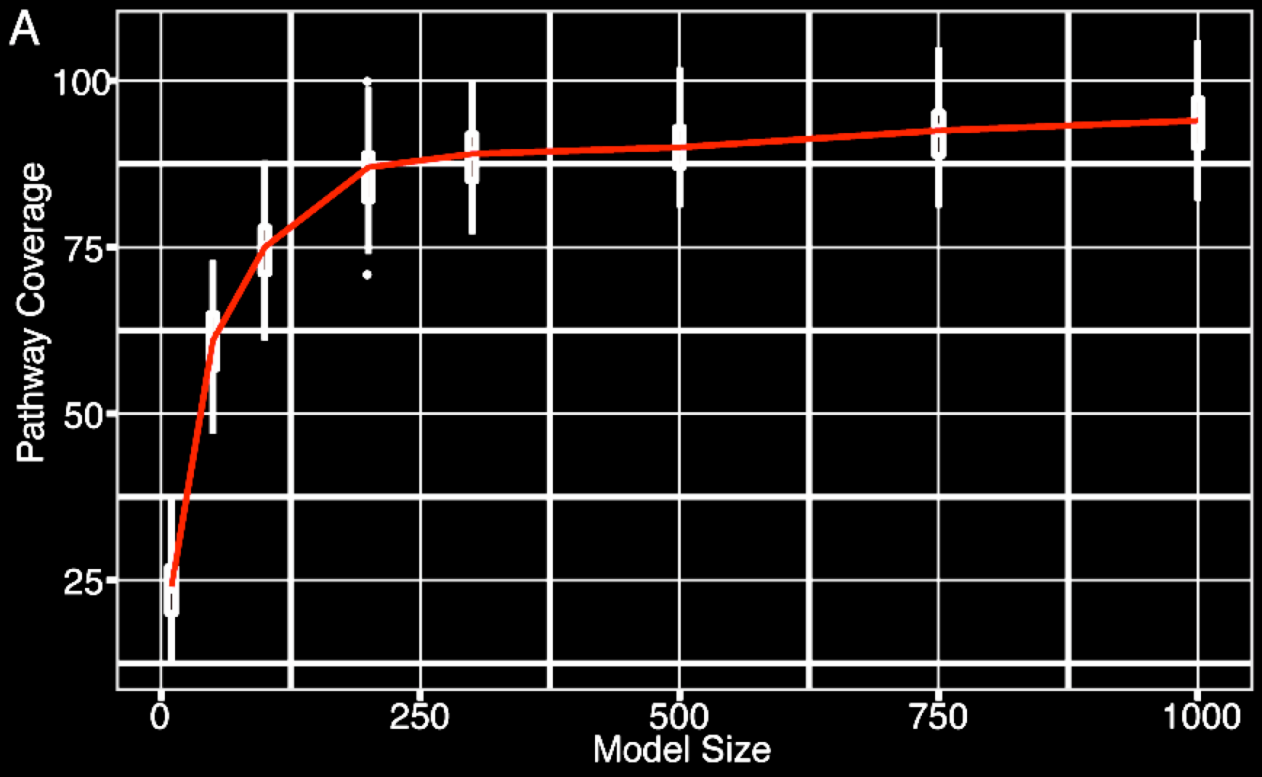
Fork me on GitHub

ADAGE analysis of publicly available gene expression data collections illuminates *Pseudomonas aeruginosa*-host interactions

bioRxiv: http://dx.doi.org/10.1101/030650
github: http://github.com/greenelab/adage
Tan, Hammond, Hogan, and Greene. *mSystems*. 2016

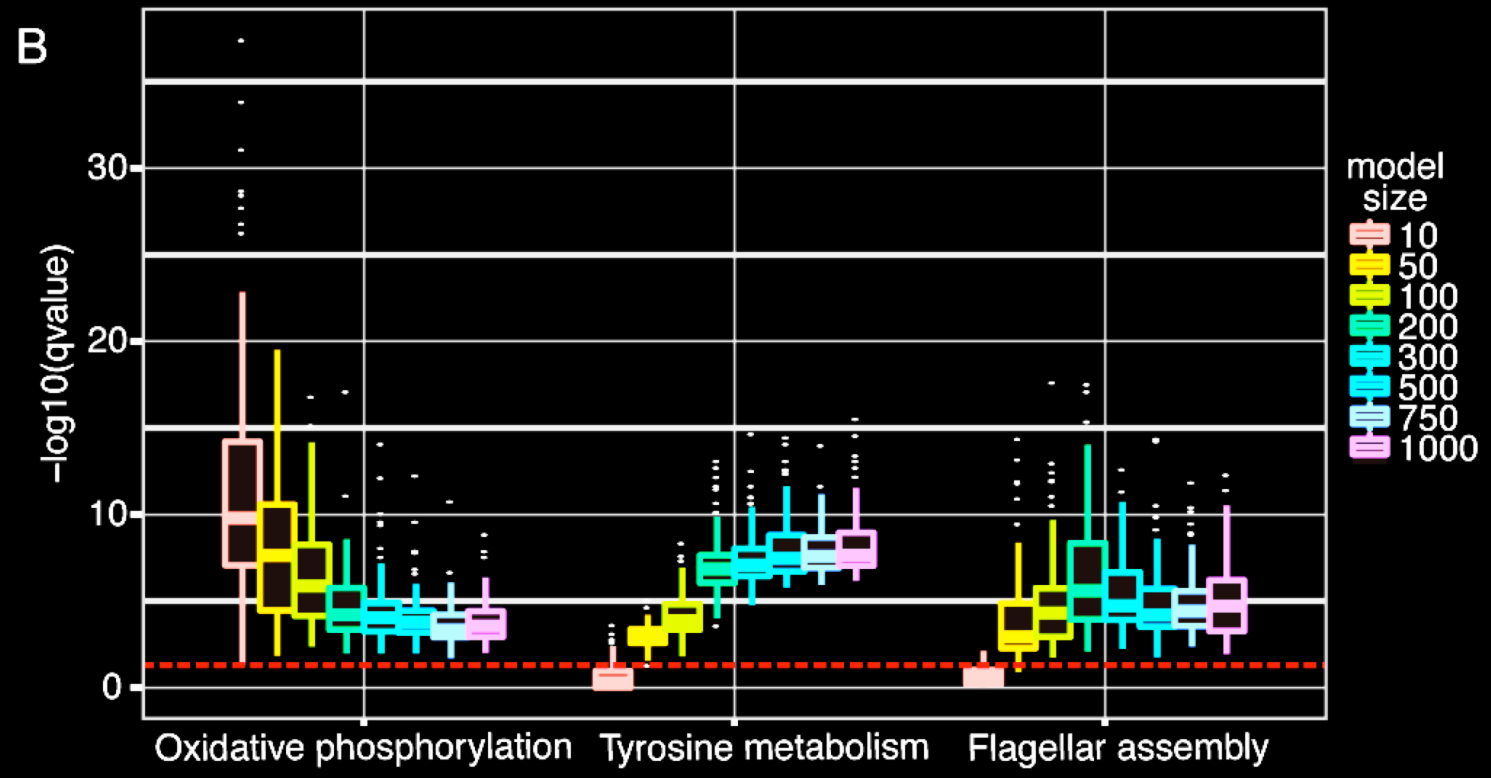


Bigger models generally find more pathways.



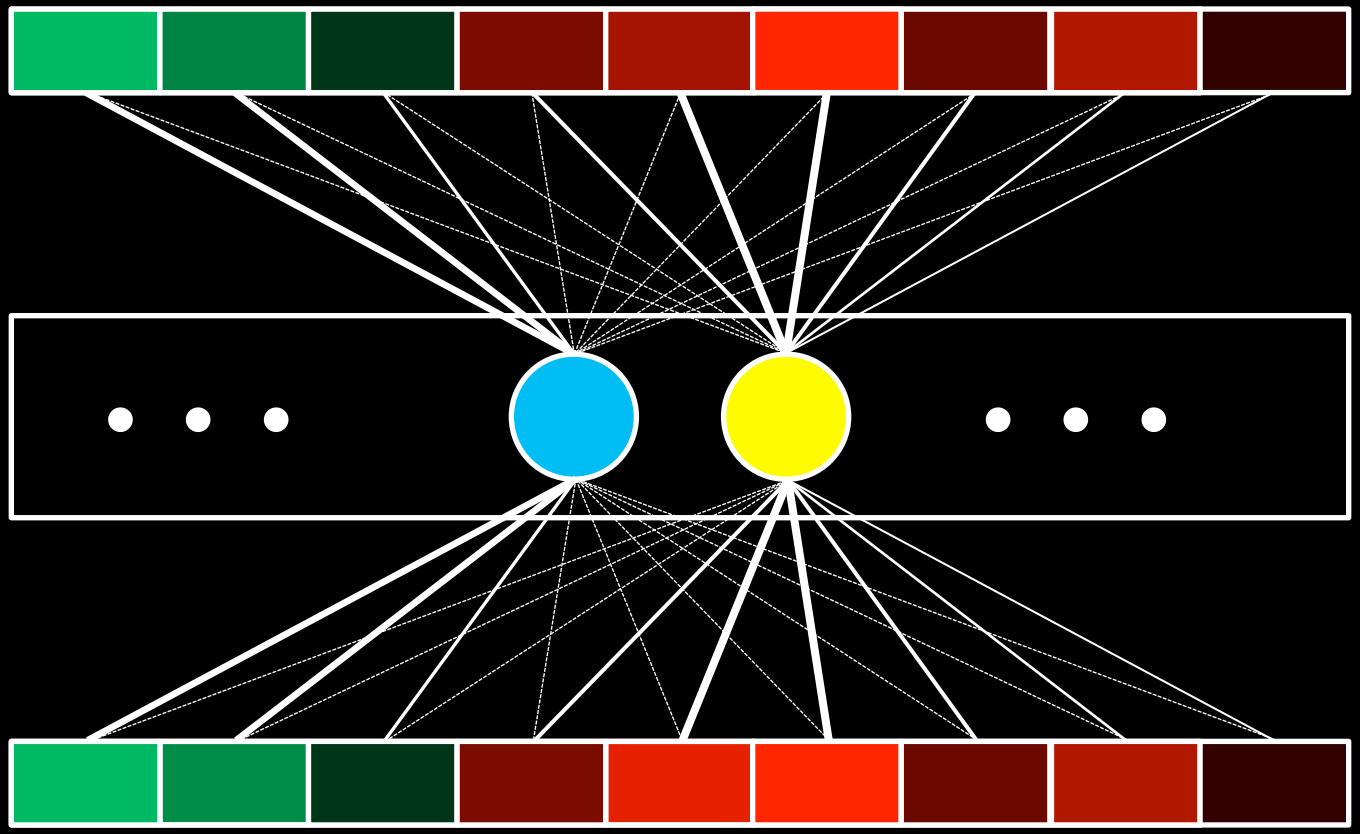
Tan et al. *in prep*.

The pathways captured by ADAGE change with model size.



Tan et al. *in prep*.

We can measure gene-gene similarity with ADAGE weights.

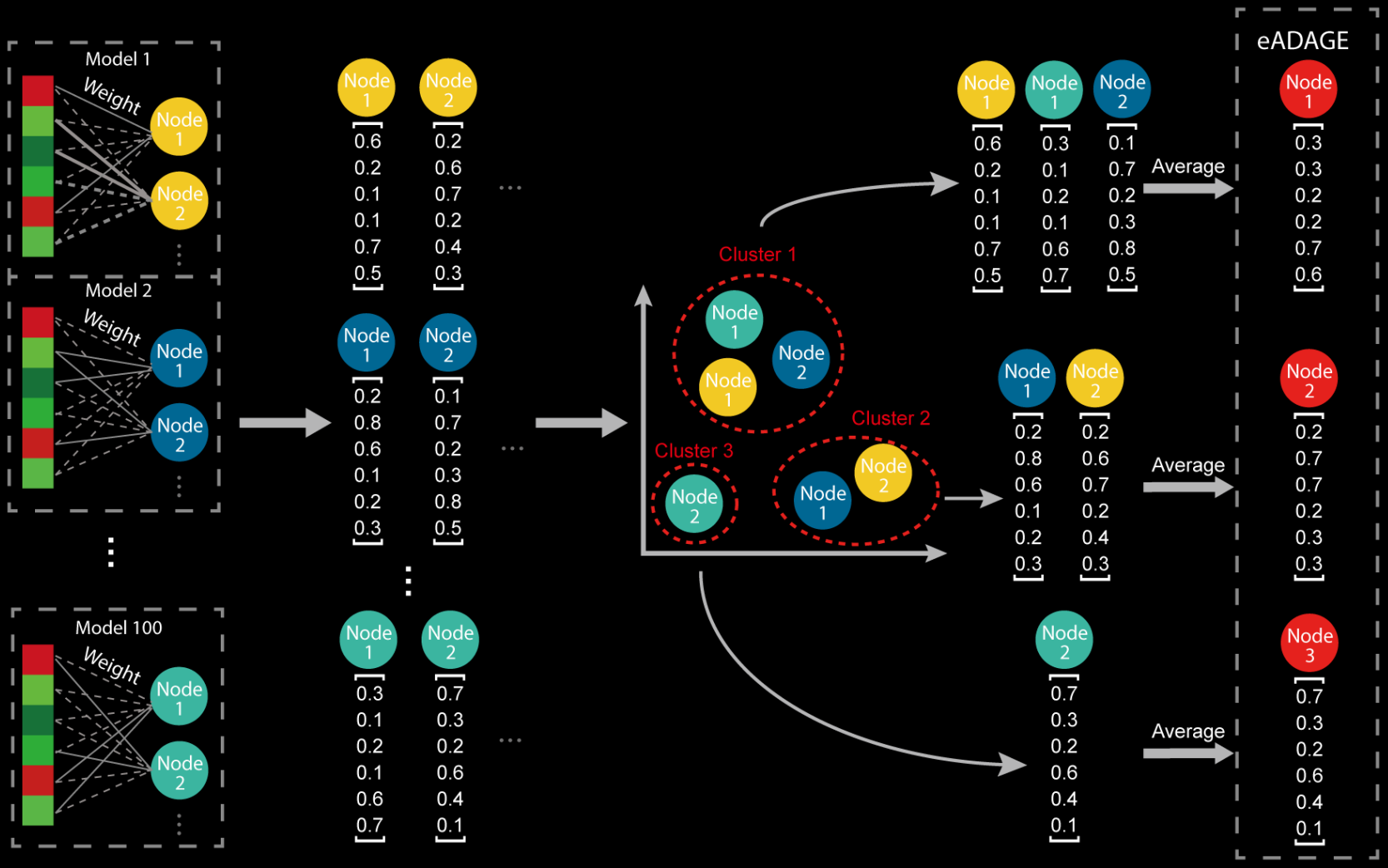


ADAGE similarity captures functional similarity.

Tan et al. *in prep*.

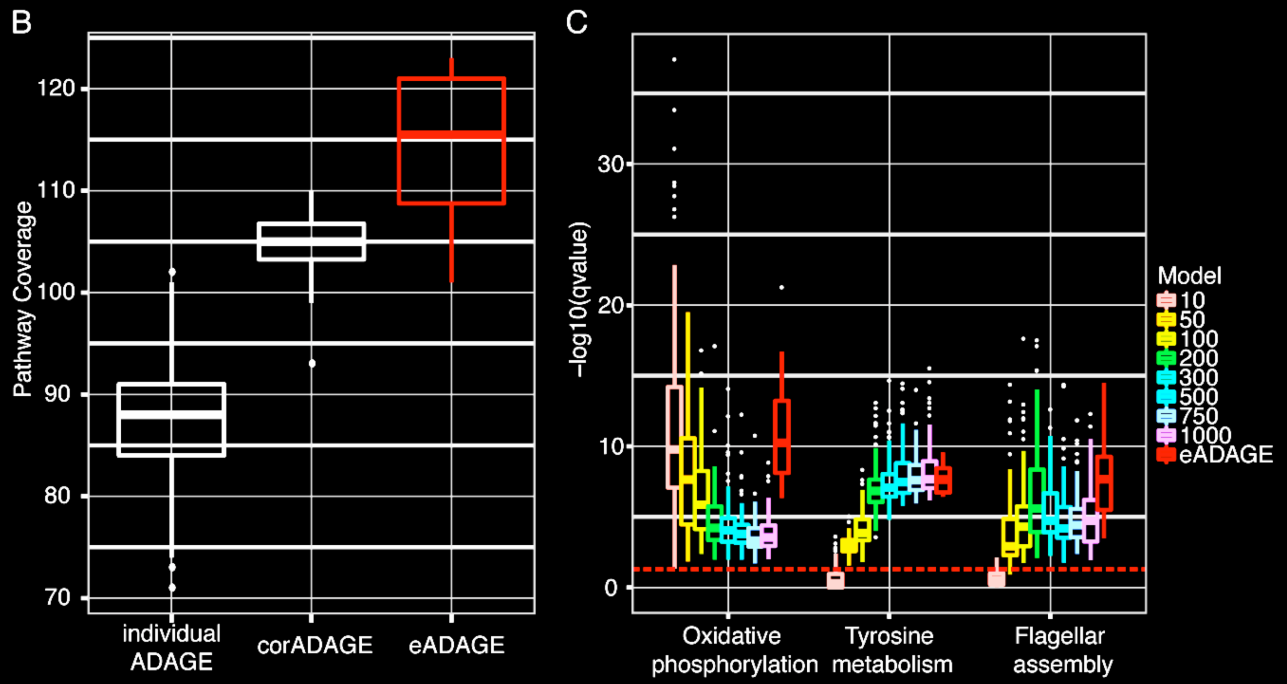
Image by Asafredell

ensemble ADAGE (eADAGE)



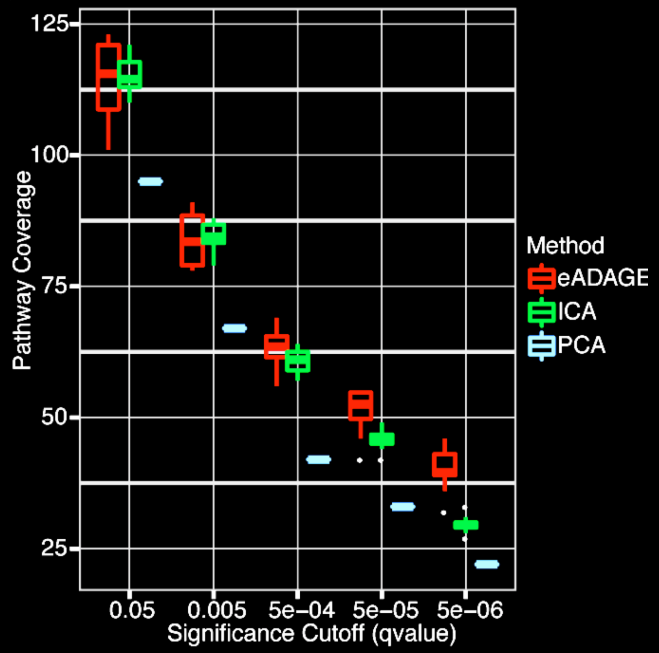
Tan et al. *in prep*.

eADAGE captures more pathways than ADAGE.



Tan et al. *in prep*.

eADAGE captures more pathways than ICA/PCA



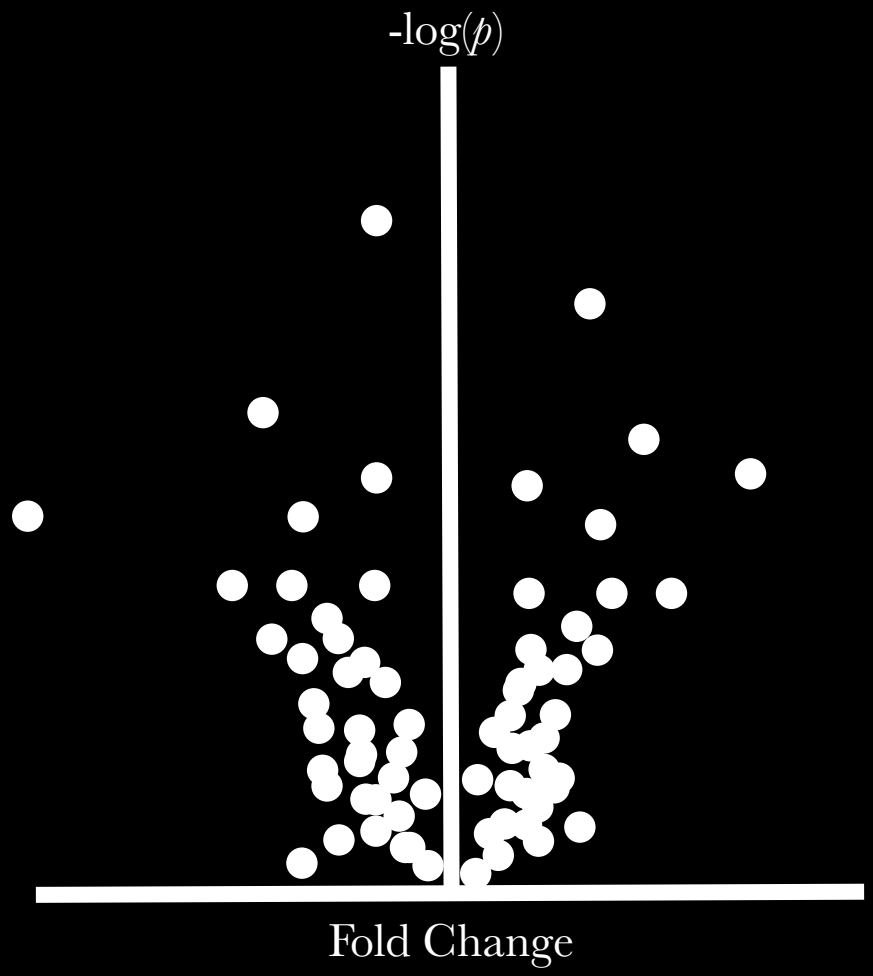
Tan et al. *in prep*.

I didn't want to just know the names of things. I remember really wanting to know how it all worked.

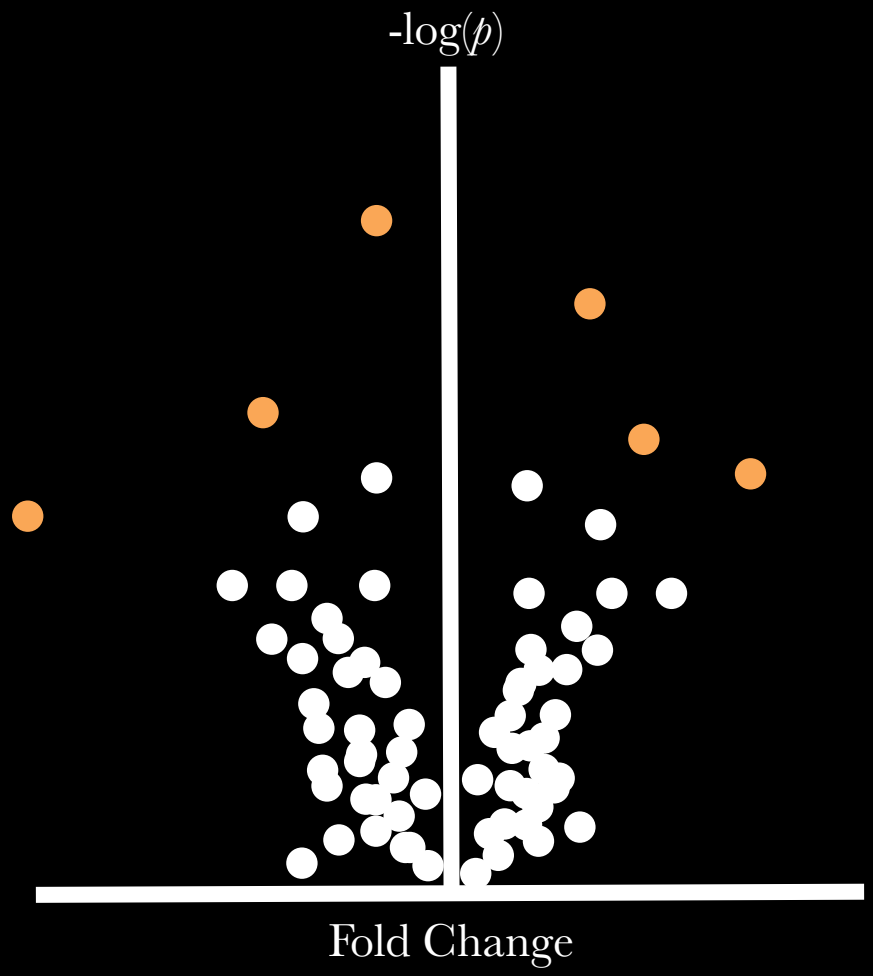
- Elizabeth Blackburn

Image: US Embassy Sweden

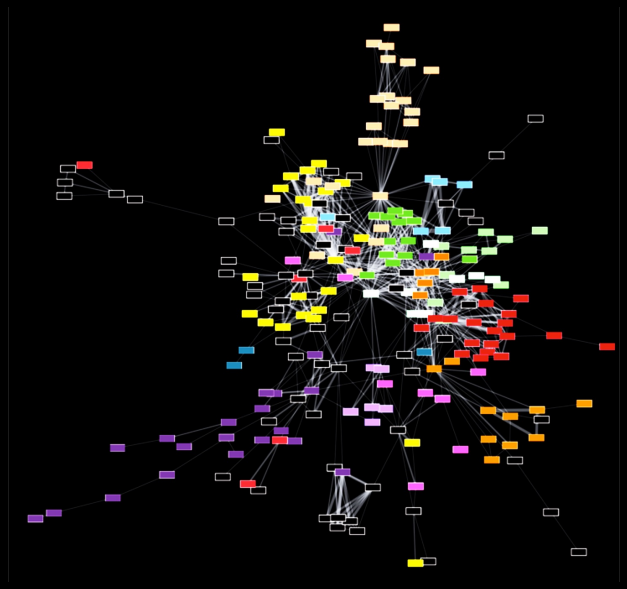
Activity volcano plot



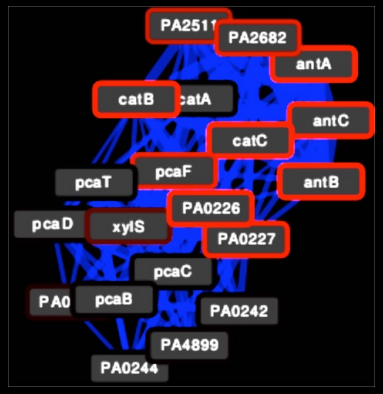
Pareto activity selection

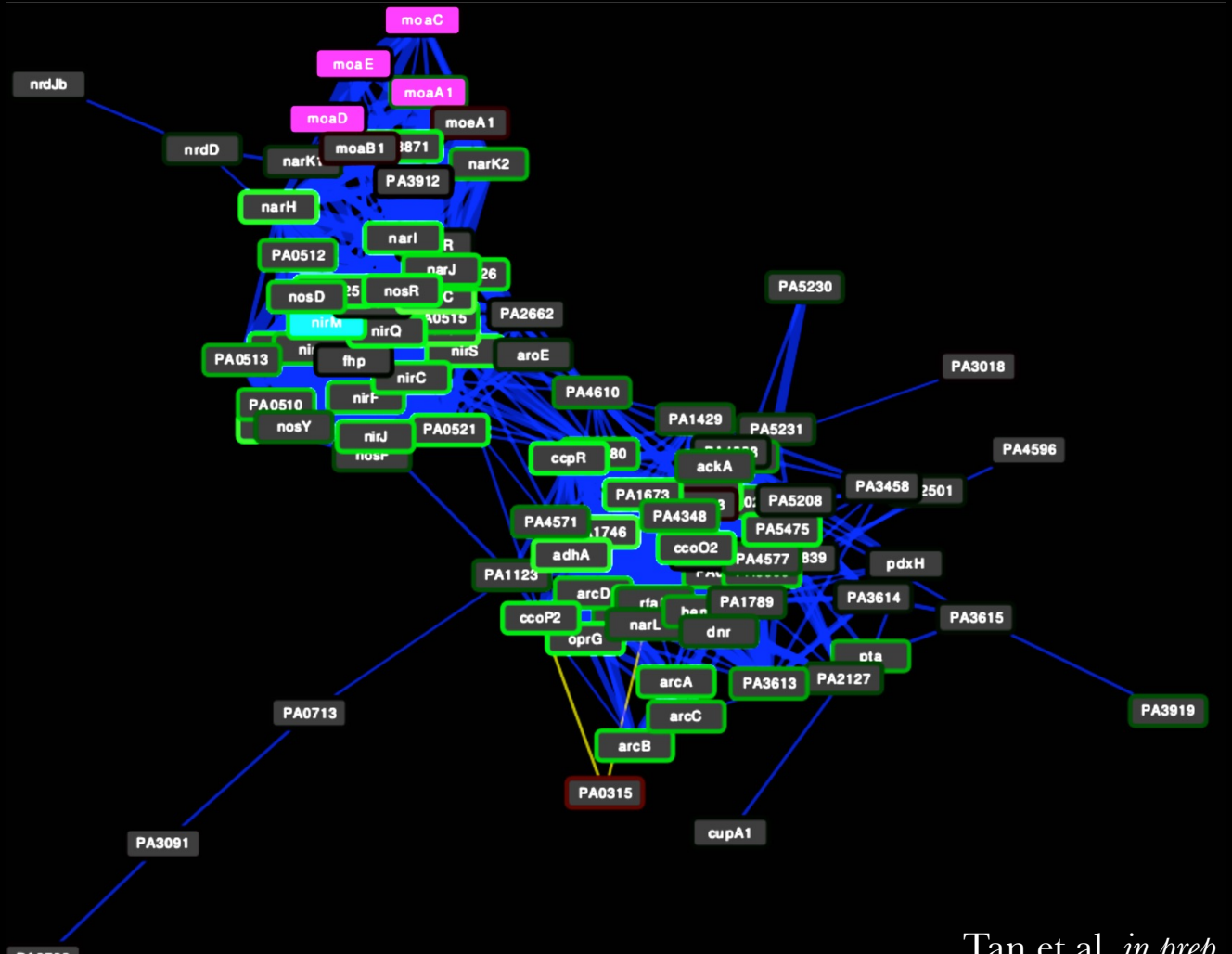


volcano + networks = pathway-style analysis



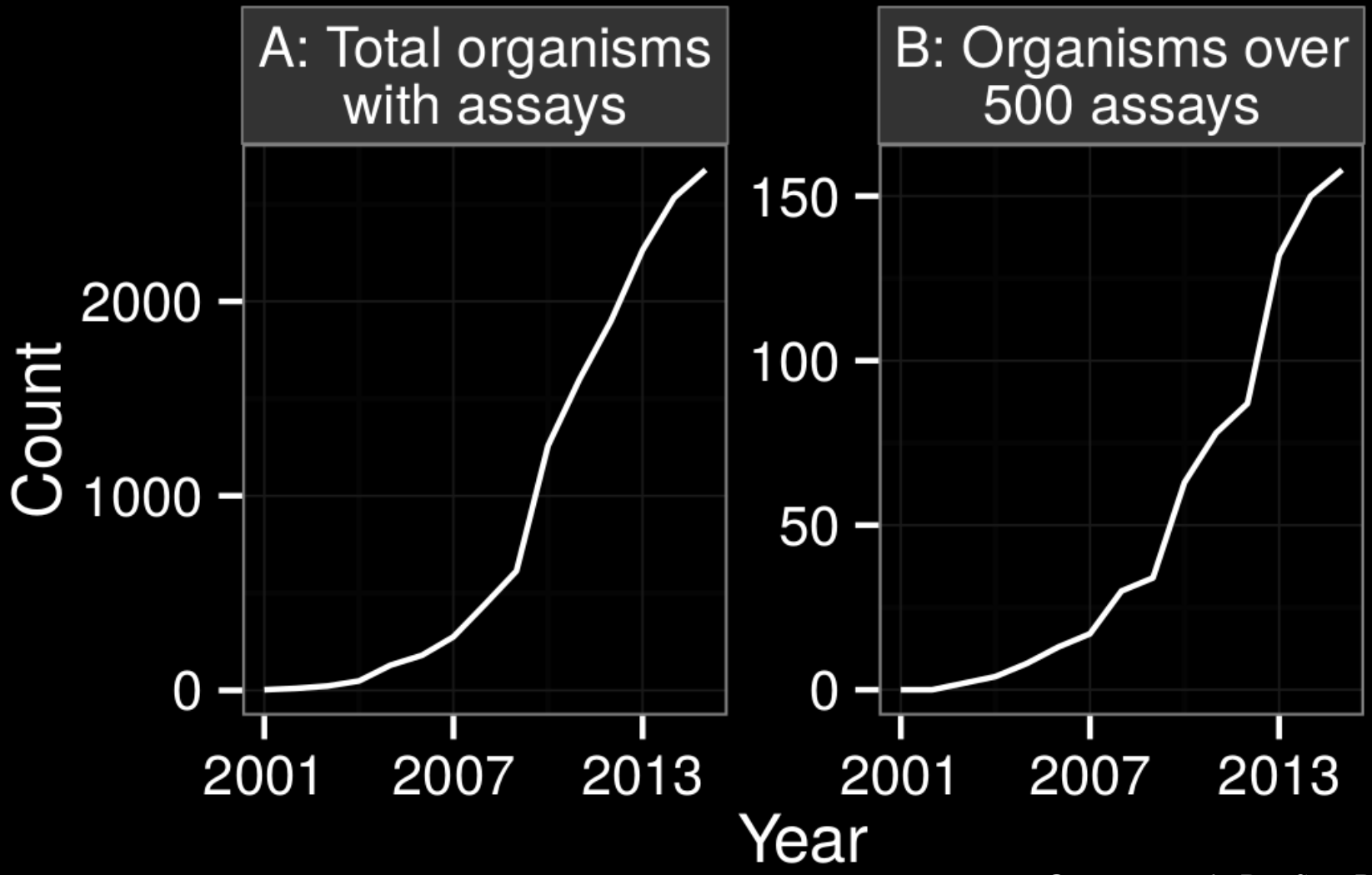
ADAGE-based pathway analysis reveals transcriptomic changes



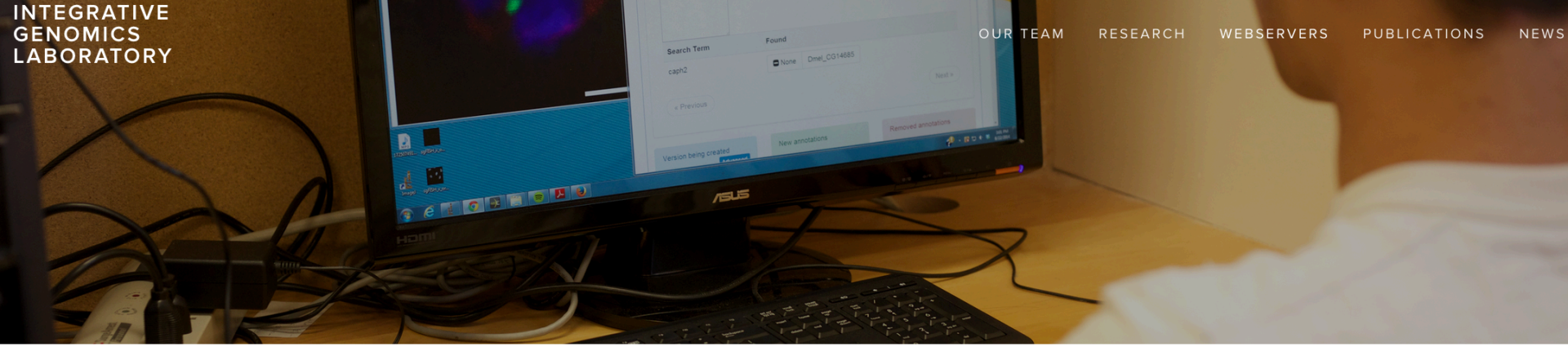


Tan et al. *in prep*.

Where do we have (enough) data?



Greene et al. *Pac Sym Bio*. 2016



As we develop new methods, we develop webservers to put these methods into the hands of biologists. Servers that we have developed or co-developed are listed below. If you want to get early access to our new servers, sign up as an early and test user.

EARLY ACCESS USER SIGNUP

Sign up here to be offered opportunities to test our webservers as we develop them. We may ask that you share your feedback about either specific or general features during development.

First Name | Last Name | Email Address | SIGN UP

ADAGE webserver coming soon!

http://www.greenelab.com/webservers

The NEW ENGLAND JOURNAL of MEDICINE

EDITORIALS

Data Sharing

Dan L. Longo, M.D., and Jeffrey M. Drazen, M.D.

Research Parasite Awards (The "Parasites")

Selection criteria for the work in question:

- Not involved the design of the experiments that generated the data.
- Published independently of the original investigators
- May have extended, replicated or disproved what the original investigators had posited.
- Provided source code and intermediate or final results in a manner that enhances reproducibility.

Research Parasite Awards (The "Parasites")

Additional selection criteria for the Junior Parasite award:

- The award is based on work described in a single manuscript.
- Must have published the work at the training stage of their career (postdoctoral, graduate, or undergraduate).
- Should not have been in an independent investigator position for more than 2 years.

Research Parasite Awards (The "Parasites")

Additional selection criteria for the Sustained Parasitism award:

- Must be an independent investigator in academia, industry or public sector.
- Based on three manuscripts.
- Must be last or corresponding author on each manuscript.
- At least a five-year period must have elapsed between the publication of the first manuscript and the final manuscript.

It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so.

- Attributed to Mark Twain

Greene Lab:

Jie Tan (Grad Student)
Gregory Way (Grad Student)
Brett Beaulieu-Jones (Grad Student)
René Zelaya (Programmer)
Matt Huyck (Programmer)
Dongbo Hu (Programmer)
Kathy Chen (Undergrad)
Mulin Xiong (Undergrad)
Tim Chang (Undergrad)

Collaborators:

Deb Hogan & Jack Hammond

Data:

All investigators who publicly release their gene expression data.

Images:

Artists who release their work under a Creative Commons license.

Funding:

Gordon and Betty Moore Foundation
National Science Foundation
Cystic Fibrosis Foundation
National Institutes of Health
American Cancer Society

Find us online:

http://www.greenelab.com
Twitter: @GreeneScientist

Greene Lab

Fork me on GitHub

Calvin and Hobbes. Bill Watterson