

Introduction to Independent Component Analysis (ICA) applied to transcriptomics data

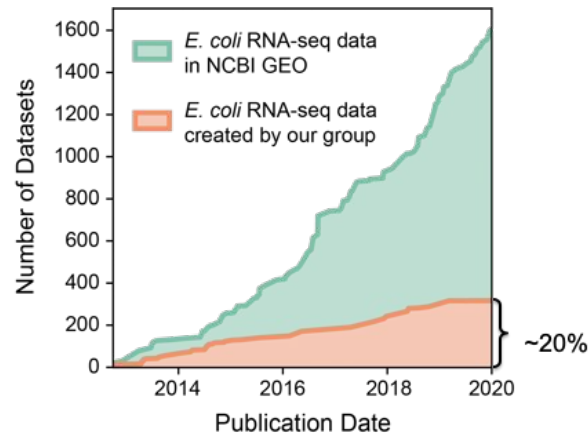
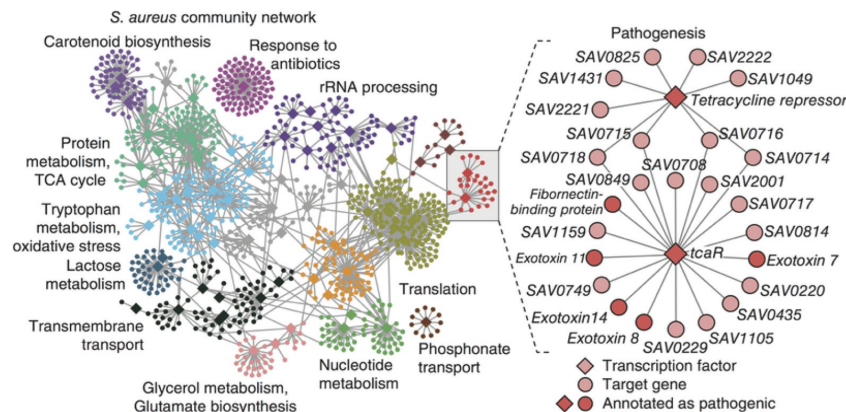
Anand Sastry
Saugat Poudel

Outline

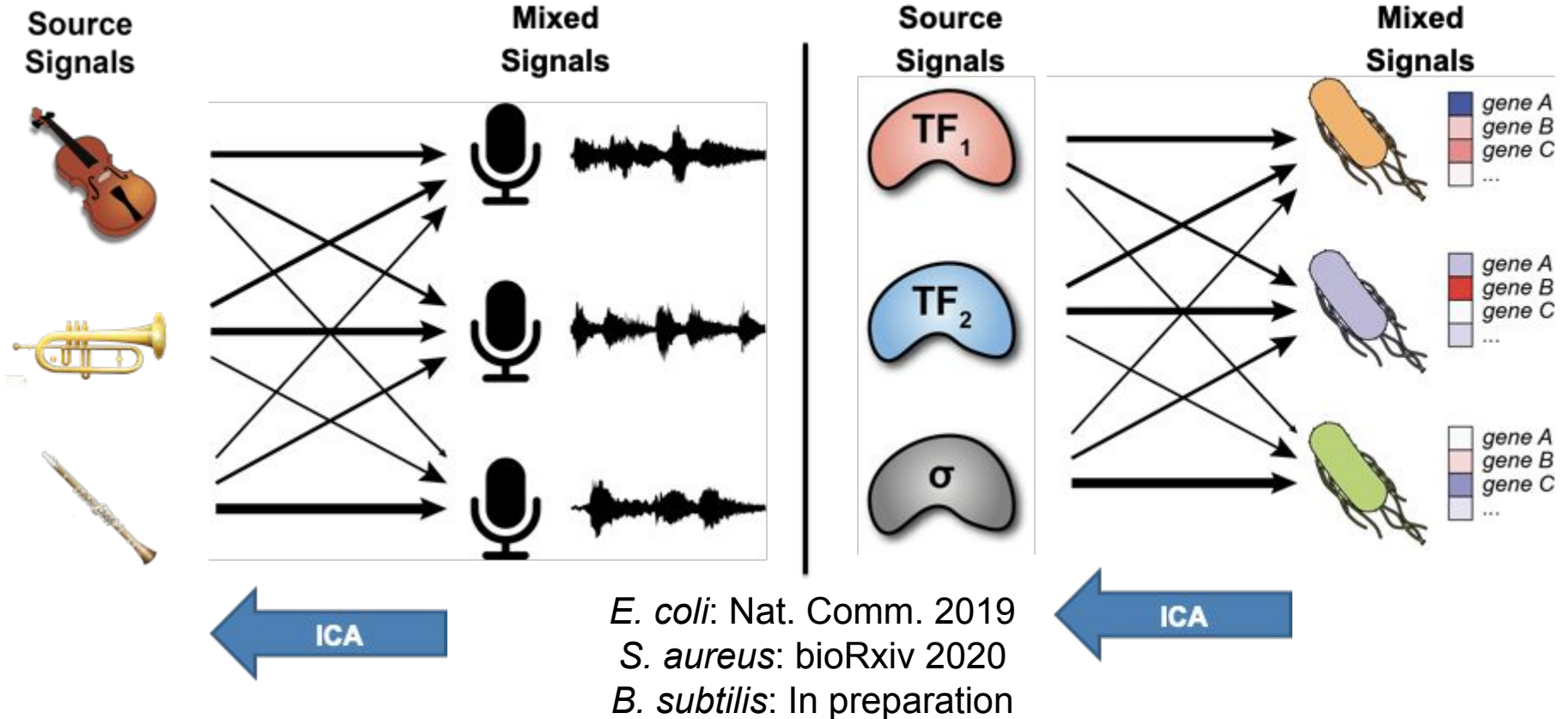
- Introduction to ICA and i-modulons
- Jupyter notebook live tutorials
 - a. Pre-processing data
 - b. Interpreting i-modulons
 - c. Applying i-modulons to new data
- Discussion

Goals of the method

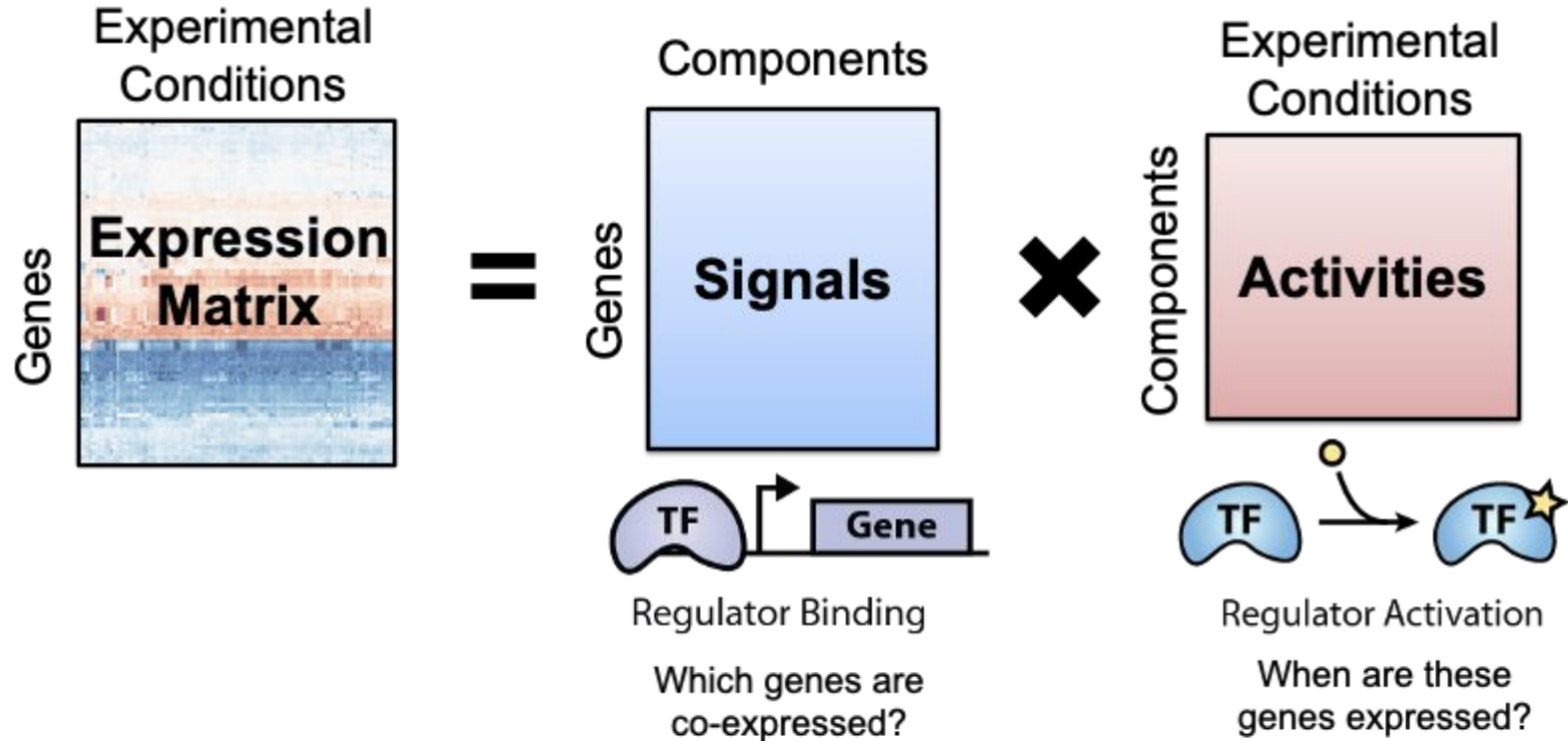
- Build a model of the Transcriptional Regulatory Network (TRN)
 - Allows us to better understand how an organism responds to changes in the environment
 - Makes it easier understand complex differential expression patterns
- Use RNA-seq data to build the model
 - Most non-model organisms lack antibodies against TF
 - But growing number of organisms have large RNA-seq data publically available



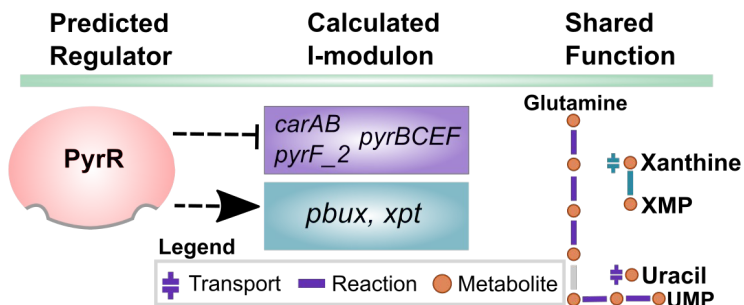
Independent Component Analysis



ICA of Gene Expression Data



Anatomy of an I-modulon

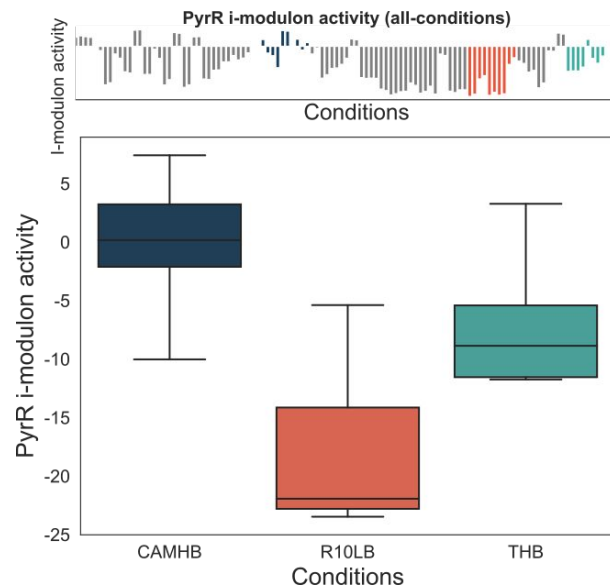


Independently modulated sets of genes (S-matrix)

Expression of genes in an i-modulon vary with each other and independent of all other genes not in the respective i-modulon.

Activity and gene expression (A-matrix)

The activity represents the strength of signal from the regulator governing the expression of genes in an i-modulon.

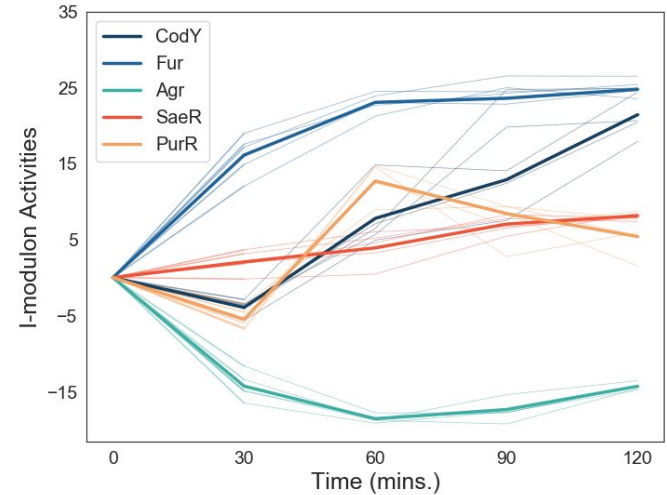


Putting it together

ICA provides both condition-invariant structure of TRN and its condition-specific activity

Applications of ICA

- Discover new interactions between different regulators
- Predict unknown regulons for regulators
- Simplify analysis of complex RNA-seq data
- Integrate it with M-model to understand interactions between metabolic and regulatory networks
- Detailed examples:
 - *E.coli*: <https://doi.org/10.1038/s41467-019-13483-w>
 - *S. aureus*: <https://doi.org/10.1101/2020.03.18.997296>



S. aureus i-modulon activities during growth in human blood

Additional resources

Tutorial Github: https://github.com/avsasttry/U01_ICA_tutorial

PRECISE-DB Github page: <https://github.com/SBRG/precise-db>

E. coli i-modulon paper: <https://www.nature.com/articles/s41467-019-13483-w>

S. aureus i-modulon paper:

<https://www.biorxiv.org/content/10.1101/2020.03.18.997296v1>

Email us at avsasttry@ucsd.edu and sapoudel@ucsd.edu for additional code.

Jupyter tutorial

Table of Contents

- [1 The *E. coli* Expression Dataset](#)
- [2 Experimental Metadata](#)
- [3 Quality Control](#)
- [4 Centering to reference condition](#)
- [5 Save centered dataset](#)
- [6 Run ICA](#)

Table of Contents

- ▼ [1 Load data](#)
 - [1.1 Load expression data and metadata](#)
 - [1.2 Load S and A matrices](#)
- ▼ [2 Exploring the S matrix](#)
 - [2.1 Example: Component #36](#)
 - [2.2 I-modulon thresholds](#)
 - [2.3 Load gene annotation](#)
 - [2.4 Load TRN](#)
 - [2.5 Regulator enrichments](#)
- [3 Naming I-modulons](#)
- ▼ [4 Exploring the A matrix](#)
 - [4.1 Fear-greed trade-off](#)
 - [4.2 The Fur i-modulons](#)
 - [4.3 Expanding to additional strains](#)