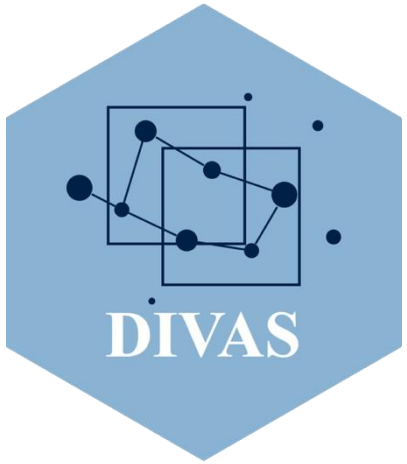
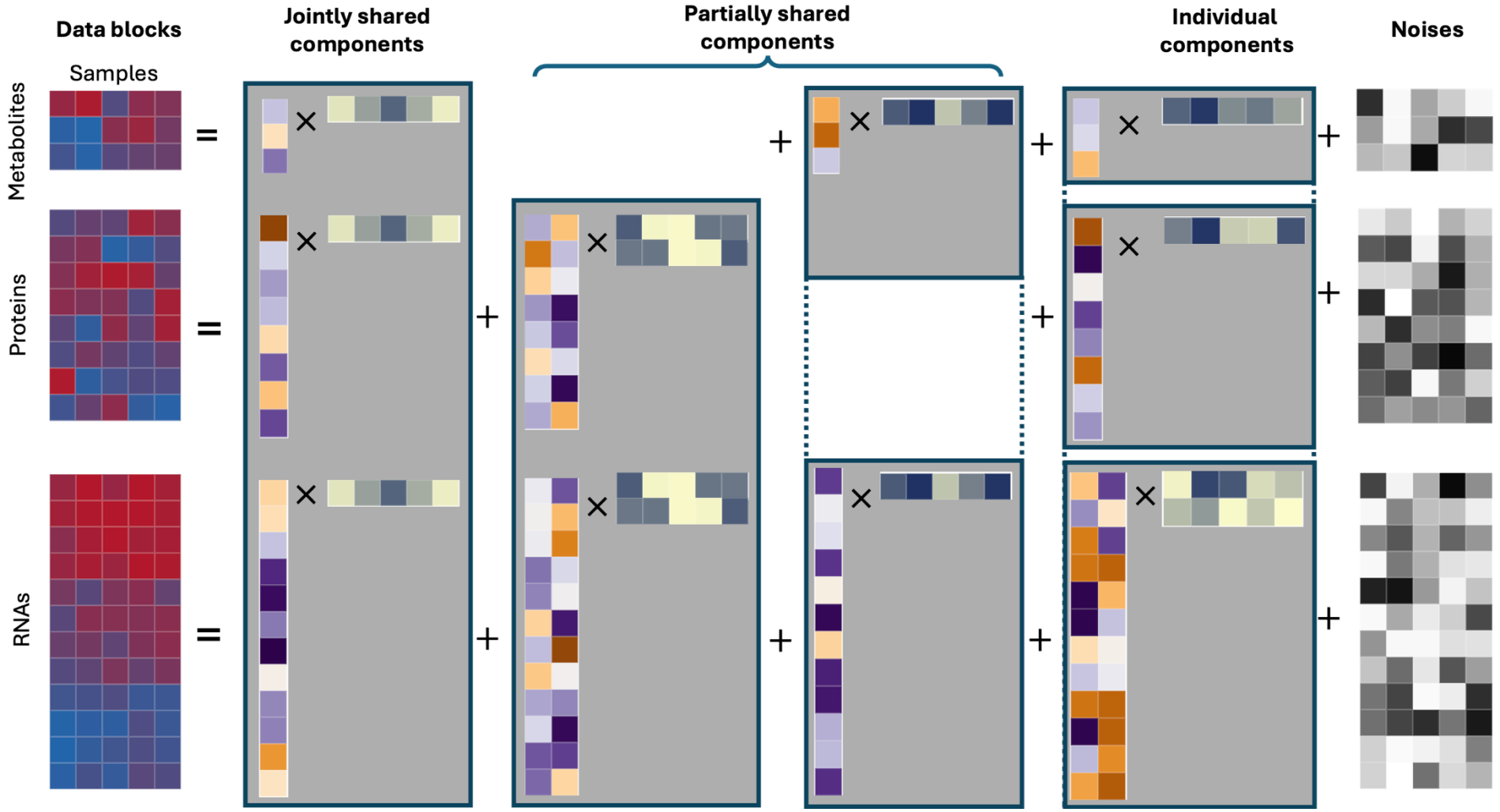
Integrating ≥ 3 omics using DIVAS

Dr Jiadong Mao
Melbourne Integrative Genomics (MIG)
University of Melbourne



Melbourne Integrative Genomics
Transforming Data Into Knowledge

Multiomics integration with matrix factorisation



Block PLS, joint NMF, ...

DIVAS

JIVE, AJIVE, MOFA, ...

Why care about what's **not** shared?

- Omic modalities can be complimentary
  - Eg H&E image + spatial omics
  - Isn't this the point of multiomics?

- Not all modalities are equally informative
  - Eg RNA + Protein (MS) + miRNA + phosphorylation
  - Particularly relevant for ≥3 omics

- Some pathways only involve a few, not all, modalities
  - Eg RNA + Protein + Metabolomics
  - Disentangle multiomic pathways

Lack of tools satisfying all these

Method	Jointly shared	Partially shared	Individual	Statistical inference	Package availability
RCCA/RGCCA	✓	✗	✗	✗	CRAN [12]
mixOmics	✓	✗	Limited[a]	✗	Bioconductor [13]
JIVE	✓	✗	✓	Limited[b]	CRAN [14]
AJIVE	✓	✗	✓	Limited[c]	GitHub [15]
MOFA/MOFA+	✓	Limited[d]	✓	✓[e]	Bioconductor [16]
SLIDE	✓	✓	✓	Limited[f]	N/A
DIVAS	✓	✓	✓	✓	GitHub [17]

Two shades of blue

Katie Hoadley, JD, Steve Marron (2023)

Photo courtesy of Lina-Sue Marron

Some Lê Cao lab members + Steve (2025)

Photo courtesy of Weichang Yu

DIVAS R package team

Rock climber
Kim-Anh Lê Cao

Caninophile
Yinuo Sun

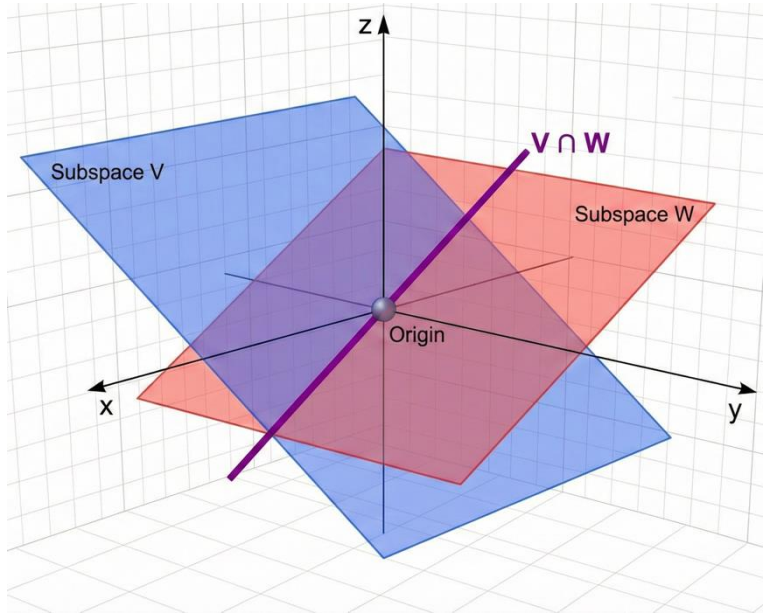
Evelyn’s second violin
Jiadong Mao

How does DIVAS work

Difficulties:

- How to identify partially shared variations
  - How many components?

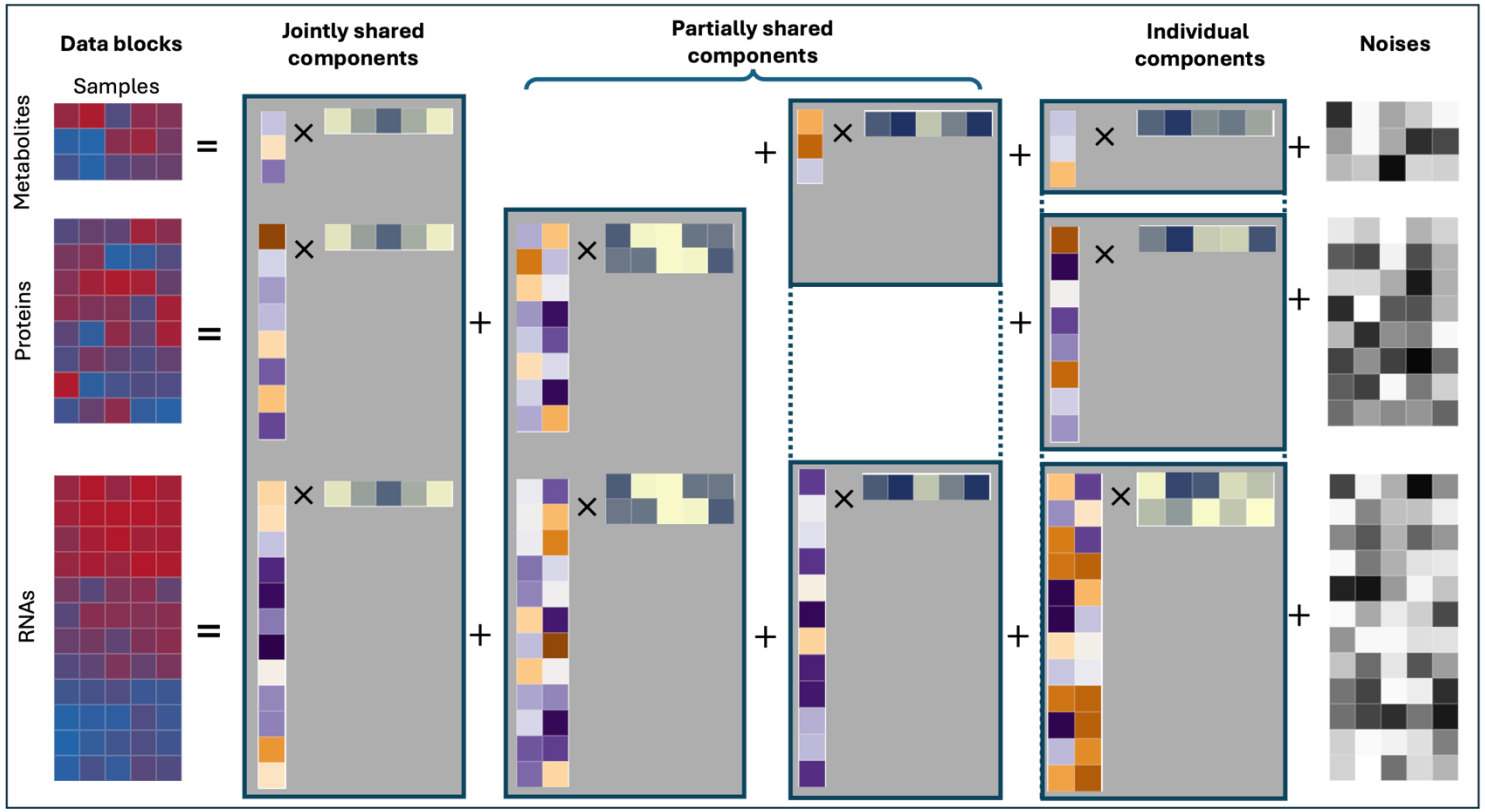
Intersection of two linear spaces

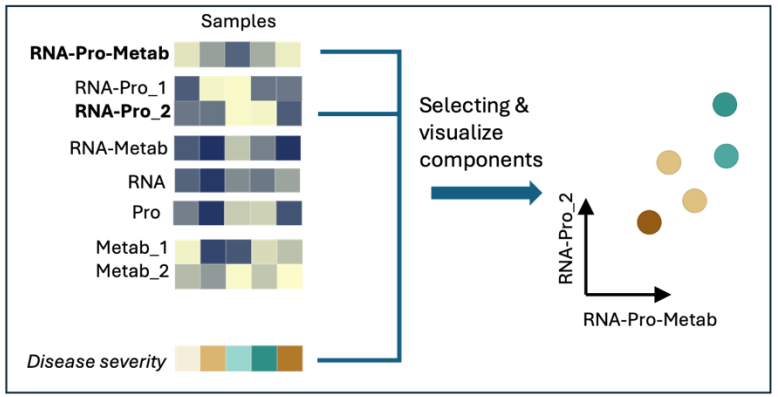


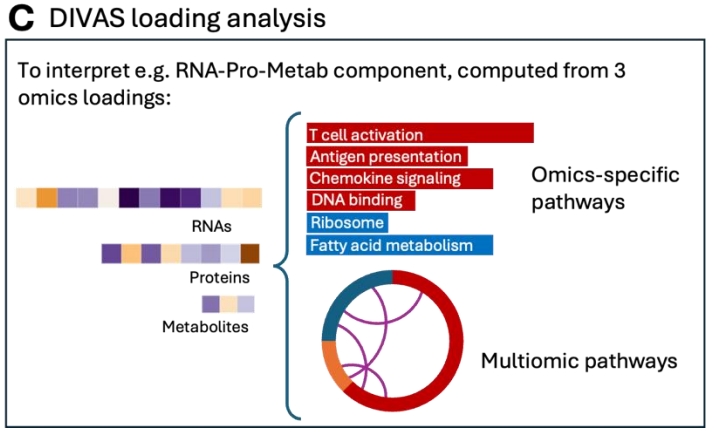
- DIVAS: Data Integration Via Analysis of Subspaces
- Each input data block defines a linear space
- **Goal**: find intersections of these omics spaces
- Define a linear space, you need vectors
  - These are score vectors of DIVAS components
- DIVAS main steps:
  - Denoise input data blocks
  - Exhaustive search for all vectors: from joint to partial to individual
  - Reconstruct loadings

Result: an algorithm that's completely data-driven, **tuning-free**

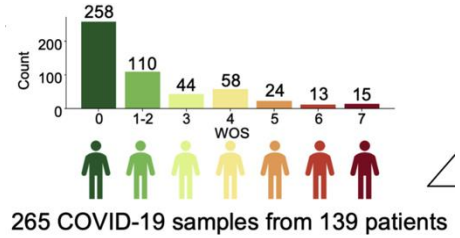
- One component = one score vec & modality specific loading vectors
- Downstream analyses using scores and loadings



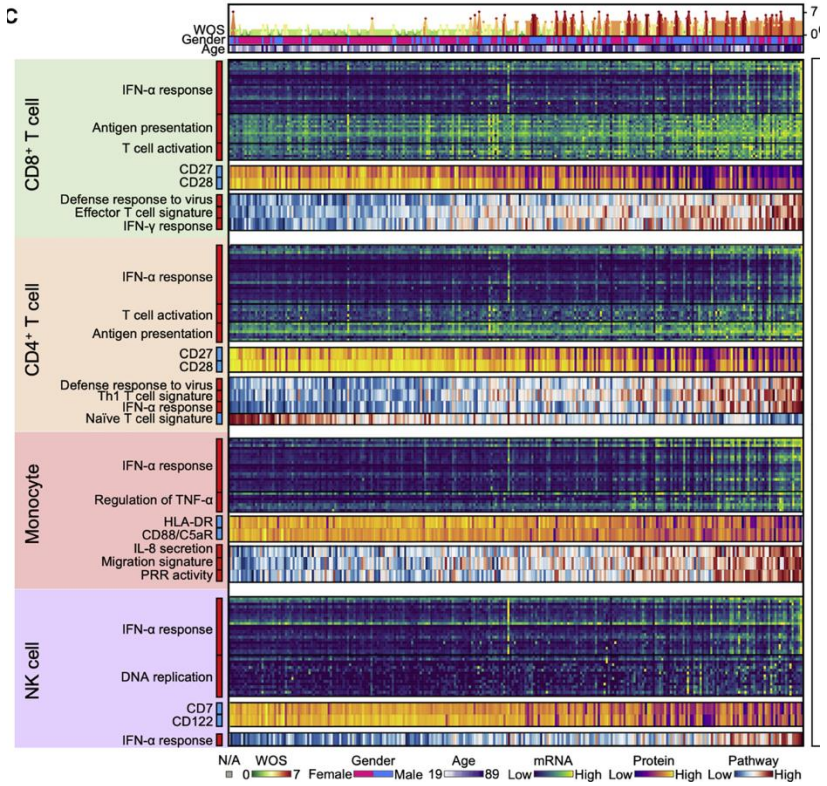




COVID multiomics

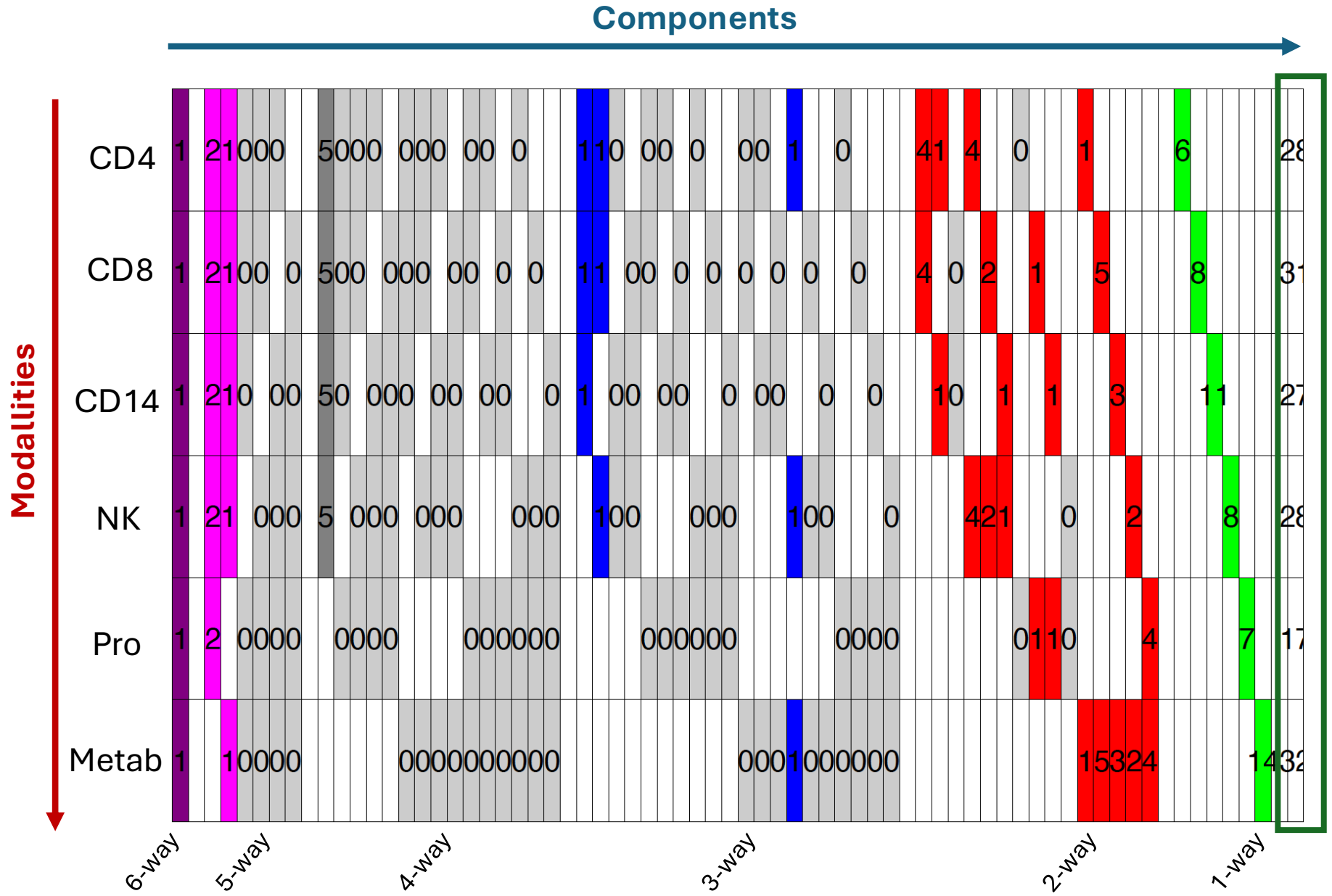


- 120 patient samples, from mild to severe
- Each has scRNA-seq + bulk Prot + bulk Metab
- Most important patient phenotype: severity score 1, …, 7
- What's done by Su et al. (2020):
  - Cell type annotation
  - See how proportions change along severity
  - Separate analysis of individual modalities

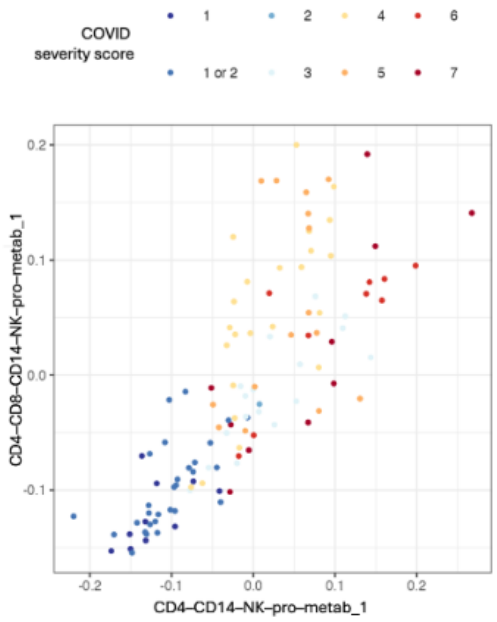


Rank breakdown

- What we do:
  - Pseudo-bulk cell types
  - Treat gene expression of each cell type as a block
  - Run DIVAS

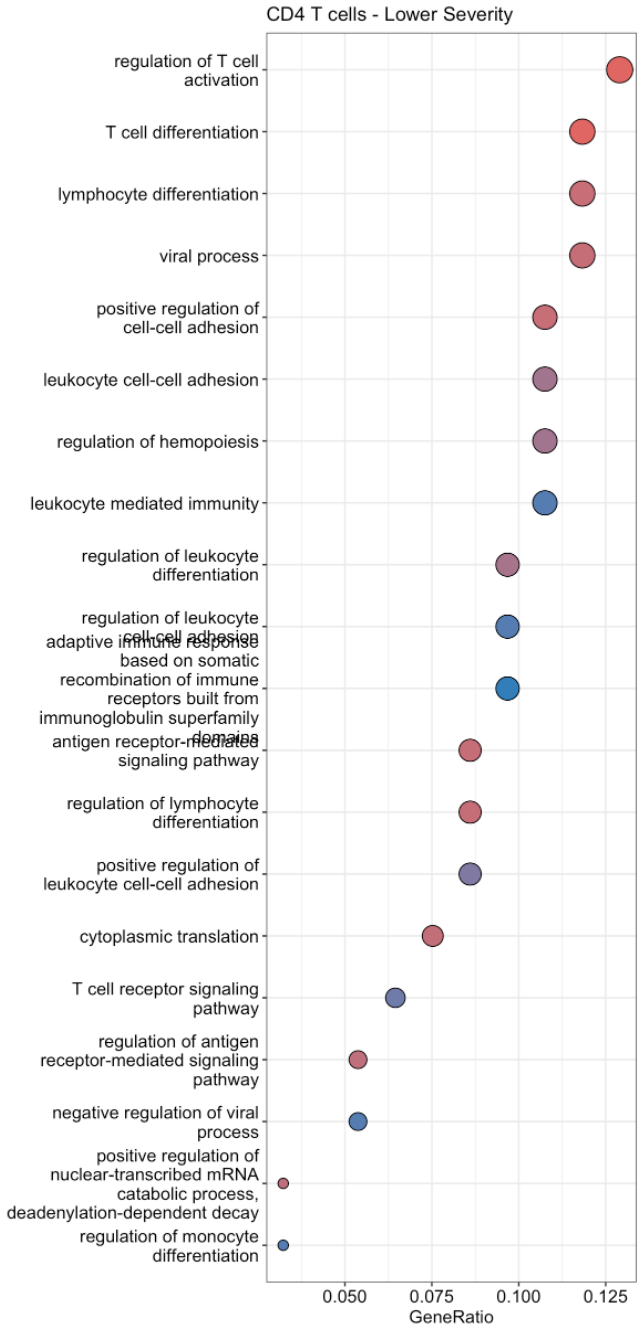


Focusing on one component



- 5-way: CD4–CD8–CD14–NK–Pro–Metab
  - Most highly correlated with severity

- Look at CD4 gene expression loadings
  - **Severe** COVID: hyperactive, dysregulated T cell differentiation and stress response
  - **Mild** COVID: coordinated 'normal' T cell activities



How about monocyte

- Milder COVID: standard antigen-presenting stuff
- Severe COVID: less antigen-presenting but more hypermetabolic and inflammatory cell states
- Ah but we do have metabolomics!

Key Mechanistic Insight

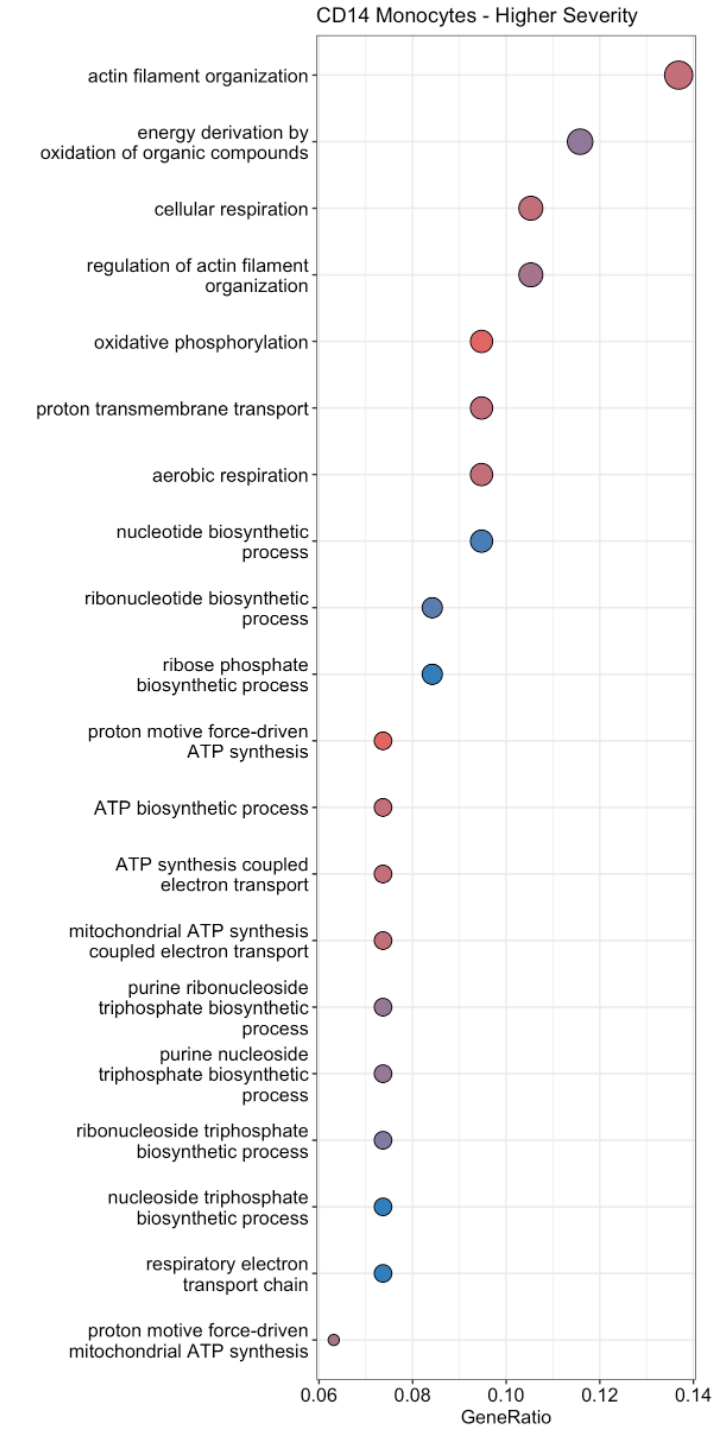
The **metabolite data bridges transcriptomics to pathophysiology**: the carnitine accumulation and altered lipid metabolism provide direct biochemical evidence that the "metabolic crisis" you saw in monocyte gene expression translates to actual mitochondrial dysfunction. The kynurenine elevation explains how hyperinflammation paradoxically leads to immunosuppression - this metabolite directly inhibits T cell function and promotes regulatory T cells.

This multi-omics integration suggests therapeutic targets should focus on:

1. **Metabolic support** (not just anti-inflammatory approaches)
2. **Kynurenine pathway modulation**
3. **Mitochondrial protection** in myeloid cells

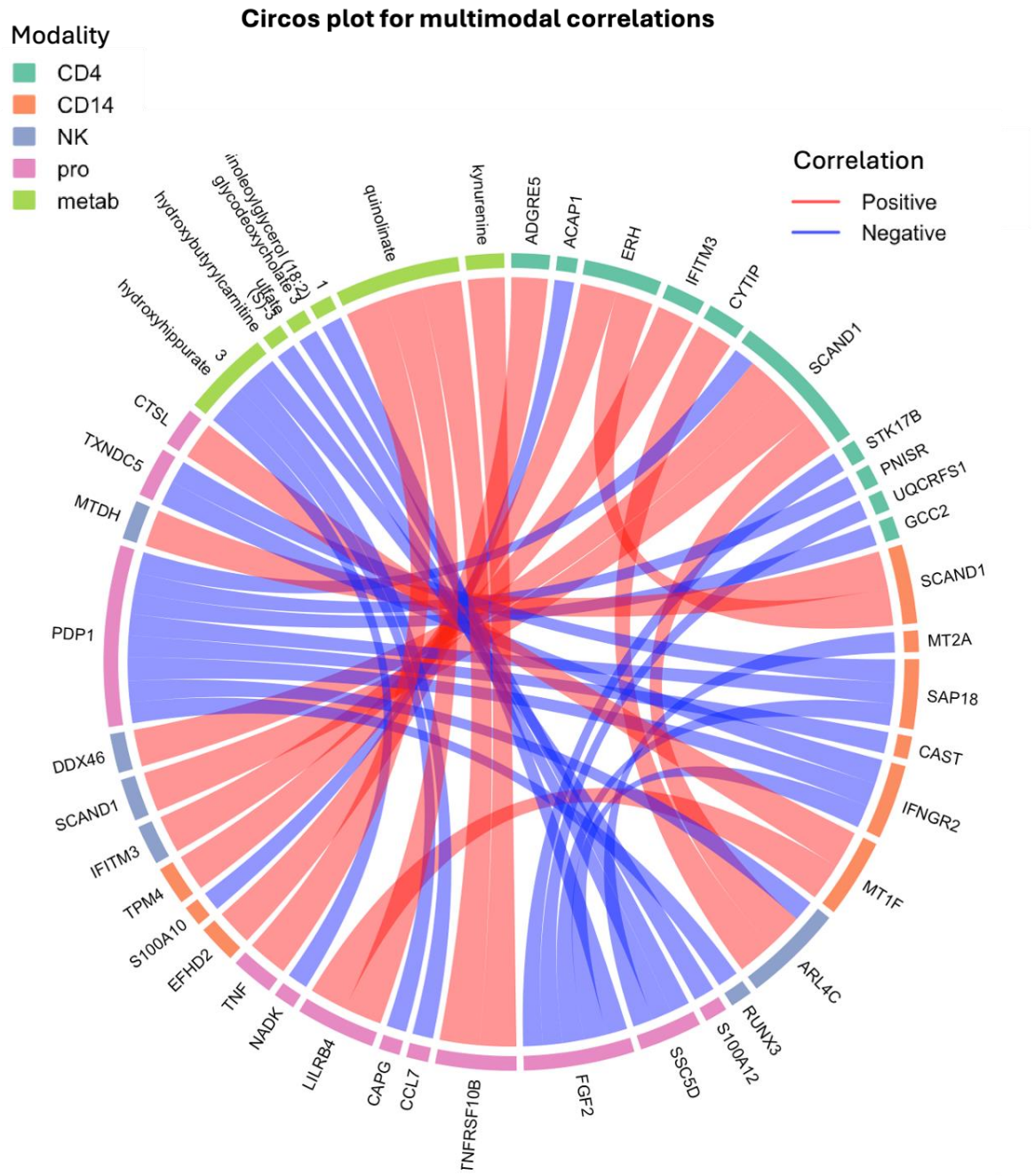
Your DIVAS component appears to have captured a fundamental axis of COVID-19 pathophysiology - the transition from coordinated immunity to metabolically-driven immunopathology.

Source: Claude Sonnet 4.5



Multiomics pathways

- Select features with top loadings
- Calculate their correlations
- Viz in circos plot



Summary

- Partially shared and individual variations also important
- DIVAS provides a statically rigorous way of finding them all

- Price to pay? Longer computation time, if many modalities and samples (pseudo-bulking)
- But number of features not a problem, eg ATAC
- Types of data also flexible, eg continuous, 0-1, SNPs (0,1,2)

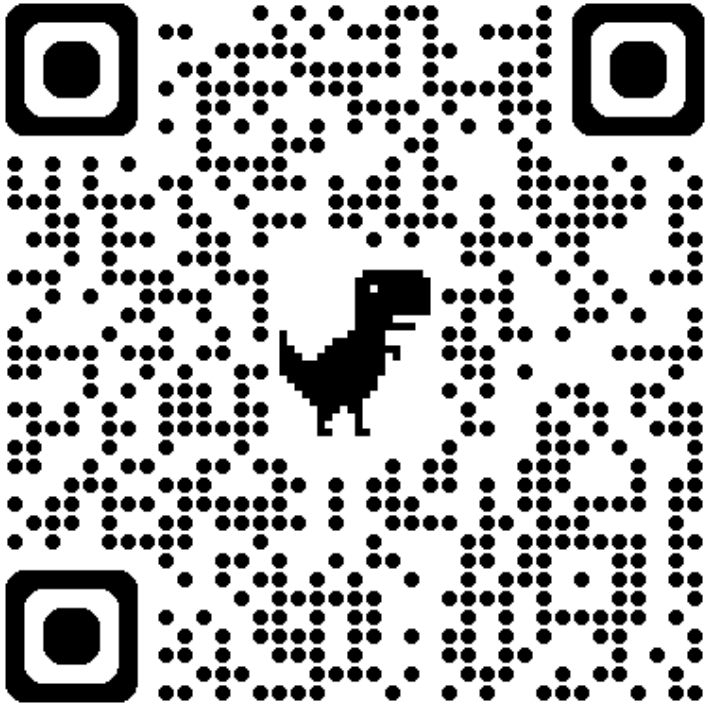
- Great as first run for a systematic search for interesting biology
- Or as a final run to select features

- Microbiome + host omics
- Image + Omics
- Clinical variables + Omics
- Time omics

- Variable selection

Sove your integration problem with DIVAS

COVID-19 Vignette



jiadong.mao@unimelb.edu.au

References

Prothero, J., ..., Marron, J. S. (2024). Data integration via analysis of subspaces (DIVAS). *Test*.

Sun, Y., Marron, J. S., Lê Cao, K.-A., Mao, J. (In preparation). DIVAS: an R package for identifying shared and individual variations of multiomics data.