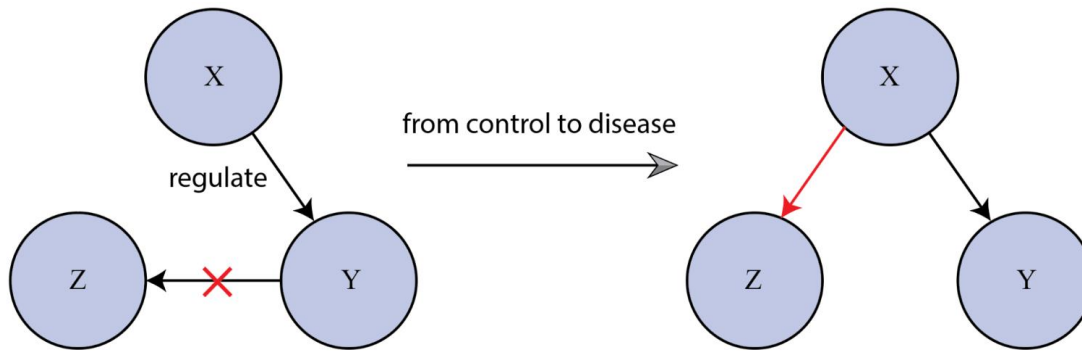


iHerd: an integrative hierarchical graph
representation learning framework to quantify
network changes and prioritize risk genes in disease

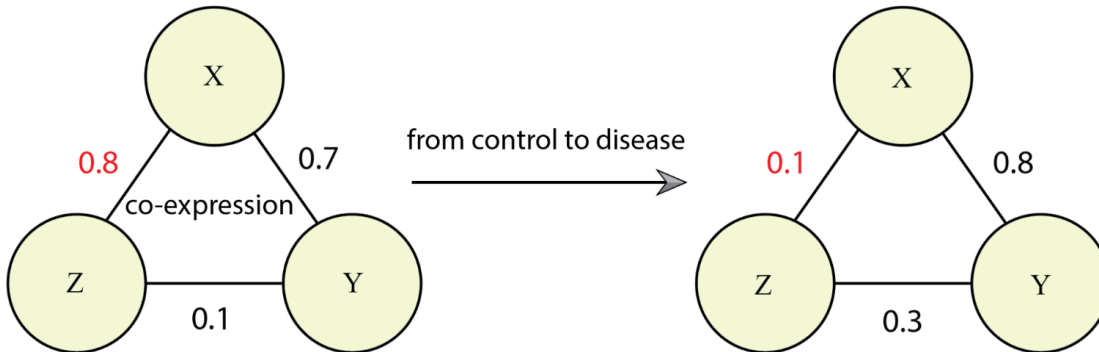
SCORCH Fall meeting
Ziheng Duan
12/05/2023

Background



Gene Regulatory Network (GRN)

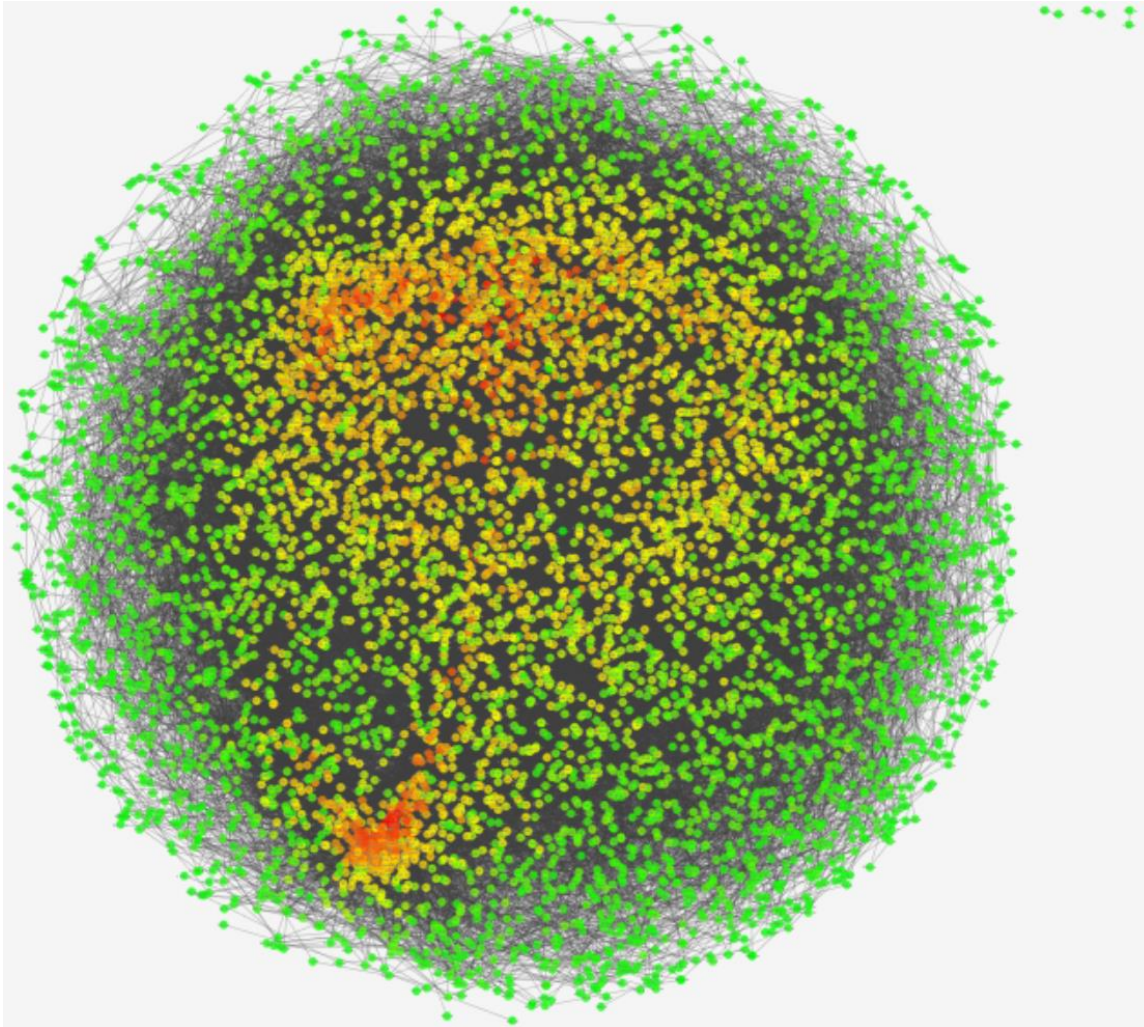
Directed Unweighted Graph



Gene Co-expression Network (GCN)

Undirected Weighted Graph

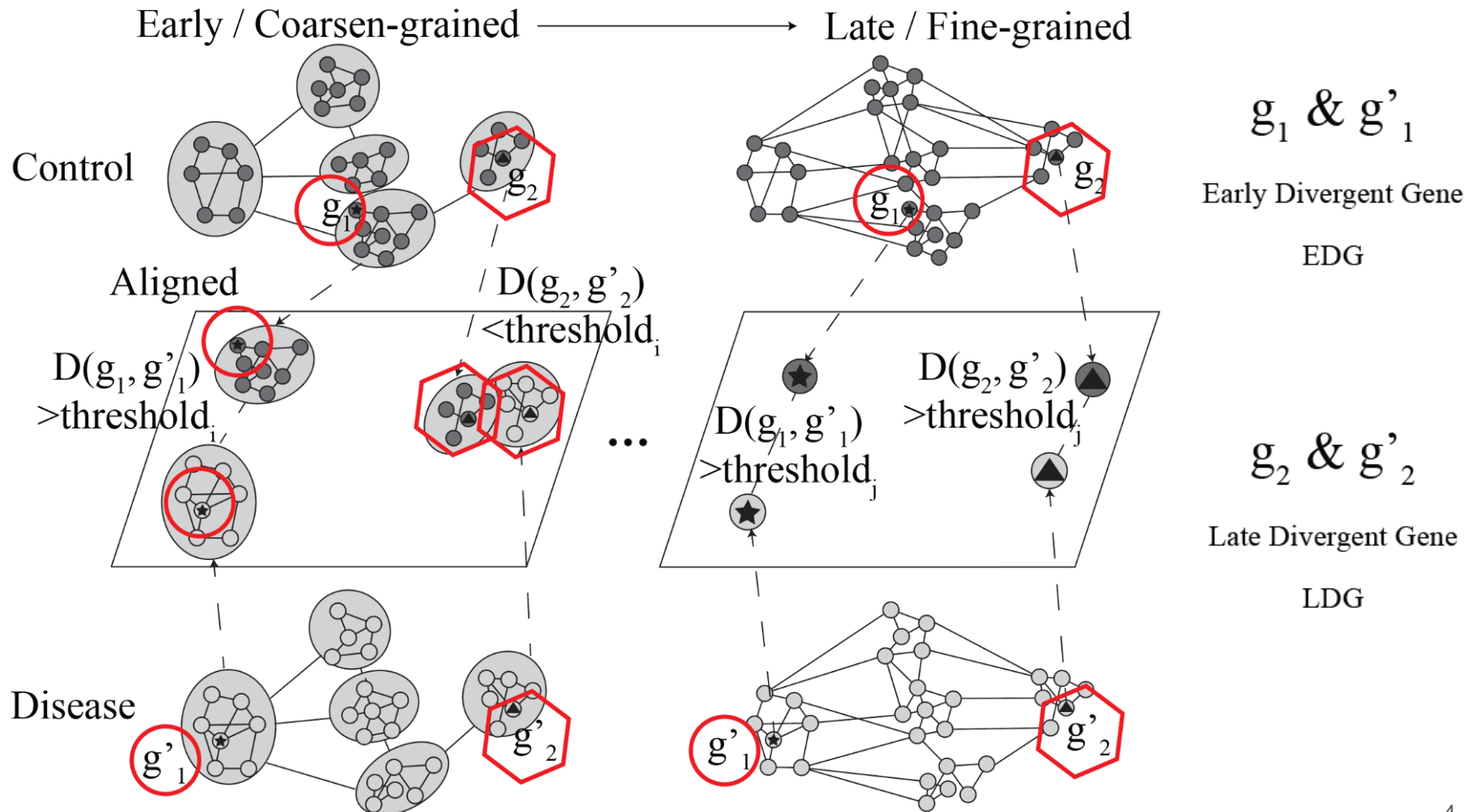
Background



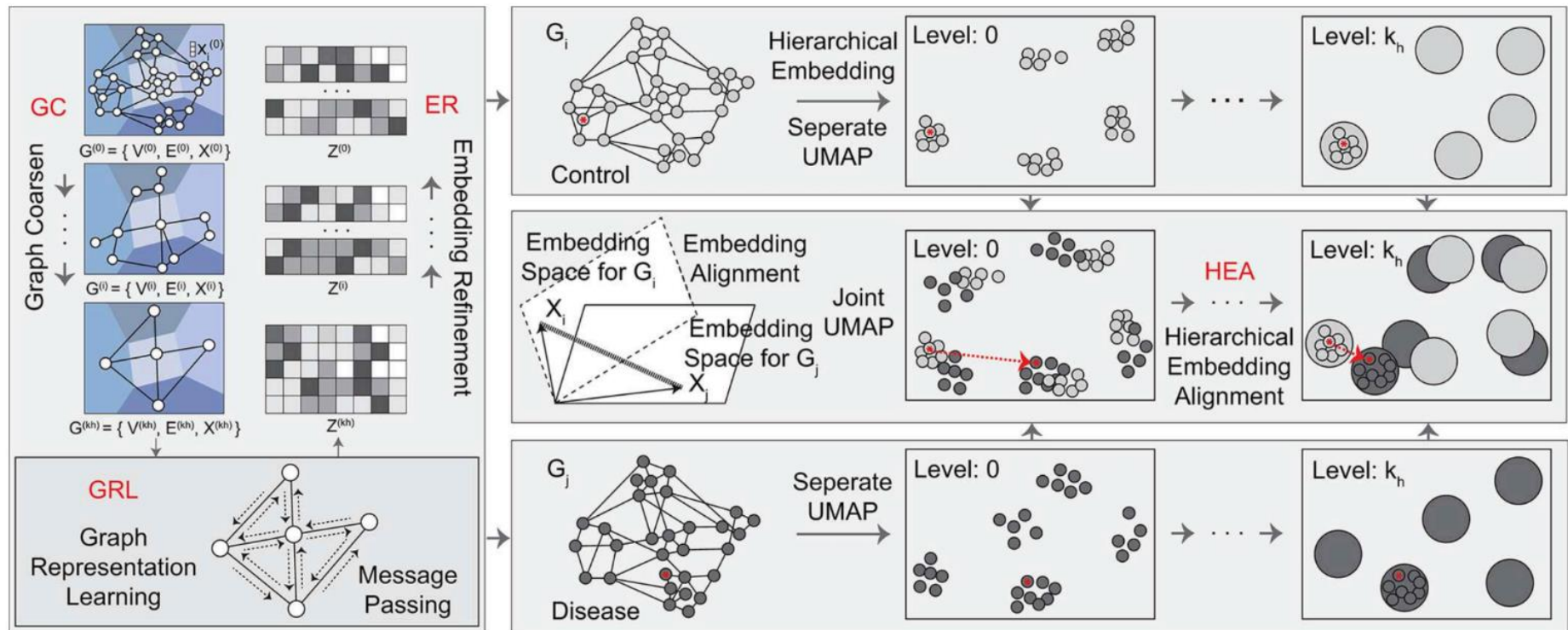
GCN with
7221 genes

Credit: S. Mohammad H. Oloomi from
https://en.wikipedia.org/wiki/Gene_co-expression_network

Background

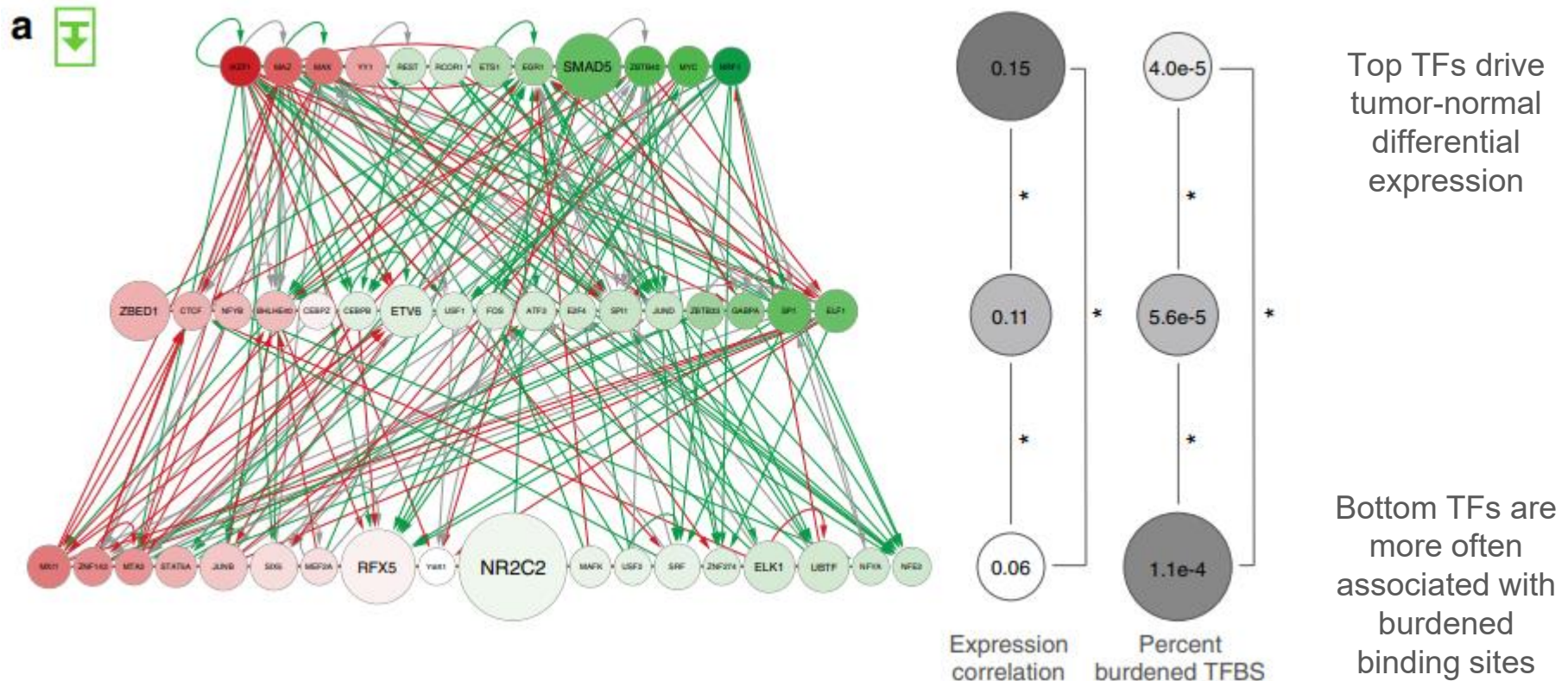


iHerd Framework

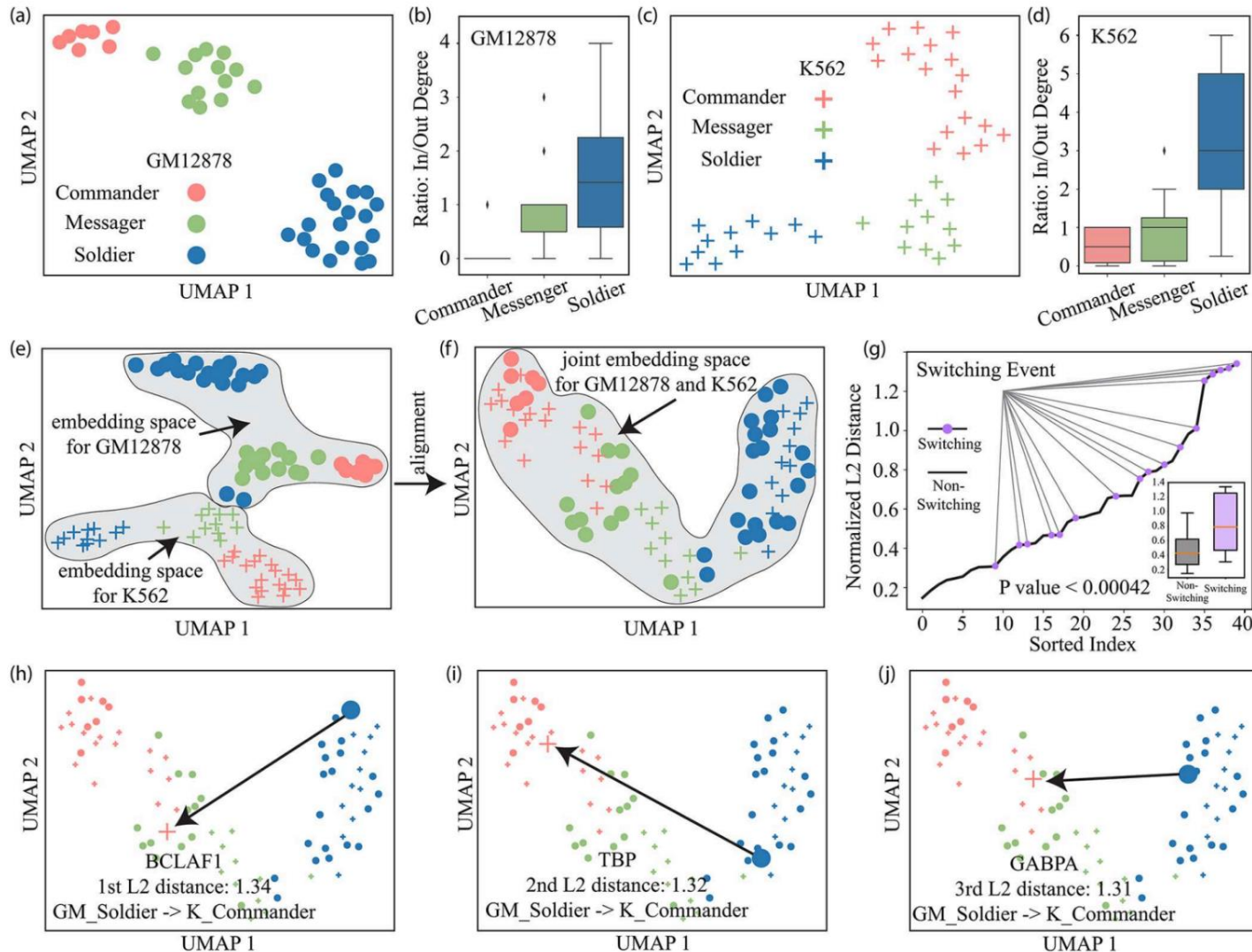


Result 1: Hierarchy in Transcript Factor (TF) GRNs

GRN for K562 and GM12878 cells



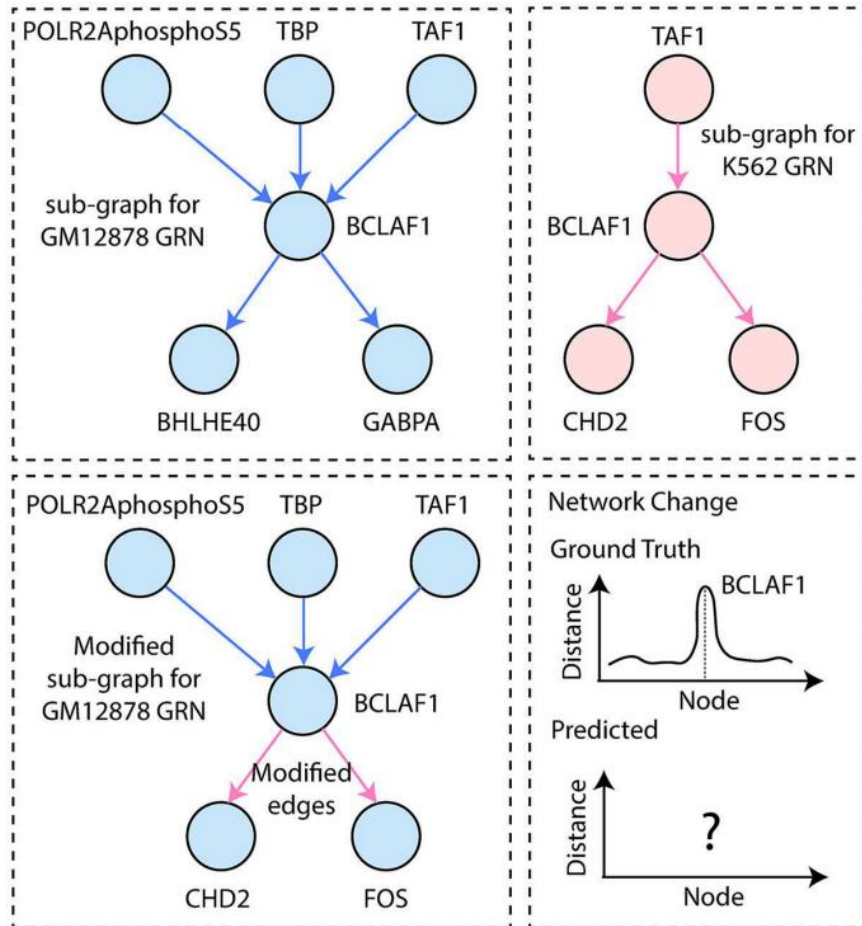
Result 1: Hierarchy in Transcript Factor (TF) GRNs



iHerd
recovers
the
hierarchy
change
within the
TF GRNs

Result 2: Simulated GRN Experiments

Data construction



Gene BCLAF1 should show the largest change from raw GRN (top left) to modified GRN (bottom left)

Result 2: Simulated GRN Experiments

iHerd outperforms four baseline methods in the TF prioritization task

Table 1. Summary of baseline methods.

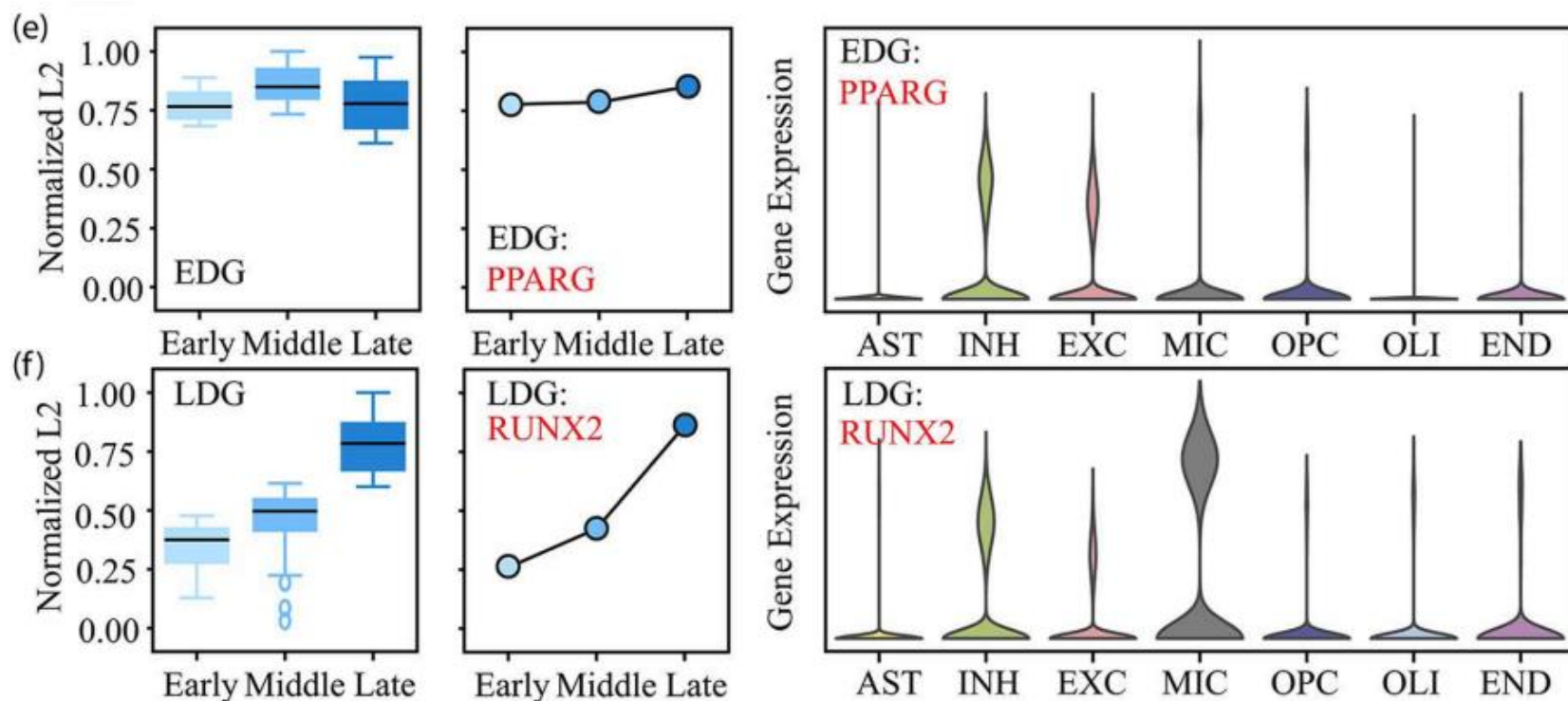
Baseline Methods	Property	Limitations
Node Degree Change	Linear	Fails when in/out degrees remain constant despite edge changes
Clustering Coefficient Change	Non-linear	Fails when clustering coefficient remains constant despite edge changes
PageRank Change	Non-linear	Struggles to identify real node changes due to global scope.
Contrastive Learning	Non-linear	Fails to detect changes when node's cluster membership remains the same.

Table 2. TF prioritization benchmarking.

Methods	R@1	R@2	R@3	R@4	R@5
Node Degree	0.068	0.091	0.114	0.159	0.182
Clustering Coefficient	0.045	0.136	0.205	0.341	0.432
PageRank	0.205	0.295	0.364	0.409	0.432
Contrastive Learning	0.008	0.030	0.068	0.083	0.121
iHerd	0.250	0.318	0.364	0.432	0.545

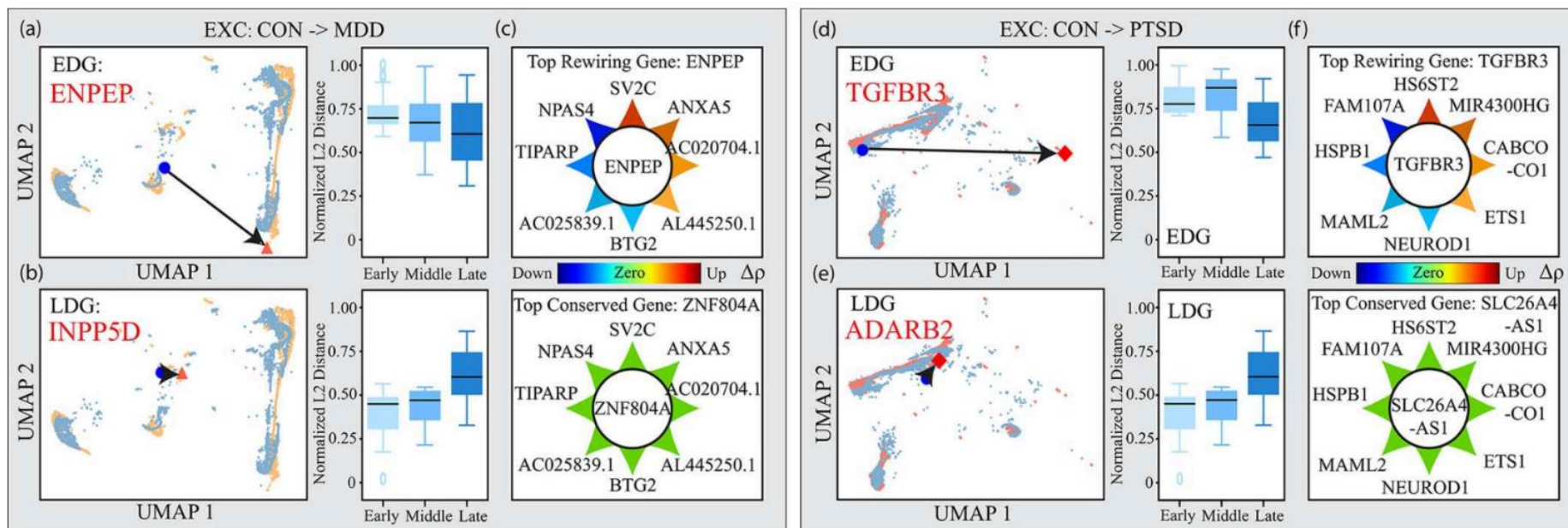
Result 3: Identify LDG/EDG for GCN

Across cell types: neuron vs non-neuron



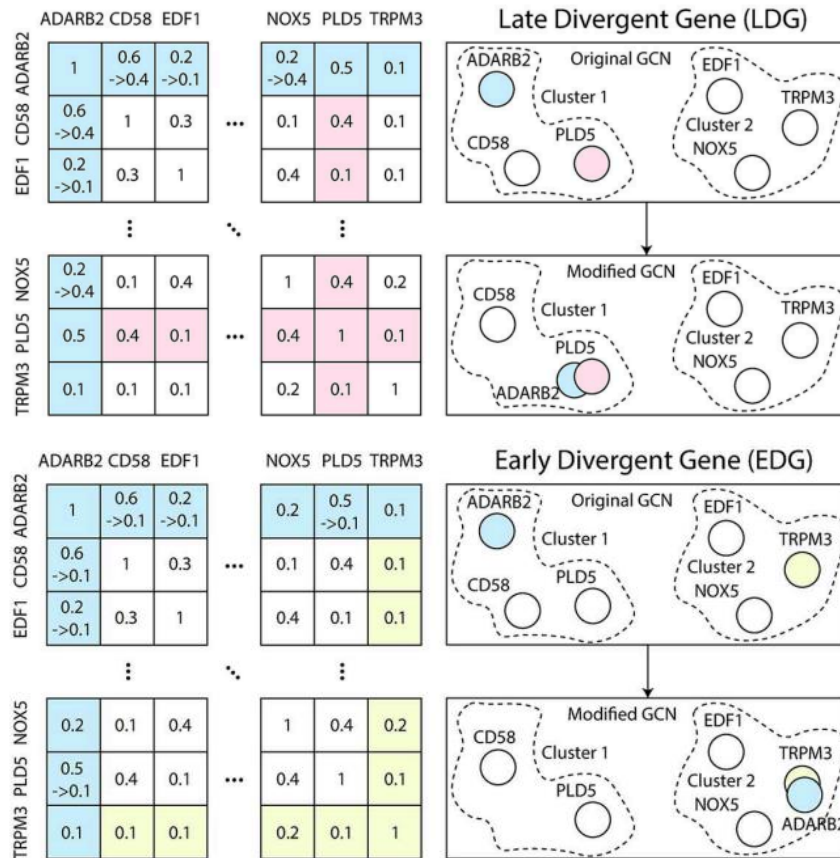
Result 3: Identify LDG/EDG for GCN

Across conditions: control -> MDD, control -> PTSD



Result 4: Simulated GCN Experiments

Data construction



LDGs remain the same clusters/pathways

EDGs change the clusters/pathways

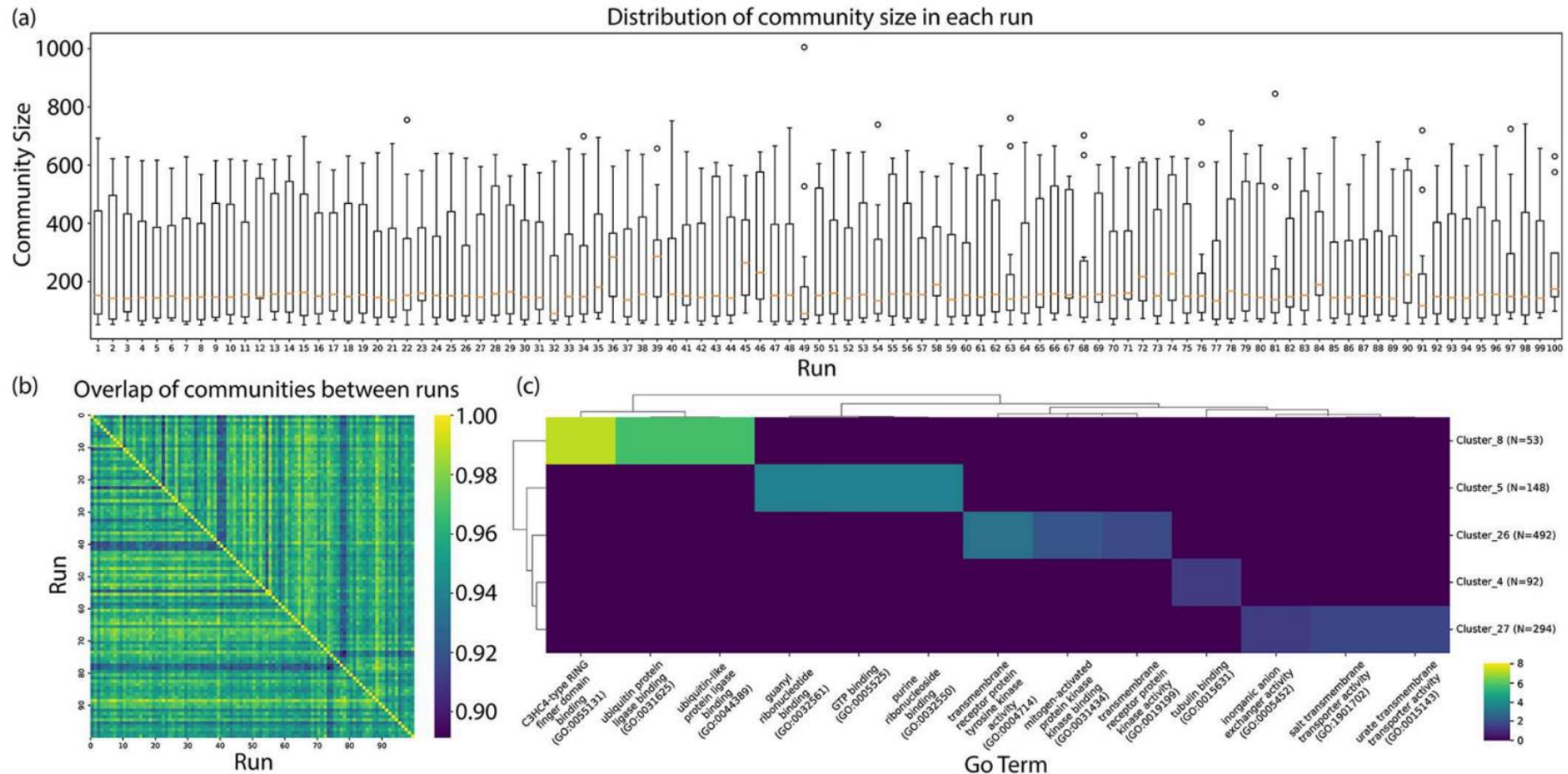
Result 4: Simulated GCN Experiments

iHerd outperforms four baseline methods in the LDG/EDG distinction task

Table 4. EDG vs LDG benchmarking.

Methods	Accuracy
Node Degree	0.672
Clustering Coefficient	0.743
PageRank	0.529
Contrastive Learning	0.856
iHerd	0.918

Result 5: Robustness & Biological Relevance



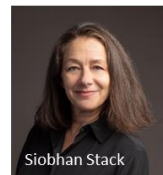
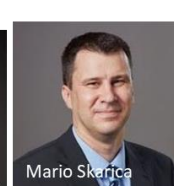
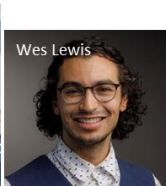
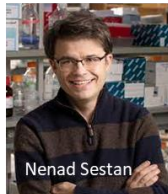
Conclusion

- iHerd can faithfully **highlight** the **changed TFs** and the **divergent genes** in each respective network across different cell types and conditions.
- iHerd is a flexible framework and it can combine different graph learning methods (focus on structural information or local neighborhood) depending on the input network. Meanwhile, how to effectively **combine the node features** into this framework would be a promising direction of future exploration.
- We have implemented iHerd as an open-source software that is freely downloadable to the public. With the exponential growth of disease data, iHerd can be a useful tool for the community to **prioritize genes** in diseases and **quantify their changes** at different stages.

Acknowledgement



M-SCORCH & Y-SCORCH group



The Zhang Lab members



Jing Zhang, Ph.D.
Assistant Professor,
Department of
Computer Science, UCI



Thanks for listening!