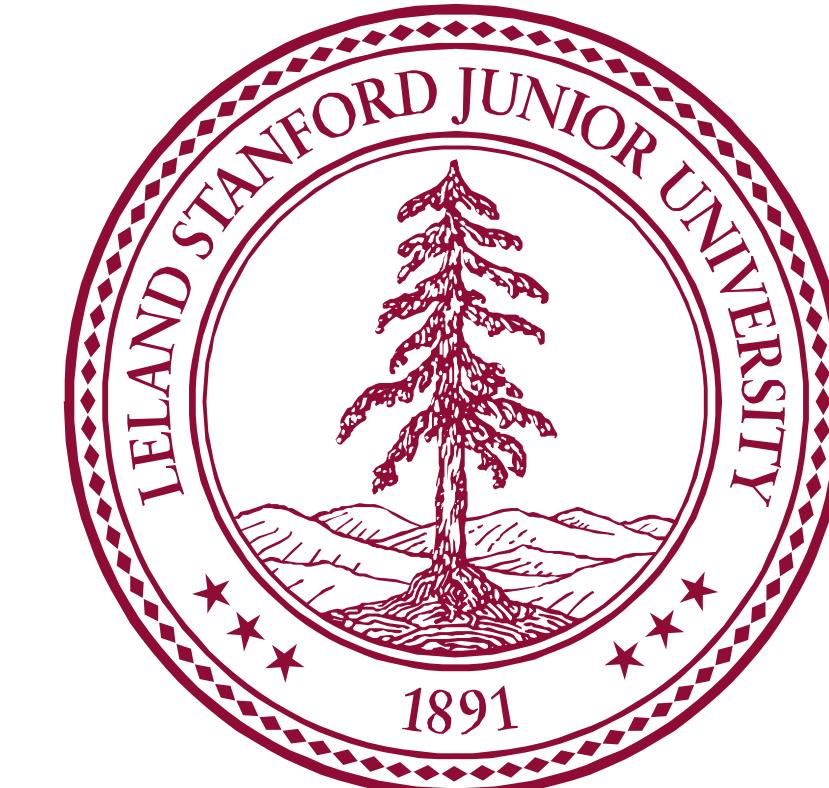


Identification of Putative Transcriptional Drivers in Asthma through Hive Plot Analysis of Gene Regulatory Networks

Authors: Brandon Guo, Abhinav Kaushik, Kari Nadeau
 Sean N. Parker Center for Allergy and Asthma Research, Stanford University School of Medicine, Stanford, California



Introduction

Asthma is a chronic lung condition, which often co-occurs with food allergies and eczema.

Canonical genetic analysis involves reconstruction of gene regulatory networks (GRNs) using graph theory. GRNs are central to cellular processes and represent directed regulatory relationships between a transcript and a regulator (transcription factor).

We propose use of hive plots that invoke topological measures to reveal otherwise occluded relationships. Analysis of hive plots reduce GRNs to a topologically relevant sub-network. Reduction of GRNs allows for more classical techniques like hypothesis testing.

We highlight the dysregulation of the MYC-EGR2 relationship in asthma through regression tests.

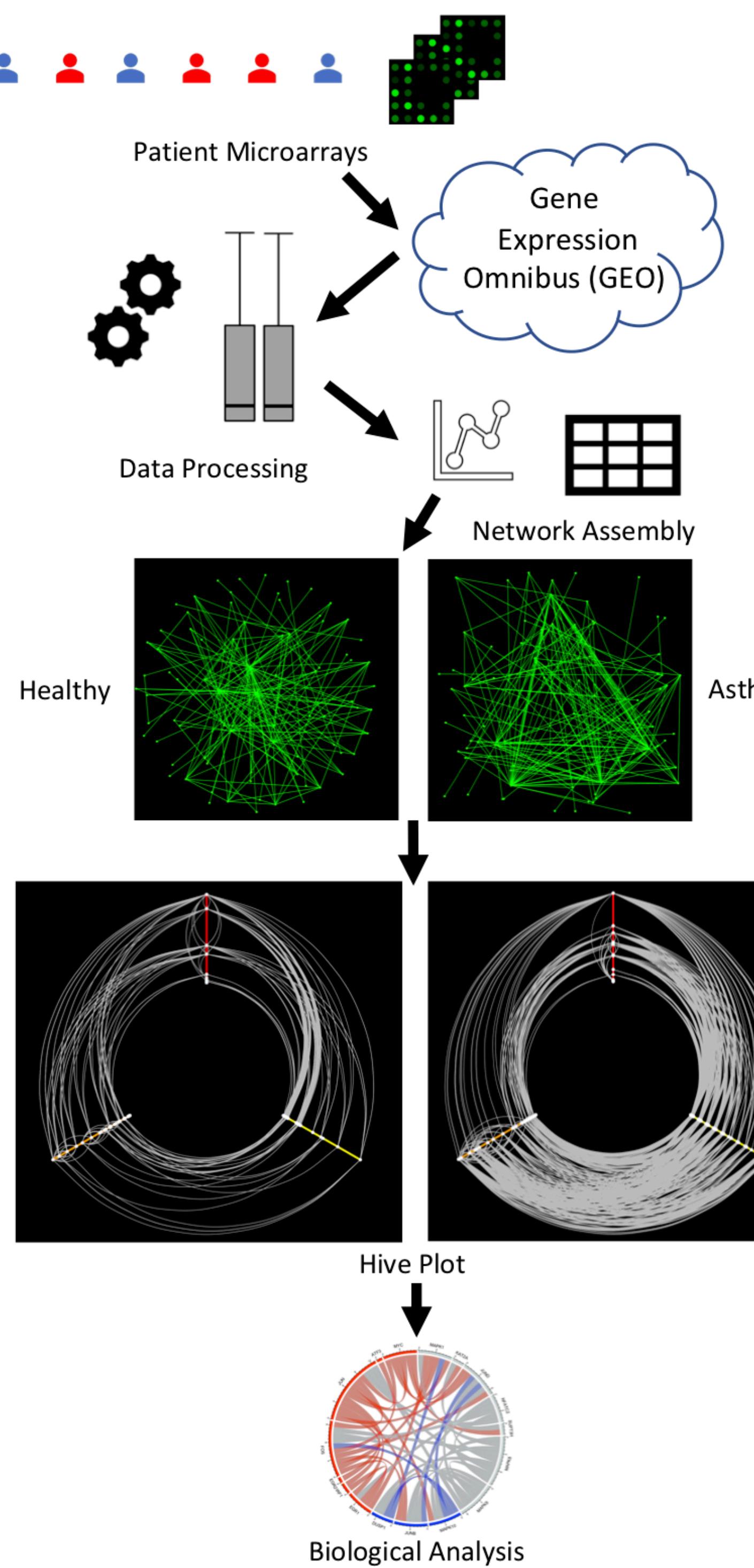


Figure 1. Analysis pipeline used. Outlines the protocols taken from raw microarray data to interpretable, meaningful biological analyses

Methods

PROCESSING:

The pipeline (Fig. 1) begins with gene expression microarray data collected from healthy ($n = 23$) and asthmatic ($n = 13$) patients. The processed data was reduced to include:

- A list of differentially expressed genes ($FDR \leq 0.05$)
- Known transcription factor-gene interactions (ENCODE)

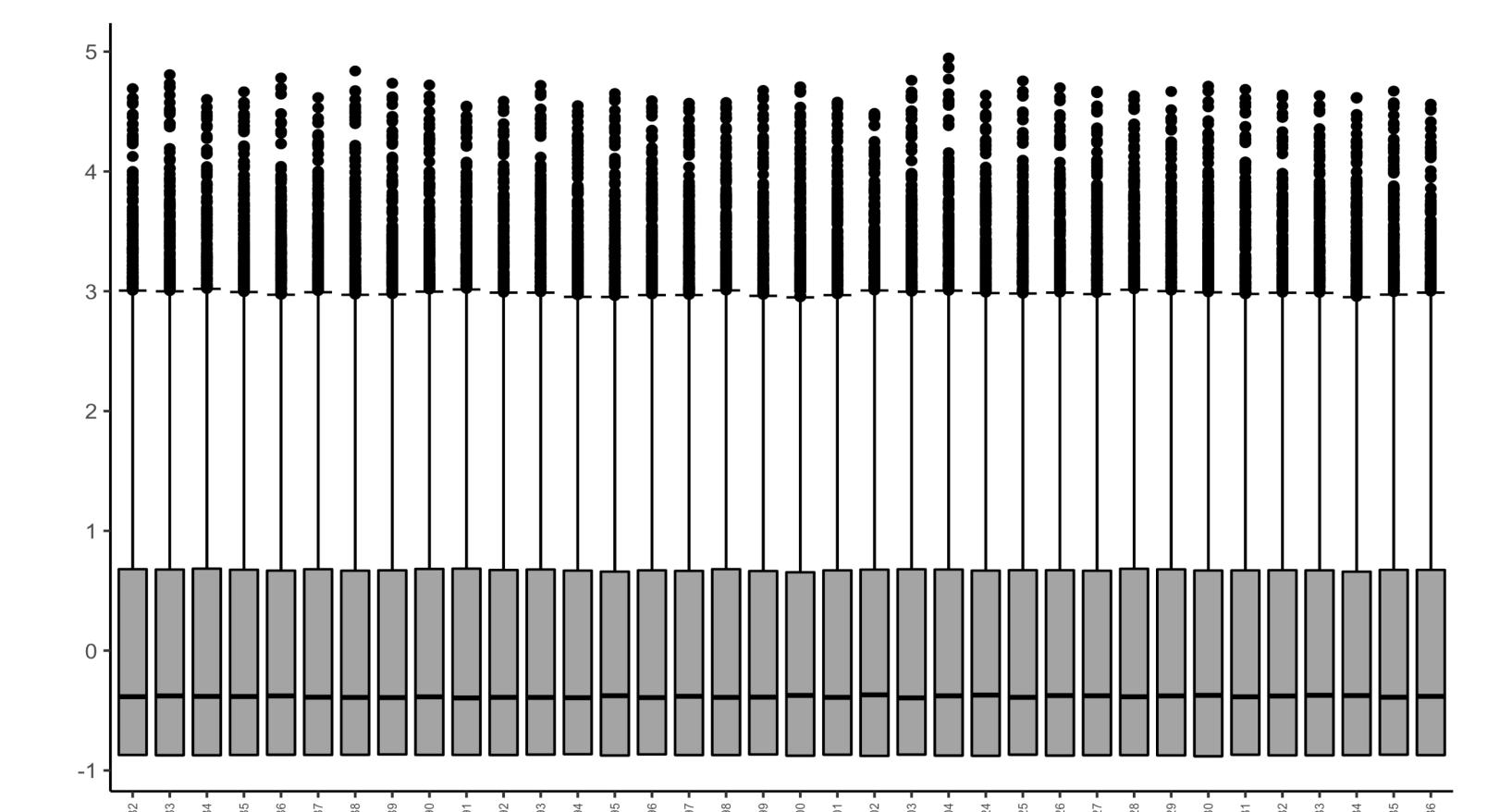


Figure 2. Box plot of gene expression ranges in the data after quantile normalization.

Correlation matrix is constructed from reduced dataset. All pairs of genes sufficiently correlated ($r > 0.75$) are mapped onto the ENCODE GRN (Fig. 3) to give directions to otherwise undirected regulatory connections.

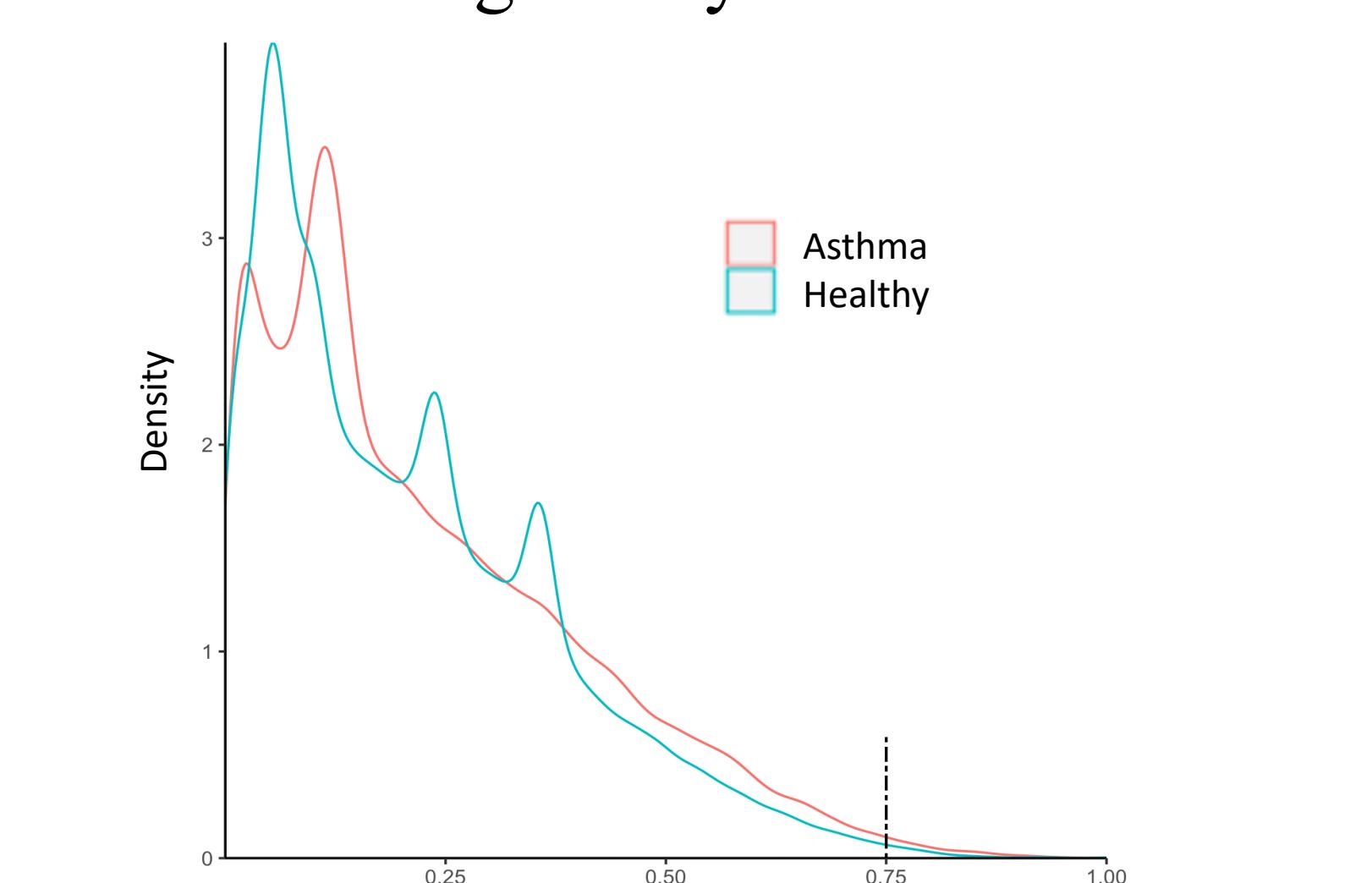


Figure 3. Density plot of correlation values found in the matrix. A cutoff of 0.75 as the minimum correlation coefficient determines significance of edges.

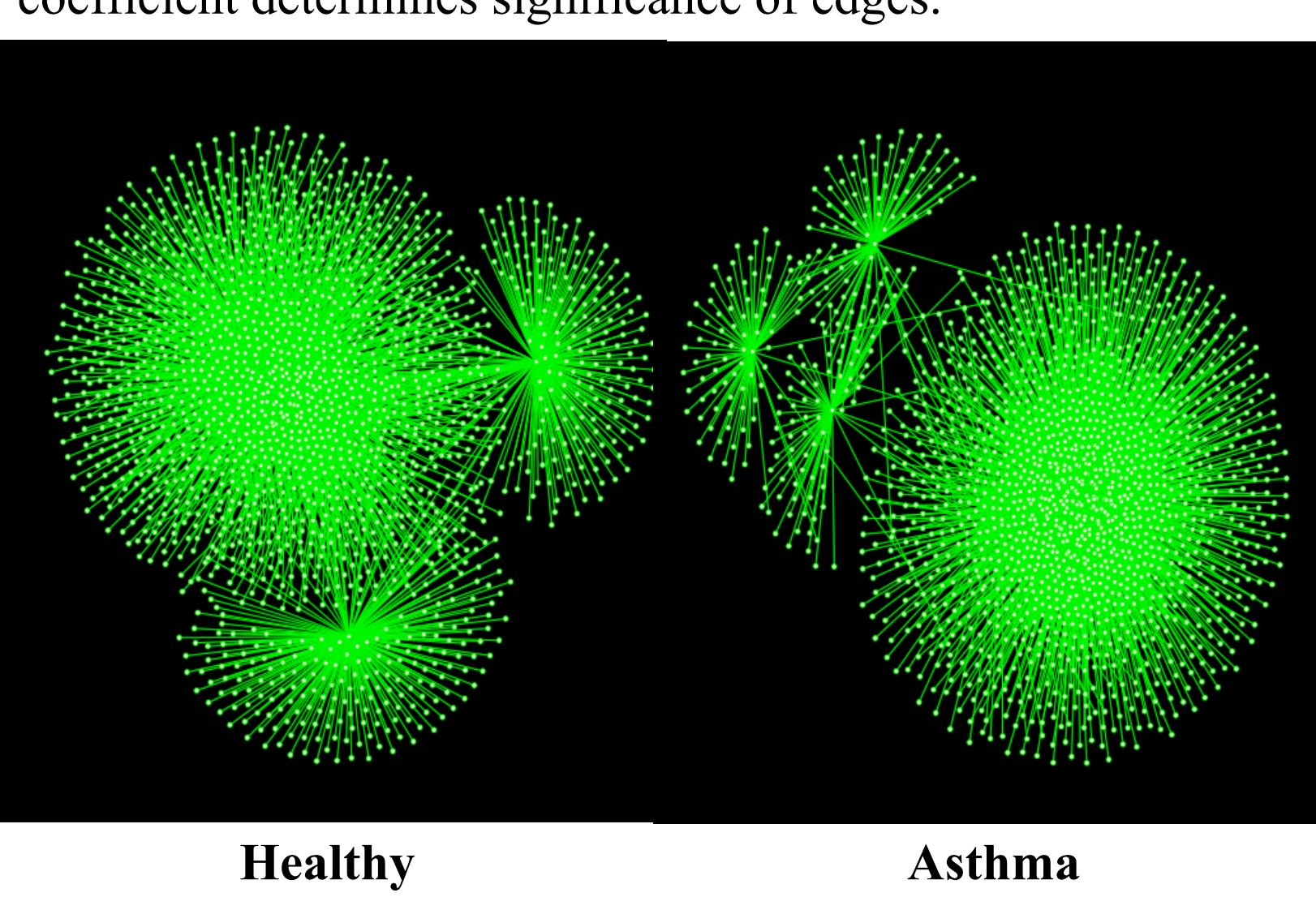


Figure 4. Classic force-based presentation of resulting GRNs.

ANALYSIS:

To overcome the challenges presented in Fig. 4, we design hive plots for four topological measures,

Table 1. Centrality measure definitions of gene $v \in G := (V, E)$.

Measure	Formal Definition	Description
Degree	# $v' \neq v : \{v', v\} \in E$	edges with v as an endpoint
Betweenness	$\sum_{a \neq b \neq v} \frac{\sigma_{ab}(v)}{\sigma_{ab}}$	fraction of global "shortest paths" through v
Eigencentrality	$\frac{1}{\lambda} \sum x_{v'} : \forall v' : \{v', v\} \in E$	weighted sum of neighbors' eigencentralities
Closeness	$(\sum_{y \in V} \text{dist}(y, x))^{-1}$	inverse of distance (in steps) to all non- v nodes

which is followed by identification of top ranked genes with differential profiles.

We take the following steps to transform our network into hive plots (Fig. 6):

1. Establish three radial axes for categorical measures (transcription factor, target gene, other gene).
2. Plot nodes categorically and quantitatively.
3. Draw edges and identify topologically important genes.

Table 2. Top ranked genes for each hive plot measure. Bolded genes are top ranked in only one condition.

Degree	Betweenness		Eigencentrality		Closeness	
	TF	Target	TF	Target	TF	Target
FOS	IER2	FOS	EGR2	FOS	DUSP1	IER2
BHLHE40	EGR2	BHLHE40	BTG2	BHLHE40	FOSB	BHLHE40
EGR1	BTG2	EGR1	GPX1	EGR1	BTG2	EGR2
JUN	EGR3	IRF1	RPL26L1	JUN	EGR3	BTG2
IRF1	MGP	GTF2B	AVPII	ATF3	AVPII	MYC
NR2F2	HOXB-AS3	MYC	HOXB-AS3	IRF1	SLT2-IT1	MYC
MYC	SLT2-IT1					

Top-ranked in Control Top-ranked in Asthma

Significant pathways are calculated through the STRING database and cross-checked with biological literature (Table 3).

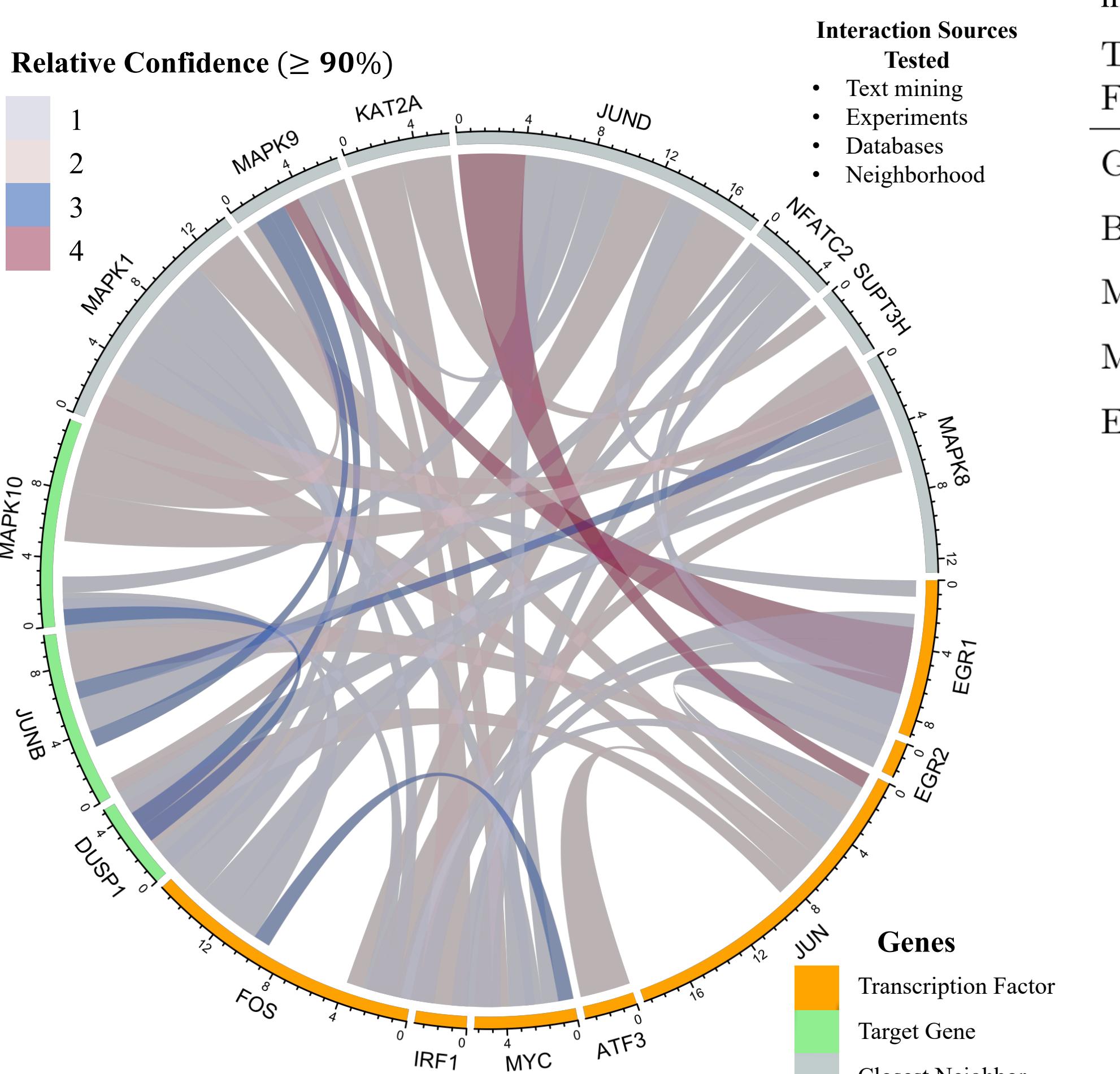


Figure 5. Circos plot for the shortlist of significant genes curated from the hive plot analysis. Edges are colored based on the number of interaction tests above 90%.

Results

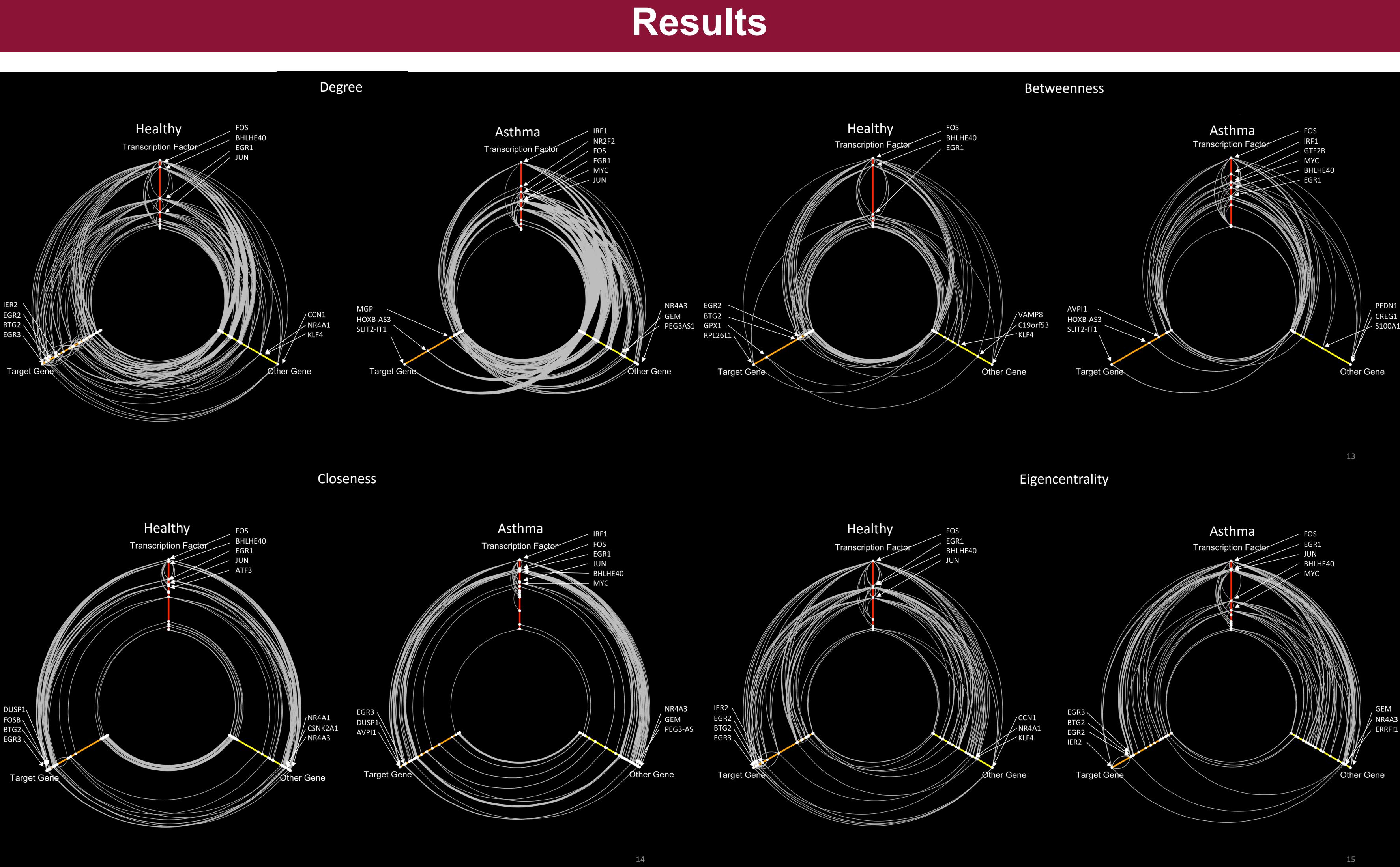


Figure 6. Hive plots for both healthy and asthmatics patients using the four centrality measures. The genes are plotted along each categorical axis (in the radially outward direction).

Further Research

- Perform replication on an independent gene expression dataset.
- Improve hive plot analysis: coloring edges and weighting nodes.
- Include topological loops, gene repression, and connections between transcription factors.
- Explore other methods of network inference (Bayesian, stochastic).

References

- Singhania A, Rupani H, Jayasekera N, Lumb S, Hales P, Gozard N, et al. (2017) Altered Epithelial Gene Expression in Peripheral Airways of Severe Asthma. *PLoS ONE* 12(1): e0168680. <https://doi.org/10.1371/journal.pone.0168680>
- Nguyen, C., Teo, J.-L., Matsuda, A., Eguchi, M., Chi, E. Y., Henderson, W. R., & Kahn, M. (2003). Chemogenomic identification of Ref-1/AP-1 as a therapeutic target for asthma. *Proceedings of the National Academy of Sciences*, 100(3), 1169–1173. <https://doi.org/10.1073/pnas.0437889100>
- Alam, R., & Gorska, M. M. (2011). Mitogen-activated protein kinase signalling and ERK1/2 bistability in asthma. *Clinical and experimental allergy: journal of the British Society for Allergy and Clinical Immunology*, 41(2), 149–159. doi:10.1111/j.1365-2222.2010.03658.x
- Alphonse, M.P., Saffar AS, Shan L, HayGlass KT, Simons FER, Goumari AS (2008) Regulation of the High Affinity IgE Receptor (FcεRI) in Human Neutrophils: Role of Seasonal Allergen Exposure and Th-2 Cytokines. *PLoS ONE* 3(4): e1921. <https://doi.org/10.1371/journal.pone.0001921>
- Barnes, P. J. (2008). The cytokine network in asthma and chronic obstructive pulmonary disease. *Journal of Clinical Investigation*, 118(11), 3546–3556. <https://doi.org/10.1172/jci36130>

Acknowledgements

- Dr. Kari Nadeau
 Abhinav Kaushik
 Stanford Institutes of Medical Research
 Nadeau Lab
 Strober Family Scholars Fund

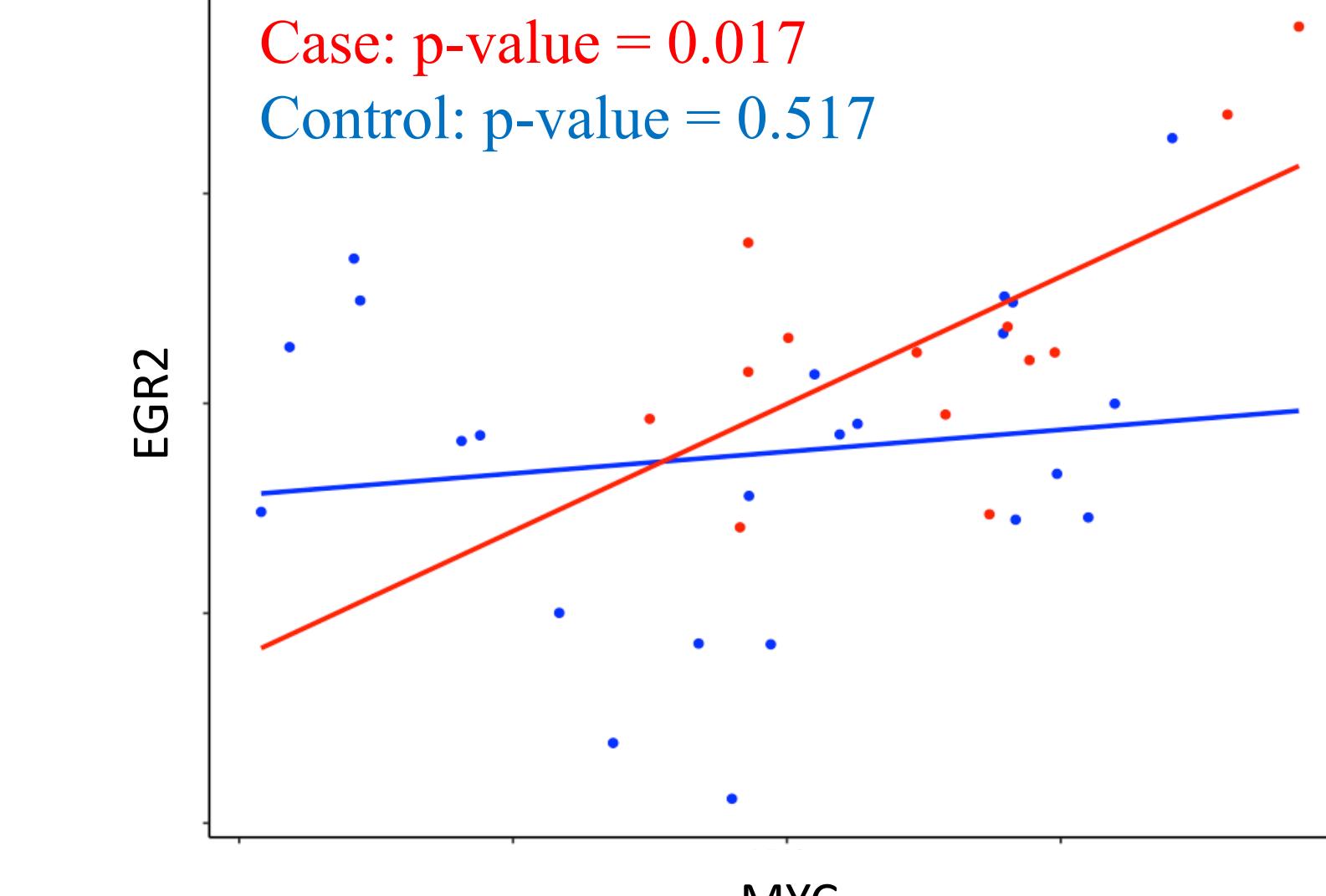


Figure 7. The gene expression values of MYC and EGR2 are plotted per sample and a Student's t-test is performed.