



Stanford
MEDICINE

Sean N. Parker Center
for Allergy & Asthma Research

A robust pipeline for multi-cohort analysis of gene co-expression networks in asthma

AMIA 2019

Brandon Guo

Monta Vista High School, Cupertino, CA

Sean N. Parker Center for Allergy and Asthma Research, Stanford, CA



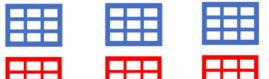
Introduction

- Asthma is an inflammatory disorder driven by complex genotypes
- “Reproducibility crisis” in bioinformatics
 - often limited to one dataset
 - “probabilistic” techniques such as bootstrapping
 - “up-regulation/down-regulation” framework is inconsistent
- Heterogeneity of data and gene-regulatory focus
- Looking for reproducible dysregulations/biomarkers
 - further biological analysis and potential drug targets
 - Goal: how to interpret results from bioinformatics data?

Healthy and Asthma Microarrays (x6)



Scale-Free Adjacency Matrix (x6)



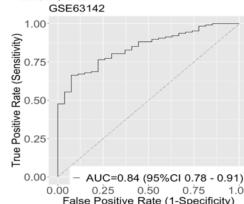
Co-expression Networks with modules



Effect Size meta-analysis using Hedge's g statistic determines a gene signature of asthma across microarrays.

$$g = \left(1 - \frac{3}{4df - 1}\right) \frac{\bar{X}_1 - \bar{X}_0}{\sqrt{\frac{(n_1-1)S_1^2 + (n_0-1)S_0^2}{n_1+n_0-2}}}$$

Independent Cohort Validation of effect size test confirms gene signature.



Dysregulated modules across conditions



Unique Modules in Asthma



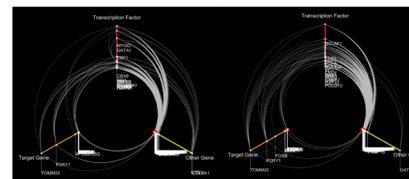
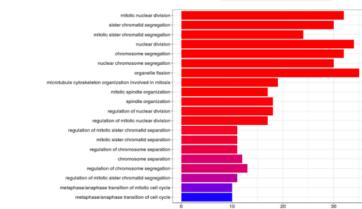
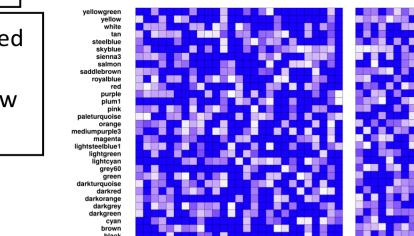
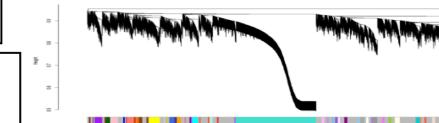
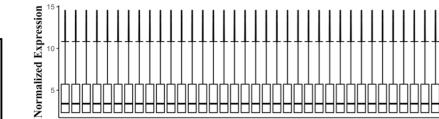
Heterogenous Epithelial Microarrays are collected and pre-processed for each dataset and condition (asthma, healthy).

WGCNA Unsupervised Clustering is used to identify modules in each microarray.

Module preservation is performed across conditions to determine asthma modules that do not show preservation in healthy.

Fisher's exact test is performed with permutations to determine *asthma-unique* modules.

Functional enrichment is performed to prioritize modules for topological analysis.



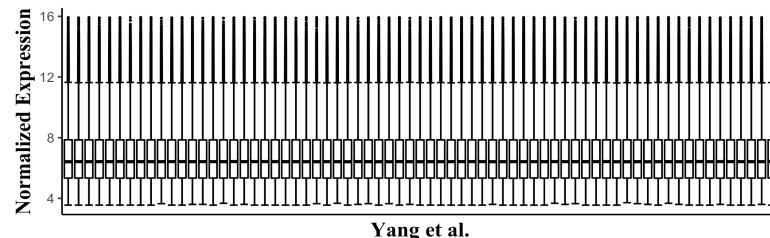
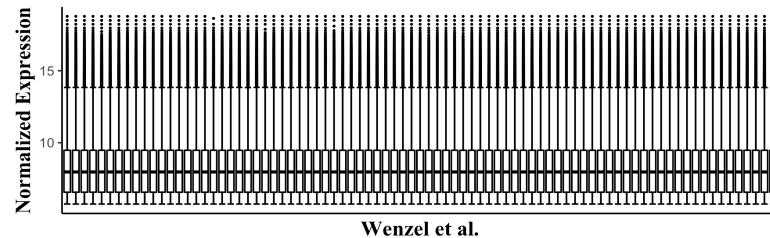
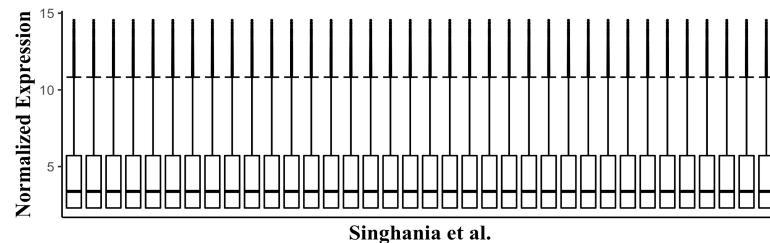
Hive plot analysis of unique modules reveals topologically important interactions.

Biomarkers, interactions, and dysregulations discovered by this pipeline are *triply robust*: they are **ontologically relevant** (functional enrichment), **statistically relevant** (WGCNA, meta-analysis), and **topologically relevant** (hive plot analysis).

QC Pipeline (Data Processing)

GEO ID	Epithelia Source	Healthy Patients	Asthma Patients	Reference
GSE64913	Bronchial	23	13	<i>Singhania et al.</i>
GSE63142	Bronchial	27	125	<i>Wenzel et al.</i>
GSE65204	Nasal	33	36	<i>Yang et al.</i>

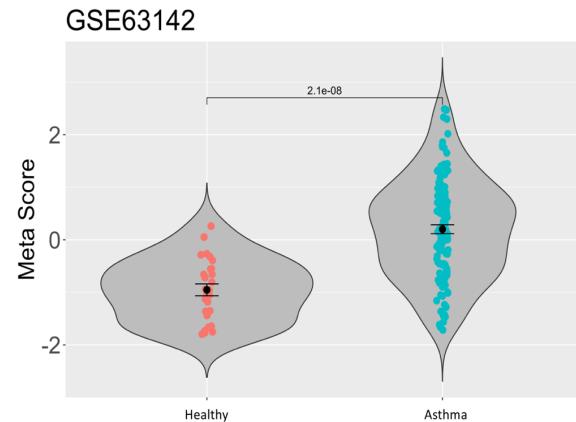
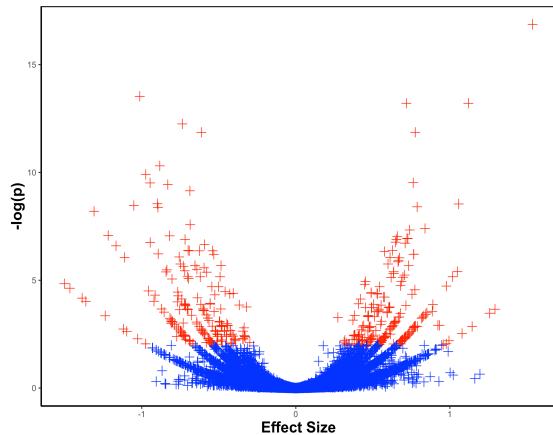
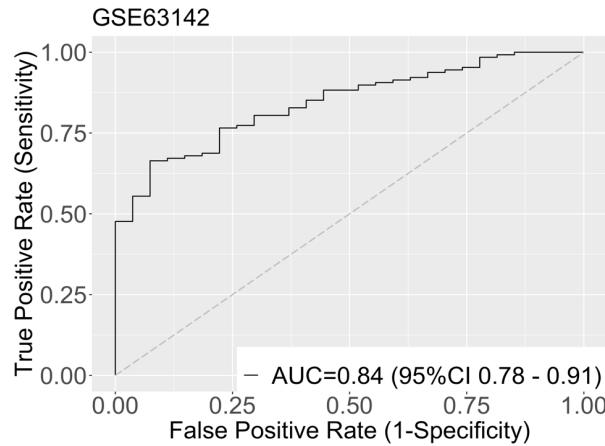
GEO ID	GPL Source	Total Probe Readings	Total Gene Readings	# Final Genes
GSE64913	GPL570	54,676	21,180	12,806
GSE63142	GPL6480	30,889	19,568	12,806
GSE65204	GPL14550	20,533	13,779	12,806



Meta-analysis

$$g = \left(1 - \frac{3}{4df - 1}\right) \frac{\bar{X}_1 - \bar{X}_0}{\sqrt{\frac{(n_1-1)S_1^2 + (n_0-1)S_0^2}{n_1+n_0-2}}}$$

$$g_s = \frac{\sum_i^n W_i g_i}{\sum_i^n W_i}$$

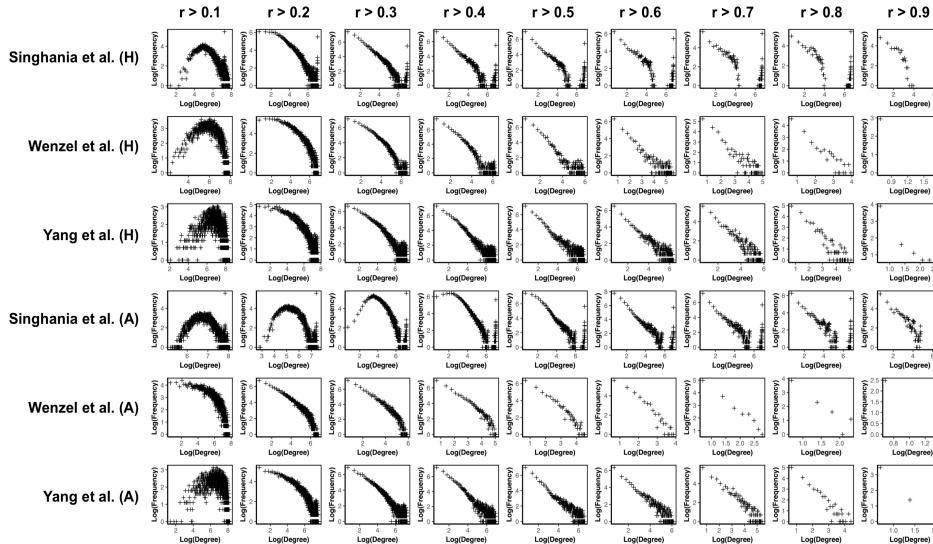


Pathway name	p-value (BH corrected)
Antigen processing and presentation	2.88e-11
Th1 and Th2 cell differentiation	2.03e-8
Asthma	5.18e-7
Th17 cell differentiation	8.56e-7

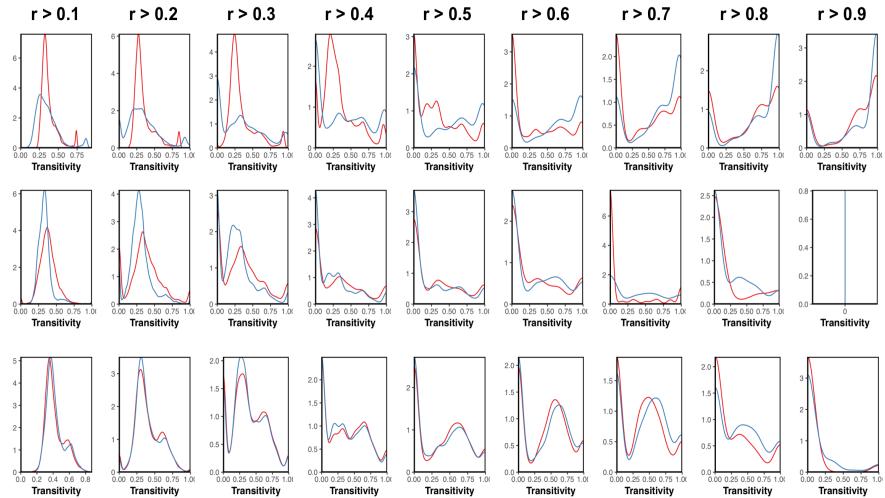
Haynes et al., Pac Symp. Biocomput.

Weighted Gene Co-expression Networks (WGCNA)

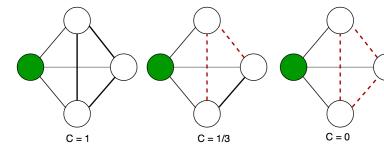
$$\deg(G) \propto k^{-\gamma}$$



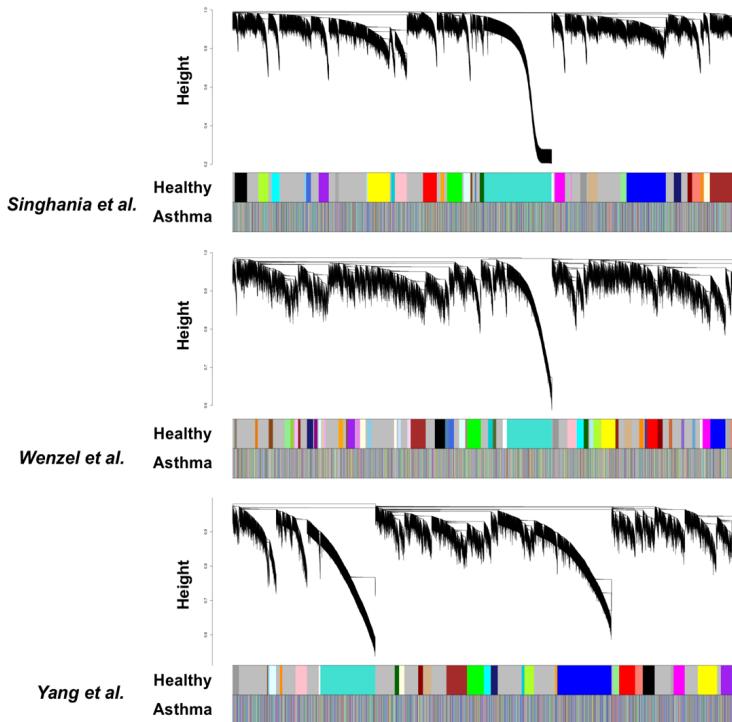
Transitivity densities of asthma (red) and healthy (blue) WGCNA sub-networks based on correlation cutoff



Langfelder et al., Bioinformatics.



Module Preservation in Networks

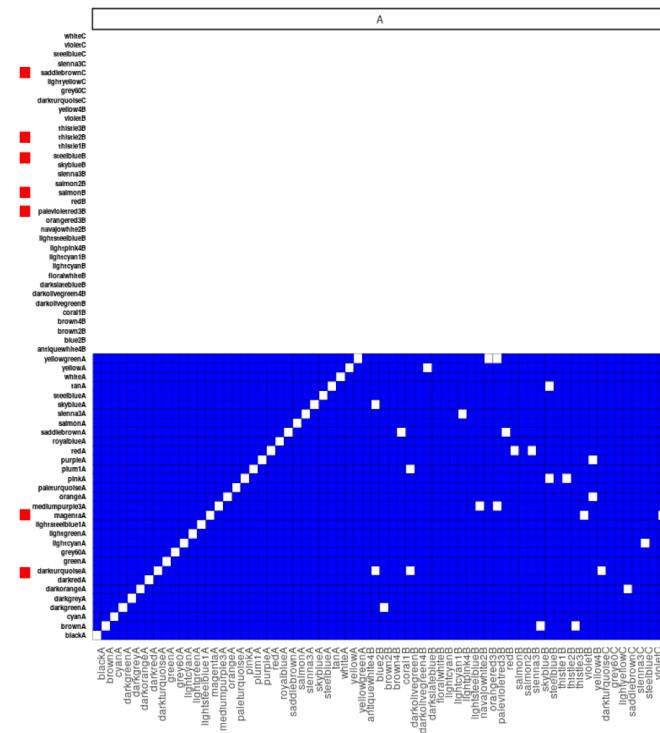


	Singhania et al.	Wenzel et al.	Yang et al.
# Healthy modules	33	55	30
# Asthma modules	42	82	38
# Dysreg. modules (Zsummary<10)	31	27	8

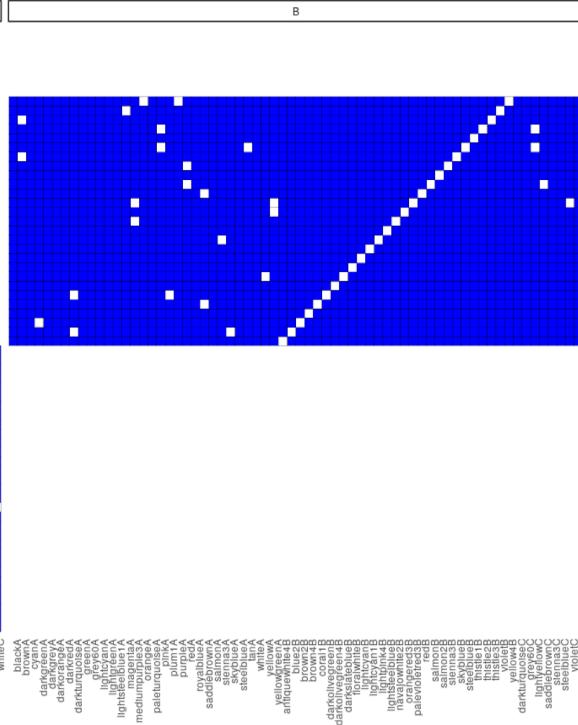
Langfelder et al., PLOS 2011.

Identification of “asthma-unique” modules

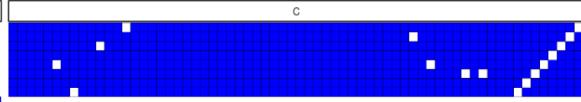
Singhania et al.



Wenzel et al.



Yang et al.



$$A, B \in S$$

$$A \cap B$$

$$B - A \cap B$$

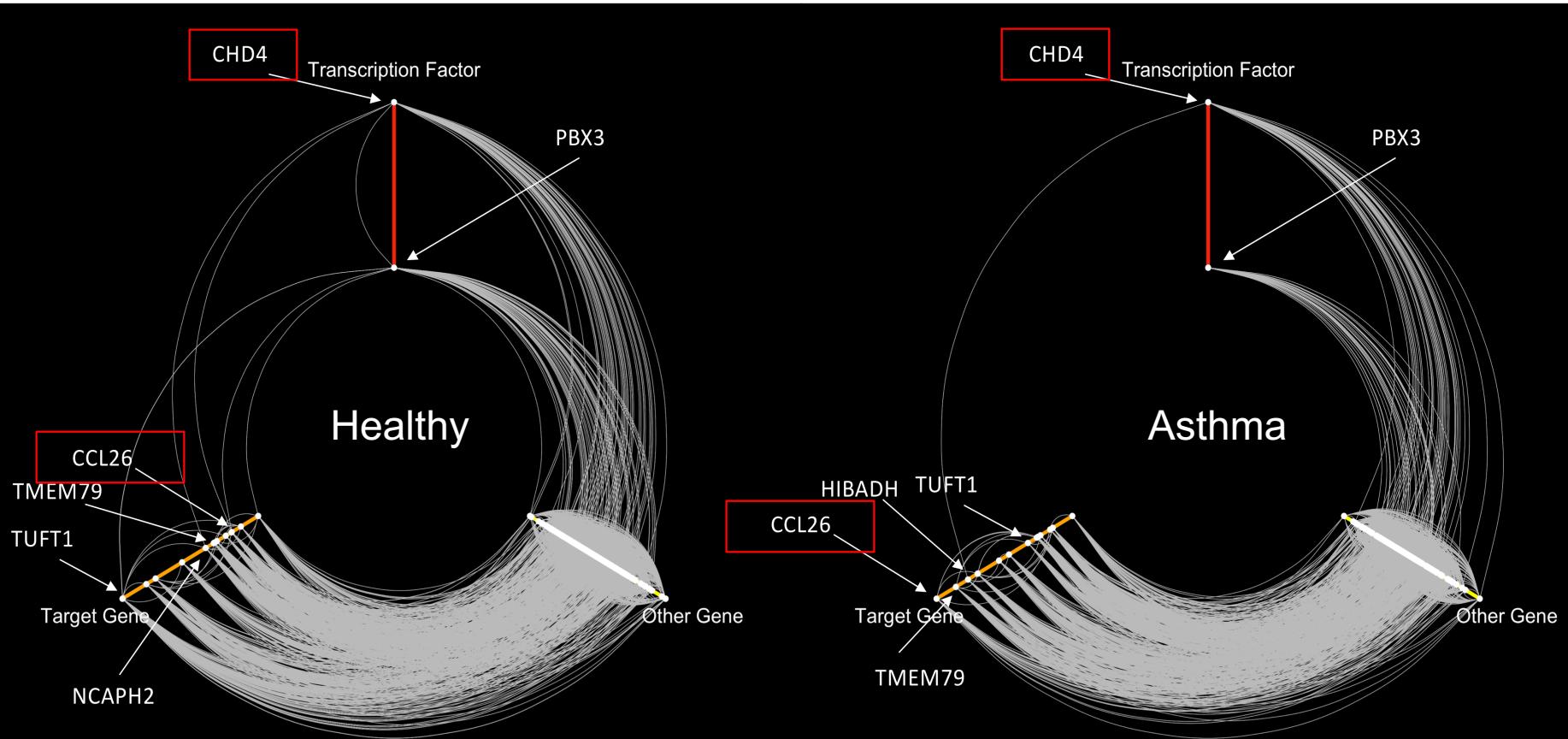
$$A - A \cap B$$

$$S = A \cup B$$

GO Analysis of asthma-unique modules

Module	Size	GO Terms
darkturquoiseA	158	protein transport, peptide transport, amide transport, protein localization, nitrogen compound transport, digestive tract development,
magentaA	274	cellular component organization, regulation of cellular process/biological process, signaling, cell communication, developmental process.
palevioletred3B	46	anatomical structure, organ morphogenesis, regulation of cellular process/glutamate secretion.
salmonB	120	thrombopoietin-mediated signaling, regulation of biological process, cellular process, regulation of biological quality, homeostatic process, cellular biosynthetic process.
steelblueB	86	acute-phase response, cellular protein metabolic process, macromolecule modification, response to wounding, regulation of phosphate/phosphorus metabolic process.
thistle2B	51	blood circulation, circulatory system process, regulation of autophagy, response to drug, cellular response to chemical stimulus.
saddlebrownC	81	DNA biosynthetic process, cellular component assembly, DNA metabolic process, regulation of protein phosphorylation.

Topological Analysis (Hive plots)



Discussion

- Overall topological dysregulation in asthma
- CHD4 – CCL26 interaction
 - transcription factor to target (DE) gene interaction
 - CCL26 recruitment of eosinophils (which control inflammation)
 - more than just “upregulated in asthma,” seems like a response
- CCL26 degree and expression not the full “biomarker”
 - How does CHD4 — CCL26 “fit in” with the established research?
 - nuance suggests instead study of the specific interaction as a biomarker

What's next?



- Reproducibility: Uploading code pipeline with built-in debugging
- Publication: submit by Jan. 2020, *BMC Bioinformatics*
- Biological Verification: PCR of key targets/modules

Acknowledgments



- Dr. Abhinav Kaushik
- Stanford Institutes of Medicine Summer Research Program (SIMR)
- Strober Family Scholars Fund
- AMIA HS Scholars Program
- Dr. Kari Nadeau and Sean. N Parker Research Lab



Thank you!

<http://bguo.us>

