Applying Weighted Differential Co-expression Analysis to Characterize Th2 High and Low Asthma Endotypes

Team Undecided
Arjun Baghela, Emma Graham, Eric Chu, Allison Tai

INTRODUCTION

Asthma, a disease characterized by chronic inflammation, affects over 235 million individuals worldwide.[1] One way to define asthma populations is by T helper cell cytokine levels: patients can have high or low levels of Th2 cytokines. Th2-high patients tend to show more severe symptoms.[2] Fortunately, gene expression biomarkers CLCA1, periostin, and serpinB2 have been shown to differentiate the asthma endotypes.[2] Our goal is to better characterize differences between the endotypes using a network based approach. Using publicly available data, we present the application of differential co-expression analysis using the transcriptomic and methylation profiles of asthmatic and control patients. Our method first identifies pairs of differentially expressed and methylated genes to subsequently investigate changes in the interactome between endotypes. Assessing pairwise differential interactions between genes may lend more insight into the disease etiology than differential expression alone.

METHODS

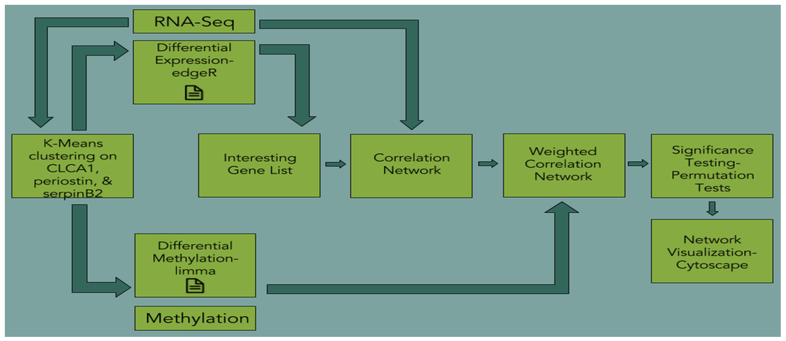

Figure 1. Differential co-expression analysis pipeline.

- RNA-Seq & methylation data was preprocessed as described in Nicodemus-Johnson et al, 2016.[3]
- Methylation probes were further filtered according to a list of non-variable probes defined by Edgar et al, 2017.[4]
- K means clustering was used to group patients into Th2 low/high patients using CLCA1, periostin, serpinB2 expression data.
- Differential expression & methylation between groups was performed using edgeR and limma, respectively.[5-6] This was performed to obtain a reduced list of interesting genes for differential co-expression analysis.
- A correlation network was constructed for each group, using expression data for DEGs. A differential correlation network was then calculated by taking the absolute differences between groups, for each corresponding gene pair. networks. Graph edges were further weighted if genes were also differential methylated ($-\log_2$FDR).
- Permutations tests (1000) were performed to derive a p-value for each observed differential co-expression.
- Significant gene pairs (top 500 genes with FDR=0) were visualized in Cytoscape.[7]

DATA

- RNA-Seq & methylation data was obtained from NCBI Gene Expression Omnibus (GEO) under accession code GSE85568.
- RNA-Seq & methylation data was performed on AECs (74 asthmatics, 41 nonasthmatics) using the Illumina HiSeq 2000 platform & Infinium Human Methylation 450K Bead Chip, respectively.
- Original findings were published in "DNA methylation in lung cells is associated with asthma endotypes and genetic risk" (Nicodemus-Johnson et al, 2016).

RESULTS

RNA-Seq & methylation data inspection

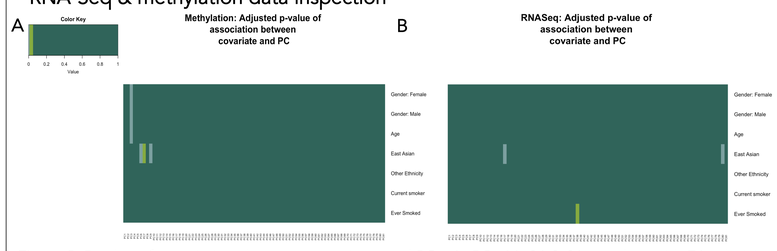

Figure 2. P-value of associations between covariates and PCs. A- PCs 16, 36 & 81 correlate with EA ethnicity and smoking status. Because PCs explain little variance, additional batch correction was not performed. B- Several of PCs 1-10 were correlated with gender, ethnicity and age. However, regressing out these PCs, which explain most of the variance, would reduce the ability to detect biological signal.

K-means clustering for patient differentiation

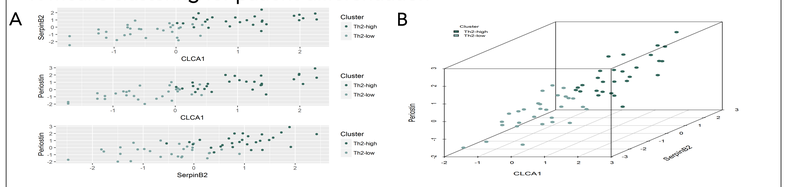

Figure 3. Asthma patients clustered into Th2 low & high groups according to expression levels of CLCA1, serpinB2, and periostin. A- Clustering in 2-dimensions: CLCA1 vs. serpinB2, CLCA1 vs. periostin, serpinB2 vs. periostin. B- Clustering in a 3-dimensional scatterplot.

Correlation Network Construction

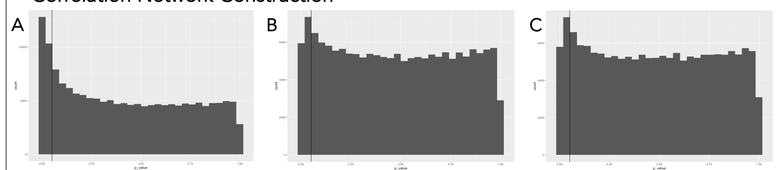

Figure 4. P-value distribution resulting from permutation tests for group comparisons. A- Control-Th2 high. B- Control-Th2 low. C- Th2 high-Th2 low.

Cytoscape Visualization

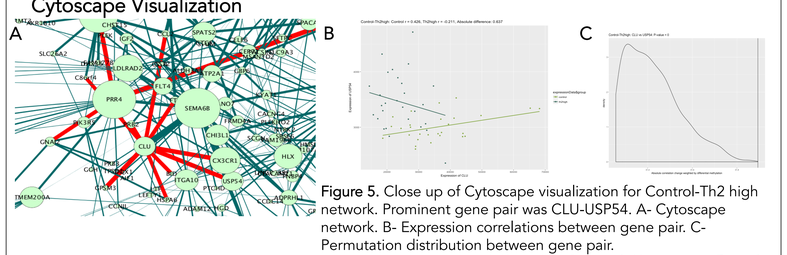

Figure 5. Close up of Cytoscape visualization for Control-Th2 high network. Prominent gene pair was CLU-USP54. A- Cytoscape network. B- Expression correlations between gene pair. C- Permutation distribution between gene pair.

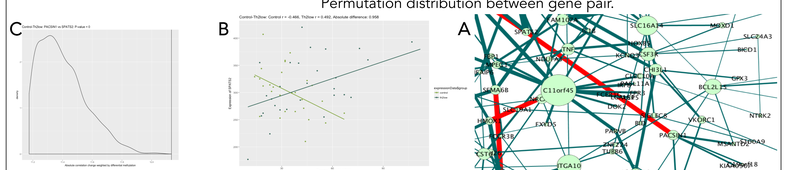

Figure 6. Control-Th2 low. A- Cytoscape network. B- Expression correlation between Pacsin1-Spats2. C- Permutation distribution.

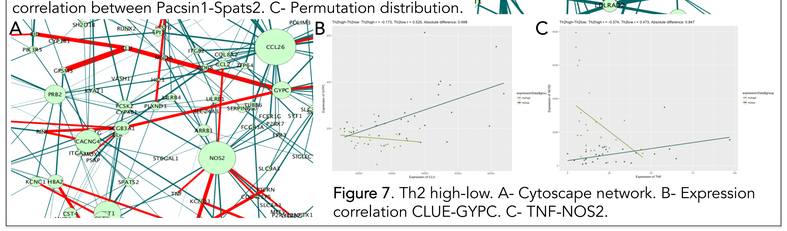

Figure 7. Th2 high-low. A- Cytoscape network. B- Expression correlation CLUE-GYPC. C- TNF-NOS2.

CONCLUSIONS

- CLCA1, serpinB2, and periostin appear to be co-expressed, showing there is validity in past findings they can be used as biomarkers for asthma endotypes.
- Between comparisons, ~7% of significant gene pairings were conserved among the top 500 gene pairings with the highest absolute differential correlation.
- This indicates a degree of conservation in differential co-expression between endotypes. However, the majority of significant gene-pairs are different, indicating a dynamic interactome.
- More differentially methylated CpGs were found when comprising asthmatics to controls, rather than between asthma subtypes. This indicates a shared methylation profile among asthmatics, regardless of subtype.
- Several gene pairs found to be differentially co-expressed and methylated are known to be involved in asthma. Other genes may be upstream or downstream regulators in asthma pathways
- In this project, we present what appears to be a novel method of identifying gene pairs which are differentially expressed and differentially methylated.
- Our source of network construction is gene expression data, however, we propose a weighting scheme to support edges which also show differential methylation.
- Integrating methylation data is appropriate when studying asthma, as epigenetic modulators are shown to be associated by asthma pathogenesis and clinical heterogeneity.[3]
- Our method suggests the importance of 500 significant gene pairs between each condition (filters can be adjusted), which can be further validated for their role in asthma pathogenesis.

REFERENCES

1. Pawankar, R. (2014). Allergic diseases and asthma: a global public health concern and a call to action. World Allergy Organization Journal, 7(1), 12.
2. Wesolowska-Andersen, A., & Seibold, M. A. (2015). Airway molecular endotypes of asthma: dissecting the heterogeneity. Current opinion in allergy and clinical immunology, 15(2), 163.
3. Nicodemus-Johnson, J., Myers, R. A., Sakabe, N. J., Sobreira, D. R., Hogarth, D. K., Naureckas, E. T., ... & Nicolae, D. L. (2016). DNA methylation in lung cells is associated with asthma endotypes and genetic risk. JCI insight, 1(20).
4. Edgar, R. D., Jones, M. J., Robinson, W. P., & Kobor, M. S. (2017). An empirically driven data reduction method on the human 450K methylation array to remove tissue specific non-variable CpGs. Clinical Epigenetics, 9(1), 11
5. Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26(1), 139-140.
6. Ritchie, ME, Phipson, B, Wu, D, Hu, Y, Law, CW, Shi, W, and Smyth, GK (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 43(7), e47.
7. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., ... & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research, 13(11), 2498-2504.

ACKNOWLEDGEMENTS

- UBC Statistics 540 teaching staff for guidance.
- Carole Ober of the University of Chicago for data interpretation support and making the data publicly available.