CHIT1 Project Update: csSAM Analysis

Brian Alexander Barron, YSM3

September 12th 2017

What is csSAM?

- csSAM ("<u>c</u>ell-<u>s</u>pecific <u>S</u>ignificance <u>A</u>nalysis of <u>M</u>icroarrays") is an R package that implements the csSAM algorithm published in Shen-Orr et al., 2010, Nature Methods.
  - The algorithm uses a least-squares regression to identify the coefficients of the deconvoluted matrix:
    - $G_{ij} = C_{ik} \cdot M_{kj}$
      - *G* is the <u>observed</u> gene expression matrix (with "i" samples and "j" genes)
      - *C* is the <u>observed</u> cell frequency matrix (with "i" samples and "k" cell types)
      - *M* is the <u>computed</u> cell-specific gene expression matrix (with "k" cell types and "j" genes)

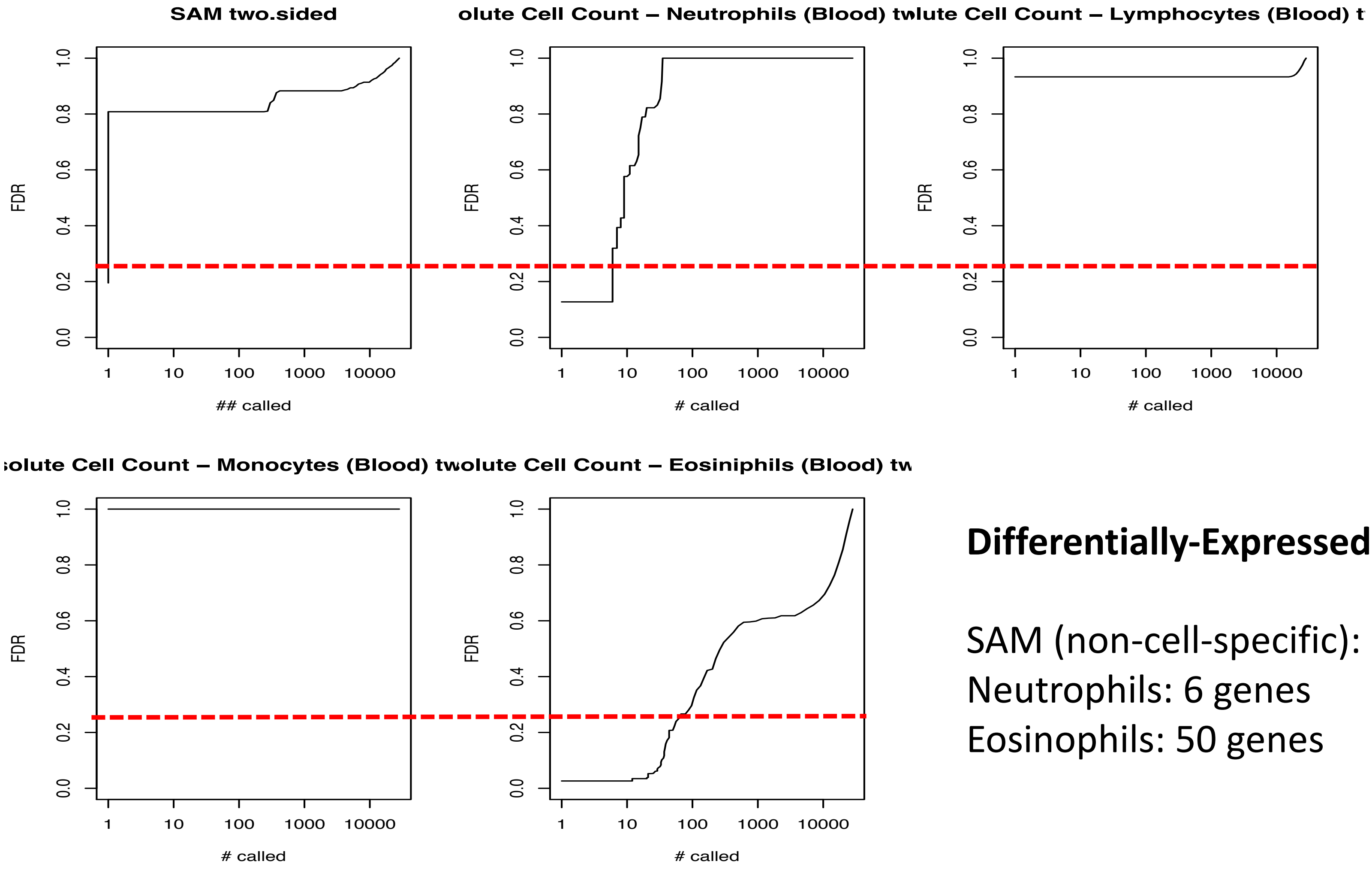
    Or

    - $\mathrm{G}_{ij} = \Sigma_{k=1}{}^{K}\, w_{ik} h_{kj} + \mathrm{e}_{ij}$
      - *w* is the <u>computed</u> cell-specific gene expression coefficient matrix
      - *h* is the the <u>observed</u> cell frequency matrix
      - e_{ij} is the error matrix

The Data Set

- Gene Expression Data – Vera
    - 86 samples (24 control, 62 asthma) isolated from blood
    - 28,203 genetic loci from GPL6244 (Affymetrix Gene 1.0 ST)
- Clinical Data
    - 208 clinical variables, including blood cell frequencies

csSAM Analysis Results



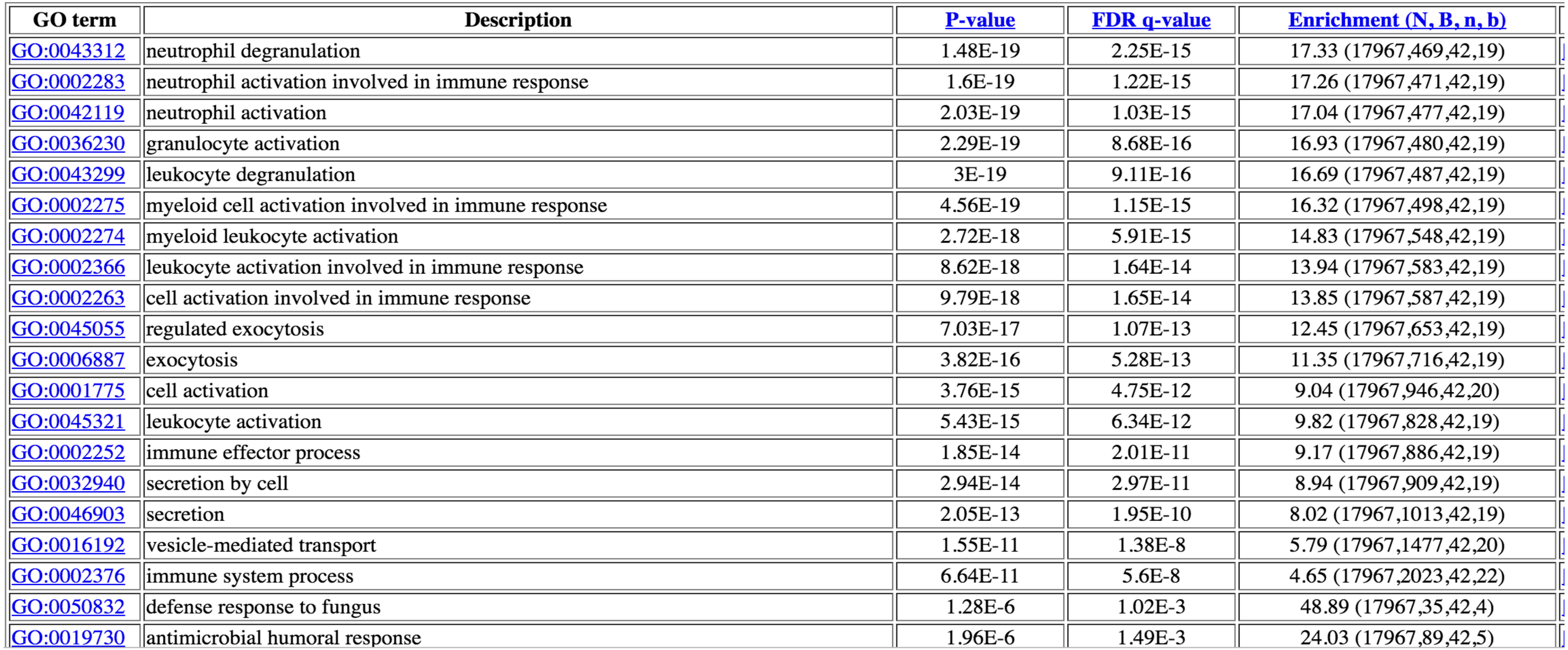
Differentially-Expressed Genes (FDR < 0.25)

SAM (non-cell-specific): 0 genes
Neutrophils: 6 genes
Eosinophils: 50 genes

Neutrophils: GO Enrichment

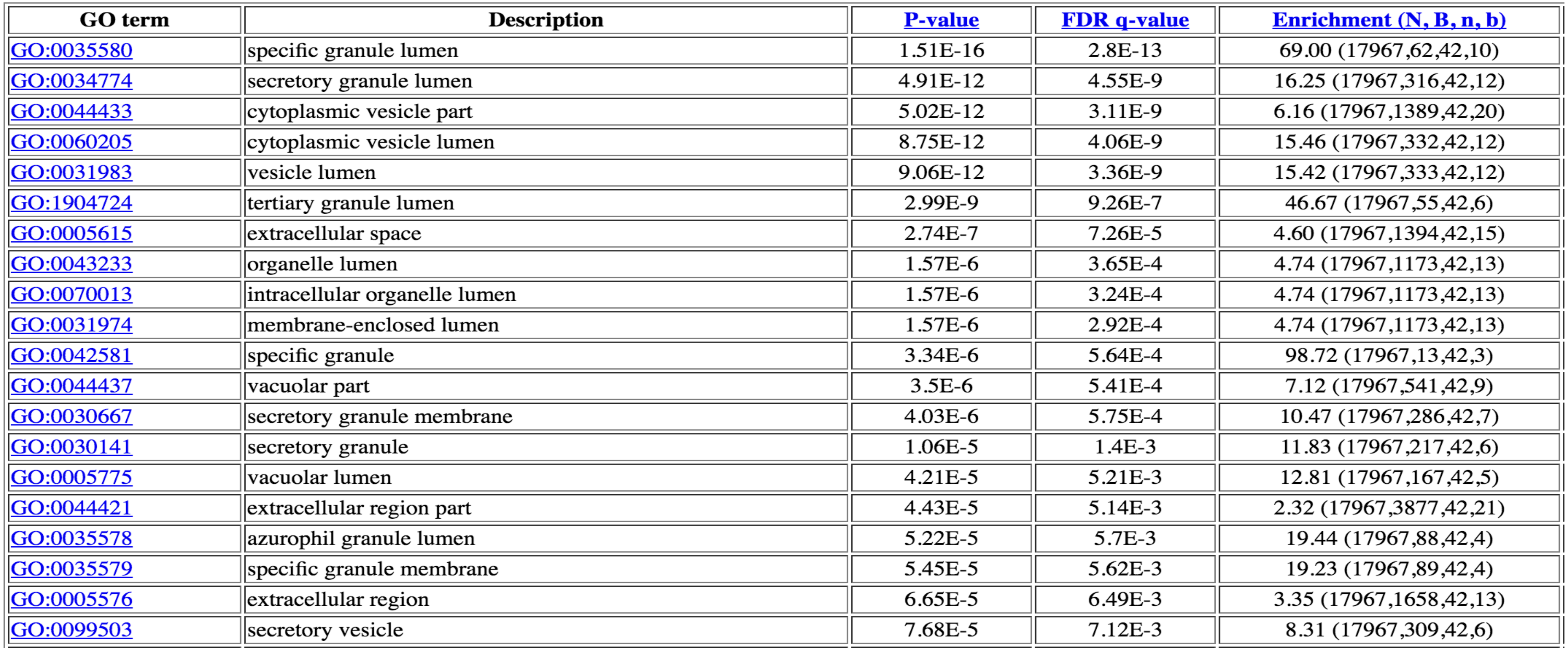
- The six differentially-expressed genes did not show any GO enrichment.
  - RHD – Rhesus D antigen
  - C4BPA – Complement Component 4 Binding Protein Alpha
  - RPS4Y1 – Ribosomal Protein S4
  - DDX3Y – Dead-Box Helicase 3, Y-linked
  - EIF1AY – Eukaryotic Translation Initiation Factor 1A, Y-linked
  - UTY – Ubiquitously Transcribed Tetratricopeptide Repeat Containing, Y-linked (histone demethylase)

Eosinophil: GO Enrichment – Process



GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)
GO:0043312	neutrophil degranulation	1.48E-19	2.25E-15	17.33 (17967,469,42,19)
GO:0002283	neutrophil activation involved in immune response	1.6E-19	1.22E-15	17.26 (17967,471,42,19)
GO:0042119	neutrophil activation	2.03E-19	1.03E-15	17.04 (17967,477,42,19)
GO:0036230	granulocyte activation	2.29E-19	8.68E-16	16.93 (17967,480,42,19)
GO:0043299	leukocyte degranulation	3E-19	9.11E-16	16.69 (17967,487,42,19)
GO:0002275	myeloid cell activation involved in immune response	4.56E-19	1.15E-15	16.32 (17967,498,42,19)
GO:0002274	myeloid leukocyte activation	2.72E-18	5.91E-15	14.83 (17967,548,42,19)
GO:0002366	leukocyte activation involved in immune response	8.62E-18	1.64E-14	13.94 (17967,583,42,19)
GO:0002263	cell activation involved in immune response	9.79E-18	1.65E-14	13.85 (17967,587,42,19)
GO:0045055	regulated exocytosis	7.03E-17	1.07E-13	12.45 (17967,653,42,19)
GO:0006887	exocytosis	3.82E-16	5.28E-13	11.35 (17967,716,42,19)
GO:0001775	cell activation	3.76E-15	4.75E-12	9.04 (17967,946,42,20)
GO:0045321	leukocyte activation	5.43E-15	6.34E-12	9.82 (17967,828,42,19)
GO:0002252	immune effector process	1.85E-14	2.01E-11	9.17 (17967,886,42,19)
GO:0032940	secretion by cell	2.94E-14	2.97E-11	8.94 (17967,909,42,19)
GO:0046903	secretion	2.05E-13	1.95E-10	8.02 (17967,1013,42,19)
GO:0016192	vesicle-mediated transport	1.55E-11	1.38E-8	5.79 (17967,1477,42,20)
GO:0002376	immune system process	6.64E-11	5.6E-8	4.65 (17967,2023,42,22)
GO:0050832	defense response to fungus	1.28E-6	1.02E-3	48.89 (17967,35,42,4)
GO:0019730	antimicrobial humoral response	1.96E-6	1.49E-3	24.03 (17967,89,42,5)

Eosinophil: GO Enrichment – Component



GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)
GO:0035580	specific granule lumen	1.51E-16	2.8E-13	69.00 (17967,62,42,10)
GO:0034774	secretory granule lumen	4.91E-12	4.55E-9	16.25 (17967,316,42,12)
GO:0044433	cytoplasmic vesicle part	5.02E-12	3.11E-9	6.16 (17967,1389,42,20)
GO:0060205	cytoplasmic vesicle lumen	8.75E-12	4.06E-9	15.46 (17967,332,42,12)
GO:0031983	vesicle lumen	9.06E-12	3.36E-9	15.42 (17967,333,42,12)
GO:1904724	tertiary granule lumen	2.99E-9	9.26E-7	46.67 (17967,55,42,6)
GO:0005615	extracellular space	2.74E-7	7.26E-5	4.60 (17967,1394,42,15)
GO:0043233	organelle lumen	1.57E-6	3.65E-4	4.74 (17967,1173,42,13)
GO:0070013	intracellular organelle lumen	1.57E-6	3.24E-4	4.74 (17967,1173,42,13)
GO:0031974	membrane-enclosed lumen	1.57E-6	2.92E-4	4.74 (17967,1173,42,13)
GO:0042581	specific granule	3.34E-6	5.64E-4	98.72 (17967,13,42,3)
GO:0044437	vacuolar part	3.5E-6	5.41E-4	7.12 (17967,541,42,9)
GO:0030667	secretory granule membrane	4.03E-6	5.75E-4	10.47 (17967,286,42,7)
GO:0030141	secretory granule	1.06E-5	1.4E-3	11.83 (17967,217,42,6)
GO:0005775	vacuolar lumen	4.21E-5	5.21E-3	12.81 (17967,167,42,5)
GO:0044421	extracellular region part	4.43E-5	5.14E-3	2.32 (17967,3877,42,21)
GO:0035578	azurophil granule lumen	5.22E-5	5.7E-3	19.44 (17967,88,42,4)
GO:0035579	specific granule membrane	5.45E-5	5.62E-3	19.23 (17967,89,42,4)
GO:0005576	extracellular region	6.65E-5	6.49E-3	3.35 (17967,1658,42,13)
GO:0099503	secretory vesicle	7.68E-5	7.12E-3	8.31 (17967,309,42,6)

Next Steps?

- The Neutrophil differentially expressed genes (DEGs) do not seem that interesting. Should we investigate them further?
- Further investigate the eosinophil DEGs.
  - However, is this relevant to our study? We seemed to focus more on the neutrophils.
- Construct new WGCNA networks using the deconvoluted cell-specific gene matrix data to see if module differences exist between the control and asthmatic samples.
  - Establish correlations between modules and clinical data.

Bibliography

- Shen-Orr, S. S., et al. (2010). "Cell type-specific gene expression differences in complex tissues." Nature Methods **7**(4): 287-289.