

META-ANALYSIS OF DIFFERENTIAL GENE CO-EXPRESSION: APPLICATION TO LUPUS

SUMIT B. MAKASHIR

*Lindner College of Business, University of Cincinnati, 2925 Campus Green Dr.,
Cincinnati, OH 45221, USA
Email: makashsb@mail.uc.edu*

LEAH C. KOTTYAN

*Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital, 3333 Burnet Ave.,
Mail Location 15021, Cincinnati, OH 45229-3039, USA
Email: Leah.Kotlyan@cchmc.org*

MATTHEW T. WEIRAUCH*

*Center for Autoimmune Genomics and Etiology, and Division of Biomedical Informatics
Cincinnati Children's Hospital, 3333 Burnet Ave.,
Mail Location 15021, Cincinnati, OH 45229-3039, USA
Email: Matthew.Weirauch@cchmc.org*

We present a novel statistical framework for meta-analysis of differential gene co-expression. In contrast to standard methods, which identify genes that are over or under expressed in disease vs controls, differential co-expression identifies gene pairs with correlated expression profiles specific to one state. We apply our differential co-expression meta-analysis method to identify genes specifically mis-expressed in blood-derived cells of systemic lupus erythematosus (SLE) patients. The resulting network is strongly enriched for genes genetically associated with SLE, and effectively identifies gene modules known to play important roles in SLE etiology, such as increased type 1 interferon response and response to wounding. Our results also strongly support previous preliminary studies suggesting a role for dysregulation of neutrophil extracellular trap formation in SLE. Strikingly, two of the gene modules we identify contain SLE-associated transcription factors that have binding sites significantly enriched in the promoter regions of their respective gene modules, suggesting a possible mechanism underlying the mis-expression of the modules. Thus, our general method is capable of identifying specific dysregulated gene expression programs, as opposed to large global responses. We anticipate that methods such as ours will be more and more useful as gene expression monitoring becomes increasingly common in clinical settings.

*Corresponding author

1. Introduction

Proper control of gene expression is vital to cellular function, and improper control is thought to play a role in many diseases. There are >50,000 datasets currently available in the Gene Expression Omnibus database [1], covering a wide variety of disease states, tissue types, and conditions. Development of methods for extracting relevant information from these rich data remains a limiting step in the longstanding goal to translate them into useful clinical information.

Differential expression analysis is by far the most popular method for identifying genes that are mis-expressed in a disease state, having been utilized in hundreds of studies to date (reviewed in [2-9]). Such analyses apply statistical methods such as t-tests to cohorts of individuals in order to identify genes with high expression levels in the disease state and low expression levels in normal conditions (or *vice versa*) (reviewed in [10]). However, these analyses often result in large lists of genes that must subsequently be further decomposed into specific dysregulated pathways and expression programs. Furthermore, differences in gene expression levels between individuals due to factors such as genotype, disease severity or stage, and age/gender/race may confound results from standard differential expression analysis methods.

In contrast, differential co-expression analysis is a relatively underutilized, complementary procedure capable of revealing subtle but important relationships missed by standard differential expression methods [11-20]. In differential co-expression, pairs of genes are identified that have correlated expression patterns in one state (e.g., disease), but not the other (e.g., normal conditions). Such methods are therefore capable of capturing relationships that might be influenced by differences between individuals within each cohort. Further, groups of genes that are co-expressed only in one state might represent regulatory programs that are specific to that state, such as pathways that are over-active in a disease.

Despite its clear utility, gene expression data contains a large amount of noise, and findings are thus not always reproducible. Moreover, differences in cohorts, treatments, and experimental methodologies can lead to disparate findings between studies. Meta-analysis (statistically combining results from multiple studies) is a particularly useful tool to address these issues and extract relevant signals from multiple related datasets. Indeed, several studies have employed meta-analysis techniques to identify genes that are consistently co-expressed across studies (e.g., [21-24]), or to identify gene pairs that are specifically co-expressed under certain conditions [17]. Notably, disease signatures are largely consistent across studies [25], emphasizing the potential benefits of applying meta-analysis-based techniques to gene expression compendia.

To our knowledge, there is no standard statistical framework for performing meta-analysis of differential co-expression. In this study, we propose a novel statistical framework, and apply it to identify gene regulatory programs that are mis-expressed in systemic lupus erythematosus (SLE), an incurable, debilitating, and potentially fatal disease affecting an estimated 3.5 million people worldwide [26]. Our results offer new insights into the etiology of SLE, highlighting the potential benefits of differential co-expression in the analysis of disease gene expression datasets.

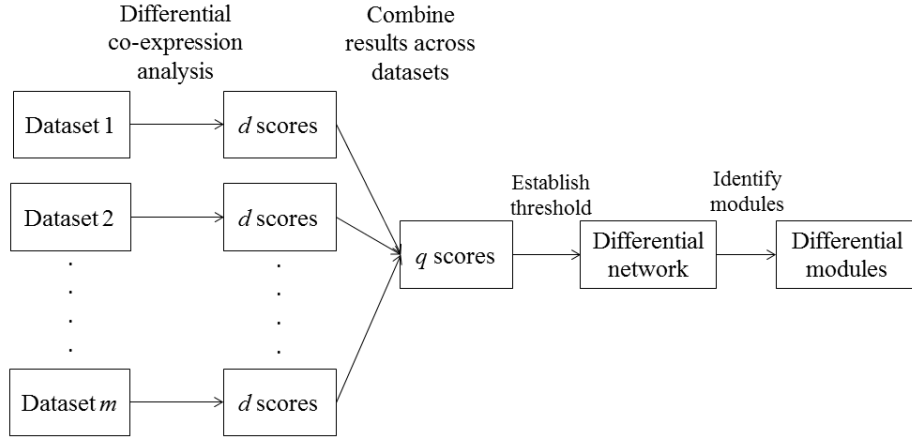


Fig 1. Schematic overview of the method. There are three major components: 1) differential co-expression analysis (e.g. disease vs. control) within each study; 2) combination of results across studies; and 3) gene network and module identification. Differential co-expression analysis of each dataset provides a d score for each gene pair. The d scores are then combined across datasets, resulting in q scores. The q scores are used to construct a differential network, which is decomposed into gene modules representing mis-expressed regulatory programs.

2. Methods

2.1. Statistical framework for meta-analysis of differential co-expression

Let X , Y denote two random variables (e.g., the expression levels of two genes across a collection of people) that follow a bivariate normal distribution under each of the two conditions A, B (e.g., diseased, healthy). Let ρ_A and ρ_B be the correlations between X and Y under conditions A and B, respectively. Let there be m independent studies measuring the two random variables X , Y under both conditions. Under these assumptions, we developed a statistical framework to test the hypothesis that there does not exist a differential relationship between X and Y . The framework takes into account data (e.g., a matrix of gene expression values) from m different studies, along with the sample size of each study. A schematic overview of the method is provided in Fig 1.

2.1.1. s score definition

Within each of the m studies, the sample Pearson correlation coefficient between the two random variables X and Y is calculated as given in Eq. (1):

$$r_{X,Y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

Note that this computation needs to be performed under each condition (A & B) separately.

The Pearson correlation coefficient, r , is then transformed to Fisher's z score using Fisher's transform as given in Eq. (2):

$$z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) \quad (2)$$

Note: $-\infty < z < \infty$.

Thus, within each study, two z scores are calculated for each pair X,Y by transforming the two sample Pearson correlation coefficients, one under each condition.

Within each study, the difference between the z scores is next calculated as given in Eq. (3):

$$d = z_A - z_B \quad (3)$$

Note: d can also be defined as: $d = z_B - z_A$.

Finally, we define the s score as the sum of the d scores from the m studies, and compute it as:

$$s = \sum_{k=1}^m d_k \quad (4)$$

2.1.2. Probability distribution of s scores

Under the assumption that two random variables X,Y follow a bivariate normal distribution with a correlation of ρ , the random variable Z is approximately normally distributed [27, 28]:

$$Z \sim N \left(\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right), \frac{1}{n-3} \right) \quad (5)$$

Note: n is the sample size (e.g., the number of conditions/individuals in the expression dataset).

The random variable D, by the d score definition given in Eq. (3), is then the difference of two random variables Z_A and Z_B . Since Z_A and Z_B are normally distributed random variables and are independent by definition, the random variable D will also follow a normal distribution, with the expected value equal to the difference of the expected values of Z_A and Z_B , and variance equal to the sum of the variances of Z_A and Z_B :

$$D \sim N \left(\frac{1}{2} \ln \left(\frac{1+\rho_A}{1-\rho_A} \right) - \frac{1}{2} \ln \left(\frac{1+\rho_B}{1-\rho_B} \right), \left(\frac{1}{n_A-3} \right) + \left(\frac{1}{n_B-3} \right) \right) \quad (6)$$

We next define a random variable S that is a linear combination of the normal random variables D_1, D_2, \dots, D_m , which are all independent by definition. Thus, using the result stated in Eq. (5), the random variable S will follow a normal distribution as given in Eq. (7).

$$S \sim N\left(m\left(\frac{1}{2}\ln\left(\frac{1+\rho_A}{1-\rho_A}\right) - \frac{1}{2}\ln\left(\frac{1+\rho_B}{1-\rho_B}\right)\right), \sum_{k=1}^m\left(\frac{1}{n_{A_k}-3} + \frac{1}{n_{B_k}-3}\right)\right) \quad (7)$$

Further, under the assumption that there does not exist a differential relationship between X and Y , ρ_A will be equal to ρ_B , and the expected value of the random variable S will be zero.

The random variable S can then be converted to a standard normal random variable, denoted Q :

$$Q = \frac{S-0}{\sqrt{\sum_{k=1}^m\left(\frac{1}{n_{A_k}-3} + \frac{1}{n_{B_k}-3}\right)}} \quad (8)$$

Since Q is a standard normal random variable, for any observed s score, the corresponding q score can be calculated using the result stated in Eq. (8). The P-value ($P > |q|$) can then be calculated easily using the standard normal distribution function. We note that this P-value provides a way to test the hypothesis that there does not exist a differential relationship between X and Y .

2.2. SLE dataset selection for meta-analysis

We selected a total of five gene expression studies after screening several based on cell type, sample size, patient treatment, etc. (Table 1). All five studies measure gene expression in Peripheral Blood Mononuclear Cells (PBMC) of SLE patients and healthy individuals. We excluded any samples from individuals that received intensive immunosuppressive therapy (e.g., greater than 20 mg/day of steroids or interferon alpha Kinoid immunization).

2.3. Data pre-processing

Data pre-processing consisted of two operations: normalization and probe ID mapping. For three datasets, only pre-normalized data were available, which we used as-is. We quantile normalized the remaining two datasets, using the *normalize.quantile* function from the *preprocessCore* R library. We mapped gene probe IDs from each study to Ensembl gene IDs. In the cases where multiple probes mapped to one Ensembl ID (e.g. for genes with multiple splice variants), we used the mean expression value of the probes. These mappings resulted in a total of 18,663 unique genes across the five studies.

Table 1. Gene expression datasets used in this study

Database	ID	Study Title	N (SLE)	N (CTRL)
GEO	GSE22098	Whole blood transcriptional profiles of patients with active tuberculosis (TB) and other inflammatory and infectious diseases	110	81
GEO	GSE20864	Human peripheral blood cells: systemic lupus erythematosus vs healthy individual	21	45
GEO	GSE39088	Down-regulation of Interferon signature in systemic lupus erythematosus patients by active immunization with Interferon alpha-Kinoid	26	46
GEO	GSE8650	Blood Leukocyte Microarrays to Diagnose Systemic Onset Juvenile Idiopathic Arthritis and Follow IL-1 blockade	13	21
Array Express	E-MTAB-145	Transcription profiling of human separated leukocyte subsets in SLE and vasculitis	13	25

2.4. Meta-analysis of differential gene co-expression in SLE

We performed differential gene co-expression meta-analysis using the framework described in section 2.1, and outlined in Fig 1. Here, the random variables X , Y represent the expression levels of two genes across a collection of individuals, and conditions A and B represent diseased (SLE) and healthy (control) conditions, respectively. Note that in applying the meta-analysis framework to SLE, we converted any negative z scores to a z score of zero, since negative correlations are non-transitive, and thus might have adverse effects on subsequent clustering analyses. This is a conservative modification, in the sense that it may lower the d score, but can never increase it; this modification thus will not result in any false positives in our network. The 18,663 unique genes translate into 174,144,453 pairwise relationships. However not all genes are present in all studies, and not all genes present in a study have the minimum number (3) of non-missing observations we required to compute correlation coefficients. In this study, we only considered gene pairs for which we could compute correlation coefficients in at least four studies, a total of 85,276,272 pairs (~49%). We used this value to adjust our final P-values for multiple testing with the Benjamini-Hochberg procedure, using *p.adjust* function in R, requiring a significance level of 0.05.

2.5. Comparison of Type I error rates and statistical power

We empirically estimated Type I error rate and statistical power using simulation-based approaches. We performed 10,000 iterations wherein we created between 2 and 10 studies, with each study consisting of between 25 and 100 samples/observations (all values were chosen uniformly). To create expression levels for two genes across the conditions, we picked random values independently from two standard normal distributions for the Type I error rate analysis, since its null hypothesis assumes no correlation. For the power analysis, which has the null hypothesis that there is some underlying correlation, we simulated expression levels by picking from a bivariate normal distribution with expected values [0, 0], variances [1, 1], and correlation (ρ) values drawn from a uniform (0,1) distribution. We calculated the sample Pearson correlation between the two genes across conditions, and performed meta-analysis using one of the three methods (ours, Fisher's, or Stouffer's). Type I error rate and power were then calculated as the fraction of meta-analysis results across the 10,000 iterations with P-value < 0.05 (i.e., $\alpha=0.05$). We repeated this process 100 times and report the mean.

2.6. Differential expression analysis

We also performed standard differential gene expression analysis. For each dataset, we used a two sample independent t-test to estimate the significance of the difference of the expression levels of each gene between SLE and control conditions, and calculated the mean across the five datasets,.

2.7. Identification and annotation of gene modules

We identified densely connected regions of the final network (gene modules) using the fastgreedy algorithm in the R *igraph* package. We annotated modules by calculating their overlap with known biological processes using Fisher's exact test in ToppGene [29]. We identified enriched transcription factor (TF) binding sites in 1,000 base upstream regions of genes in each module using a large collection of human TF binding motifs [30] and the Pscan algorithm [31]. We identified TFs with enriched ChIP-seq binding peaks in the promoters of module genes using data obtained from UC Santa Cruz [32] and PAZAR [33], and a novel statistical method. Briefly, for each ChIP-seq dataset, we calculated its observed overlap with the promoter regions of the genes of the module. We then built a distribution of expected overlap values from 1,000 randomly selected promoters. The distribution of the overlap scores from the randomized data resembles a normal distribution (not shown), which we used to generate a Z-score and P-value for the observed number of peaks that overlap each module. P-values were corrected using Bonferroni's method.

2.7. Availability

Source code is available at <https://tf.cchmc.org/pubs/makashir2014/>. Full results are available from MTW upon request.

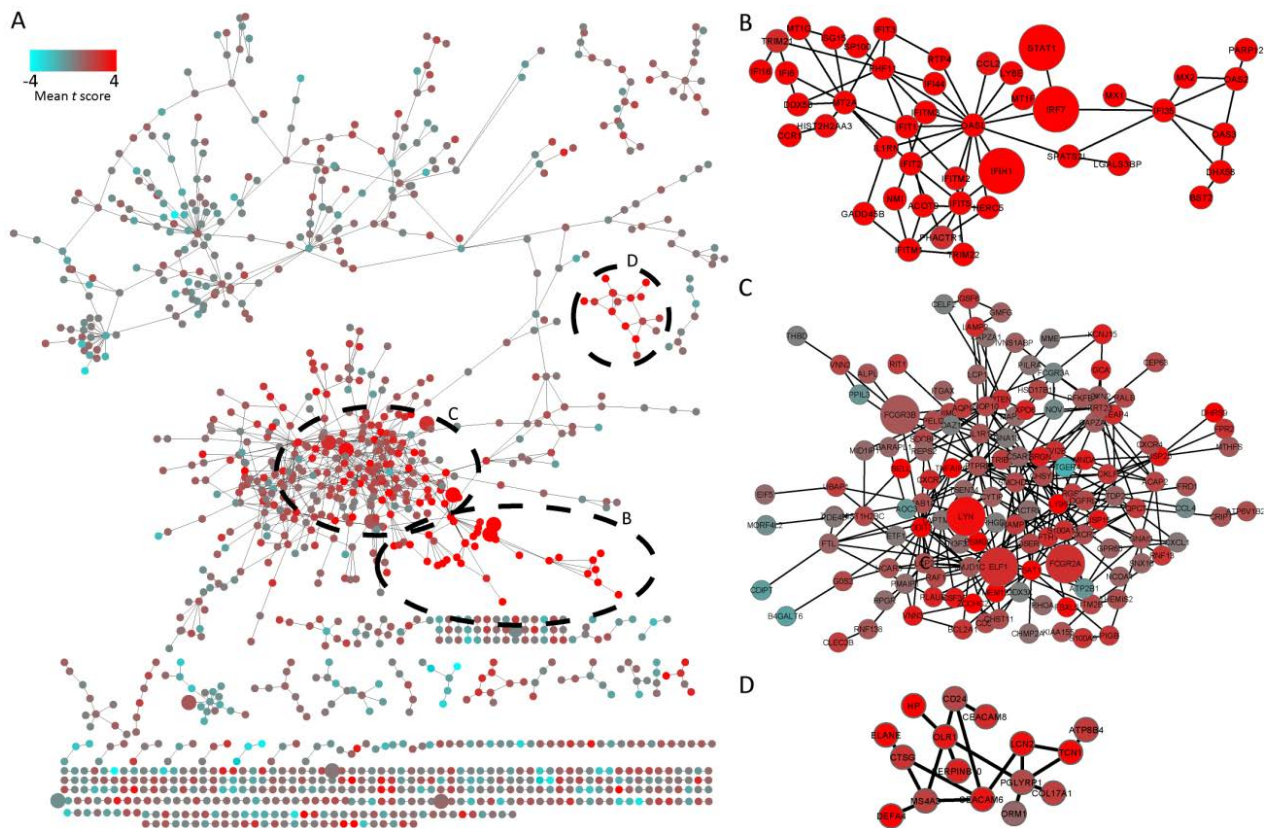


Fig.2. Global view of the SLE-specific gene co-expression network. Each node represents a gene, with edges between genes indicating co-expression relationships specific to SLE. Nodes are colored based on standard differential expression analysis; red indicates over-expression, while blue indicates under-expression in SLE, compared to healthy controls (see key, upper left). Large nodes indicate genes genetically associated with SLE. The network has been laid out such that highly connected groups of nodes are located near each other in 2D space. A. Full network. Dashed circles indicate gene modules highlighted in panels B-D and discussed in the text.

3. Results

3.1. A network of genes specifically co-expressed in SLE

Using simulations, we empirically estimated the Type I error rate of our method to be 0.0496, which is comparable to that of two standard statistical meta-analysis procedures, Fisher's method (0.0494), and Stouffer's method (0.0496) (see Methods). Further, our method has significantly greater statistical power than these methods (0.8798 vs. 0.8527 and 0.8534 for Fisher and Stouffer, respectively; $P < 10^{-16}$, paired t-test with Bonferroni multiple testing correction; see Methods).

We applied our procedure to identify pairs of genes specifically co-expressed in blood samples from SLE patients, relative to controls. The resulting network, which contains 1,250 nodes and 1,310 links, is depicted in Fig 2A. Edges in the network indicate pairs of genes that are strongly co-expressed in SLE patients, but not in normal individuals, and thus might indicate regulatory relationships specific to the disease state.

Twelve genes contained in our network are genetically associated with SLE, a 2.8-fold enrichment over random expectation ($P \sim 0.001$), and a near identical number to that obtained from a list of the top 1,250 differentially expressed genes (14). Interestingly, only seven of these genes are identified by both methods, highlighting the complementary nature of differential expression and co-expression based approaches. We note that we used a strict threshold to generate the network, resulting in a largely sparse network. Notwithstanding, there is a clear modularity to the network, with one large component containing mostly down-regulated genes (top), a second large component containing mostly up-regulated genes (middle), and several smaller components. We therefore used a graph-theoretic algorithm to extract gene modules from the network (see Methods), which might correspond to specific regulatory programs that are mis-expressed in SLE.

3.2. Gene co-expression modules specific to SLE

3.2.1. Type I interferon response

The most strongly up-regulated genes in the network comprise a single gene module enriched for genes involved in the Type I interferon response ($P < 10^{-29}$), a process with a well-established history of elevated levels in the sera of SLE patients [34, 35]. Specifically, the module contains many interferon inducible genes, including OAS family members, and *IFIT1*, *IFIT5*, *IFIH1*, and *MX1* [36, 37] (Fig 2B). Intriguingly, this module also contains three SLE-associated genes (*STAT1*, *IRF7*, and *IFIH1*), all of which have risk variants associated with an increase in Type I IFN activity [37-39]. Given that *STAT1* and *IRF7* are transcription factors (TFs) with well-known roles in the interferon response [40, 41] and established elevated expression and activity in SLE [42, 43], we hypothesized that their heightened activity in SLE might play a role in the elevated expression of Type I interferon genes in SLE. Indeed, two lines of evidence strongly support this possibility. First, the most strongly enriched TF binding motifs in the promoter regions of the genes of this module are for IRF3 (highly related to IRF7) and *STAT1* ($P < 10^{-15}$ and 10^{-11} , respectively, see Methods). Second, *STAT1* and *IRF1* ChIP-seq binding peaks are over 200-fold enriched in the promoters of these genes ($P < 10^{-186}$ and 10^{-177} , respectively), covering 25 and 17 of the 44 genes in the module, respectively (see Methods) (we could not locate ChIP-seq data for *IRF7*; *IRF1* is highly related, and recognizes a highly related DNA binding motif). In summary, our data suggest that this Type I interferon response gene module is specifically over-expressed in SLE patients as a consequence of the elevated activity of *STAT1* and *IRF7*.

3.2.2. Cell movement and response to wounding

Our method also identified a module strongly enriched for genes involved in response to wounding ($P < 10^{-7}$) and leukocyte cell migration ($P < 10^{-6}$) (Fig 2C). Four of the genes contained in this module are encoded in regions genetically associated with SLE (*ELF1*, *LYN*, *FCGR2A*, and *FCGR3B*). ELF1 is a TF, and we again found that ELF1's DNA binding motif is significantly enriched in the promoters of the genes in this module ($P < 10^{-3}$). Strikingly, ELF1 ChIP-seq binding events in B cells are present in the promoters of 64 of the 134 genes in this module (>12-fold enrichment, $P < 10^{-80}$). Although the precise role of cell migration in SLE has yet to be fully elucidated, recent studies in mice have demonstrated that blocking chemotaxis of immune effector cells into organs can inhibit immune-mediated damage such as nephritis [44]. Further analysis of the genes in this module will likely reveal the role played by ELF1 in the dysregulated co-expression of mediators of cell migration, and the contribution of this pathway to the pathoetiology of SLE.

3.2.3. Immune defense against extracellular organisms

Finally, we identified a gene module containing genes with elevated expression levels involved in immune defense against extracellular organisms such as response to bacterium, fungus, and symbiotic cells ($P < 10^{-4}$) (Fig 2D). In particular, twelve out of sixteen (75%) of these genes have established roles in neutrophil maturation and extracellular trap formation (including *DEFA4*, *CTSG*, *ELANE*, *PGYRP1*, *OLR1*, and *CD24*) [45-47]. Among these, *DEFA4* (a defensin involved in neutrophil extracellular traps) is particularly intriguing, since it is known to impact some of the immune-mediated tissue damage that occurs in response to autoantibody deposition in SLE [48]. Further, both *DEF4A* and *CTSG* (a serine protease that is also released by neutrophils during neutrophil extracellular trap formation) are expressed at higher levels in SLE patients with active disease [49]. Together, these results support recent studies suggesting a role for neutrophil extracellular traps in SLE etiology [50] and, importantly, identify specific members of this pathway whose regulatory programs might be compromised in SLE patients.

4. Discussion

Despite numerous advantages to differential co-expression-based approaches, methods such as ours remain largely underutilized. Here, we present a novel general differential co-expression meta-analysis framework, and show how it can be used in conjunction with differential expression techniques, transcription factor binding analysis, and genetic association information to identify dysregulated gene expression programs in SLE. Notably, our method identifies several regulatory programs not captured in a previous meta-analysis study of pathway differential expression in SLE

(both capture the interferon response) [51]. We anticipate that integrated methods such as ours will have increasing utility as data from genome-scale technologies become more and more prevalent. Notably, results from our method offered considerable biological insights, despite data limitations relating to differing study designs, differing array platforms, and heterogeneous cell populations. It is likely that applications to well-controlled datasets (e.g., single population or single cell RNA-seq data) will afford even higher resolution views into regulatory programs that are disrupted in human diseases.

5. Conclusions

We developed a novel statistical methodology for differential gene co-expression meta-analysis. Our method is complementary to traditional differential expression methods, and is capable of identifying specific dysregulated expression programs, as opposed to large global responses. Our framework is flexible and general, and thus can be applied to other diseases (e.g. cancer), or other genome-wide data types (e.g. RNA-seq). Here, application to SLE captured a significant number of genetically associated genes, and provided new insights into SLE etiology. Specifically, our data implicate the STAT1 and IRF7 TFs in the elevated Type I Interferon response in SLE, and suggest a role for the ELF1 TF in a novel dysregulated cell movement pathway. Importantly, these predictions are strongly supported by enrichment of binding sites and ChIP-seq binding events for these TFs in the promoters of the associated gene modules. Further, we have revealed a new potential role for the dysregulation of specific neutrophil extracellular trap genes in SLE. As technologies for measuring gene expression continue to mature, we anticipate that meta-analysis methods such as ours will play a vital role in obtaining high-confidence biological insights into the gene expression programs of many human diseases.

5. Acknowledgments

We thank Xiaoting Chen and Siddharth Dixit for computational support, and Josh Stuart for helpful discussions.

References

1. Barrett, T., et al. *Nucleic Acids Res.* **41**(Database issue):D991-5 (2013).
2. Jakobs, T.C. *Cold Spring Harb Perspect Med.* **4**(7) (2014).
3. Ruppert, V. and B. Maisch. *Herz.* **37**(6):619-26 (2012).
4. Bacher, U., A. Kohlmann, and T. Haferlach. *Cancer Treat Rev.* **36**(8):637-46 (2010).

5. Kinter, J., T. Zeis, and N. Schaeren-Wiemers. *Int MS J.* **15**(2):51-8 (2008).
6. Fehrmann, R.S., et al. *Oncologist.* **12**(8):960-6 (2007).
7. Nanni, L., et al. *J Mol Cell Cardiol.* **41**(6):934-48 (2006).
8. Lemmer, E.R., S.L. Friedman, and J.M. Llovet. *Semin Liver Dis.* **26**(4):373-84 (2006).
9. Mandel, M. and A. Achiron. *Lupus.* **15**(7):451-6 (2006).
10. Cui, X. and G.A. Churchill. *Genome Biol.* **4**(4):210 (2003).
11. Kostka, D. and R. Spang. *Bioinformatics.* **20 Suppl 1**:i194-9 (2004).
12. Choi, J.K., et al. *Bioinformatics.* **21**(24):4348-55 (2005).
13. Wang, B.D., et al. *Mol Cancer.* **9**:98 (2010).
14. Fu, S., X. Pan, and W. Fang. *Mol Med Rep.* **10**(2):713-8 (2014).
15. Lai, Y. *Bioinformatics.* **24**(5):666-73 (2008).
16. de Jong, S., et al. *PLoS One.* **7**(6):e39498 (2012).
17. Gillis, J. and P. Pavlidis. *BMC Bioinformatics.* **10**:306 (2009).
18. Fang, G., et al. *Pac Symp Biocomput.* 145-56 (2010).
19. Li, K.C. *Proc Natl Acad Sci U S A.* **99**(26):16875-80 (2002).
20. Oldham, M.C., S. Horvath, and D.H. Geschwind. *Proc Natl Acad Sci U S A.* **103**(47):17973-8 (2006).
21. Kim, S.K., et al. *Science.* **293**(5537):2087-92 (2001).
22. Lee, H.K., et al. *Genome Res.* **14**(6):1085-94 (2004).
23. Guan, Y., et al. *PLoS Comput Biol.* **4**(9):e1000165 (2008).
24. Stuart, J.M., et al. *Science.* **302**(5643):249-55 (2003).
25. Dudley, J.T., et al. *Mol Syst Biol.* **5**:307 (2009).
26. Danchenko, N., J.A. Satia, and M.S. Anthony. *Lupus.* **15**(5):308-18 (2006).
27. Fisher, R.A. *Biometrika.* **10**:507-521 (1915).
28. Fisher, R.A. *Metron* **1**:3-32 (1921).
29. Chen, J., et al. *Nucleic Acids Res.* **37**(Web Server issue):W305-11 (2009).
30. Weirauch, M.T., et al. *Cell.* **158**(6):1431-43 (2014).
31. Zambelli, F., G. Pesole, and G. Pavesi. *Nucleic Acids Res.* **37**(Web Server issue):W247-52 (2009).
32. Rosenbloom, K.R., et al. *Nucleic Acids Res.* **41**(Database issue):D56-63 (2013).
33. Portales-Casamar, E., et al. *Nucleic Acids Res.* **37**(Database issue):D54-60 (2009).
34. Hooks, J.J., B. Detrick-Hooks, and A.I. Levinson. *J Am Vet Med Assoc.* **181**(10):1111-4 (1982).
35. Hooks, J.J., et al. *N Engl J Med.* **301**(1):5-8 (1979).
36. Ghodke-Puranik, Y. and T.B. Niewold. *Int J Clin Rheumatol.* **8**(6) (2013).
37. Kariuki, S.N., et al. *J Immunol.* **182**(1):34-8 (2009).
38. Mavragani, C.P., et al. *Front Immunol.* **4**:238 (2013).
39. Niewold, T.B. *J Interferon Cytokine Res.* **31**(12):887-92 (2011).
40. Darnell, J.E., Jr., I.M. Kerr, and G.R. Stark. *Science.* **264**(5164):1415-21 (1994).
41. Marie, I., J.E. Durbin, and D.E. Levy. *EMBO J.* **17**(22):6660-9 (1998).
42. Dong, J., et al. *Lupus.* **16**(2):101-9 (2007).
43. Lin, L.H., P. Ling, and M.F. Liu. *J Rheumatol.* **38**(9):1914-9 (2011).
44. Bignon, A., et al. *J Immunol.* **192**(3):886-96 (2014).
45. Parlato, M., et al. *J Immunol.* **192**(5):2449-59 (2014).
46. Ghosh, A., et al. *Invest Ophthalmol Vis Sci.* **50**(9):4185-91 (2009).
47. Martinelli, S., et al. *J Biol Chem.* **279**(42):44123-32 (2004).
48. Villanueva, E., et al. *J Immunol.* **187**(1):538-52 (2011).
49. Tamiya, H., et al. *Rheumatol Int.* **27**(2):147-52 (2006).
50. Bouts, Y.M., et al. *Autoimmunity.* **45**(8):597-601 (2012).
51. Arasappan, D., et al. *BMC Med.* **9**:65 (2011).