Cell Systems

Supplemental Information

Master Regulators of Infiltrate Recruitment in Autoimmune Disease Identified through Network-Based Molecular Deconvolution

James C. Chen, Jane E. Cerise, Ali Jabbari, Raphael Clynes, and Angela M. Christiano

Supplemental Figure Legends

S1, Enriched pathways in the AAGS, Figure1. Supplemental Ingenuity Pathway Analysis shows enrichment of immune and cytotoxic signaling cascades for both infiltrating populations and end organ processes within the AAGS.

S2, AD and Ps disease gene signatures, Figure6. Unsupervised hierarchical clustering of lesional and unaffected patient samples using gene expression. Patients cleanly segregate by clinical presentation in both psoriasis (A) and atopic dermatitis (B) using the associated gene expression signatures. Sample dendrograms are provided here for reference for the heatmaps provided in figure 6.

S3, Cytotoxicity assays, Figure4. Optimizations of PBMC concentration (A) and time window (B) for cytotoxicity assays identify a PBMC:target ratio of 100:1 and a time of at least 6 hours to achieve optimal separation.

Supplemental Tables

S1, AAGS functional categories, Figure1. This table shows all genes that are in the original AAGS. Genes that survived de-convolution are marked with an asterisk. P-values and fold-changes comparing AA patients vs controls for each gene are provided.

S2, AAGS signature, Figure2 These tables synopse the statistically enriched GO pathways (bold titles) and the genes in the AAGS that comprise each category. Immune-specific category enrichments are lost following de-convolution.

S3, Directional modeule, Figure 5. This table details the IKZF1 and DLX4 edges of the ARACNe regulatory network (MI values provided) that were validated by exogenous expression of the MRs according to the criteria detailed in Methods.

S4, AD/PS signatures, Figure6. This table lists the deconvolved psoriatic and atopic dermatitis signatures used in the comparative master regulator analysis in figure 6 as EntrezIDs.

Extended Experimental Methods

Gene expression studies

A total of 122 samples from 96 patients were profiled on the Affymetrix U133 2Plus array consisting of 28 AAP patients, 32 AT/AU patients, and 36 unaffected controls. The remaining 26 samples correspond to patient-matched non-lesional biopsies from the AAP cohort. These non-lesional samples were not included in the inference of an initial signature, but used later (below). RNA from these patient biopsies was isolated and processed on the Affymetrix U133 2Plus array. Data post-processing was done via R using MAS5 normalization(Giorgi, Bolger, Lohse, & Usadel, 2010) with standard

packages available through Bioconductor. These data are available at the Gene Expression Omnibus as GSE68801. This dataset was broken into two sets for training and validation.

An initial panel of gene markers was identified by two differential expression analyses comparing (1) AA vs unaffected and (2) lesional vs non-lesional in the training set. A threshold was set for differential expression at $p<0.05$ and a fold change>25%. This relatively lax threshold was implemented because the network analyses are based on consensus. The analysis is not primarily concerned with candidate ranks, but instead relies on having enough molecular information to infer TF activity. This approach is also necessarily more robust to noise that could be introduced by a more relaxed threshold, since the addition of noise would be applied across the entire dataset and normalized out of the consensus by both ARACNe and master regulator analysis (see below) (Margolin, Nemenman, et al., 2006a; Margolin, Wang, et al., 2006b). All X- and Y-linked genes were additionally removed to remove any possible gender bias in the ranking and clustering of differentially expressed genes.

Gene Set Enrichment Analysis

GSEA is a method for measuring nonparametric statistical enrichment in the differential expression of a defined panel of genes(Subramanian et al., 2005). A default differential expression analysis between experimental and control cohorts done, and genes are rank-sorted by differential expression with no threshold (all genes included). This can be done according to any user-specified criteria (fold-change, p-value, etc).

This enrichment score is then compared to an empirically generated null distribution by shuffling sample labels, *i.e.,* by randomizing case and control samples and repeating the analysis. This is repeated over 1000 iterations to generated a null distribution of Enrichment Scores, which the observed score can be compared against to generate a p-value.

Cloning

Each primer pair provided below was used in PCR reactions with the Accuprime Taq PCR mixes according to manufacturer protocols on cDNAs derived from HEK293T cells. cDNAs were generated from cultured cells using the SuperScript First-Strand Synthesis System from Invitrogen. PCR products were run out by gel electrophoresis, and any isoforms present were separately excised using the Qiagen Gel Extraction Kit.

mRNA fidelity was verified via sequencing from Genewiz, and correct sequences were digested with the appropriate enzymes (SPEI and ASCI) from New England Biosystems in SmartCut buffer for 2 hours. The pLOC-RFP vector was digested in parallel, and the cut backbone was excised by gel extraction. After purification of the backbone and inserts, each insert was ligated into the cut pLOC vector using the RapidLigation Kit from Roche, according to manufacturer protocols and transformed into DH5α cells for amplification.

Successful transformations were validated for sequence fidelity via colony PCR and sequencing (Genewiz). Correct constructs were amplified and purified by Maxiprep (Qiagen) for experiments

Primers used to clone genes for insertion into the pLOC vector are provided below in the following format, 5' to 3': *spacer-enzyme-mRNAsequence.*

IKZF1.1

Forward GGC-ACTAGT-ATGGATGCTGATGAGGGTCAA

Reverse ATT-GGCGCGCC-TTAGCTCATGTGGAAGCGGT

IKZF1.2

Forward GGC-ACTAGT-ATGGATGCTGATGAGGGTCAAG

Reverse ATT-GGCGCGCC-TTAGCTCATGTGGAAGCGGT (identical to 1.1)

DLX4

Forward GGC-ACTAGT-ATGAAACTGTCCGTCCTACCCC

Reverse ATT-GGCGCGCC-TCATTCACACGCTGGGGCTGG

Cell culture and transfections

Both huDP and HK cells were kept in standard conditions for growth: DMEM 10%FBS at 37C and 5%CO2. huDP cells are cultured primary human dermal papillae that were microdissected from human skin samples. For the experiments in this work, only huDP and HK cells with a passage number <6 were used.

Cells were transformed with pLOC expression constructs using the JetPRIME transfection reagent according to manufacturer protocols. Transfections were allowed to carry overnight using a 2:1 concentration of reagent (ul) to DNA (ug).

Microarrays of MR rescue

Transfections of IKZF1 and DLX4 into HK and huDP cells were carried out as described above in cells cultured in 10cm plates. 36 hours post-transfection these cells were harvested in PBS with a cell scraped, then lysed and processed for purified RNA using the RNeasy kit from Qiagen following manufacturer protocols. RNA quality control was done using a spectrometer and submitted for processing on the Affymetrix human U133 2Plus array by the Columbia facility (Pathology Department). Array data was again normalized and processed using MAS5 normalization through the Bioconductor package in R.

qPCRs

Quantitative PCR reactions were performed on cDNAs extracted from an independent cohort of eight primary lesional biopsies (one was found to be degraded and was excluded from the study), four unaffected controls, and five pairs of patient-matched lesional and non-lesional samples. Reaction mixes using SYBR Green were made in 25ul volumes according to manufacturer protocols and analyzed on a 7300 series Real Time PCR Machine from Applied Biosystems. Primers for each gene are provided at the end of this section.

All samples were tested in technical triplicates in stamp-plate format (each replicate was performed on one plate, with all samples and controls prepared at once, repeated three times). Data from these replicates was analyzed via the δδCT method, normalizing all experimental series to the average normalized values of the control tissues. The SEM was derived across the comparisons using standard statistical error propagation.

Primers for assaying transcripts by qPCR are provided below, 5' to 3'. The primers for full-length amplification of DLX4 were used because the transcript is ~300 bp (the optimal transcript length for our provided protocol is 200-300 bp).

IKZF1

Forward ACTCCGTTGGTAAACCTCAC

Reverse CTGATCCTATCTTGCACAGGTC

DLX4

same as cloning primers

ACTB

Forward GAAGGATTCCTATGTGGGCGAC

Reverse GGGTCATCTTCTCGCGGTTG

Isolating fresh Peripheral Blood Mononuclear Cells

Fresh PBMCs were isolated from whole blood draws the evening before the intended cytotoxicity assays. PBMCs were separated from whole blood using the Histopaque-1077 reagent (Ficoll) by diluting 8-ml aliquots of whole blood in sterile PBS 1:1, and layering that solution over Ficoll at a final volumetric ratio of 2:1. This solution was centrifuged at 1200 rpm for 45 minutes. The monocyte-bearing interface layer was isolated, diluted in 5x volumes of sterile PBS and centrifuged again for 15 minutes at 1500 rpm. Supernatant was discarded, and the pellet was resuspended in 3ml of DMEM 10%FBS. Cell count was performed with a hemocytometer and the solution was diluted to a final concentration of $1x10^6$ cells per ml with DMEM 10%FBS. This was stored overnight at 37C and 5% CO2 for the experiments next-morning.

REFERENCES

Giorgi, F. M., Bolger, A. M., Lohse, M., & Usadel, B. (2010). Algorithm-driven artifacts in median Polish summarization of microarray data. *BMC Bioinformatics*, *11*(1), 553. http://doi.org/10.1186/1471-2105-11-553

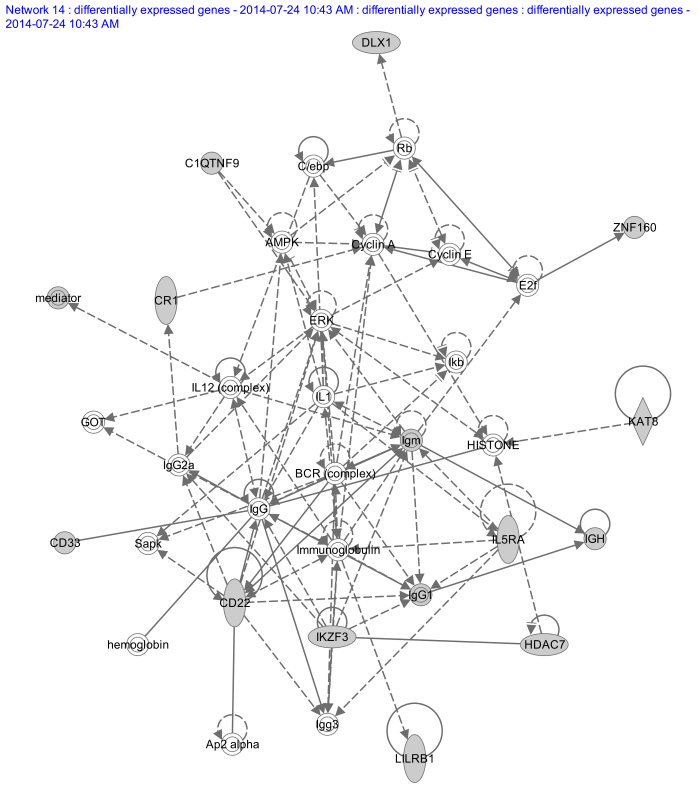
Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla-Favera, R., & Califano, A. (2006a). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, *7 Suppl 1*(Suppl 1), S7. http://doi.org/10.1186/1471-2105-7-S1-S7

Margolin, A. A., Wang, K., Lim, W. K., Kustagi, M., Nemenman, I., & Califano, A. (2006b). Reverse engineering cellular networks. *Nature Protocols*, *1*(2), 662–671. http://doi.org/10.1038/nprot.2006.106

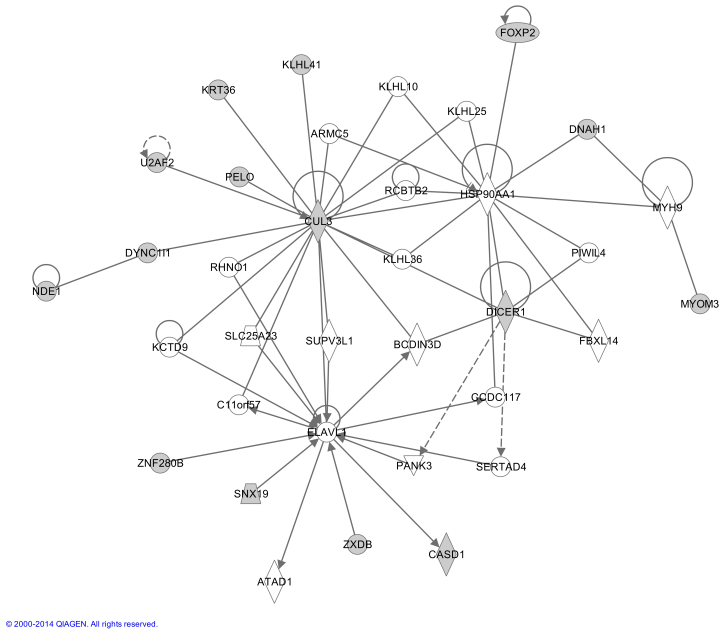
Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550. http://doi.org/10.1073/pnas.0506580102

Supplemental Figure 1. Differentially expressed genes regulated by MRs include many membrane-bound, cell death- and Immune-associated proteins

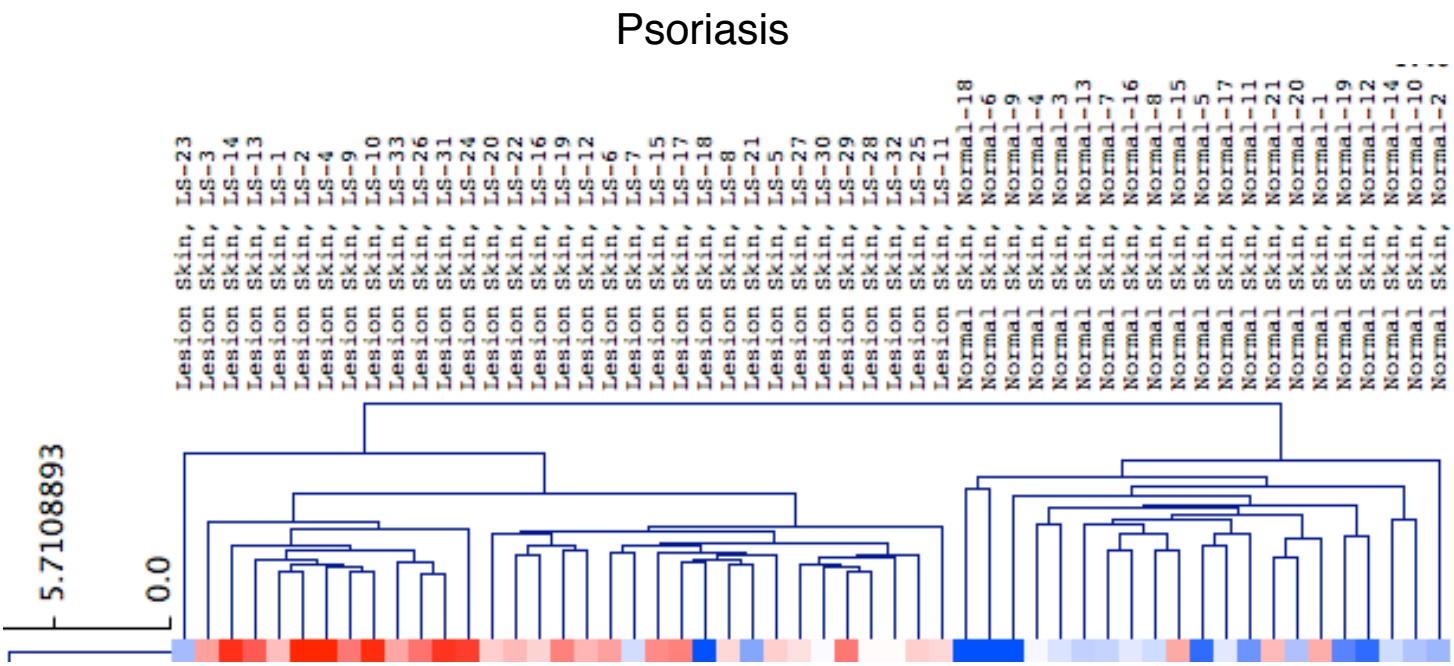
Immune interaction network

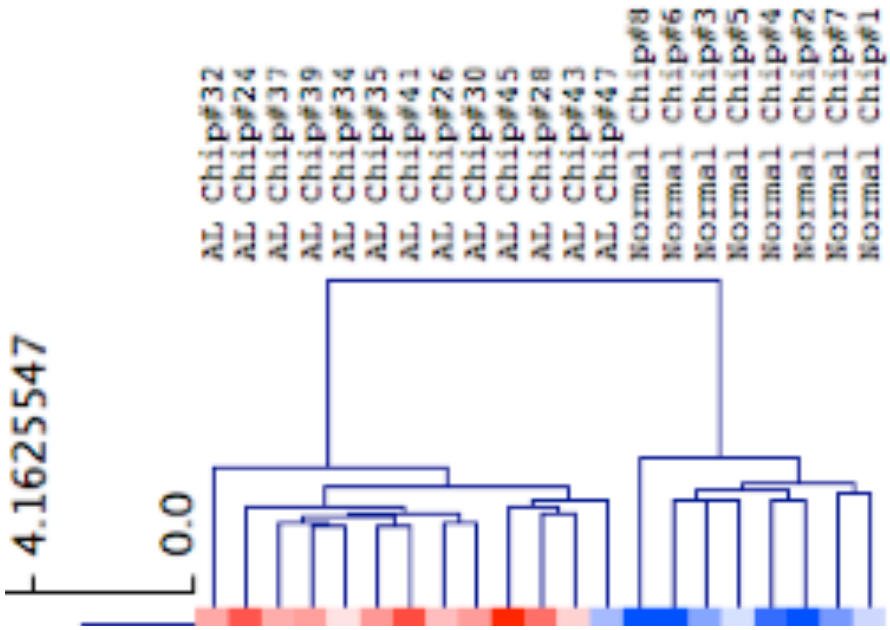


Cytotoxic interaction network



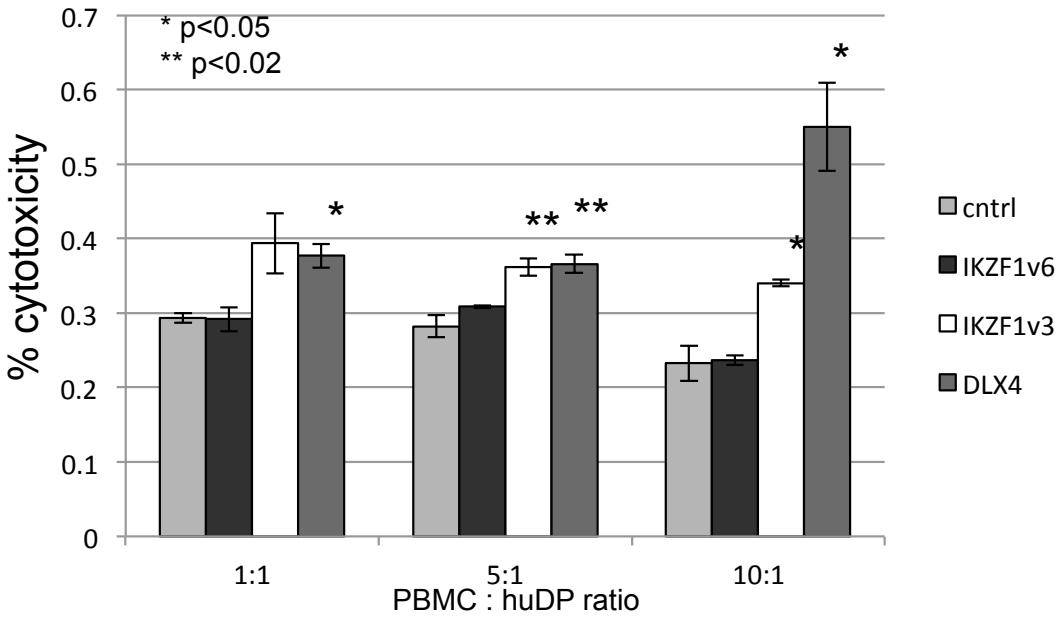
Supplemental Figure 2. Psoriasis and Atopic Dermatitis cohorts have gene expression signatures that clearly delineate patients from unaffected controls





Supplemental Figure 3. Optiization of PBMC-dependent cytotoxicity assays for PBMC concentration and time of exposure to cultured human dermal papillae

A



B