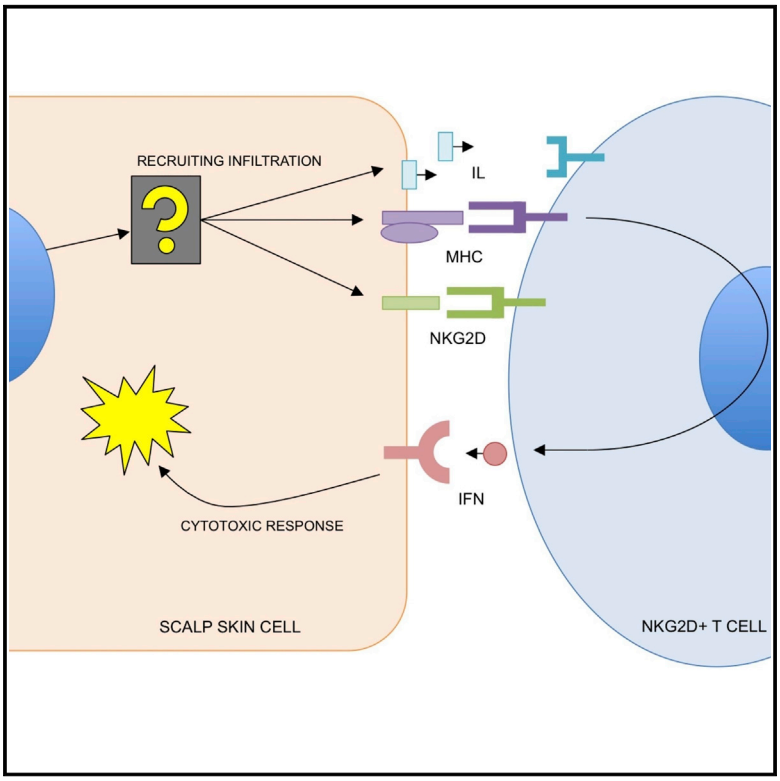
Article

Cell Systems

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

Graphical Abstract



Authors

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

Correspondence

amc65@columbia.edu

In Brief

Modeling complex diseases with pathology involving interacting tissues is difficult when using reverse-engineered regulatory networks due to the presence of different networks in a biopsy sample containing multiple cell types. Chen et al. introduce an approach for deconvolving regulatory modules that originate uniquely in one tissue and apply it to identify IKZF1 and DLX4 as key transcriptional regulators in Alopecia Areata.

Highlights

- Deconvolution of mixed tissue signatures using context-specific regulatory networks
- IKZF1 and DLX4 are master regulators of Alopecia Areata pathogenesis
- IKZF1 and DLX4 induce Alopecia Areata-like immune infiltration in cultured cells
- This framework is applicable to other autoimmune diseases

Chen et al., 2015, Cell Systems *1*, 326–337
November 25, 2015
http://dx.doi.org/10.1016/j.cels.2015.11.001

CellPress

Cell Systems
Article

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

James C. Chen,[1,2] Jane E. Cerise,[1] Ali Jabbari,[1] Raphael Clynes,[1] and Angela M. Christiano[1,3,*]
[1]Department of Dermatology, Columbia University, 1150 Saint Nicholas Avenue, New York, NY 10032, USA
[2]Department of Systems Biology, Columbia University, 1130 Saint Nicholas Avenue, New York, NY 10032, USA
[3]Department of Genetics and Development, Columbia University, 701 West 168th Street, New York, NY 10032, USA
*Correspondence: amc65@columbia.edu
http://dx.doi.org/10.1016/j.cels.2015.11.001

SUMMARY

Network-based molecular modeling of physiological behaviors has proven invaluable in the study of complex diseases such as cancer, but these approaches remain largely untested in contexts involving interacting tissues such as in autoimmunity. Here, using Alopecia Areata (AA) as a model, we have adapted regulatory network analysis to specifically isolate physiological behaviors in the skin that contribute to the recruitment of immune cells in autoimmune disease. We use context-specific regulatory networks to deconvolve and identify skin-specific regulatory modules with IKZF1 and DLX4 as master regulators (MRs). These MRs are sufficient to induce AA-like molecular states in vitro in three cultured cell lines, resulting in induced NKG2D-dependent cytotoxicity. This work demonstrates the feasibility of a network-based approach for compartmentalizing and targeting molecular behaviors contributing to interactions between tissues in autoimmune disease.

INTRODUCTION

Systems-level analysis using reverse-engineered regulatory networks is an emerging computational discipline that has demonstrated great promise in the study of complex diseases such as cancer and Alzheimer's disease (Aubry et al., 2015; Chen et al., 2014; Shelanski et al., 2015; Zhang et al., 2013). This approach enables the modeling of complex physiological behaviors as modules of genes (subsets of differentially expressed genes that associate with disease) that are controlled by master regulators (MRs). MRs represent the minimal number of transcription factors (TFs) that are predicted to specifically activate or repress a target module and, by extension, the associated physiological behavior (Carro et al., 2010; Lefebvre et al., 2010). They can be regarded as molecular "switches" that regulate physiological behaviors. The inference of MRs is made possible through the reverse engineering of context-specific regulatory networks using computational algorithms such as ARACNe (Margolin et al., 2006a).

These MRs are validated biologically and serve as targetable "hubs" governing disease pathology. These approaches have proven highly effective for the study of cell autonomous behaviors in diseases such as cancer. Physiological behaviors such as mesenchymal transformation in glioblastoma (Carro et al., 2010) and oncogenesis in B cell lymphoma or breast cancer (Chen et al., 2014; Mani et al., 2008), as well as onset of Alzheimer's disease (Zhang et al., 2013) have been functionally linked to a relatively small number of MRs, which in turn become the "bottleneck" that can be used to infer driver mutations in patients (Chen et al., 2014) or become the targets of drug screens for treatment (Shelanski et al., 2015).

However, this type of computational approach is only starting to be implemented to target pathogenic, non-cell autonomous interactions between different tissues such as autoimmune disease. In particular, inferring MRs cannot be done directly using typical ARACNe-based analysis because of fundamental assumptions made during the generation of a regulatory network: (1) that the samples used are relatively pure or represent the one underlying transcriptional network; and (2) the underlying molecular behavior of a data set exists at a steady state such that each sample can be treated as a "snapshot" of regulatory dependency within the overall network (Basso et al., 2005; Califano et al., 2012; Margolin et al., 2006b). A contaminated sample, particularly by a tissue that exhibits a different context-specific regulatory network, can impair the accuracy of regulatory predictions. Further, when pathogenesis is dependent on the interaction between the two tissues, there will always be an artifact correlation between contaminant gene signatures and the molecular modules that recruit them, but are expressed in the other tissue. This makes it difficult to clearly define modules exclusive to one tissue or the other when analyzing gene expression data generated from a mixture of the two tissues.

Alopecia Areata (AA) provides an ideal model for such a study since it is characterized by cytotoxic T cells actively infiltrating the hair follicles and scalp skin that are typically absent in normal skin (Xing et al., 2014). AA typically presents as loss of distinct, random patches of hair that can spread to the entire scalp (alopecia totalis) or the entire body (alopecia universalis) (Olsen et al., 1999). Previous research from our lab and others has directly implicated immune genes in AA (Martinez-Mir et al., 2007; Petukhova et al., 2010), many of which are shared with other autoimmune diseases such as type 1 diabetes, celiac disease, and

rheumatoid arthritis (Betz et al., 2015; Farh et al., 2014; Petukhova et al., 2010). Previous studies have identified infiltration of cytotoxic CD8-positive, NKG2D-positive T cells into the skin of AA mice (Xing et al., 2014), and the pathology of AA involves IFN-gamma-dependent signaling pathways, which are frequently disrupted in association with immune evasion in cancer (Dunn et al., 2002; Sato et al., 1998).

Little work has been done to determine if there are intrinsic factors in the "end organ" (the tissue that suffers autoimmune attack) that contribute to the disease, such as scalp skin in AA, making this molecular component a prime target for the analysis. We predict that pathogenic changes in the molecular profile of the scalp skin will contain genes that mediate interactions with the infiltrating T cells. As a corollary, identifying the MRs will grant regulatory control over the modules that are sufficient to induce immune recruitment. To study this, we leverage context-specific regulatory networks for the *regulatory* deconvolution of a mixed-signature gene expression profile of AA patients. The goal of this work was to develop a framework capable of separating mixed AA tissue biopsy gene expression data into skin-specific modules of AA pathology and infiltrate recruitment.

We identified a molecular profile of AA that includes the genetic modules of infiltrate recruitment in the scalp skin by filtering genes that do not accurately map to a skin-specific network. This scalp skin signature allowed the subsequent identification of two MRs of scalp skin contribution to infiltration: IKZF1 and DLX4. These two genes are expressed in primary scalp biopsies and are sufficient to induce an AA-like molecular signature and NKG2D-dependent cytotoxicity in independent, wild-type cellular contexts, allowing for direct genetic induction of immune-mediated cytotoxicity.

RESULTS

Initial Definition of a Pathogenic Expression Signature in AA Reveals the Presence of Local Scalp Skin and Infiltrating Immune Signals

First, we created a molecular signature comparing AA patients to controls to generate a molecular representation of AA. We analyzed a training set of microarray studies of patient biopsies from an initial cohort of 34 unique biopsy samples: 21 AA patients of varying clinical presentations and 13 unaffected controls. We additionally had patient-matched, nonlesional scalp biopsies for 12 of the 21 AA patients. These 34 patients were gathered as the first of two cohorts totaling 96 patients, the remainder of which was saved for validation studies.

We created an overall gene expression signature by comparing patients of two distinct clinical presentations, patchy AA (AAP) and totalis and universalis (AT/AU) all against unaffected controls. To account for artifacts in the signature associated with secondary effects of infiltration such as hair loss, we then performed hierarchical clustering using this gene signature on a set of patient-matched lesional (symptomatic skin with hair loss) and nonlesional (asymptomatic hair-bearing skin) samples. This analysis identified gene clusters that were differentially expressed between these samples and those that were systemically equivalent across lesional and nonlesional samples. We subsequently removed from the first expression set any genes that fell in clusters correlating with lesional versus nonlesional states. This primarily removed a significant number (but not all) of the keratin and keratin-associated proteins from the signature.

The resulting gene expression signature, the Alopecia Areata Gene Signature (AAGS), consisted of a total of 136 unique genes (Table S1) and provided sufficient information to cluster the entire training cohort into two appropriate superclusters corresponding to the control and disease states (Figure 1A). Clustering these genes by co-expression also revealed two distinct modules of genes, with greater diversity of co-expression in the genes upregulated in the disease state (Figure 1B). As a qualitative measure of the genes differentially expressed between affected and unaffected patients, we analyzed them for functional annotation enrichments. The analysis revealed the presence of HLA genes, immune response elements, and inflammatory and cell death pathway gene expression in the affected patient samples (Figure 1C). The two most significant superclusters of the AAGS were transmembrane signaling peptides ($p = 2.8 \times 10^{-11}$) and secreted cell-cell signaling peptides ($p = 2.1 \times 10^{-10}$). As expected, this list also contains several antigen-presenting elements and immune response elements that are associated with AA and autoimmune disease (Figure S1; Table S2). These results indicate that there are significant alterations of multiple biological processes in AA-presenting cells. We postulate that some subset of these genes originate from the scalp skin and are required to induce infiltration recruitment.

There is also significant evidence for immune-related genes originating from infiltrating immune cells that must be filtered beforehand; or else they could confound the identification of skin-specific molecular programs. Gene markers associated with immune cells or immune response were detected as part of the AAGS including *CD8a*, *CXCL9/10*, and *CCL5/18/20/26* (a full list can be found in the Supplemental Information). In primary patient biopsy samples, defining skin-specific molecular behaviors contributing to AA is a difficult task due to the presence of infiltrating T cells and secondary response pathways in AA skin samples.

Leveraging Regulatory Networks to Deconvolve Skin and Immune Signatures in the AAGS into Regulatory Modules

With clear definitions of the disease signature, we sought to deconvolve the skin molecular program in the AAGS from the molecular program originating in infiltrating immune cells in a systemic, unbiased manner. Rather than using GO pathway enrichment or other annotation-based methods that rely on a priori knowledge and potentially ambiguous annotations, we instead utilize our inferred regulatory networks under the hypothesis that we can filter nonskin (immune infiltrate) gene expression by identifying the genes that cannot be mapped to a skin-specific regulatory network.

A transcriptional regulatory network of the scalp skin was generated using the ARACNe algorithm and associated software suites (see Experimental Procedures). Specifically, to generate the network, we included a cohort of 106 primary scalp skin samples consisting of normal (unaffected) whole skin biopsies and several samples of primary cultured dermal fibroblasts and dermal papilla cells, which contain few or no T cell infiltrates. This network represents the regulatory network in uninfiltrated skin-derived tissues

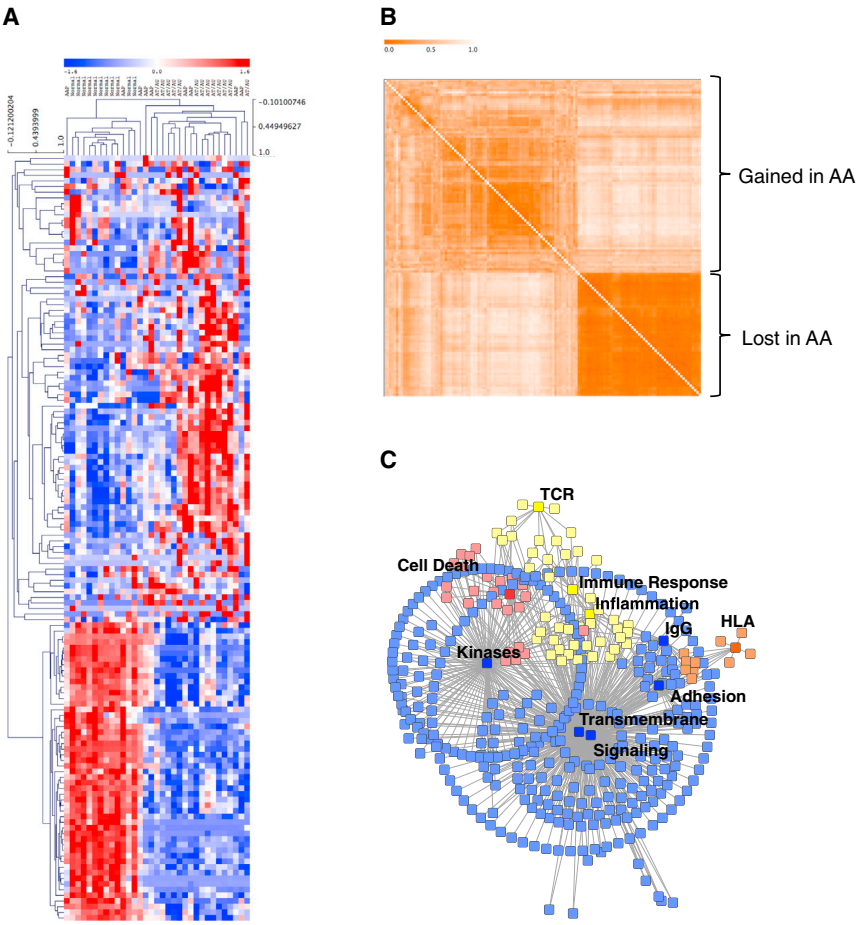


Figure 1. Gene Expression Analysis Identifies Mixed-Tissue Gene Signatures

(A) Unsupervised hierarchical clustering of a cohort of AAP, AT/AU, and unaffected controls (Normal) using the AAGS (blue, underexpression and red, overexpression).

(B) Gene co-similarity matrix showing gene clusters. The stronger orange indicates lower dissimilarity in gene expression. The clusters over- and under-expressed in AA are indicated.

(C) Graphical representation of genes in the signature and the statistically enriched functional categories associated with them. The blue indicates signaling pathways; the yellow indicates immune/inflammation pathways; the orange indicates HLA; and the red indicates cell death pathways. The pathways at $p < 0.05$ FDR corrected were kept for this analysis.

and serves as the cornerstone of the deconvolution, which occurs in two primary steps as detailed in Figure 2.

For deconvolution of regulatory modules, the genes in the AAGS are directly mapped to the regulatory network (Figure 2A; see Experimental Procedures for details). A gene in the AAGS is only retained if there is a direct regulatory interaction between it and a TF using the regulatory logic of a skin ARACNe network (red, solid edges). Any genes that come uniquely from infiltrating immune cells will not have significant representation in the ARACNe network, and are subsequently removed from the AAGS (black, dotted edges) for skin, and added to an Immune Gene Signature (IGS).

The IGS was used as a "negative control" signature, adapted from previous work in characterizing cancer immune infiltrates (Bindea et al., 2013). The signatures were defined as a set of genes that are specifically expressed in each immune cell type, including T cells, B cells, mast cells, and macrophages. This step iteratively re-defines the AAGS and IGS by separating those genes whose regulation can be accounted for by an uninfiltrated regulatory network (AAGS) from those that cannot (IGS). By extension, we expected the filtered AAGS to be enriched enough in skin gene expression to generate accurate skin-specific regulons.

As indicated in Figure 2B, 13 infiltrate-specific genes were removed from the AAGS (9.5% of the total signature) when passed through the skin-specific regulatory network. These genes are also listed in Table S1. This resulted in two mutually exclusive gene modules (no overlapping genes, $p = 1.77 \times 10^{-4}$), the AAGS and the IGS. A subsequent pathway enrichment analysis further confirmed loss of statistical enrichment of the "T cell activation" and "Immune response" categories (see Table S2), while retaining the other clusters including known skin immune response elements (such as the HLA genes). This left a total of 123 genes in the AAGS that we interpret to represent all end-organ programs associated with AA pathology, including end-organ-initiated immune recruitment and immune response (Table S1, starred entries).

Note that we have made the distinction between annotations associated with immune cells (e.g., *CD8a*) and annotations associated with immune response genes (e.g., HLA). The former are removed by the regulatory network as unrepresented in a skin regulatory network. The latter are signature genes that we aim to keep, as they represent the response elements in the skin and are relevant for the pathology of the disease.

Clustering the filtered AAGS revealed two distinct molecular modules that define the transition from unaffected patients (Figure 2B, second) to an AA disease state (Figure 2B, third). Each node represents a gene in the signature, and its size represents the relative expression in each state (larger means higher expression). We labeled these gene groups: (1) genes whose expression is increased when transitioning into the disease state, and (2) genes whose expression is lost in the transition. This filtered AAGS reflects end-organ-specific gene modules and served as the input to our MR analysis.

IKZF1 and *DLX4* Are MRs of the Skin AAGS and, by Extension, Infiltrate Recruitment

The next step is the most important in identifying end-organ-specific MRs. We performed MR analyses on both the deconvolved AAGS and the IGS independently and in parallel using the scalp skin regulatory network (Figure 2C, first, red outline). Using only regulatory interactions represented in skin, we identified the

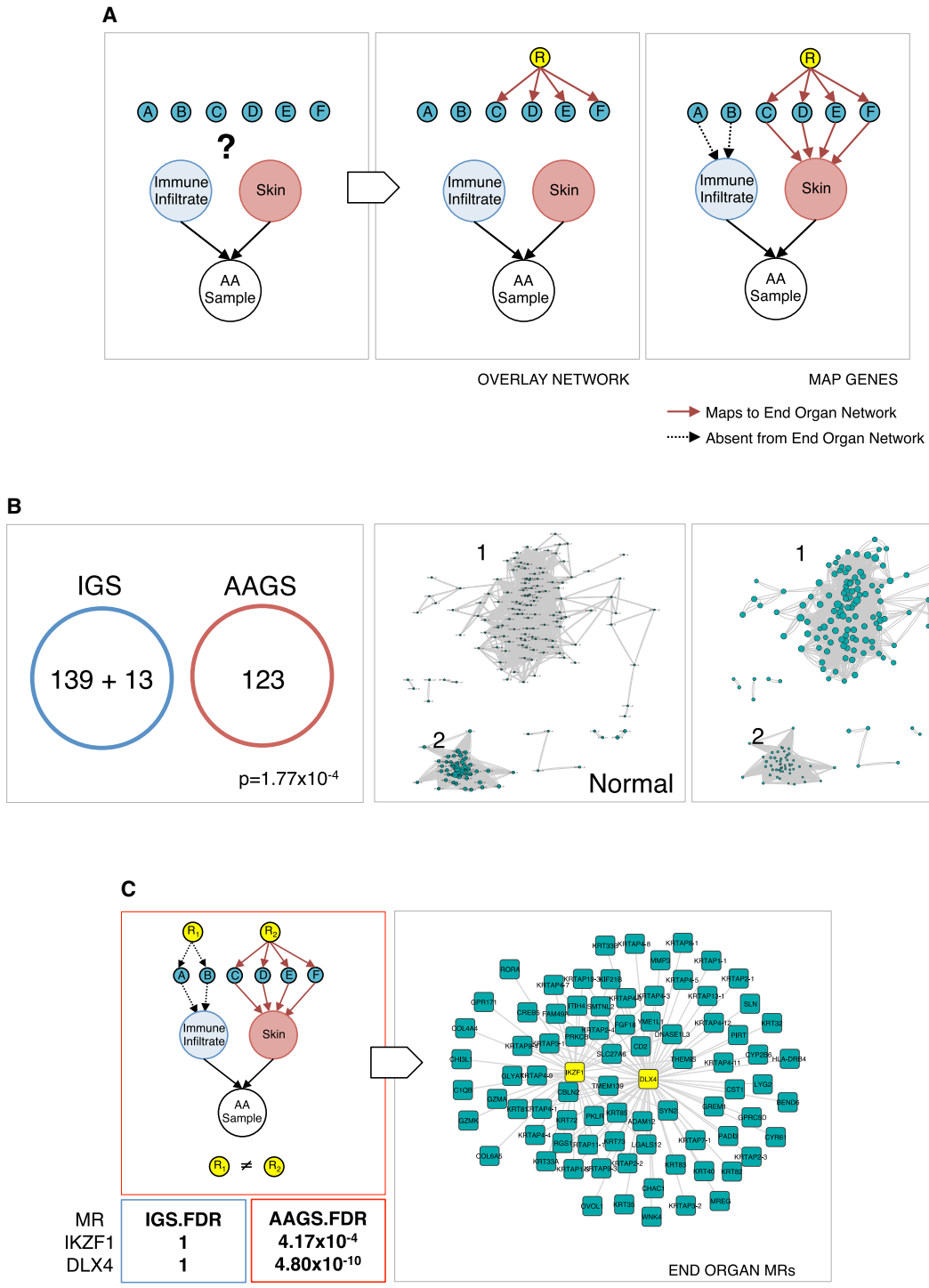


Figure 2. Identification of *IKZF1* and *DLX4* as MR

An overall flow of the pipeline used to deconvolve regulators of genes expressed in the end organ (skin) from those expressed in infiltrating tissue (immune cells). (A) Genes (aqua nodes labeled A–F) measured from a complex primary tissue sample are assigned to either end-organ (red, AAGS) or infiltrate (blue) based on whether or not they can be mapped to regulators in the skin network (R). Only the genes mapped to the red node are considered for MR analysis. The genes mapped to the blue node are pruned away.

(B) The resulting pruning of the AAGS provides an end-organ-enriched gene expression signature (aqua nodes) that is mutually exclusive with an IGS, $p = 1.77 \times 10^{-4}$, that is overexpressed (cluster 1, node sizes proportional to fold change) and suppressed (cluster 2) in AA.

(legend continued on next page)

transcriptional regulators that had the highest specificity for the deconvolved AAGS (red arrows) and repeated the analysis for the IGS (black arrows). This step compares the AAGS against the IGS in terms of *regulatory logic* in the scalp skin, as opposed to direct coverage of gene expression. This analysis assays which TFs are the best candidates for the deconvolved AAGS (and *not* for the IGS) using a molecular regulatory network specific to the skin. We identify skin-specific candidate MRs by keeping only the candidates that were both significant in AAGS coverage and insignificant for IGS coverage.

Of the significant candidate MRs specifically for the AAGS, we employed a greedy sort to identify the fewest number of regulators needed to maximize the coverage of the AAGS. We found that two MRs were sufficient to cover >60% of the AAGS: *IKZF1* and *DLX4*. Any additional candidates boosted the coverage by a statistically insignificant margin (<5%). We conclude that the maximum AAGS fidelity (most faithful recreation of the expression signature) and efficiency (fewest necessary regulators) could be achieved through these two genes (*IKZF1* $p = 4.17 \times 10^{-4}$ and *DLX4* $p = 4.8 \times 10^{-10}$ FDR-corrected).

An equivalent MR analysis conducted on the IGS modules failed to generate any statistically significant or meaningful MRs when using the scalp-skin regulatory network. Specifically, the best candidates for the AAGS, *IKZF1* and *DLX4*, fall to statistical irrelevance (falling from first and second to 159th and 210th, respectively, FDR = 1) (Figure 2C, IGS.FDR). Conducting the MR analysis on the AAGS without deconvolution fails to generate MR candidates at the threshold that is typically expected (both in p value and signature coverage) due to the presence of contaminating genes in the signature which cannot accurately be mapped to a MR, but nonetheless count against enrichment in the analysis.

These two candidates represent the minimum number of regulators required to recreate the AAGS using regulatory interactions derived from a specific tissue context (scalp skin), distinct from any immune-specific regulatory modules that were deconvolved away using this method. *IKZF1* and *DLX4* therefore represent a genetic regulatory module in the scalp skin that contributes to AA pathogenesis (Figure 2C, last) and may be sufficient to induce infiltration recruitment in an AA-like manner.

The identification of *IKZF1* was unexpected, since it is a well-established T cell differentiation factor, though it is not without precedent that IKZF1 may have a role in cells outside the immune system (Javierre et al., 2011). However, it is important to note that this analysis does not imply that a MR such as *IKZF1* has no role in T cells contributing to AA pathogenesis, but rather, that there is significant evidence that *IKZF1* additionally functions in the scalp skin to mediate the interactions between the tissues.

Expression of *IKZF1* and *DLX4* Induces an AAGS-like Signature in Normal Hair Follicle Dermal Papillae and Human Keratinocytes

To validate our MR predictions with functional studies, we exogenously overexpressed IKZF1 and DLX4 in skin-derived cell lines and cultured cells to test for sufficiency in influencing expression of the AAGS. We cloned *DLX4* and two isoforms of *IKZF1* for exogenous expression in cultured cells. The active *IKZF1* isoform served as the experimental arm of the study, while the isoform that lacks a DNA binding domain was included as a negative control (IKZF1$^{\delta}$). We expressed these genes in cultured primary human hair follicle dermal papillae (huDP) and human keratinocytes (HK). This experimental system allowed us to directly test two distinct, but related, hypotheses: (1) *IKZF1* and *DLX4* can induce AA-like recruitment of immune cells, and (2) they do so through expression in the skin (not the immune infiltrates).

We identified a set of genes that were significantly differentially expressed in the same direction in *IKZF1* and *DLX4* transfections across both cell types. Unsupervised hierarchical clustering of all samples based on these transcripts reveals clean co-segregation of *IKZF1* and *DLX4* transfections from IKZF1$^{\delta}$ and RFP (red fluorescent protein) controls (Figure 3A). Furthermore, we observed that the subclustering within these supergroups was not biased based on cell type used (HK did not cluster with HK, and DP did not cluster with DP), supporting that we have identified context-independent effects of MR overexpression. Interestingly, we observed that *DLX4* transfections resulted in increased levels of *IKZF1* transcript and protein, whereas the *IKZF1* transfections did not influence *DLX4* expression (Figures 3B and 3C).

We subsequently interrogated the expression data for enrichment of the AAGS genes using gene set enrichment analysis (GSEA). We performed two differential gene expression studies comparing the *IKZF1* transfections versus RFP controls and *DLX4* transfections versus RFP controls. The results show that the ectopic expression of the MRs is followed by significant enrichment in the induction of the AAGS (*IKZF1* $p = 0.012$ and *DLX4* $p = 2.08 \times 10^{-4}$; Figures 3D and 3E).

IKZF1 and DLX4 Expression Are Sufficient to Induce NKG2D-Mediated Cytotoxicity in Normal Cultured Skin

IKZF1 and *DLX4* overexpression suggest that these two genes are MRs capable of mediating the AAGS when applied to HK and huDP. However, the functional relevance of these MRs to autoimmunity and immune infiltration is whether or not their expression is sufficient to induce a targeted autoimmune response. In order to investigate this ex vivo, we performed experiments measuring the level of cytotoxic cell death in HK and huDP cells when exposed to peripheral blood mononuclear cells (PBMCs).

We again transfected both HK and huDP cells with one of four expression constructs: *IKZF1*, *DLX4*, RFP (negative control), or IKZF1$^{\delta}$ (negative control). At 24 hr post-transfection, these cells were incubated with fresh, purified PBMCs. We additionally cultured human dermal fibroblasts and autologous healthy donor PBMCs. The PBMCs were obtained from a healthy control subject with no history of AA or any other autoimmune disease.

(C) To deconvolve the scalp skin regulators, we performed MR analysis on the AAGS and the IGS, yielding candidate regulators of each signature. The true MRs in the skin (R2) will only appear when using the AAGS and not in the IGS. The infiltrate regulators (R1) will not be detected using the AAGS. The *IKZF1* and *DLX4* only have significant FDR values when using the AAGS and are insignificant (FDR = 1) when using the IGS, left. This analysis establishes *IKZF1* and *DLX4* as AAGS (aqua squares) and MRs (yellow squares) in the skin (right).

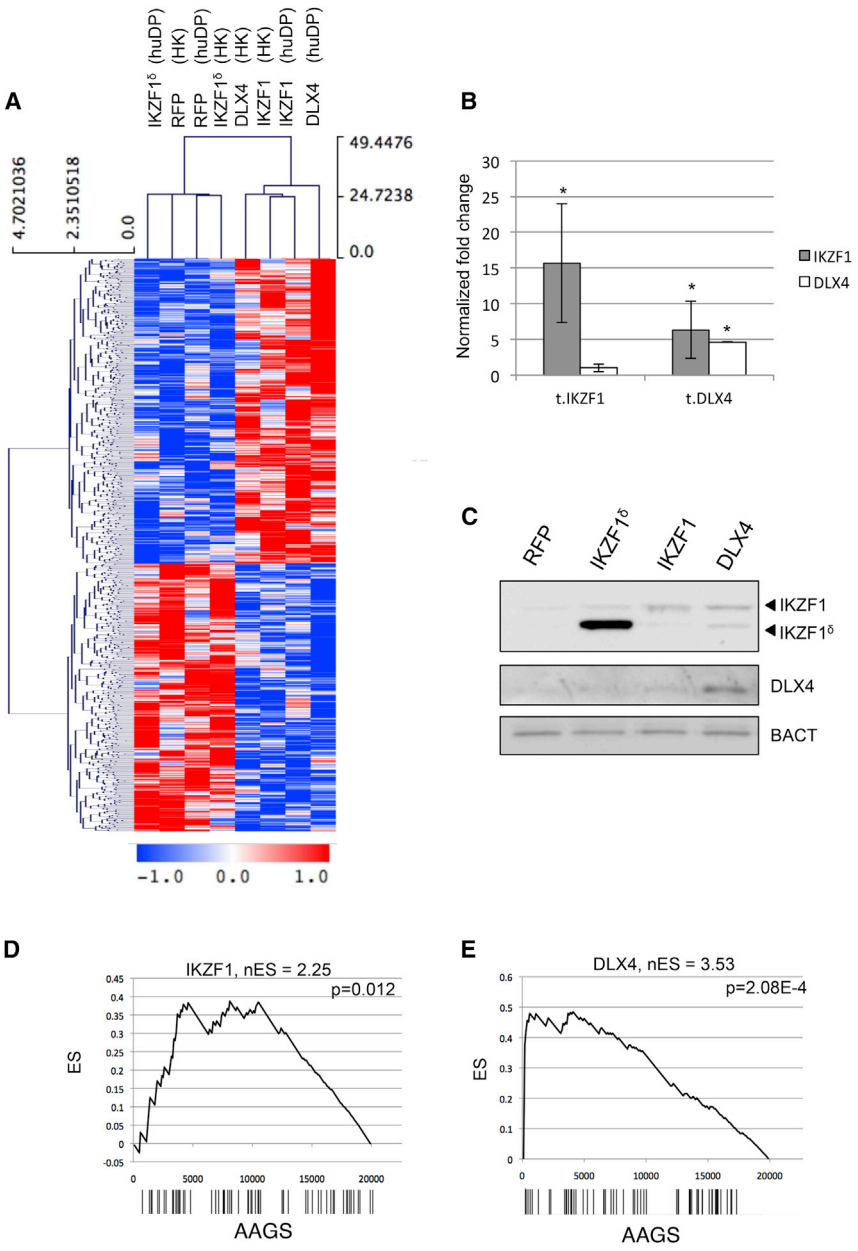


Figure 3. Exongeous Expression of *IKZF1* and *DLX4* Induces a Context-Independent AA-like Gene Expression Signature

(A) 2D hierarchical clustering of gene expression measured in huDP and HK transfected with plasmid vectors expressing *IKZF1*, *DLX4*, or controls expressing RFP and IKZF$^{\delta}$, an isoform lacking DNA binding domains. The treatment type and cell type for each experiment are indicated at the top of the heatmap. The blue indicates decreased expression and the red indicates overexpression.

(B) Analysis of *IKZF1* and *DLX4* mRNA expression in transfected cells in quadruplicate, represented as average ± SEM, normalized to B-actin.

(C) Western blot confirming IKZF1 and DLX4 proteins. The GSEA plots measuring the specificity of AA-like response assayed by differential expression of the AAGS following (D) *IKZF1* or (E) *DLX4* overexpression. The genes are ranked left to right from most- to least-differentially expressed on the x axis and barcodes represent the positions of *IKZF1* and for *DLX4* signature genes. The Enrichment Score (ES) is shown in the plot, and the normalized Enrichment Score (nES) is displayed at the top. The nES is derived from the ES at the "leading edge" of the plot, that is, the first maximal ES peak obtained. The p value is computed for the nES compared against a randomized null distribution.

In all comparisons, we observed a statistically significant increase in PBMC-dependent cytotoxicity for the *IKZF1* and *DLX4* transfections compared to RFP and IKZF1$^{\delta}$ controls (Figure 4, center columns, total bar height). The patient-matched PBMCs and RFP-control transfected fibroblasts exhibited no evidence of cytotoxic interactions, as expected in healthy target cells (Figure 4A, center). However, the introduction of *IKZF1* and *DLX4* were both sufficient to induce an interaction between these previously non-interacting cells, resulting in significant increase of total cytotoxicity. In a similar fashion, both huDP (Figure 4B, center) and HK cells (Figure 4C, center) showed a significant increase above background levels in cytotoxic sensitivity to the PBMCs.

Since we previously showed that the likely pathogenic immune cells in AA are CD8+ NKG2D+ activated T cells, we also performed all treatments with the addition of an NKG2D-blocking antibody (see Experimental Procedures) to prevent NKG2D-dependent interactions. In all cases, we observed that blocking NKG2D suppressed the cytotoxicity in both *IKZF1* and *DLX4* treatments to levels comparable to controls (Figure 4, center, gray bars). From the difference between the inhibitor-treated and untreated cells, we can infer the cytotoxicity that is NKG2D-dependent (Figure 4, center, white bars), which can be normalized to that observed in controls for a relative fold change analysis. From the NKG2D blockade, we observed a statistically significant increase specifically in *IKZF1* and *DLX4* transfections across all trials (Figure 4, right). There was a large (>50-fold) increase in patient-matched cytotoxicity compared to the control transfection, which again showed no significant cytotoxicity. There was approximately a 2- to 8-fold increase in NKG2D-dependent cytotoxicity compared to both controls, despite a statistically significant, but small (<10%), increase in NKG2D-independent cytotoxicity. We conclude from these experiments that *IKZF1* and *DLX4* are capable of inducing NKG2D-dependent interactions with normal PBMCs that result in toxicity for the transfected cells irrespective of the exact tissue type.

Importantly, these experiments establish the cell autonomous function for *IKZF1* and *DLX4* in the scalp skin as opposed to infiltrating cells, since the exogenous modification was done strictly

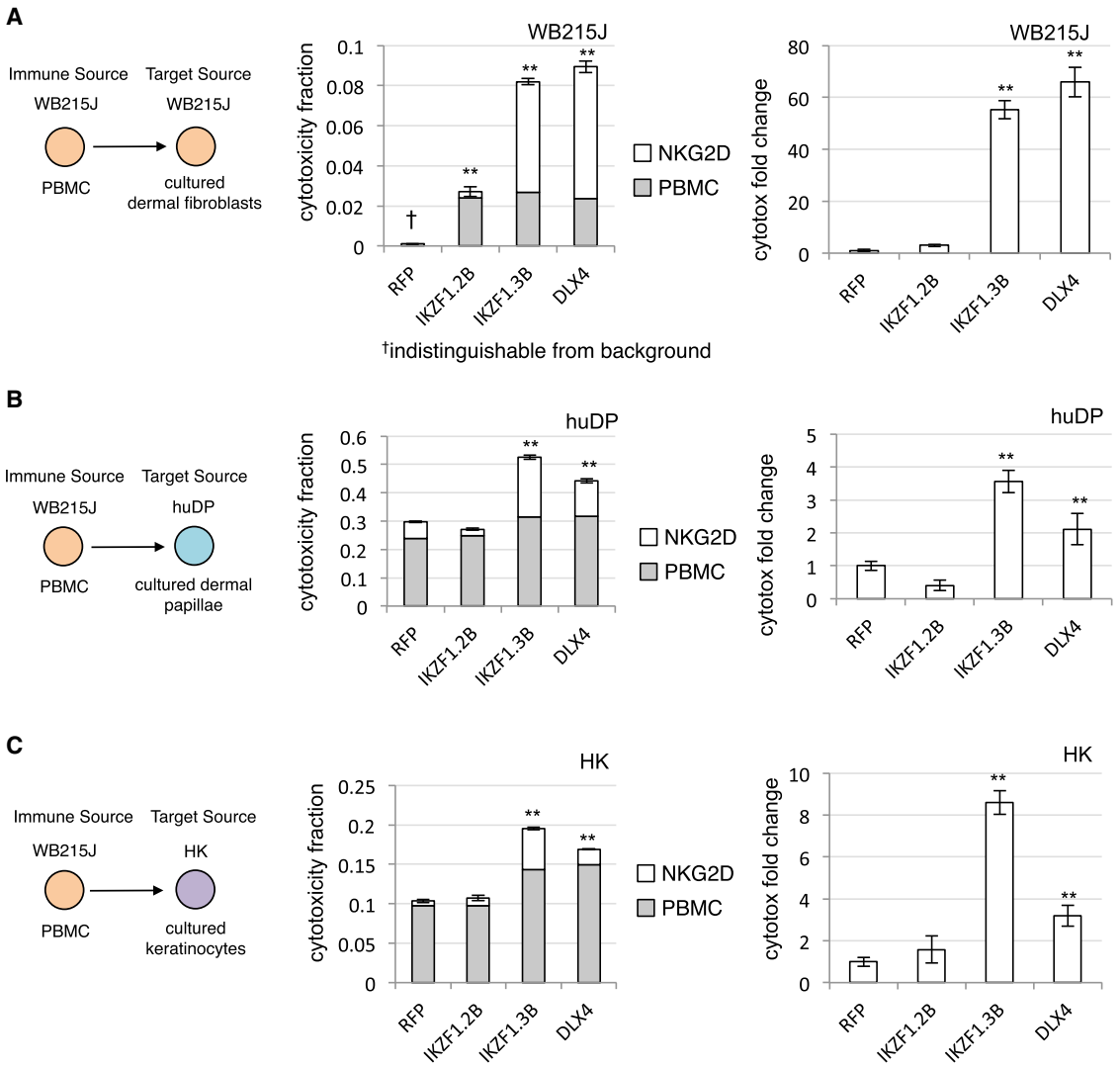


Figure 4. Exogenous Expression of IKZF1 and DLX4 Induces Increased NKG2D-Dependent PBMC-Associated Cytotoxicity in Three Cultured Cell Types

The schematic on the left of each row describes the tissues introduced to PBMCs for cytotoxicity assays (in triplicate). The colors indicate host sources (matching colors indicate host-matched tissues). The middle bar graphs present the cytotoxicity values obtained after either 6 hr of incubation (total bar height) or the cytotoxicity observed after 6 hr with the addition of human anti-NKG2D monoclonal antibody (gray bar). The NKG2D-dependent cytotoxicity is the difference between the two (white bar). The right bar graphs report the changes in NKG2D-dependent cytotoxicity normalized to the RFP controls. IKZF1.2B indicates cells transfected with the IKZF1$^{\delta}$ vector, and IKZF1.3B indicates the full-length transcript. The y axis reports cytotoxicity measured as a fraction of maximum cytotoxicity (total cell count). All error bars report ± SEM. ** indicate statistically significant difference from RFP control at FDR <0.05.

(A) Dataseries corresponding to WB215J PBMCs and WB215J fibroblasts.

(B) WB215J PBMCs against cultured huDP.

(C) WB215J PBMCs against cultured HK.

on normal cultured cells and exposed to healthy PBMCs from a source with no history of autoimmune disease.

MR Expression Permits Reconstruction of a Directional Skin-Specific MR Module of Infiltration Recruitment

After establishing that *IKZF1* and *DLX4* are sufficient to induce the AAGS, we sought to use this data to fully reconstruct the AA MR module. ARACNe is capable of detecting direct transcriptional dependencies between a TF and nonregulatory genes that are potential targets (T) because we can infer that the regulation is TF → T. ARACNe cannot infer directional interactions between TF-TF pairs and subsequently cannot infer secondary T of MRs due to the regulatory equivalence of TFs (Figure 5A, first). However, since we have directly perturbed HKs and huDPs with specific MRs (Figure 5A, asterisks), we can use the gene expression data to infer directionality. If TF_B is a T of our MR (TF_A), then overexpression of TF_A will result in the differential expression of TF_B and we can infer that $TF_A \rightarrow TF_B$. Subsequently, any marker genes in the signature associated with TF_B can be linked to MR as secondary T $TF_A \rightarrow TF_B \rightarrow T$ (Figure 5A, top). If TF_B functions

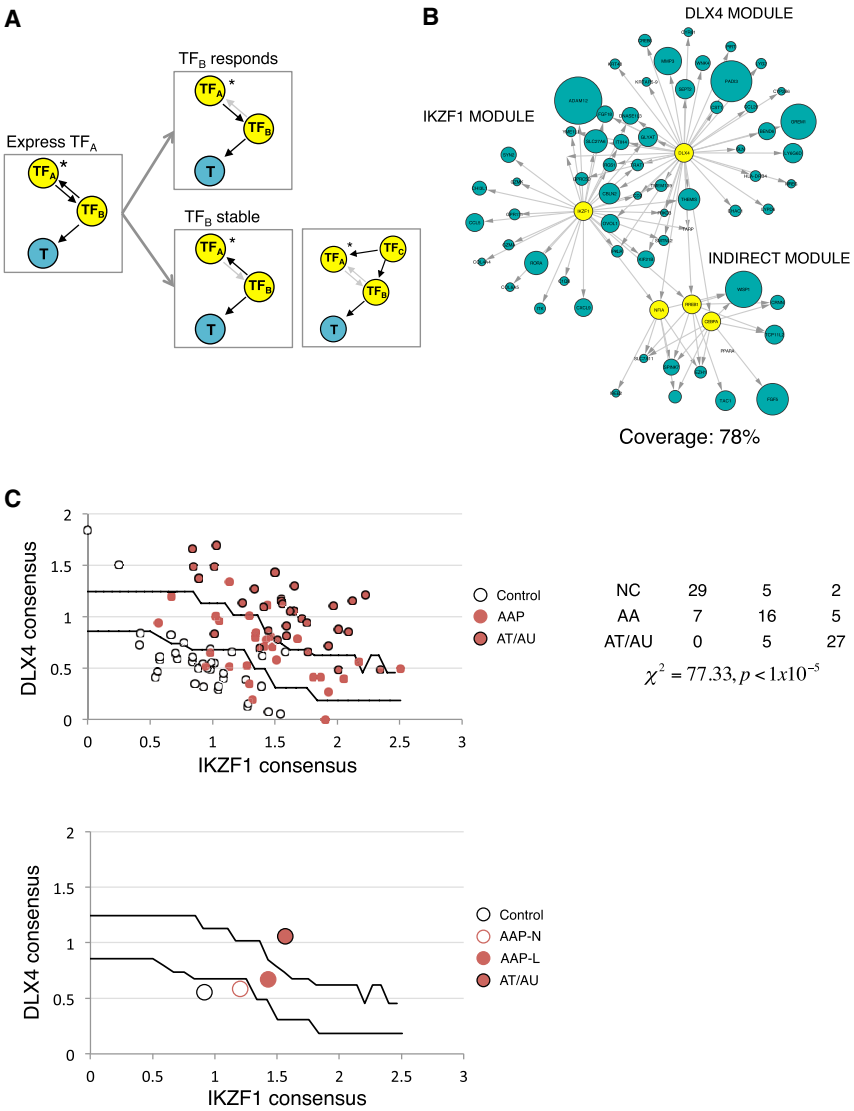


Figure 5. The Fully Reconstructed Master Regulator Module Predicts Both Immune Infiltration and Severity

(A) Using the exogenous expression data, it is possible to infer both direct transcriptional MR T (MR → T), as well as T regulated by TFs that are T of the MR (MR → TF → T). Any TF (TF_B) that is paired with MRs *IKZF1* or *DLX4* (TF_A) and that exhibits changes in expression upon overexpression of the TF_A is regulated by the TF_A. Subsequently, any genes (T) in the AAGS that are linked to TF_B are secondary T of TF_A (TF_B responds). Any TF_B that does not respond to transfection of TF_A is not regulated by the TF_A, so either TF_B regulates TF_A (TF_B stable, left) or both are co-regulated by a third, TF_C (TF_B stable, right).

(B) Using this approach, 78% of AAGS can be mapped to *IKZF1* or *DLX4* within one indirect TF_B. The blue nodes represent AAGS genes that respond to *IKZF1* or *DLX4* expression, the size of nodes scaled to the fold change experimentally observed (only nodes having at least 25% change are shown).

(C) Using these T, we generate single numeric scores of IKZF1 and DLX4 transcriptional activity and used them to create classifiers for AA severity. The AA samples are then imposed over the search space to assess accuracy (top chart). The table provides quantitation and statistics for separation of presentations across territories in the search space (unaffected: NC; patchy AA: AAP; and totalis/universalis: AT/AU). The centroid representations can be used to show how populations transition into disease states by moving across the trained boundaries (bottom chart; nonlesional: AAP-N and lesional: AAP-L).

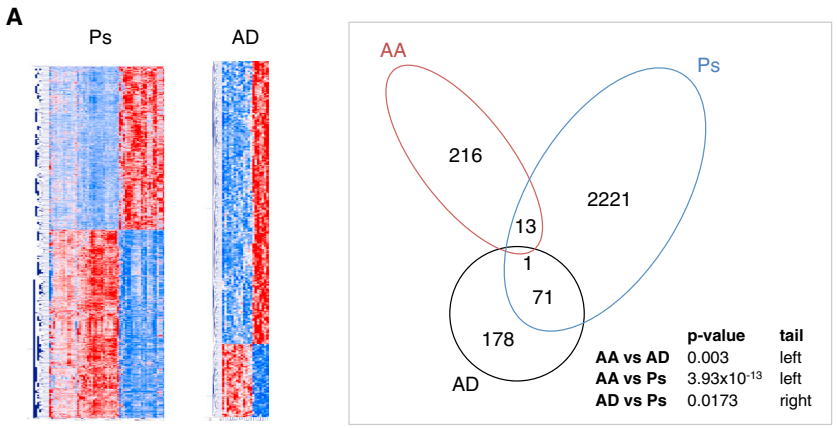
upstream of or in parallel with MR then the expression of TF_B and T will not be affected by overexpression of TF_A (Figure 5A, bottom).

Using this logic, we reconstructed the regulatory module to measure the full extent of the coverage obtained by overexpressing *IKZF1* and *DLX4* in these cellular contexts. We mapped any downstream T of TFs that both (1) respond to *IKZF1*/*DLX4* expression in the experiments, and (2) are predicted to have mutual information with the expressed MR by ARACNe to the regulatory module. We found that 78% of the responding AAGS are within 2° of downstream separation from the MRs *IKZF1* and *DLX4* based on these criteria (Figure 5B; full module listed in Table S4).

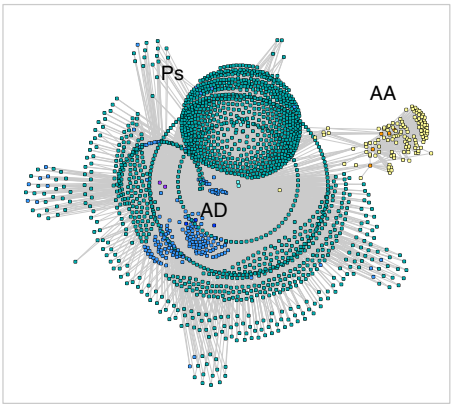
IKZF1 and DLX4 Can Be Used to Predict Both Immune Infiltration and Disease Severity in an Independent Cohort

As validation of this module, we returned to our original AA array cohort and performed a machine-learning analysis. We attempted to classify a validation AA set into control and affected samples using only the inferred IKZF1 and DLX4 activity. Using the earlier training set from Figure 1, we arrayed the samples into a search space of two dimensions: the consensus activity of IKZF1 (x axis), and the consensus activity of DLX4 (y axis) (see Experimental Procedures). From the training set, we generated a topographical map of the consensus activity space to define ranges of IZKF1 and DLX4 activity associated with control samples, patchy AA, and AT/AU samples (Figure 5C, black lines). The region in Figure 5C closest to the origin of the plot represents the lowest combined IKZF1 and DLX4 activity; its upper bound (the lower black line) is the support vector machine (SVM) margin that maximizes the difference between control and all AA patients. The next upper bound (the upper black line) represents the SVM margin that maximizes the separation of AT/AU patients from AAP.

Using these measures of MR activity, we turned to the validation set and tested for the predictive power of these parameters in separating patients and controls. We observed a strong ability to separate samples into disease and control states, in addition to clinical severity (Figure 5C, top, $p < 1 \times 10^{-5}$). A centroid map of each patient subgroup more clearly reveals how the transition of patient groups from Control (NC) to AAP and AT/AU is reflected by relative IKZF1 and DLX4 activity (Figure 5C, bottom).



B



psoriasis	Ps p-value	IGS p-value
POL2R2	2.02×10^{-4}	1
SOX15	2.07×10^{-5}	1
HTLF†	8.36×10^{-5}	1
SMAD2*†	5.06×10^{-5}	1
TGFBR1*	4.75×10^{-5}	1

*studied in published work in psoriasis

atopic.derm	AD p-value	IGS p-value
HLTF†	1.78×10^{-7}	0.944
SMAD2†	1.96×10^{-5}	0.990
ETS1	4.13×10^{-5}	0.817
RB1	4.16×10^{-5}	0.707
CIR1	7.42×10^{-5}	0.320

†common MR in psoriasis and AD

Figure 6. Deconvolved Regulatory Modules Can Be Generated for AA, Ps, and AD Using the Same Naive Framework

(A) Disease-associated gene expression signatures for Ps and AD can be clearly defined by differential expression. The comparison of these signatures to the AA gene signature reveals that the AA signature is statistically distinct from both Ps and AD signatures (Fisher's exact test), whereas there is statistical evidence for some sharing between the Ps and AD signatures.

(B) Translating these signatures into regulatory modules reveals entirely different MRs governing AD and Ps compared to AA. The yellow nodes = AA gene signature; the blue nodes = AD gene signature; the aqua nodes = Ps gene signature; the orange nodes = AA MR; the dark blue nodes = AD MR; and the cyan nodes = Ps MR. The list of top five AD and Ps MRs are provided, ranked by coverage of the corresponding signature. Also provided are the p values of each MR without deconvolution (IGS p value) (* indicates published regulators and † indicates an MR common to AD and Ps).

For comparison, we also included a centroid for the AAP nonlesional sample biopsies, which were not included in the training set.

Deconvolution Applied to Independent Inflammatory Skin Diseases Identifies Known Genes

For comparison, and to provide proof-of-concept for the generalizability of our approach, we downloaded publicly available gene expression data sets for atopic dermatitis (AD) (Suárez-Fariñas et al., 2011) and psoriasis (Ps) (Yao et al., 2008). We generated gene expression signatures for each disease by comparing lesional biopsies to unaffected biopsies, similar to our AA analysis (Figures 6A and S2; Table S3). A direct comparison of the genes within the AA, AD, and Ps signatures revealed statistically significant evidence that the AAGS is distinct from both AD and Ps ($p = 0.003$ and 3.93×10^{-13}, respectively). By contrast, comparison of the AD and Ps signatures to each other revealed statistically significant evidence for shared molecular signatures and, by extension, possible shared molecular pathology ($p = 0.0173$).

These two signatures were applied to our pipeline. Figure 6B reports the top five MRs identified after the analysis, ranked by their total coverage of the appropriate disease signatures (Ps or AD). Also provided are the ranks of the MRs using the corresponding deconvolved IGS. The results indicate that the key regulatory hubs associated with AA (specifically *IKZF1* and *DLX4*) are unique to AA. Each disease was assigned its own unique list of MRs, but there additionally was overlap of two candidate MRs in AD and Ps: *SMAD2* and *HLTF*. *SMAD2* and *TGFBR1* are TFs with published evidence of involvement in Ps, and our pipeline was able to identify them with no a priori evidence, using a basic definition of a Ps gene expression signature (Doi et al., 2003; Gambichler et al., 2013). These results demonstrate the effectiveness of modeling complex genetic behaviors as regulatory modules to differentiate mechanisms of pathology.

DISCUSSION

Systemic generation and analysis of gene regulatory networks and gene expression data capitalizing on genome-wide profiling has proven to be instrumental in the study of complex diseases. Integrative projects to interrogate functional interactions have recently been leveraged in genome-wide expression signature deconvolution (Bindea et al., 2013; Newman et al., 2015) and cross-tissue interactions (MacNeil et al., 2015) in diabetes and atherosclerosis (Hecker et al., 2009; Keller et al., 2008). These studies have been invaluable in identifying infiltrating gene signatures, which provide insight into the types of pathogenic immune infiltrates associated with disease. They have also helped identify driver genes from eQTLs and other genomic association tests, similar to the systematic algorithms being developed in cancer and Alzheimer's disease research (Chen et al., 2014; Zhang et al., 2013) by providing significant genome-level coverage of regulatory activity and tissue-level gene panels of interacting tissues.

However, particularly in contexts such as AA, little has been done to characterize the modular *regulation* of discrete pathogenic molecular behaviors within a gene expression profile and

how they translate to physiological interactions between tissues of the disease. Modeling physiological traits as genetic programs controlled by MRs provides a uniquely powerful perspective in the study of complex disease. The approach canalizes large gene expression signatures into a relatively few number of selected MRs that subsequently become the T of manipulation via gene therapies or drugs and small molecules.

Here, we extend the application of regulatory networks to interrogate the complex molecular state of a mixed sample of end organ (scalp skin) and infiltrating (immune infiltrates) tissue in AA by comparing regulatory networks of different skin contexts (infiltrated and normal). We establish that in addition to their typical use for identifying the key regulatory hubs governing molecular phenotype switches, these networks can be used to isolate and compartmentalize molecular behaviors that originate from different tissues based on whether or not they are accurately represented in an independent context-specific network. This allows for more precise identification of tissue-specific molecular programs from a mixed sample that contribute to an integrated, interactive physiological behavior such as immune infiltration. Using this pipeline, we were able to reconstruct the MRs mediating infiltration from the skin not only in the context of AA, but our analysis of Ps and AD provides additional candidates for the genetic regulation of inflammatory skin diseases in general, and demonstrates the general applicability of the approach.

Aside from the direct implications in AA pathology, this work provides the proof-of-principle for two key, generalizable notions: (1) a complex interaction between two tissues can be modeled as quantifiable, molecular gene expression modules; and (2) these modules and their regulators can be extracted from expression data, compartmentalized to a tissue, and co-opted to induce the associated interaction in normal cell types. This was evidenced by our ability to recapitulate the AAGS upon ectopic expression of MRs IKZF1 and DLX4 and to subsequently induce enhanced cytotoxicity in non-AA cell lines using normal (non-AA) PBMCs solely via the manipulation of *IKZF1* and *DLX4* expression within the end organ itself (no genetic manipulation of the PBMCs).

Specifically, our analysis identified MRs that are sufficient to induce interactions with immune cells when expressed solely in scalp skin. Even in a patient-matched context with samples from a healthy, AA-unaffected patient, *IKZF1* and *DLX4* expression were sufficient to induce aberrant NKG2D-depedent interactions between dermal fibroblasts and PBMCs resulting in cytotoxicity. These interactions were not present in control transfections and they were repeated in two other (nonpatient-matched) cell types, indicating that the expression of *IKZF1* or *DLX4* is sufficient to induce interactions with normal immune cells irrespective of the specific tissue or host matching. The identification of *IKZF1* and *DLX4* would have been impossible without our network-based deconvolution, since the significant presence of infiltrating signature in the original AAGS would have prevented any accurate identification of candidate MRs. Instead, network-based deconvolution identified MRs that are capable of inducing specific molecular interactions in any of several molecular contexts that are completely independent of AA itself.

The identification of *IKZF1* was unexpected, since *IKZF1* is widely studied in the context of T cell differentiation (Kleinmann et al., 2008). However, its identification came solely from using a deconvolved AA signature, and not the IGS, using regulatory logic derived from skin. Had we relied on public databases, previous literature, or GO annotations to filter our gene expression data, we would have disregarded and removed *IKZF1* entirely due to extensive annotation as a T cell differentiation factor. Instead, by turning to regulatory networks, we were able to identify the possibility that local expression of *IKZF1* could have a pathogenic relevance independent of its established role directly in immune cells.

While *IKZF1* is well characterized in the context of immune cells, a role for *IKZF1* outside of immune cells is not without precedent in the literature. The losses of *IKZF1* and *DLX4* loci are also associated with oncogenesis in colorectal, lung, and breast cancers, and low-grade squamous intraepithelial lesions (Javierre et al., 2011; Sakane et al., 2015; Tomida et al., 2007). These studies obtain their genomic information directly from tumor masses, indicating that somatic losses of these two loci can contribute to cancer pathophysiology as end organ genomic alterations. Our studies into *IKZF1* and *DLX4* as MRs inducing immune infiltration support these results and raise the possibility that the loss of these loci may contribute to immune evasion in cancer. Further, these observations, and the identification of *IKZF1* and *DLX4* as MRs of immune infiltration recruitment, provide support that there is a function for *IKZF1* outside of its role as a T cell-specific differentiation factor and raises support for the hypothesis that autoimmunity in AA and tumor immune-evasion exist at opposite extremes of normal immune interactions. The loss of the MRs of immune infiltration is associated with cancer, and their overexpression is associated with the onset of autoimmune disease in AA.

We have shown that systems biology and network analysis can be used to model the molecular mechanisms mediating interactions between two distinct tissues, identify the key regulators, and use them to re-create the interactive trait in other contexts. While the output for the validation of these MRs was ultimately induction of cell death, the function of these MRs in the context of autoimmune disease is to induce a molecular profile that ultimately signals to and recruits immune infiltrates. Up to this point, applications of systems biology have mainly been to identify ''breakpoints'' in cell-autonomous molecular behaviors of cancers. The controlled induction of cross-tissue interactions, particularly those involving the immune system, invites potentially significant avenues for modeling complex genetic traits with regulatory networks that has previously not been feasible. We provide a proof-of-concept framework that can be used to actively compartmentalize molecular behaviors for study even in complex diseases involving interactions between different tissues.

EXPERIMENTAL PROCEDURES

This section contains a description of the less common or unique methods implemented in this study. The remaining methods are detailed in Supplemental Information.

ARACNe

To generate a context-specific transcriptional interaction network for scalp skin, we employed the ARACNe algorithm (Margolin et al., 2006a) on a set

on of 128 microarray experiments independent of the analytic cohorts in this study. These experiments represent platform-matched (Affymetrix U133 2plus) data acquired on whole skin samples from a mixture of normal whole skin biopsies, AA patient biopsies, microdissected dermal papillae, and separated dermis and epidermis samples. These samples collectively provide the heterogeneity required for accurate detection of transcriptional dependencies in the scalp skin. The experiments were pooled and post-processed as described above and a standard ARACNe analysis was performed. The ARACNe software suite is available from the Califano lab website, http://wiki.c2b2.columbia.edu/califanolab/index.php/Software.

MR Analysis

MRs for a specific gene expression signature were defined as TFs whose direct ARACNe-predicted T (regulon) are statistically enriched in the gene expression signature. Each TF's regulon was tested for enrichment of the AAGS using Fisher's exact test, FDR = 0.05. This analysis allows for the ranking and determination of the minimum number of TFs required to specifically cover a gene expression module associated with a physiological trait. http://wiki.c2b2.columbia.edu/califanolab/index.php/Software/MARINA

MR Activity Classifiers

The ARACNe-predicted T of IKZF1 and DLX4 were integrated with the exogenous gene expression studies to identify all genes in the AAGS that could be mapped as T of IKZF1 and DLX4. This was done by intersecting the ARACNe regulons of IKZF1 and DLX4 with the AAGS. The intersection of these two sets was then screened in the expression studies for any genes that responded with at least 25% fold change. This set of genes was used to construct a consensus "meta-activity" for the IKZF1 and DLX4 loci. The rank-normalized change of each gene across the AA patient cohort was integrated into an average as a consensus measure of the relative activity of the parent MR.

These values were subsequently used to define a 2D search space, $(X \times Y)$, where X = IKZF1 meta-activity and Y = DLX4 meta-activity, to classify each of the patients in the AA training set. The meta-activity vectors were rank transformed such that the minimum values were bound to the origin of the search space (0,0) and such that activity measures were positive. This transformation has no influence on the results other than projecting the search space into a more intuitive grid for display purposes, in which both axes are bound between [0,n], where n is positive.

Classification in this space was done using a modified nonlinear, soft-margin SVM algorithm. The algorithm is formalized:

$$X = ranksort(activity_{IKZF1})$$
$$Y = ranksort(activity_{DLX4})$$
$$define(A \times B) : \forall a \in X, \arg\max_{b \in Y} f(a,b) = \left(\frac{p(S_j, Q_I) \times p(S_i, Q_{III})}{p(S_i, Q_I) \times p(S_j, Q_{III})}\right)$$

The algorithm defines a vector set $(A \times B)$, which exists within the search space $(X \times Y)$, such that every given pair (a,b) maximizes the likelihood ratio $f(a,b)$. This function is defined such that S_j is the next order of disease severity to S_i and Q_I and Q_{III} are the quadrants I and III of the grid created by the hyperplanes $(a \times R)$ and $(R \times b)$. Samples in the training set are mapped to each grid with known molecular subtypes and the likelihood ratio is computed for the segregation of subtypes defined by S. The severity ranking used for S was Normal < Mild < Severe. Each coordinate set in $(A \times B)$ therefore defines the points to a nonlinear plane that maximizes the separation between samples of different molecular classes in the IKZF1/DLX4 meta-activity space.

Cytotoxicity Assay

PBMC-dependent cytotoxicity was measured using the CytoTox 96 Nonradioactive Cytotoxicity Assay available through Promega. For the processing of samples and solutions, we followed manufacturer protocols. The optimization for PBMC:T was done as below, but using variable concentrations (1:1, 5:1, and 10:1) (Figure S3).

Cytotoxicity experiments were set up in 96-well format, with each treatment done in triplicate. Transfections were done 36 hr prior to the experiment. The day of the experiment, HK and huDP cells were trypsinized and diluted with Dulbecco's modified Eagle's medium (DMEM) into working stocks. The T concentration per well was 80,000 cells in 50 μl DMEM, combined with 800,000 PBMCs. The NKG2D inhibitor was the Human NKG2D MAb (clone 149810) from R&D Systems (Cat. MAB139), used at a final concentration of 20 μg/ml. Each transfection was allocated in triplicate according to manufacturer instructions.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cels.2015.11.001.

AUTHOR CONTRIBUTIONS

A.M.C. and J.C.C. conceived the overall framework of this study. A.M.C. and J.C.C. wrote the manuscript. J.C.C. designed and implemented all computational analysis and experimental work and analyzed the data. R.C. provided his guidance and expertise in cancer and autoimmunity for the project. A.J. acquired sample biopsies and processed them for microarray analysis with analysis performed by J.E.C.

ACKNOWLEDGMENTS

We would like to thank Julian Mackay-Wiggan and the Alopecia Areata Registry sites for patient samples. We thank Drs. Zhenpeng Dai, Annemieke de Jong, and Lynn Petukhova for their stimulatory discussions. We thank the Department of Systems Biology and Andrea Califano for use of their resources and guidance. J.C.C. and J.E.C. were supported by the Kirschstein NRSA postdoctoral training grant in Medical Genetics, T32GM082271 (to A.M.C.). Funding for the project was provided by the Alopecia Areata Initiative R01AR056016 (to A.M.C.) and R21AR061881 (to A.M.C. and R.C.) and the Locks of Love Foundation. A.J. was supported by the Dermatology Foundation Physician Scientist Career Development Award and by the Louis V. Gerstner Jr. Scholars Program.

Received: May 4, 2015
Revised: August 19, 2015
Accepted: November 3, 2015
Published: November 25, 2015

REFERENCES

Aubry, S., Shin, W., Crary, J.F., Lefort, R., Qureshi, Y.H., Lefebvre, C., Califano, A., and Shelanski, M.L. (2015). Assembly and interrogation of Alzheimer's disease genetic networks reveal novel regulators of progression. PLoS ONE *10*, e0120352.

Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. Nat. Genet. *37*, 382–390.

Betz, R.C., Petukhova, L., Ripke, S., Huang, H., Menelaou, A., Redler, S., Becker, T., Heilmann, S., Yamany, T., Duvic, M., et al. (2015). Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci. Nat. Commun. *6*, 5966.

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A.C., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., et al. (2013). Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. Immunity *39*, 782–795.

Califano, A., Butte, A.J., Friend, S., Ideker, T., and Schadt, E. (2012). Leveraging models of cell regulation and GWAS data in integrative network-based association studies. Nat. Genet. *44*, 841–847.

Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne, S.L., Doetsch, F., Colman, H., et al. (2010). The transcriptional network for mesenchymal transformation of brain tumours. Nature *463*, 318–325.

Chen, J.C., Alvarez, M.J., Talos, F., Dhruv, H., Rieckhof, G.E., Iyer, A., Diefes, K.L., Aldape, K., Berens, M., Shen, M.M., and Califano, A. (2014). Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks. Cell *159*, 402–414.

Doi, H., Shibata, M.-A., Kiyokane, K., and Otsuki, Y. (2003). Downregulation of TGFbeta isoforms and their receptors contributes to keratinocyte hyperproliferation in psoriasis vulgaris. J. Dermatol. Sci. *33*, 7–16.

Dunn, G.P., Bruce, A.T., Ikeda, H., Old, L.J., and Schreiber, R.D. (2002). Cancer immunoediting: from immunosurveillance to tumor escape. Nat. Immunol. *3*, 991–998.

Farh, K.K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J., Shishkin, A.A., et al. (2014). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature *518*, 337–343.

Gambichler, T., Terras, S., and Skrygan, M. (2013). TGFβ/Smad signalling in psoriatic epidermis models exposed to salt water soaks and narrowband ultraviolet B radiation. Cytokine *64*, 35–38.

Hecker, M., Goertsches, R.H., Engelmann, R., Thiesen, H.-J., and Guthke, R. (2009). Integrative modeling of transcriptional regulation in response to antirheumatic therapy. BMC Bioinformatics *10*, 262.

Javierre, B.M., Rodriguez-Ubreva, J., Al-Shahrour, F., Corominas, M., Graña, O., Ciudad, L., Agirre, X., Pisano, D.G., Valencia, A., Roman-Gomez, J., et al. (2011). Long-range epigenetic silencing associates with deregulation of Ikaros targets in colorectal cancer cells. Mol. Cancer Res. *9*, 1139–1151.

Keller, M.P., Choi, Y., Wang, P., Davis, D.B., Rabaglia, M.E., Oler, A.T., Stapleton, D.S., Argmann, C., Schueler, K.L., Edwards, S., et al. (2008). A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility. Genome Res. *18*, 706–716.

Kleinmann, E., Geimer Le Lay, A.-S., Sellars, M., Kastner, P., and Chan, S. (2008). Ikaros represses the transcriptional response to Notch signaling in T-cell development. Mol. Cell. Biol. *28*, 7465–7475.

Lefebvre, C., Rajbhandari, P., Alvarez, M.J., Bandaru, P., Lim, W.K., Sato, M., Wang, K., Sumazin, P., Kustagi, M., Bisikirska, B.C., et al. (2010). A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. Mol. Syst. Biol. *6*, 377.

MacNeil, L.T., Pons, C., Arda, H.E., Giese, G.E., Myers, C.L., and Walhout, A.J.M. (2015). Transcription factor activity mapping of a tissue-specific in vivo gene regulatory network. Cell Syst. *1*, 152–162.

Mani, K.M., Lefebvre, C., Wang, K., Lim, W.K., Basso, K., Dalla-Favera, R., and Califano, A. (2008). A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. Mol. Syst. Biol. *4*, 169.

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

Margolin, A.A., Wang, K., Lim, W.K., Kustagi, M., Nemenman, I., and Califano, A. (2006b). Reverse engineering cellular networks. Nat. Protoc. *1*, 662–671.

Martinez-Mir, A., Zlotogorski, A., Gordon, D., Petukhova, L., Mo, J., Gilliam, T.C., Londono, D., Haynes, C., Ott, J., Hordinsky, M., et al. (2007). Genomewide scan for linkage reveals evidence of several susceptibility loci for alopecia areata. Am. J. Hum. Genet. *80*, 316–328.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457.

Olsen, E., Hordinsky, M., McDonald-Hull, S., Price, V., Roberts, J., Shapiro, J., and Stenn, K.; National Alopecia Areata Foundation (1999). Alopecia areata investigational assessment guidelines. J. Am. Acad. Dermatol. *40*, 242–246.

Petukhova, L., Duvic, M., Hordinsky, M., Norris, D., Price, V., Shimomura, Y., Kim, H., Singh, P., Lee, A., Chen, W.V., et al. (2010). Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. Nature *466*, 113–117.

Sakane, J., Taniyama, K., Miyamoto, K., Saito, A., Kuraoka, K., Nishimura, T., Sentani, K., Oue, N., and Yasui, W. (2015). Aberrant DNA methylation of DLX4 and SIM1 is a predictive marker for disease progression of uterine cervical low-grade squamous intraepithelial lesion. Diagn. Cytopathol. *43*, 462–470, http://dx.doi.org/10.1002/dc.23256.

Sato, M., Goto, S., Kaneko, R., Ito, M., Sato, S., and Takeuchi, S. (1998). Impaired production of Th1 cytokines and increased frequency of Th2 subsets in PBMC from advanced cancer patients. Anticancer Res. *18* (5D), 3951–3955.

Shelanski, M., Shin, W., Aubry, S., Sims, P., Alvarez, M.J., and Califano, A. (2015). A systems approach to drug discovery in Alzheimer's disease. Neurotherapeutics *12*, 126–131.

Suárez-Fariñas, M., Tintle, S.J., Shemer, A., Chiricozzi, A., Nograles, K., Cardinale, I., Duan, S., Bowcock, A.M., Krueger, J.G., and Guttman-Yassky, E. (2011). Nonlesional atopic dermatitis skin is characterized by broad terminal differentiation defects and variable immune abnormalities. J. Allergy Clin. Immunol. *127* (4), 954–964.e1–4, http://dx.doi.org/10.1016/j.jaci.2010.12.1124.

Tomida, S., Yanagisawa, K., Koshikawa, K., Yatabe, Y., Mitsudomi, T., Osada, H., and Takahashi, T. (2007). Identification of a metastasis signature and the DLX4 homeobox protein as a regulator of metastasis by combined transcriptome approach. Oncogene *26*, 4600–4608.

Xing, L., Dai, Z., Jabbari, A., Cerise, J.E., Higgins, C.A., Gong, W., de Jong, A., Harel, S., DeStefano, G.M., Rothman, L., et al. (2014). Alopecia areata is driven by cytotoxic T lymphocytes and is reversed by JAK inhibition. Nat. Med. *20*, 1043–1049.

Yao, Y., Richman, L., Morehouse, C., de los Reyes, M., Higgs, B.W., Boutrin, A., White, B., Coyle, A., Krueger, J., Kiener, P.A., and Jallal, B. (2008). Type I interferon: potential therapeutic target for psoriasis? PLoS ONE *3*, e2737.

Zhang, B., Gaiteri, C., Bodea, L.-G., Wang, Z., McElwee, J., Podtelezhnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., et al. (2013). Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell *153*, 707–720.

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

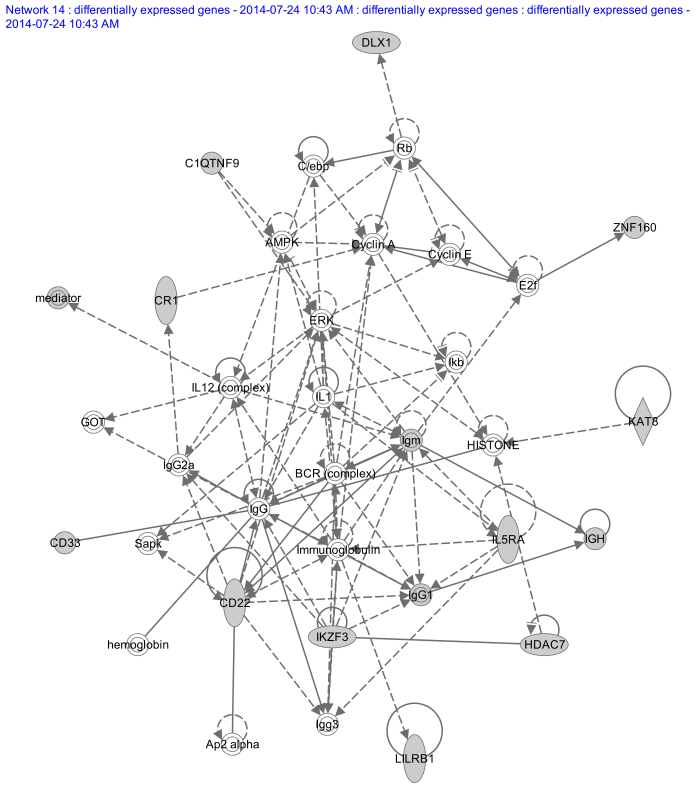
Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla-Favera, R., & Califano, A. (2006a). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, *7 Suppl 1*(Suppl 1), S7. http://doi.org/10.1186/1471-2105-7-S1-S7

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

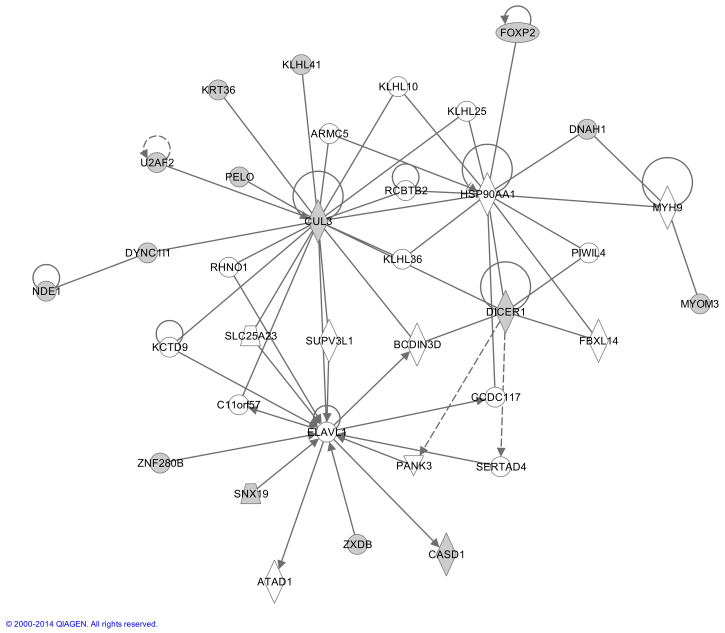
Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43),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

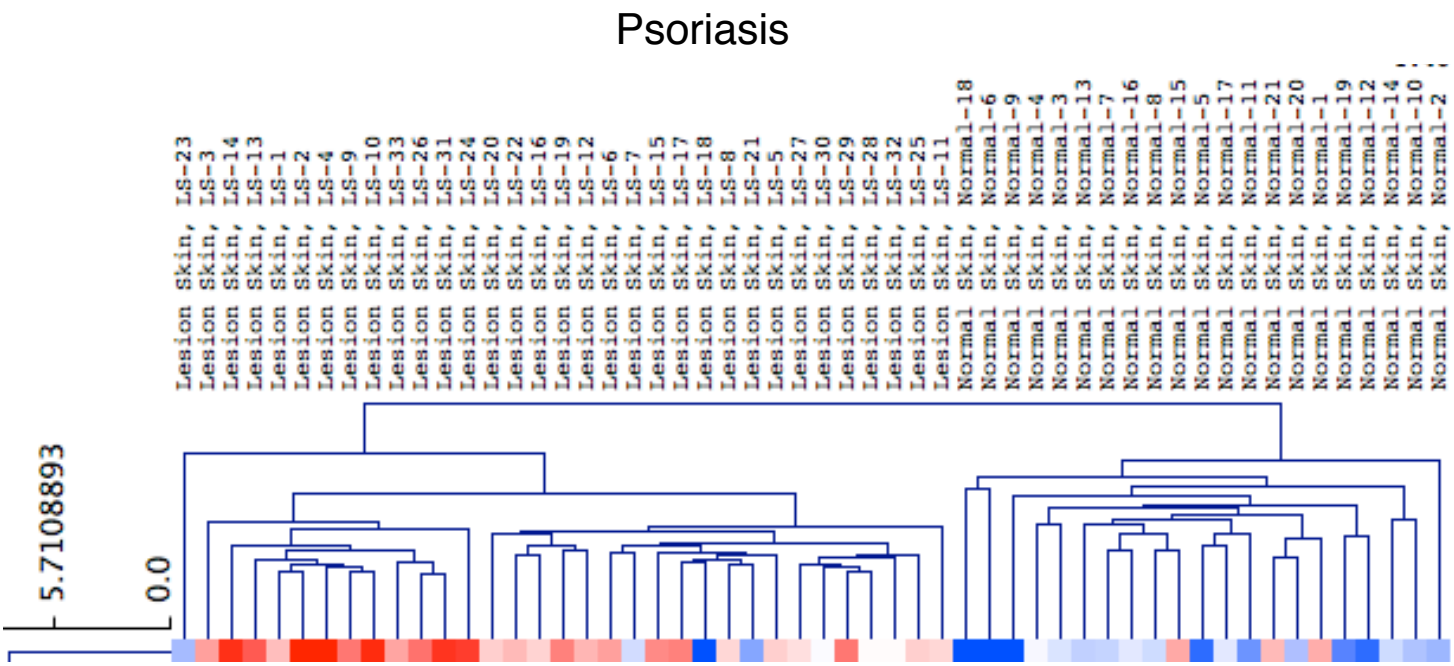
Immune interaction network

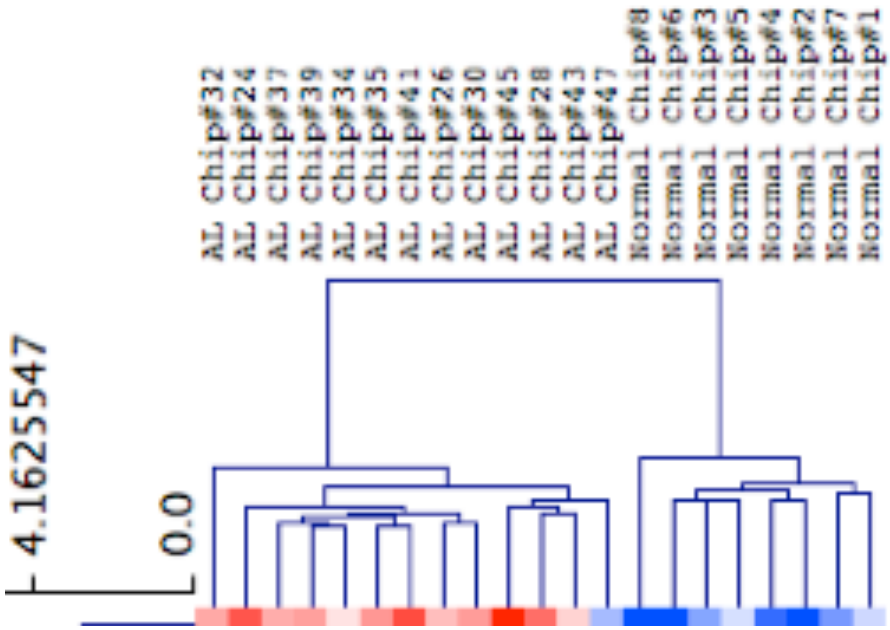


Cytotoxic interaction network



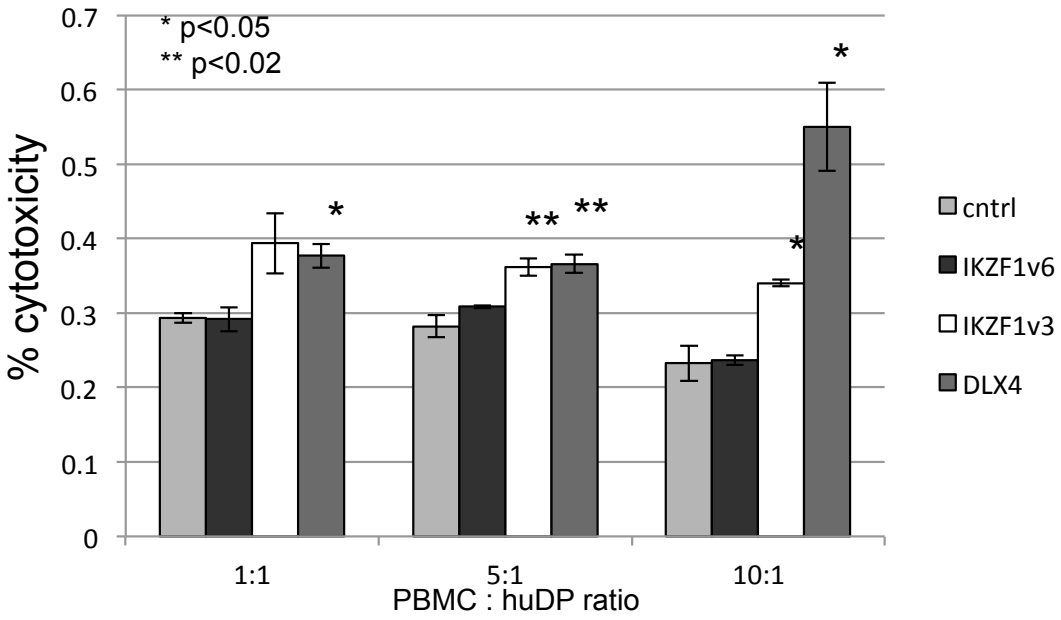
Supplemental Figure 2. Psoriasis and Atopic Dermatitis cohorts have gene expression signatures that clearly delineate patients from unaffected controls





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

A



B