Domino: reconstructing intercellular signaling dynamics with transcription factor activation in model biomaterial environments

Christopher Cherry[1], Patrick Cahan[2], Lana X Garmire[3], Jennifer H Elisseeff[1,4*]

[1]Translational Tissue Engineering Center, Wilmer Eye Institute and the Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD;

[2]Department of Biomedical Engineering and Institute for Cell Engineering, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[3]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor. MI 48105

[4]Bloomberg~Kimmel Institute for Cancer Immunotherapy and Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD;

[*]To whom correspondence should be addressed: jhe@jhu.edu

Intercellular communication is crucial for function of cells but current methods for reconstruction of intercellular signaling from single cell data can be challenging to connect with biological outcomes. Here, we present Domino (github.com/chris-cherry/domino), a computational tool which reconstructs intercellular signaling coupled with transcription factor activation. Matching signaling with transcription factor activation allows for immediate connection of results to known biological phenotypes. We first apply Domino to model biomaterial environments in a muscle wound where different chemical compositions produce distinct tissue microenvironments. While traditional single cell RNA sequencing analyses captured few differences between conditions, Domino identified previously unknown signaling pathways distinct to conditions. We then demonstrate Domino on a previously published human Alzheimer's data set to discover new signaling pathways associated with health and disease.

Introduction

Single cell technologies provide unprecedented capabilities to rapidly assess the cellular content of tissues during development and in health and disease. Programs such as the Human Cell Atlas (Regev et al., 2017) and the Tabula Muris Consortium (Schaum et al., 2018) continue to increase the number of data sets available for analysis that cover a broad range of conditions. While extensive information is achieved from single cell sequencing, there are still limitations in the depth of sequencing and capture of rare cell populations with significant biological implications that can limit understanding (Rizzetto et al., 2017; Svensson et al., 2017). New strategies to uncover more biology from each cell experimentally are developing, including mult-iomics (Angermueller et al., 2016; Hou et al., 2016) and spatial analysis (Rodriques et al., 2019; Vickovic et al., 2019; Xia et al., 2019). However, new computational techniques can also be leveraged to enable greater biological understanding and therapeutic targets from new and existing single cell sequencing datasets (Lähnemann et al., 2020). A key area where computational techniques can increase biological understanding is cell-cell communication which is central to many processes in tissue homeostasis and dysfunction.

Cell-cell communication depends primarily on ligands (L), secreted by a signaling cell, that diffuses to and activates receptors (R) on target cells which subsequently leads to transcriptional activation and phenotypic changes. Computational tools available today correlate expression of ligands and receptors across groups of cells (Efremova et al., 2020; Noël et al., 2020; Tsuyuzaki et al., 2019; Wang et al., 2019a; Wang et al., 2019b). These methods do not account for activation of the receptor and downstream signaling which are influenced by multiple factors beyond ligand and receptor expression. These confounding factors reduce the ability of ligand and receptor mRNA counts to accurately predict cell-cell communication. Furthermore, some receptors may be activated by multiple ligands and receptor activation in some cases can lead to different transcriptional activation. NicheNet (Browaeys et al., 2020) addresses this issue by first identifying activation of a phenotypic feature and then receptors and ligands related to that feature, increasing the probability that the ligand-receptor pairs are activated and inform the associated phenotype with that activation. This program requires a user-defined set of genes associated with a feature to screen for expression of ligand-receptor (LR) pairs. Unfortunately, previously defined gene sets relevant to specific experimental conditions do not always exist. In such cases interpretation of results can be challenging.

To address the challenge of defining cell-cell communication from single cell data sets in an unbiased manner without prior knowledge of LR pairs, we developed Domino. Domino reconstructs intracellular communication based on transcription factor (TF) activation. We use transcription factor activation scores derived from gene regulatory network analysis to identify upstream receptors, so a user-defined gene set is not required. Pearson correlation is used to link receptors with transcription factor and finally identify ligands capable of activating those receptors. This analysis is performed globally with individual cells. Signaling networks from differing experimental conditions can then be compared to identify condition specific transcription factor-receptor correlations, independent of clustering methodology. The signaling network can also be used to generate cluster-cluster communications and investigate signaling specific to clusters. We demonstrate the ability of Domino to identify unique cell-cell communication and activation in a new aggregated single cell data set from biomaterial tissue environments with similar clusters via standard algorithms yet with known differences in immune profile and physiological outcomes. Finally, we apply Domino to a previously published data set from Alzheimer's patients (Grubman et al., 2019) and identify new signaling pathways specific to diseased or healthy brain.

Results

Generation of intracellular and intercellular signaling networks

Intercellular communication can be modeled as the result of four biological steps (Fig. 1A). Ligands are produced by signaling cells, ligands bind to and activate receptors on target cells, receptor activation triggers a signaling cascade, and transcription factors at the end of the cascade initiate transcription of specific genes. These transcriptional changes then lead to phenotypic changes in the target cell population. Domino reconstructs cell communication events in reverse to identify signaling connected with specific phenotypic changes (Fig. 1B).

To reconstruct cell communication, we first use the SCENIC (Aibar et al., 2017) gene regulatory network analysis pipeline to generate transcription factor activation scores from raw counts data. Second, transcription factor scores are connected with receptors in order to identify receptors potentially upstream of transcription factor activation using Pearson correlation. To ensure correlation is not due to transcription factor targeting of the receptor, receptors found in individual transcription factor modules are excluded from correlation. Finally, a publicly available ligand-receptor database (Efremova et al., 2020) is queried to identify potential ligands for receptors.

The L-R-TF linkages assembled from the cell communication reconstruction form a global signaling network for the data set irrespective of any clustering. This process allows for unsupervised exploration of L-R-TF activation in single cell data sets. This method of reverse assembly allows for identification of receptors correlated with transcription factors. Because transcription factor activation in specific cell types is associated with phenotypic changes, it allows for prediction of downstream biological changes. Further, condition-specific signaling networks can be calculated and then compared to identify transcription factors and receptors unique to experimental conditions or disease state.

The global signaling network is independent of clustering but can be integrated with cluster labels to investigate intercluster signaling dynamics. To identify signaling relevant for specific clusters, we first identify transcription factors enriched by cluster (Fig. 1C). The global signaling network is then pruned for nodes disconnected with the cluster-specific transcription factors to generate a cluster-specific signaling network. This process selects for ligands not only capable of binding to receptors expressed by the target cluster but also enriches ligands likely to be responsible for activation of transcription factors. These sets of ligands are then used to generate a cluster-cluster signaling network based on global expression patterns of the ligand sets (Fig. 1D).

A single cell atlas and connectome of the biomaterial tissue microenvironment

Biomaterials are a component of medical devices and implants (Christman, 2019). They are also used to create specialized cell and tissue microenvironments to probe biological questions (Dye et al., 2020; Xie and Murphy, 2019). In this case, biomaterials can serve as a model for characterization of cell populations in controlled local environments. While many aspects of the physiological response to biomaterials have been studied over the years, single cell studies are limited. Application of single cell analyses is poised to transform our understanding of the biomaterial response and thus biomaterial design.

We applied clinically relevant biomaterials that induce divergent immune phenotypes and physiological outcomes in a muscle wound. Specifically, we used a biological scaffold derived from urinary bladder extracellular matrix (ECM) to induce a pro-regenerative tissue environment and a type 2 immune response characterized by interleukin 4 (IL4) (Sadtler et al., 2016; Wolf et al., 2019) and the synthetic material polycaprolactone (PCL) that promotes fibrosis, the development of senescent cells and a type 3 immune environment characterized by IL17 (Anderson et al., 2008; Chung et al., 2020).

To create a single cell atlas of the biomaterial response, we assembled data sets from multiple experiments with ECM and PCL implanted in a murine volumetric muscle loss wound in

young and old animals for 1 and 6 weeks. Fibroblasts were sorted as CD45-CD31-CD29+ and processed for single cell RNA sequencing as previously described (Macosko et al., 2015). The compiled dataset from sorted fibroblasts, sorted macrophages (Sommerfeld et al., 2019), and $CD45^{+}$ enriched whole tissue (Han et al., in preparation) includes a total of 42,156 cells (after filtering) with an average of 198,000 reads per cell. After integration with Harmony and shared-nearest neighbor based clustering, we identified 18 clusters of cells characterized by expression of signature genes (Fig. 2A-C, Supplementary Table 2).

Despite known immunological and physiological differences in the outcomes, cell numbers did not differ significantly between clusters (Supplementary Figure 2) and differential expression between conditions within each cluster found few strong condition specific expression signatures (Supplementary Figure 3). Because of the gap between known divergent biological outcomes and the single cell clustering and differential expression, we hypothesized that differences in signaling between the similar cell populations may be responsible for the distinct outcomes that could not be captured with standard analysis.

Domino identifies signaling patterns associated with biomaterial conditions and physiological outcomes.

To identify signaling patterns that were specific to a treatment condition, we developed a method of cross-comparison of signaling networks between conditions (Fig. 2D). We applied Domino on the data set to determine global signaling networks for the ECM (Fig. 3A) and PCL (Fig. 3D) tissue environments. In both networks, receptors and transcription factors self-assembled into modules enriched in fibroblasts, immune cells, or tissue-specific cells when visualized with a force-directed layout. To identify condition specific signaling components, we performed set comparisons between the gene members of the two networks. These transcription factors, receptors, and ligands represent possible signaling pathways specific to each condition.

ECM specific transcription factors were enriched in fibroblasts, tissue-specific cells, and immune cells according to their modules as predicted (Fig. 3B). Many of their predicted receptors were specific to ECM as well. Taken together, these transcription factors and their receptors represent signaling pathways which may be active specifically in the ECM biomaterial environment that contribute to wound healing and resolution (Fig. 3C). For example, we show *Esrra* as a downstream target of *Il4ra* in myeloid cells in ECM. *Esrra* activates anti-inflammatory macrophages (Yuk et al., 2015), which are induced by IL4 and have been shown to be active in response to ECM. We further show activation of *Sox17* in endothelial cells linked to expression of the receptor *Osmr. Sox17* has been shown as critical for endothelial cell regeneration in response to injury (Liu et al., 2019) but not identified in ECM response. Finally, we also identify activation of *Ctcf*, a muscle specific transcription factor (Delgado-Olguín et al., 2011), in fibroblasts linked to *Tnfrsf12a* or *TweakR*, recently shown to improve burn wound healing (Liu et al., 2018) but connections with ECM response are not known. These three findings identify IL4, OSM, and TWEAK signaling pathways as potential therapeutic targets to promote regenerative wound healing.

The PCL specific transcription factors and their predicted receptors were also organized by enrichment in specific cell subsets (Fig. 3E). In contrast to ECM, the PCL-specific transcription factors associated with tissue-specific cell subsets had modest to low expression levels. Instead, the cell specific subsets shared expression patterns similar to fibroblasts which may suggest suppression of tissue specific regenerative programs. Using Domino's signaling predictions, we identified a linkage between *Tgfbr2* and both *Sox11* and *Sox4* (Fig. 3F). Related to these findings, Bhattaram et al. demonstrated that *Sox4* and *Sox11* induce transformation of synoviocytes to fibroblast-like synoviocytes in rheumatoid arthritis (Bhattaram et al., 2018). These cells are thought to be responsible for the inflammatory environment that drives RA disease but their involvement in biomaterials response is unknown. Transforming growth factor

beta (TGF-β) signaling is active in fibrotic environments like the foreign body response (Gerarduzzi and Di Battista, 2017). While it is known that TGF-β signaling is active in fibrosis, these findings suggest that induction of an inflammatory phenotype in fibroblasts driven by *Sox4* and *Sox11* may regulate TGF-β driven fibrosis. We also identify *Pirb* in the PCL myeloid population correlated with pro-inflammatory transcription factor *Irf4*. Interestingly, *Pirb* is an immune checkpoint regulating anti-inflammatory effects in myeloid populations (Bashirova et al., 2014). This suggests that *Pirb* may be a novel target to reduce chronic inflammation associated with fibrosis in response to PCL.

Comparison of transcription factor and gene set based methods

To compare results from transcription factor (Domino) and gene set programs, we applied NicheNet on the same biomaterials data set. We used genes differentially expressed between ECM and PCL conditions (Supplementary Figures 4-6). NicheNet defines overrepresentation of ligand regulatory targets in a given set of genes. Because we did not have specific gene sets associated with biomaterials response *a priori*, we used sets of differentially expressed genes between ECM and PCL. As described earlier (Supplementary Fig. 3), differentially expressed genes by condition had few differences in relative expression (Supplementary Figures 4C, 5C, and 6C). We also ran gene set analysis using PANTHER (Mi et al., 2018) with the Reactome Pathways (Jassal et al., 2019) gene sets (Supplementary Table 3). Collagen genes present in the fibroblast and tissue subsets drove enrichment of pathways regulating collagen and extracellular matrix. Pathways provided by PANTHER in the immune cells were unrelated to immune function.

In both fibroblasts and tissue specific cells *Tgfb1* is predicted as an important signaling ligand connected to the differentially expressed collagens. Expression levels for the collagens are similar in both ECM and PCL (Supplementary Figures 4C, 5C). Because of this the relationship between *Tgfb1* signaling to collagens and biomaterials environment is unclear. Domino connected *Tgfbr2* to an inflammatory program in the fibroblasts driven by *Sox4* and *Sox11*. Other predicted ligands from NicheNet were difficult to interpret. The ligand targets were not associated with specific enriched pathways and the relevance of individual genes was not clear due to the lack of clear expression differences between conditions.

The immune cells had stronger differences in expression signatures between ECM and PCL (Supplementary Fig 6C). NicheNet found *Tgfb1* as a potential regulator of some cytokines and chemokines including *Il1b, Ccl12,* and *Ccl5*, all of which were upregulated in PCL. In contrast, Domino identified *Ccr2* and *Csf2ra* as regulators of transcription factors *Ascl2* and *Tfeb,* promoters of inflammatory response in the PCL environment, in the myeloid population. It also identified a myeloid immune checkpoint *Pirb* as a potential target in response to PCL.

While NicheNet findings clearly connect ligands with potential regulatory targets, the results were frequently difficult to interpret in our data. There is no clearly defined gene set associated with components of biomaterials response, so we generated sets by comparing conditions with differential expression. In order to interpret importance of ligands regulating these genes, the importance of the gene set must be known. We attempted gene set analysis to determine this, but results were broad and, in the case of immune cells, irrelevant to immune function (Supplementary Table 3). In some cases, knowledge about individual genes was sufficient to determine their significance to biomaterials response but in many cases, it was not, making interpretation of ligand importance impossible. With good quality gene sets *a priori* NicheNet provides precise analysis of signaling connected to context specific cellular function.

Domino generates new hypotheses from a publicly available Alzheimer's data set

To test applicability of Domino to other data sets and tissue environments, we analyzed a publicly available data set of healthy brain and brain with Alzheimer's disease (AD) (Grubman et al., 2019). We split the data by disease status and repeated the analysis pipeline to identify

signaling pathways specific to each condition. The original publication includes gene regulatory network analysis, so we used the clustering provided to ensure identical comparison groups in order to determine whether our analysis provided any new insights.

Similar to the different biomaterials environments, both Alzheimer's and healthy signaling networks organized into modules associated with cell types (Fig. 4A, D). Additionally, condition specific transcription factors were connected almost entirely with condition specific receptors (Fig 4B, E). The results for AD indicated IL17RB as a potential receptor upstream of FOS (Fig. 4C). In support of this finding, IL17 neutralizing antibody has recently been shown to improve Alzheimer's disease (Cristiano et al., 2019). Our data suggest that it may be due to FOS activation in astrocytes which has been associated with Alzheimer's (Anderson et al., 1994). We also found that neurotrophin-3 signaling through NTRK3 may activate BHLHE40 in astrocytes, a transcription factor associated with autoimmunity (Chih-Chung, 2017) suggesting further links to IL17 signaling (Zhu and Qian, 2012). While BHLHE40 was not identified as a relevant transcription factor in the dataset's authors' analysis, FOS did appear in a list of active transcription factors in brains from Alzheimer's patients but was not discussed further.

Domino also identified several transcription factors and correlated receptors specific to the healthy brain (Fig. 4E). For example, we identified brain derived neurotrophic factor signaling through *NTRK2* targeting *PAX6* in astrocytes uniquely in healthy brain (Fig 4F). Polymorphisms in *NTRK2* have been correlated with Alzheimer's disease (Zeng et al., 2013) and *PAX6* activation has been shown to activate neurogenesis in astrocytes (Sakurai and Osumi, 2008), suggesting that interference in *NTRK2* function may lead to reduced *PAX6* activation and reduced neurogenesis in Alzheimer's. We also show activin signaling through *ACVR1C* that may induce *SOX10* activation in oligodendrocytes, a transcription factor necessary for survival of myelin-producing oligodendrocytes (Takada et al., 2010). Like the previously identified AD-associated signaling pathways, these findings were not found in the original analysis.

Discussion

In summary, Domino provides a method to evaluate signaling networks using transcription factor activation without need for predefined gene sets. Creating an "atlas" with data from a large number of cells may not be adequate to accurately define physiological properties or therapeutic targets. Domino generated unique signaling networks and activated cell populations in a large single cell data set from different biomaterial microenvironments that had minimal differential gene expression or cell clustering distribution. We believe Domino provides a new resource for unsupervised exploration of condition specific signaling patterns and generation of biologically testable signaling hypotheses.

Author Contributions

C.C. and J.H.E. conceptualized and drafted figures and manuscript, contributed to experimental design, and interpretation of findings. C.C. developed software. All authors contributed to computational methodology and revised manuscript and figures.

Acknowledgements

The authors gratefully acknowledge financial support from the Morton Goldberg Chair, NIH Directors Pioneer Award, Bloomberg-Kimmel Institute for Cancer Immunotherapy. Lana Garmire is supported by grants K01ES025434 awarded by NIEHS through funds provided by the trans-NIH Big Data to Knowledge (BD2K) initiative (www.bd2k.nih.gov), R01 LM012373 and LM012907 awarded by NLM, and R01 HD084633 awarded by NICHD. We would also like to thank Elana Fertig for their suggestions on some of the computational methods used to develop Domino.

Methods

Data availability

All raw data, processed files and commented code are made available at GEOXXXXXXXXXX.

Data preprocessing

Seurat was used for most processing steps where other software is not specified (Satija et al., 2015). All cell counts were pruned of cells with UMI counts below 250, cells with more than 10% mitochondrial genes, and genes expressed in fewer than 0.1% of cells. We then normalized and scaled the data with regression on UMI count and percent mitochondrial genes and calculated principle components using the top 2000 most variable genes. For muscle data sets we then corrected the principle components for batch effect using Harmony (Korsunsky et al., 2019). UMAP and shared nearest neighbor graph construction with subsequent Louvain clustering was then run on principle components.

Phenotypic assignment of clusters

Differential expression testing for clusters was run using Mann-Whitney U tests. Each cluster was compared against all other clusters. The resulting gene expression profiles were examined to determine cluster phenotype. In many cases, the source publications for the sorted data sets identified similar clusters which we used to assist in phenotypic assignment of clusters.

Gene regulatory network analysis

We used the SCENIC (Aibar et al., 2017) analysis pipeline to identify modules of genes targeted by transcription factors and calculate cell level enrichment scores. Genome ranking databases and *cis*-regulatory motif annotations were obtained from cisTarget Databases. First, we used Arboreto to fit a stochastic gradient boost machine using transcription factor counts to predict gene counts. Modules of genes targeted by transcription factor were then formed from the adjacencies including genes with feature importances greater than the 95 percentile. The modules were then pruned cross-referencing the motif annotations and ranking databases to remove modules with less than 80% of genes mapping to regions near binding sites for transcription factors or with less than 20 gene targets. Finally, enrichment for these modules was calculated using AUCell to identify cells with enrichment of genes targeted by transcription factors.

Construction of a global signaling network

A list of human ligands, receptors, and their signaling relationships was obtained from CellphoneDB2 (Efremova et al., 2020). We then used biomaRt (Durinck et al., 2009) to convert genes from HGNC to MGI symbols, taking all conversions for each gene when multiple were found. Prior to calculating signaling relationships, counts matrices were pruned for genes expressed in fewer than 2.5% of cells. Pearson correlations were calculated between transcription factor activation scores and normalized, z-scored expression for identified receptors across all cells. Correlation between receptors which were determined as transcription factor targets by gene regulatory network analysis were then set to zero. This prevents targets of transcription factors, which would be correlated with transcription factor activation, from being interpreted as upstream of its transcription factor. Finally, transcription factors and receptors were considered signaling connections if Pearson correlation was greater than 0.3 with a maximum of ten receptors per transcription factor. For all receptors connected with transcription factors, ligand signaling partners were identified from the CellphoneDB2 database. Ligands that were not found in the data set were excluded.

Comparison of condition specific global signaling networks

In order to identify intracellular signaling patterns associated with an independent variable, we first split the single cell data by that variable. More specifically, we split the volumetric muscle data set by treatment with ECM or PCL and the Alzheimer's data set by disease status of the patients. We then constructed a global signaling network for each of the separate data sets with identical parameters. Finally, we identified network items specific to each condition by set subtraction. Transcription factors or receptors only present in one condition's signaling network were considered as potentially condition-specific.

Cluster specific subnetwork identification
In order to identify intracellular signaling patterns within a cluster together with ligands responsible for their activation, we first identified active transcription factors by cluster using Mann-Whitney U tests. For each cluster, the top over-expressed genes were selected based on p-value with positive log fold change as compared to all other clusters. Transcription factors with p-values below .001 were included with up to 10 transcription factors per cluster. We then generated a signaling subnetwork for each cluster by pruning all network items not connected to the cluster enriched transcription factors.

Prediction of intercellular signaling networks
Using the cluster specific signaling subnetworks, we identified ligands most likely to be responsible for activation of cellular phenotype for each cluster specifically which we deem incoming ligands. It's important to note that in the data sets we've analyzed there tend to be many ligands below detection threshold, so we don't use signaling pathways without expressed ligands in construction of intercellular signaling networks. To calculate intercluster signaling with ligands found in the data set, we first averaged z-scored ligand expression by cluster. We then generated a signaling score by summing the averaged z-scores by cluster. The signaling represents whether a particular cluster is over- or under-expressing the ligands predicted to activate the target cluster. These values are used as directed, weighted edges between clusters as nodes to construct an intercellular signaling network.

NicheNet analysis
NicheNet (Browaeys et al., 2020) was run using recommended settings according to vignettes available on github. In order to compare results directly to Domino, we grouped cells as fibroblasts, tissue-specific cells, or immune cells. Mast, DC, Ma/Mo/Fib Ifn, Ma/Mo IL36y, T/NK, Mac R2, Ma/Mo Inflam, and Mac pre clusters were considered immune cells. Muscle, Neuro, Endo, and Satellite clusters were considered tissues-specific. Fib peri, Fib cart, Fib immune, Fib pre, and Fib generic were considered fibroblasts. To generate sets as input for NicheNet, we ran differential expression comparing ECM and PCL within each of these subsets. The resulting gene sets were run using NicheNet with default parameters. In order to classify the regulatory targets of NicheNet we used PANTHER (Mi et al., 2018) with the Reactome pathway database. As background we used all genes present in both the data set as well as the NicheNet global regulatory set.

References

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J.*, et al.* (2017). SCENIC: single-cell regulatory network inference and clustering. Nature Methods *14*, 1083-1086.

Anderson, A.J., Cummings, B.J., and Cotman, C.W. (1994). Increased Immunoreactivity for Jun- and Fos-Related Proteins in Alzheimer's Disease: Association with Pathology. Experimental Neurology *125*, 286-295.

Anderson, J.M., Rodriguez, A., and Chang, D.T. (2008). Foreign body reaction to biomaterials. Seminars in Immunology *20*, 86-100.

Angermueller, C., Clark, S.J., Lee, H.J., Macaulay, I.C., Teng, M.J., Hu, T.X., Krueger, F., Smallwood, S.A., Ponting, C.P., Voet, T.*, et al.* (2016). Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nature Methods *13*, 229-232.

Bashirova, A.A., Apps, R., Vince, N., Mochalova, Y., Yu, X.G., and Carrington, M. (2014). Diversity of the human LILRB3/A6 locus encoding a myeloid inhibitory and activating receptor pair. Immunogenetics *66*, 1-8.

Bhattaram, P., Muschler, G., Wixler, V., and Lefebvre, V. (2018). Inflammatory Cytokines Stabilize SOXC Transcription Factors to Mediate the Transformation of Fibroblast-Like Synoviocytes in Arthritic Disease. Arthritis Rheumatol *70*, 371-382.

Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. Nature Methods *17*, 159-162.

Chih-Chung, L. (2017). The Role of Bhlhe40 in Autoimmune Neuroinflammation and Mycobacterial Infection. Arts & Sciences Electronic Theses and Dissertations *1125*.

Christman, K.L. (2019). Biomaterials for tissue repair. Science *363*, 340.

Chung, L., Maestas, D.R., Lebid, A., Mageau, A., Rosson, G.D., Wu, X., Wolf, M.T., Tam, A.J., Vanderzee, I., Wang, X.*, et al.* (2020). Interleukin 17 and senescent cells regulate the foreign body response to synthetic material implants in mice and humans. Science Translational Medicine *12*, eaax3799.

Cristiano, C., Volpicelli, F., Lippiello, P., Buono, B., Raucci, F., Piccolo, M., Iqbal, A.J., Irace, C., Miniaci, M.C., Perrone Capano, C.*, et al.* (2019). Neutralization of IL-17 rescues amyloid-β-induced neuroinflammation and memory impairment. British Journal of Pharmacology *176*, 3544-3557.

Delgado-Olguín, P., Brand-Arzamendi, K., Scott, I.C., Jungblut, B., Stainier, D.Y., Bruneau, B.G., and Recillas-Targa, F. (2011). CTCF promotes muscle differentiation by modulating the activity of myogenic regulatory factors. J Biol Chem *286*, 12483-12494.

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nature Protocols *4*, 1184-1191.

Dye, B.R., Youngblood, R.L., Oakes, R.S., Kasputis, T., Clough, D.W., Spence, J.R., and Shea, L.D. (2020). Human lung organoids develop into adult airway-like structures directed by physico-chemical biomaterial properties. Biomaterials *234*, 119757.

Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nature Protocols *15*, 1484-1506.

Gerarduzzi, C., and Di Battista, J.A. (2017). Myofibroblast repair mechanisms post-inflammatory response: a fibrotic perspective. Inflammation Research *66*, 451-465.

Grubman, A., Chew, G., Ouyang, J.F., Sun, G., Choo, X.Y., McLean, C., Simmons, R.K., Buckberry, S., Vargas-Landin, D.B., Poppe, D.*, et al.* (2019). A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. Nature Neuroscience *22*, 2087-2097.

Han, J., Cherry, C., *et al.* (in preparation). Aged immunity driven by IL-17 signaling prevent effective healing response to biomaterials.

Hou, Y., Guo, H., Cao, C., Li, X., Hu, B., Zhu, P., Wu, X., Wen, L., Tang, F., Huang, Y.*, et al.* (2016). Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. Cell Research *26*, 304-319.

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R.*, et al.* (2019). The reactome pathway knowledgebase. Nucleic Acids Research *48*, D498-D503.

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-r., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nature Methods *16*, 1289-1296.

Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D.J., Hicks, S.C., Robinson, M.D., Vallejos, C.A., Campbell, K.R., Beerenwinkel, N., Mahfouz, A.*, et al.* (2020). Eleven grand challenges in single-cell data science. Genome Biology *21*, 31.

Liu, J., Peng, L., Liu, Y., Wu, K., Wang, S., Wang, X., Liu, Q., Xia, Y., and Zeng, W. (2018). Topical TWEAK Accelerates Healing of Experimental Burn Wounds in Mice. Frontiers in Pharmacology *9*.

Liu, M., Zhang, L., Marsboom, G., Jambusaria, A., Xiong, S., Toth, P.T., Benevolenskaya, E.V., Rehman, J., and Malik, A.B. (2019). Sox17 is required for endothelial regeneration following inflammation-induced vascular injury. Nature Communications *10*, 2126.

Macosko, Evan Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, Allison R., Kamitaki, N., Martersteck, Emily M.*, et al.* (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202-1214.

Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P.D. (2018). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Research *47*, D419-D426.

Noël, F., Massenet-Regad, L., Carmi-Levy, I., Cappuccio, A., Grandclaudon, M., Trichot, C., Kieffer, Y., Mechta-Grigoriou, F., and Soumelis, V. (2020). ICELLNET: a transcriptome-based framework to dissect intercellular communication. bioRxiv, 2020.2003.2005.976878.

Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M.*, et al.* (2017). The Human Cell Atlas. eLife *6*, e27041.

Rizzetto, S., Eltahla, A.A., Lin, P., Bull, R., Lloyd, A.R., Ho, J.W.K., Venturi, V., and Luciani, F. (2017). Impact of sequencing depth and read length on single cell RNA sequencing data of T cells. Scientific Reports *7*, 12781.

Rodriques, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R., Welch, J., Chen, L.M., Chen, F., and Macosko, E.Z. (2019). Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science *363*, 1463.

Sadtler, K., Estrellas, K., Allen, B.W., Wolf, M.T., Fan, H., Tam, A.J., Patel, C.H., Luber, B.S., Wang, H., Wagner, K.R.*, et al.* (2016). Developing a pro-regenerative biomaterial scaffold microenvironment requires T helper 2 cells. Science *352*, 366.

Sakurai, K., and Osumi, N. (2008). The Neurogenesis-Controlling Factor, Pax6, Inhibits Proliferation and Promotes Maturation in Murine Astrocytes. The Journal of Neuroscience *28*, 4604.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nature Biotechnology *33*, 495-502.

Schaum, N., Karkanias, J., Neff, N.F., May, A.P., Quake, S.R., Wyss-Coray, T., Darmanis, S., Batson, J., Botvinnik, O., Chen, M.B.*, et al.* (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature *562*, 367-372.

Sommerfeld, S.D., Cherry, C., Schwab, R.M., Chung, L., Maestas, D.R., Laffont, P., Stein, J.E., Tam, A., Ganguly, S., Housseau, F.*, et al.* (2019). Interleukin-36γ–producing macrophages drive IL-17–mediated fibrosis. Science Immunology *4*, eaax4783.

Svensson, V., Natarajan, K.N., Ly, L.-H., Miragaia, R.J., Labalette, C., Macaulay, I.C., Cvejic, A., and Teichmann, S.A. (2017). Power analysis of single-cell RNA-sequencing experiments. Nature Methods *14*, 381-387.

Takada, N., Kucenas, S., and Appel, B. (2010). Sox10 is necessary for oligodendrocyte survival following axon wrapping. Glia *58*, 996-1006.

Tsuyuzaki, K., Ishii, M., and Nikaido, I. (2019). Uncovering hypergraphs of cell-cell interaction from single cell RNA-sequencing data. bioRxiv, 566182.

Vickovic, S., Eraslan, G., Salmén, F., Klughammer, J., Stenbeck, L., Schapiro, D., Äijö, T., Bonneau, R., Bergenstråhle, L., Navarro, J.F.*, et al.* (2019). High-definition spatial transcriptomics for in situ tissue profiling. Nature Methods *16*, 987-990.

Wang, S., Karikomi, M., MacLean, A.L., and Nie, Q. (2019a). Cell lineage and communication network inference via optimization for single-cell transcriptomics. Nucleic Acids Research *47*, e66-e66.

Wang, Y., Wang, R., Zhang, S., Song, S., Jiang, C., Han, G., Wang, M., Ajani, J., Futreal, A., and Wang, L. (2019b). iTALK: an R Package to Characterize and Illustrate Intercellular Communication. bioRxiv, 507871.

Wolf, M.T., Ganguly, S., Wang, T.L., Anderson, C.W., Sadtler, K., Narain, R., Cherry, C., Parrillo, A.J., Park, B.V., Wang, G.*, et al.* (2019). A biologic scaffold–associated type 2 immune microenvironment inhibits tumor formation and synergizes with checkpoint immunotherapy. Science Translational Medicine *11*, eaat7973.

Xia, C., Fan, J., Emanuel, G., Hao, J., and Zhuang, X. (2019). Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. Proceedings of the National Academy of Sciences *116*, 19490.

Xie, A.W., and Murphy, W.L. (2019). Engineered biomaterials to mitigate growth factor cost in cell biomanufacturing. Current Opinion in Biomedical Engineering *10*, 1-10.

Yuk, J.-M., Kim, Tae S., Kim, Soo Y., Lee, H.-M., Han, J., Dufour, Catherine R., Kim, Jin K., Jin, Hyo S., Yang, C.-S., Park, K.-S.*, et al.* (2015). Orphan Nuclear Receptor ERRα Controls Macrophage Metabolic Signaling and A20 Expression to Negatively Regulate TLR-Induced Inflammation. Immunity *43*, 80-91.

Zeng, F., Zou, H.-Q., Zhou, H.-D., Li, J., Wang, L., Cao, H.-Y., Yi, X., Wang, X., Liang, C.-R., Wang, Y.-R.*, et al.* (2013). The relationship between single nucleotide polymorphisms of the NTRK2 gene and sporadic Alzheimer's disease in the Chinese Han population. Neuroscience Letters *550*, 55-59.

Zhu, S., and Qian, Y. (2012). IL-17/IL-17 receptor system in autoimmune disease: mechanisms and therapeutic potential. Clinical Science *122*, 487-511.

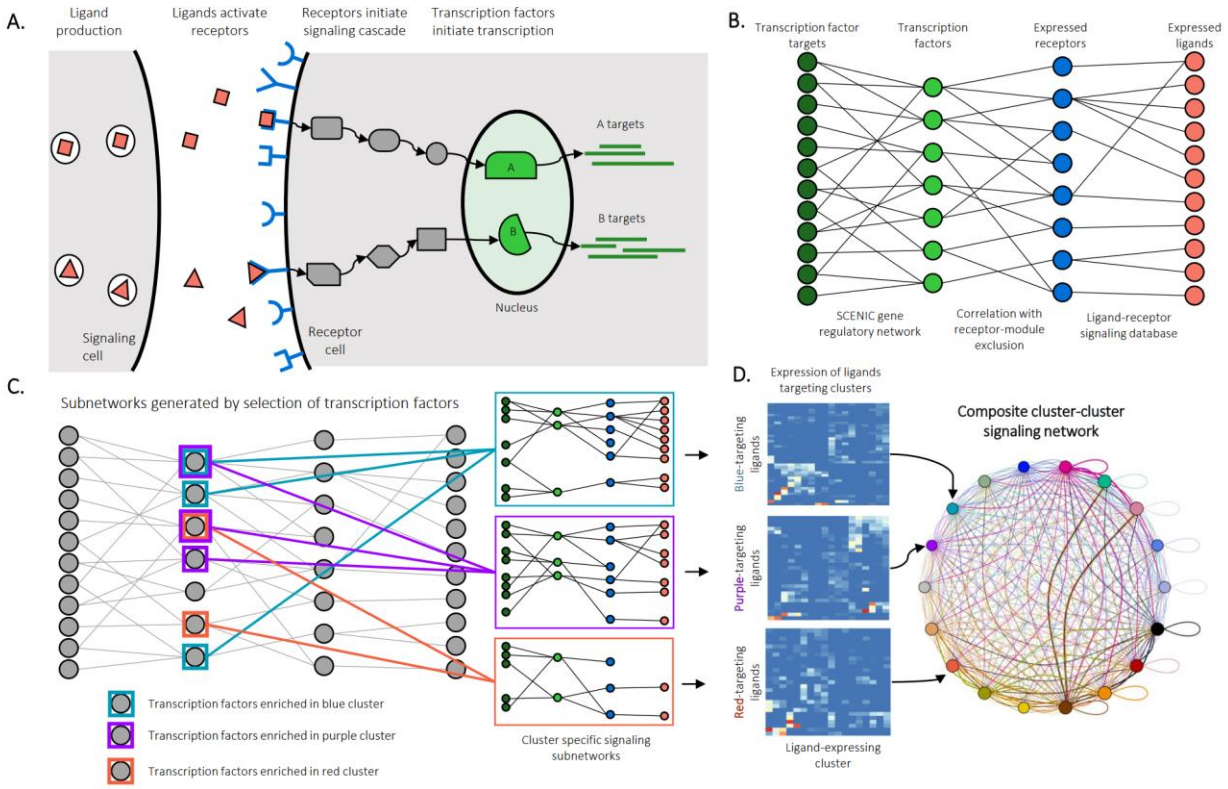


Figure 1: Generation of intercellular and intercluster signaling networks.
A). A model of biological ligand-receptor signaling. Ligands expressed in a signaling cell bind and activate receptors on a target cell. Subsequent protein-protein signaling triggers activation of transcription factors in the nucleus and expression of target genes.

B). Reconstruction of a dataset-wide signaling network. SCENIC is used to estimate transcription factor gene regulatory modules as well as transcription factor activation scores on a cell-by-cell level. Receptors expression levels are correlated with transcription factor activation scores across the entire data set with exclusion of receptors present in the transcription factor modules. Public receptor-ligand databases and then queried to identify ligands capable of activation receptors.

C). Identification of cluster-specific signaling subnetworks. Transcription factors enriched by cluster are identified by Wilcoxon rank sum and networks pruned for disconnected nodes to generate signaling subnetworks relevant for biological activation of clusters.

D). Calculation of intercluster signaling networks. Once phenotypically relevant receptors are identified by cluster specific signaling subnetworks, cluster-cluster signaling scores are calculated by cluster averaged scaled expression of ligands present in cluster-specific subnetworks. Every potential cluster-cluster combination is scored, and these weights used to generate an intercluster signaling network.

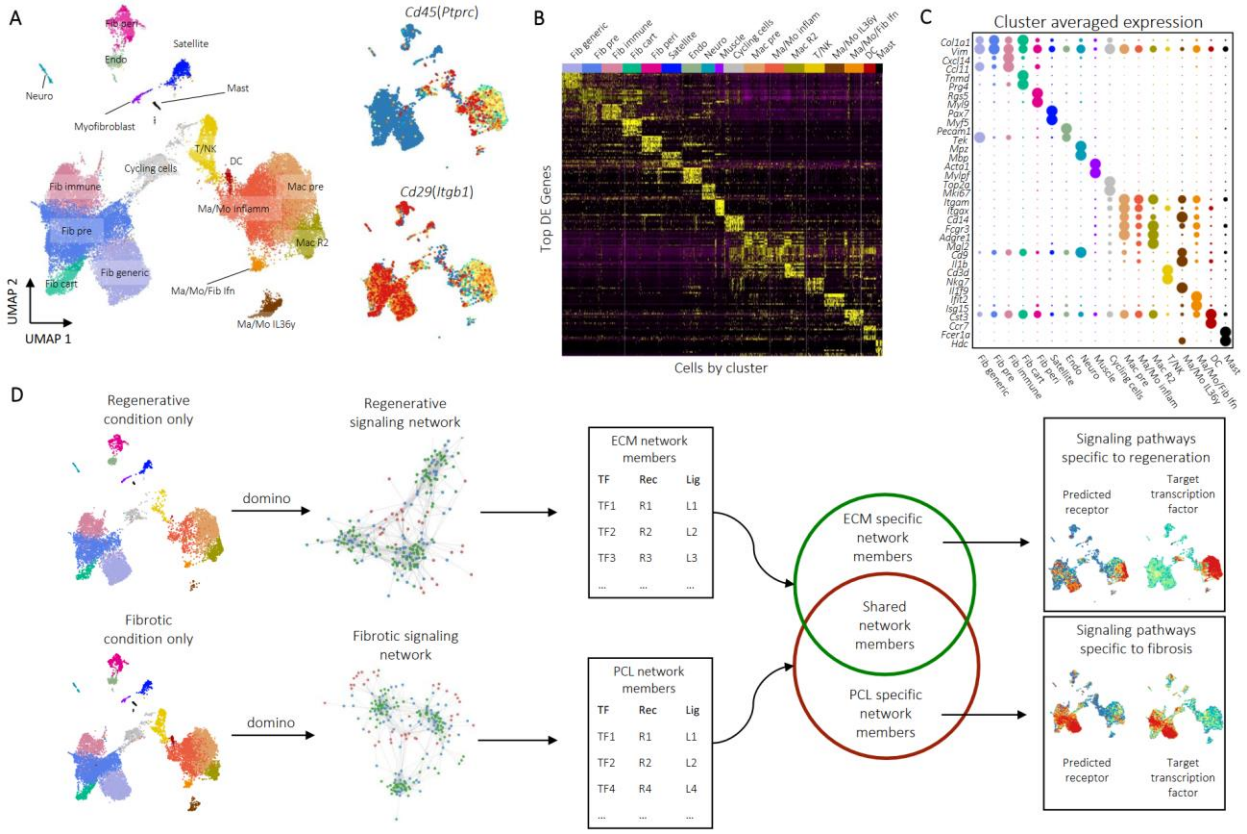


Figure 2: A Single Cell Atlas of the Biomaterials Immune Microenvironment.
A). Overview of cell clusters identified in the composite data set. UMAP plots with cluster labels (left) and expression levels of *Cd45* and *Cd29* (right) are shown.

B). Heatmap of up to 10 differentially expressed genes with highest log fold-change from each cluster. Cells are ordered and labeled by cluster with random sampling of up to 100 cells per cluster.

C). Gene markers for single cell subsets. The dotplot shows expression of genes associated with cluster identity. Cluster averaged gene expression values after normalization to the maximum averaged expression are shown.

D). Schematic for identification of condition-specific signaling pathways. The data set is first split into fibrotic (PCL) and regenerative (ECM) conditions and domino run on each subset. The resulting signaling networks are cross-referenced to identify transcription factors and receptors specific to each condition. Finally, results are examined and constructed into phenotypically relevant, condition specific individual signaling pathways.

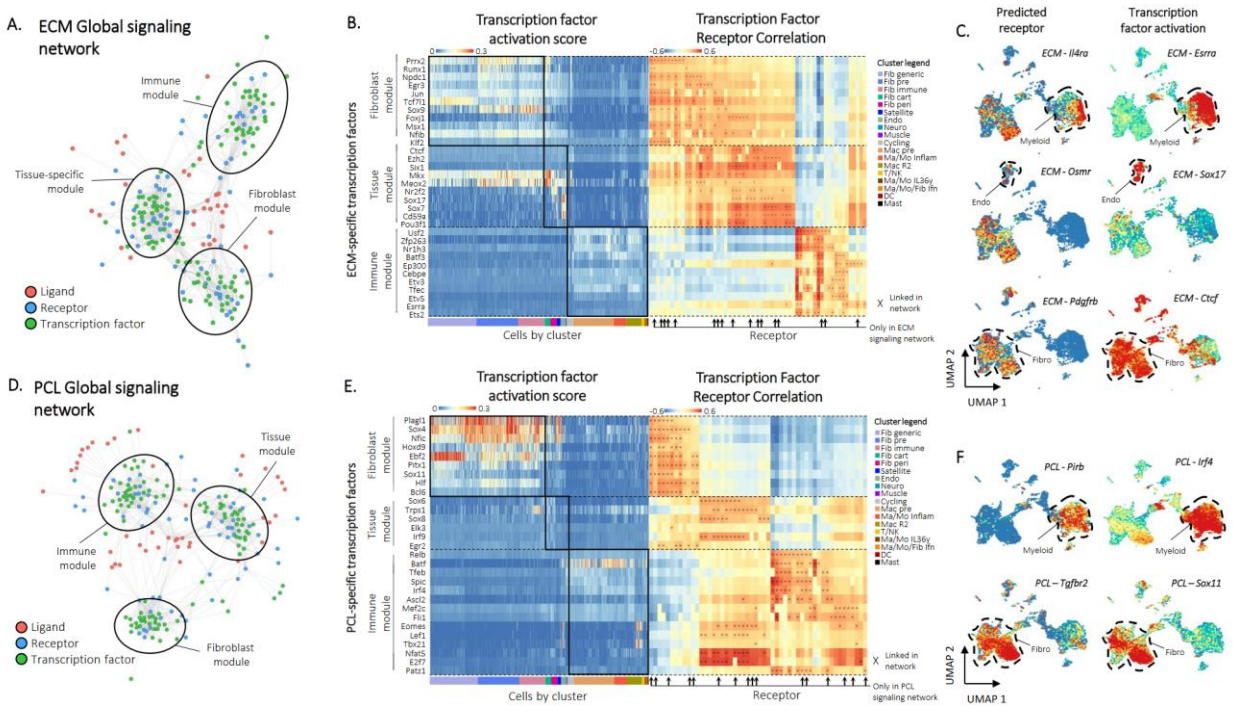


Figure 3: Domino identifies condition specific signaling pathways.
A). The ECM global signaling network. Three modules of receptors and transcription factors are readily apparent and labeled based on enrichment of transcription factors by cluster.

B). Heatmaps of transcription factor activation score for ECM-specific transcription factors (left) and correlation of transcription factor activation score with receptor expression (right). Transcription factors are binned according to their membership to the fibroblast, tissue, or immune modules from the ECM global signaling network. Cells are ordered and colored according to their cluster. Receptors found only in the ECM condition are marked with arrows. Connections between receptor and transcription factors are marked with an ‘x’ on the correlation heatmap.

C). Example feature plots of gene expression and activation scores for specific receptor-transcription factor pairs identified by domino in the ECM condition.

D). The PCL global signaling network. Three modules of receptors and transcription factors are readily apparent and labeled based on enrichment of transcription factors by cluster.

E). Heatmaps of transcription factor activation score for PCL-specific transcription factors (left) and correlation of transcription factor activation score with receptor expression (right). Transcription factors are binned according to their membership to the fibroblast, tissue, or immune modules from the PCL global signaling network. Cells are ordered and colored according to their cluster. Receptors found only in the PCL condition are marked with arrows. Connections between receptor and transcription factors are marked with an ‘x’ on the correlation heatmap.

F.) Example feature plots of gene expression and activation scores for specific receptor-transcription factor pairs identified by domino in the fibrotic (PCL) condition.

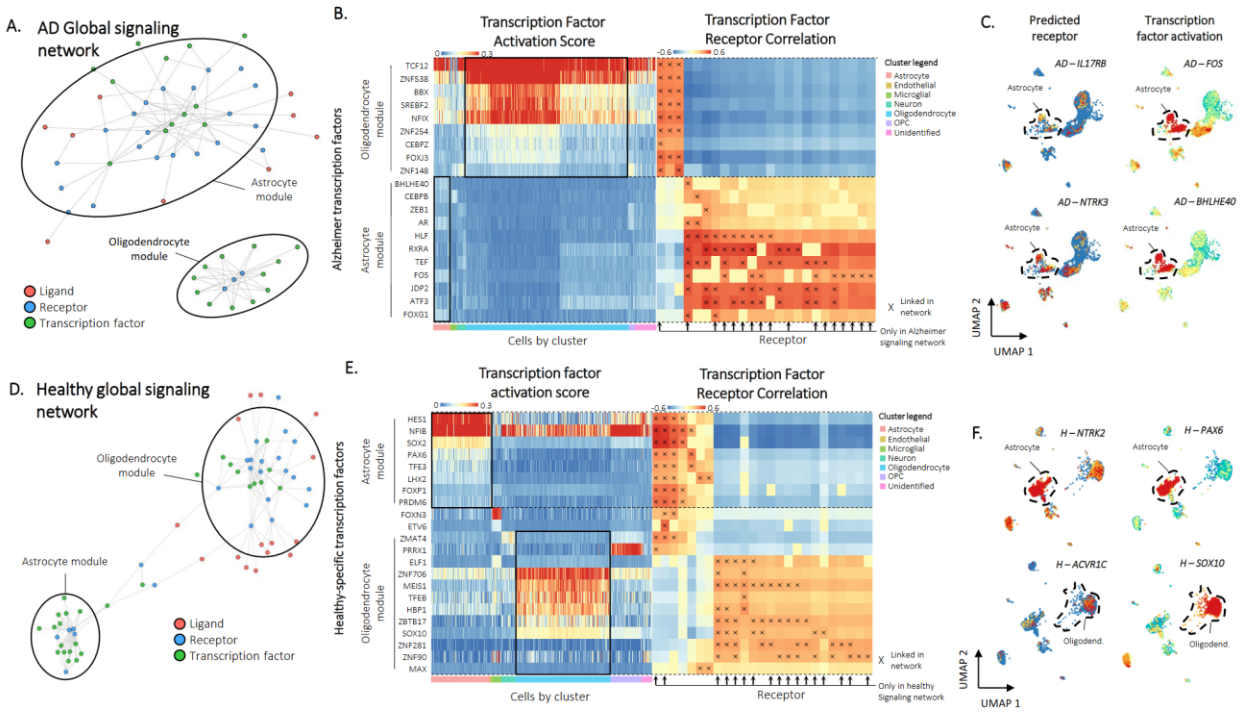


Figure 4: Pathological signaling found from a public data set of Alzheimer's Disease.
A). The Alzheimer's disease (AD) global signaling network. two modules of receptors and transcription factors are readily apparent and labeled based on enrichment of transcription factors by cluster.

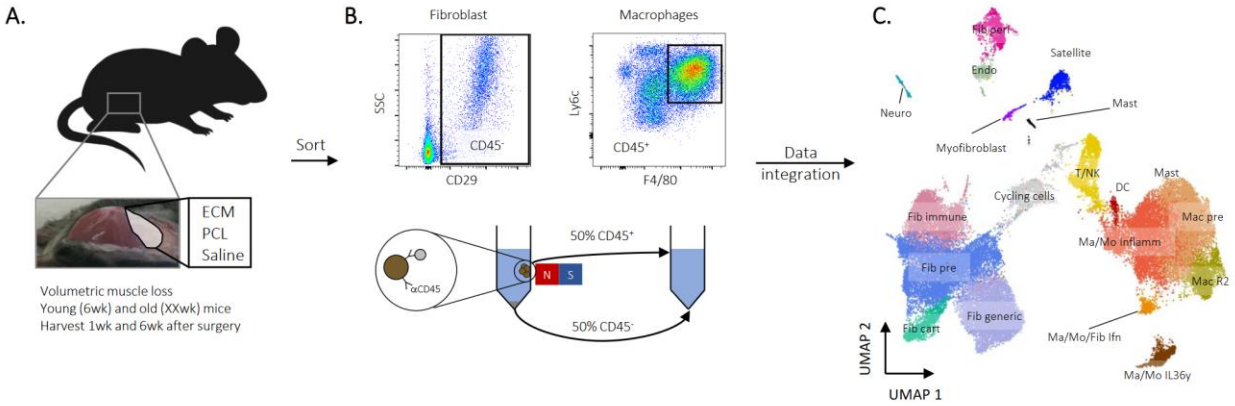
B). Heatmaps of transcription factor activation score for AD-specific transcription factors (left) and correlation of transcription factor activation score with receptor expression (right). Transcription factors are binned according to their membership to the astrocyte or oligodendrocyte modules from the AD global signaling network. Cells are ordered and colored according to their cluster. Receptors found only in AD are marked with arrows. Connections between receptor and transcription factors are marked with an 'x' on the correlation heatmap.

C). Example feature plots of gene expression and activation scores for specific receptor-transcription factor pairs identified by domino in the AD condition.

D). The healthy global signaling network. Two modules of receptors and transcription factors are readily apparent and labeled based on enrichment of transcription factors by cluster.

E). Heatmaps of transcription factor activation score for healthy-specific transcription factors (left) and correlation of transcription factor activation score with receptor expression (right). Transcription factors are binned according to their membership to the astrocyte or oligodendrocyte modules from the global signaling network. Cells are ordered and colored according to their cluster. Receptors found only in the healthy signaling network are marked with arrows. Connections between receptor and transcription factors are marked with an 'x' on the correlation heatmap.

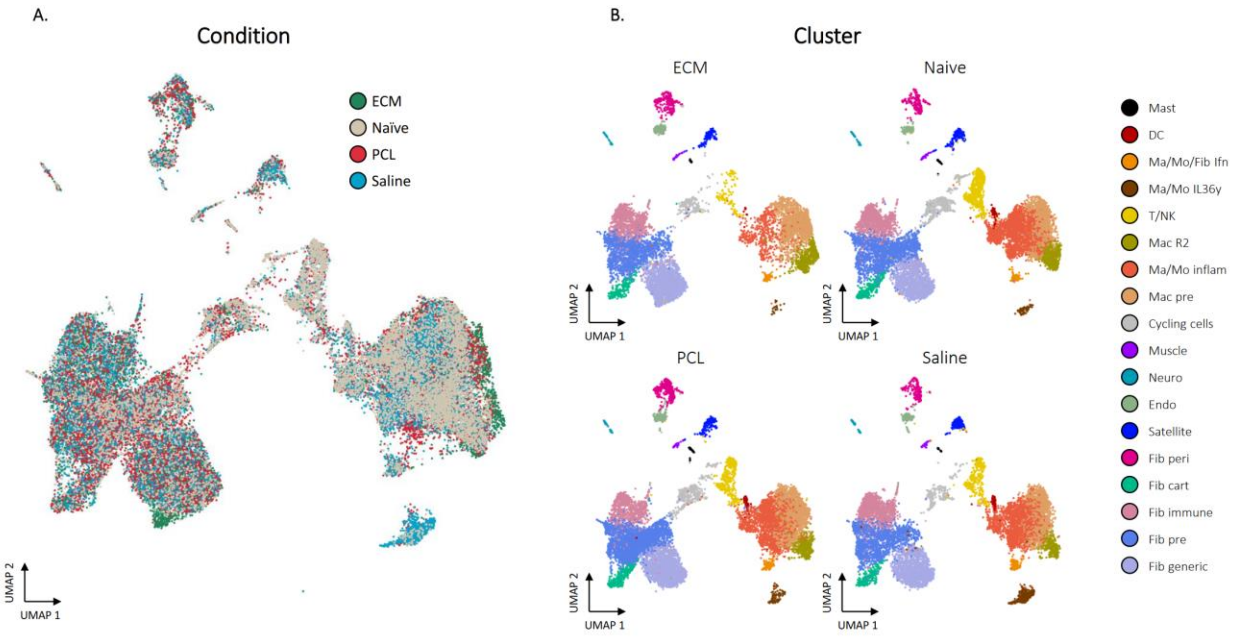
F.) Example feature plots of gene expression and activation scores for specific receptor-transcription factor pairs identified by domino in healthy cells.



Supplementary Figure 1: Experimental Overview of Assembled Data Sets.
A). All data sets were taken from mice after volumetric muscle loss treatment. After surgical excision of a large portion of the quadriceps, the wound site was filled with a biomaterial or saline control and stapled shut. Mice were then harvested 1 or 6 weeks after surgery. Young (6 week) or aged (104 week) old animals were used.

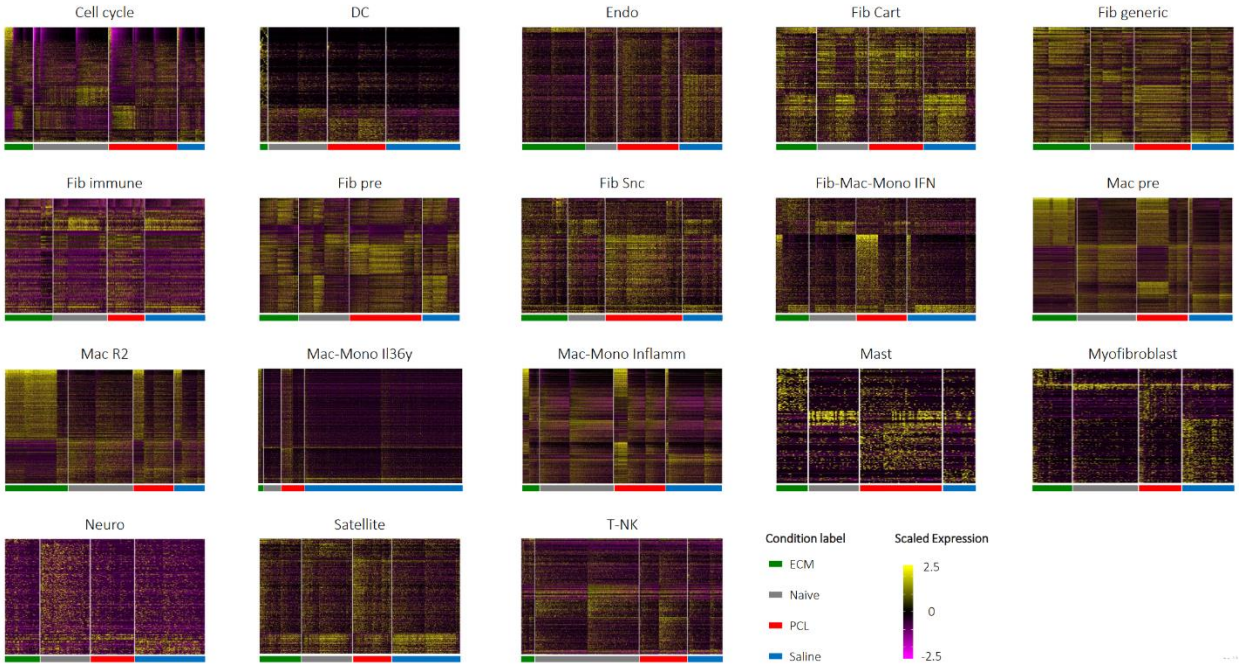
B). At time of harvest, cells were isolated one of three ways after digestions. For macrophages, cells were sorted as $CD45^{+}F4/80^{Hi}Ly6c^{+}$, for fibroblasts cells were sorted as $CD45^{-}CD29^{+}$, and for the all-cell data set $CD45^{+}$ cells were enriched to ~50% using MACS beads.

C). Data sets were integrated for analysis using Harmony. A complete summary of available data sets is given in Supplementary Table 1.



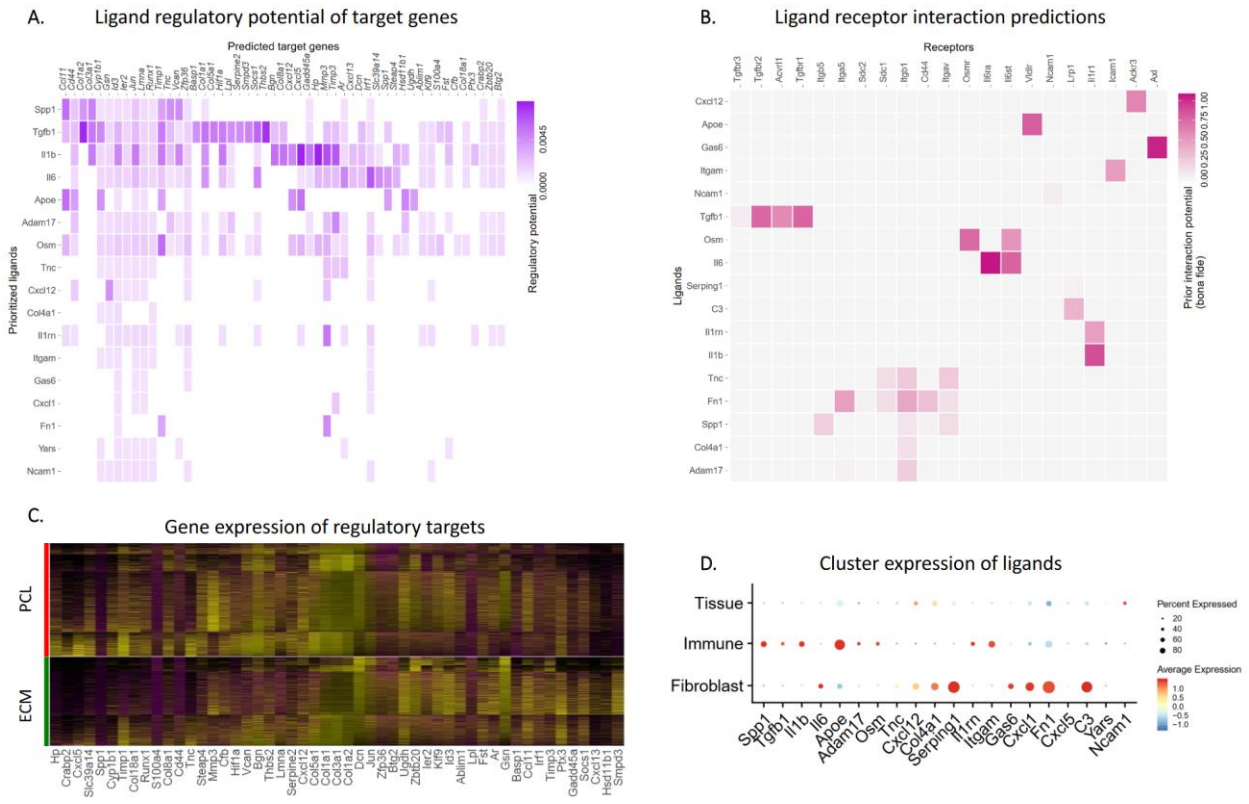
Supplementary Figure 2: Cluster Compositions by Treatment.
A). UMAP of all cells labeled by condition.

B). UMAP of cells labeled by cluster split by condition. Each UMAP displays only cells from the given condition.



Supplementary Figure 3: Visualization of Differentially Expressed Genes Between Condition within Cluster.

Heatmaps of z-scored expression data for cells belonging to each cluster are shown. Cells are ordered and labeled by membership to condition. Within each condition cells are ordered by batch. Z-scored expression values are shown for genes differentially expressed by one of the conditions compared to the other conditions with a minimum log foldchange of 0.25.

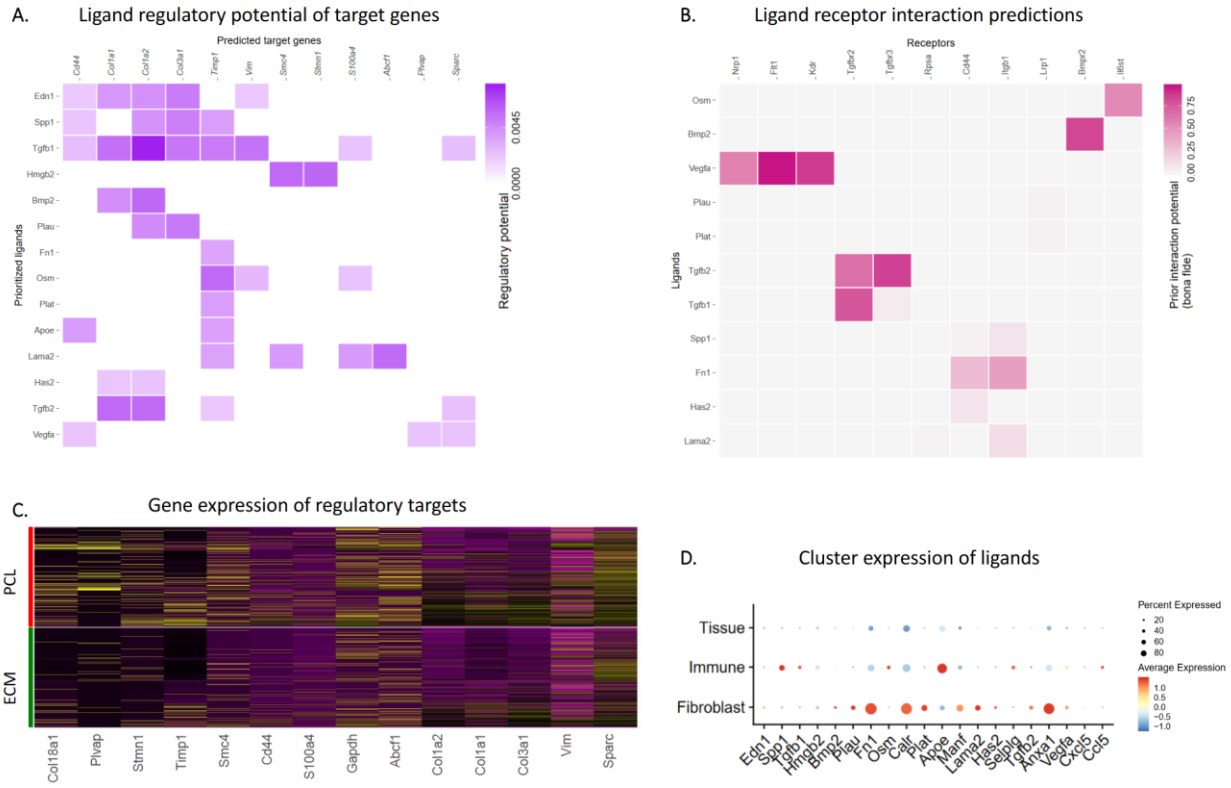


Supplementary Figure 4: NicheNet Analysis of Fibroblasts in Wound Healing.
A). Ligand regulatory potential scores for ligands and their predicted targets by NicheNet. ECM and PCL were compared following NicheNet guidelines using all non-fibroblast cells as potential signaling cells. Because NicheNet uses both genes with both positive and negative fold-change, the results represent predicted ligand signaling targets in both ECM and PCL.

B). Active ligand's receptors by NicheNet's network. The interaction potential represents connections between ligands and receptors according to bona fide literature sources (i.e. no computationally predicted interactions).

C). Heatmap of predicted target genes comparing ECM and PCL in fibroblasts. Z-scored expression values are plotted with genes ordered by fold change.

D). Cluster averaged expression of predicted ligands. Averaged z-scored expression values are shown. Larger dots represent lower dropout rates.

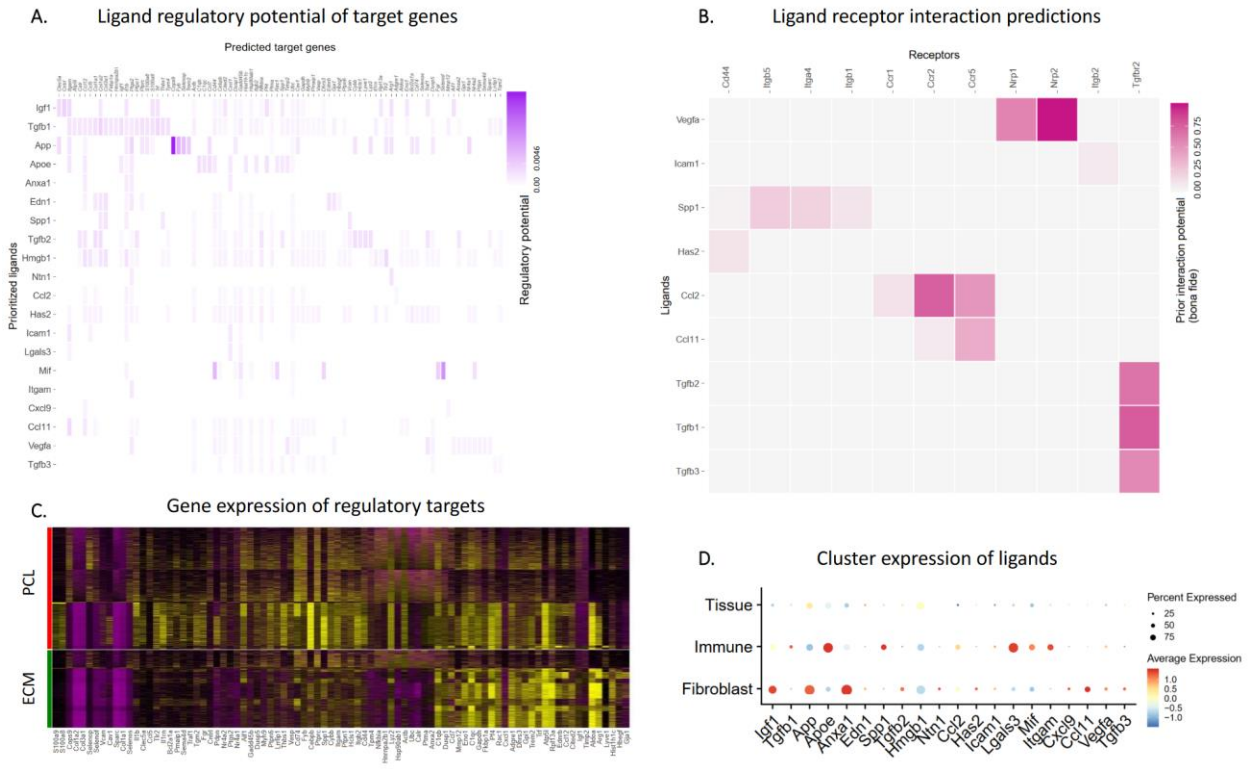


Supplementary Figure 5: NicheNet Analysis of Immune Cells in Wound Healing.
A). Ligand regulatory potential scores for ligands and their predicted targets by NicheNet. ECM and PCL were compared following NicheNet guidelines using all non-tissue cells as potential signaling cells targeting the immune population. Because NicheNet uses both genes with both positive and negative fold-change, the results represent predicted ligand signaling targets in both ECM and PCL.

B). Active ligand's receptors by NicheNet's network. The interaction potential represents connections between ligands and receptors according to bona fide literature sources (i.e. no computationally predicted interactions).

C). Heatmap of predicted target genes comparing ECM and PCL in immune cells. Z-scored expression values are plotted with genes ordered by fold change.

D). Cluster averaged expression of predicted ligands. Averaged z-scored expression values are shown. Larger dots represent lower dropout rates.



Supplementary Figure 6: NicheNet Analysis of Immune Cells in Wound Healing.
A). Ligand regulatory potential scores for ligands and their predicted targets by NicheNet. ECM and PCL were compared following NicheNet guidelines using all non-immune cells as potential signaling cells targeting the immune population. Because NicheNet uses both genes with both positive and negative fold-change, the results represent predicted ligand signaling targets in both ECM and PCL.

B). Active ligand's receptors by NicheNet's network. The interaction potential represents connections between ligands and receptors according to bona fide literature sources (i.e. no computationally predicted interactions).

C). Heatmap of predicted target genes comparing ECM and PCL in immune cells. Z-scored expression values are plotted with genes ordered by fold change.

D). Cluster averaged expression of predicted ligands. Averaged z-scored expression values are shown. Larger dots represent lower dropout rates.