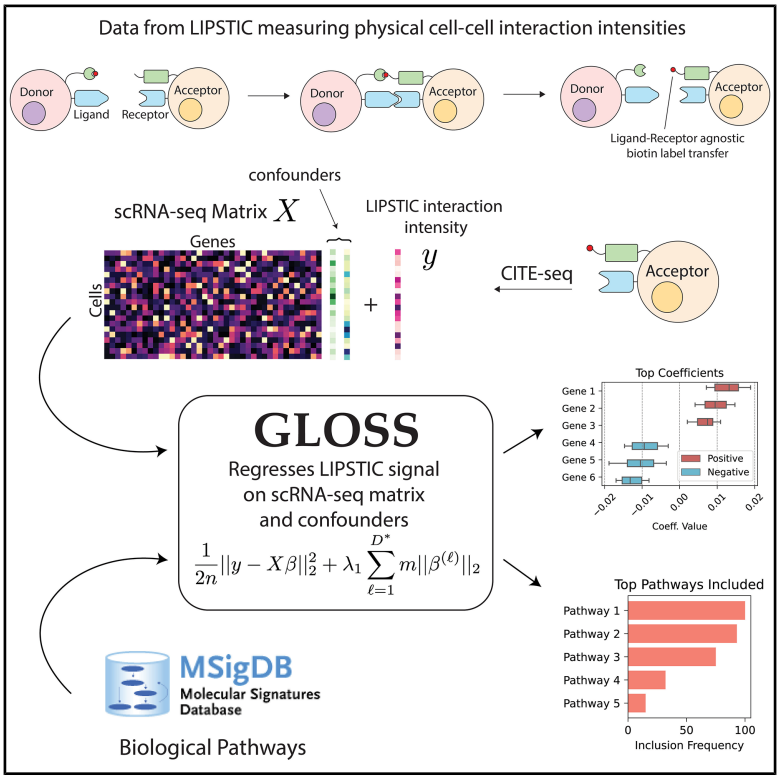
Cell Reports
Methods

Article

Predictive modeling of molecular activity underlying physical cell-cell interactions

Graphical abstract



Highlights

- Gloss is a predictive framework for analyzing physical cell-cell interactions
- The method links gene and pathway activity to LIPSTIC-measured interaction intensity
- Gloss characterizes myeloid-T cell interactions in anti-Ctla4 cancer immunotherapy
- Gloss analysis reveals shared molecular programs of T cell-T cell interactions

Authors

Tamjeed Azad, Sarah K. Walker, Gabriel D. Victora, Yuri Pritykin

Correspondence

pritykin@princeton.edu

In brief

Azad et al. introduce Gloss, a computational framework that links intracellular gene programs measured with scRNA-seq to physical cell-cell interactions measured by LIPSTIC. Using a group lasso approach, Gloss robustly identifies the molecular drivers of immune communication in models of cancer and viral infection.

Azad et al., 2026, Cell Reports Methods 6, 101301
February 23, 2026
https://doi.org/10.1016/j.crmeth.2026.101301

CellPress

Cell Reports Methods

CellPress
OPEN ACCESS

Article

Predictive modeling of molecular activity underlying physical cell-cell interactions

Tamjeed Azad,[1] Sarah K. Walker,[2] Gabriel D. Victora,[3,4] and Yuri Pritykin[1,2,5,*]
[1]Department of Computer Science, Princeton University, Princeton, NJ, USA
[2]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA
[3]Laboratory of Lymphocyte Dynamics, Rockefeller University, New York, NY, USA
[4]Howard Hughes Medical Institute, New York, NY, USA
[5]Lead contact
*Correspondence: pritykin@princeton.edu
https://doi.org/10.1016/j.crmeth.2026.101301

MOTIVATION Current methods for studying cell-cell interactions typically rely on indirect inference from spatial proximity or ligand-receptor co-expression, which cannot guarantee physical contact. While the new labeling immune partnerships by sortagging intercellular contacts (LIPSTIC) technology combined with single-cell RNA sequencing (scRNA-seq) now captures direct, quantitative interaction signals, computational frameworks to systematically model these data are lacking. We developed group lasso on scRNA-seq (Gloss) to bridge this gap, providing a predictive model that links intracellular gene and pathway activity to physical interaction intensity to identify the molecular drivers of cell communication.

SUMMARY

Interactions between cells are central to tissue organization and function in health and disease. Labeling immune partnerships by sortagging intercellular contacts (LIPSTIC) quantitatively measures direct physical cell-cell interactions. Combined with single-cell RNA sequencing (scRNA-seq), it jointly profiles cell interaction intensities and intracellular transcriptomes. Here, we present group lasso on scRNA-seq (Gloss), a predictive modeling framework that systematically links gene and pathway activity to LIPSTIC-measured interaction strength. Across multiple datasets and benchmarks, Gloss outperforms correlation-based and standard regression approaches while remaining interpretable. We apply Gloss to characterize molecular features of myeloid-T cell interactions during anti-Ctla4 immunotherapy in mouse tumors and to describe interactions between different T cell subpopulations during viral infection. Gloss provides a general computational framework for analyzing LIPSTIC+scRNA-seq data and prioritizing genes and pathways driving cellular communication.

INTRODUCTION

Tissue organization and function in health and disease emerge from the interplay between cell-intrinsic functional specialization and intercellular communication. Identifying direct physical cell-cell interactions and linking them to the molecular state of each cell is, therefore, essential for understanding the tissue, organ, and organismal phenotypic complexity. Computational approaches have typically inferred interactions indirectly, from spatial proximity in spatial transcriptomics or from ligand-receptor co-expression in bulk or single-cell RNA sequencing (scRNA-seq).[1–11] However, spatial data often lack single-cell resolution or genome-wide coverage and may require deconvolution or imputation, introducing additional assumptions, and cell spatial proximity does not guarantee functional contacts.[2,3,12,13] Likewise, ligand-receptor gene co-expression provides only indirect evidence for the engagement of the corresponding surface proteins and in modest cohorts, may yield method-dependent or inconsistent results.[2,3]

A new experimental technology, universal labeling of immune partnerships by sortagging intercellular contacts (uLIPSTIC), addresses these limitations by enabling quantitative recording of direct physical cell-cell contacts *in vivo* at unprecedented scale and resolution. In this approach, cells of interest within a specific niche or microenvironment (e.g., intestinal epithelial cells) are engineered to serve as "donors" of a transferrable biotinylated peptide label. During a uLIPSTIC experiment, this label is enzymatically transferred to physically contacting "acceptor" cells (e.g., immune cells in the intestine), accumulating in proportion to interaction intensity between donor and acceptor cells, and can be measured with flow cytometry.[13] uLIPSTIC generalizes the earlier LIPSTIC technology,[12] which was originally designed

to record interactions mediated by the CD40-CD40L receptor-ligand pairing, typically between antigen-presenting cells and CD4 T cells, respectively. In what follows, we refer to this earlier version as LIPSTICv1 and use LIPSTIC as a general term encompassing both LIPSTICv1 and uLIPSTIC.

Importantly, LIPSTIC can be coupled with scRNA-seq for joint high-throughput profiling of both cell-cell interaction intensities and intracellular transcriptomes.[13,14] The LIPSTIC signal is captured through sequencing of barcoded antibody-derived tags (ADTs) and can be profiled simultaneously with cell hashtags distinguishing distinct biological samples, as in cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq).[15,16] This creates a multimodal dataset that enables cataloging cellular interactomes, i.e., describing interacting and non-interacting cell types and cell states, as well as studying the molecular activity underlying cell interactions. Thus, LIPSTIC+scRNA-seq joins a growing family of multimodal single-cell technologies that extend scRNA-seq with additional molecular readouts, such as protein expression (CITE-seq), chromatin accessibility (single-cell assay for transposase-accessible chromatin using sequencing [scATAC]+RNA-seq), and CRISPR perturbations (Perturb-seq), to provide a more integrated view of cellular and molecular biology across experimental systems.[15–22]

The quantitative LIPSTIC signal in LIPSTIC+scRNA-seq, coupled with genome-wide single-cell transcriptomes, enables modeling of intracellular molecular activity underlying physical cell-cell interactions. Here, we present group lasso on scRNA-seq (Gloss), a predictive modeling framework that systematically associates gene and pathway activity with LIPSTIC interaction intensity. Gloss uses an overlapping group lasso regression that incorporates both single-gene features and curated biological pathways and after training, yields interpretable coefficients at both gene and pathway levels. Across multiple LIPSTIC datasets, curated ground-truth annotations, and semi-simulated benchmarks, Gloss outperforms correlation-based rankings and standard regularized linear models as well as non-linear models, while remaining computationally efficient and interpretable. As a demonstration of its utility, we apply Gloss to dissect molecular programs underlying myeloid-T cell interactions in tumors during anti-Ctla4 immunotherapy[14] and to reveal previously underappreciated T cell-T cell interactions during the early response to viral infection.[13]

In conclusion, Gloss provides a broadly applicable computational framework for multi-modal LIPSTIC + scRNA-seq data analysis across biological systems and for prioritizing candidate genes and molecular pathways underlying cellular interactions. Gloss is released as open-source software with reproducible analysis notebooks.

RESULTS

Gloss overview

Gloss is a new computational method for systematic analysis and interpretation of LIPSTIC+scRNA-seq data (Figure 1). For a given LIPSTIC+scRNA-seq dataset, we formulate a regularized regression problem where the quantitative LIPSTIC signal in single cells is predicted from single-cell gene expression (see STAR Methods for details). Gloss implements an overlapping group lasso regression,[23] which applies regularization both to individual genes and to groups of genes belonging to curated biological pathways (e.g., MSigDB or Kyoto Encyclopedia of Genes and Genomes [KEGG]), allowing genes to participate in multiple overlapping pathways. Potential confounders, such as library size (total read count) for the RNA modality and sample hashtag count for the ADT modality, are explicitly included as additional covariates in the regression (Figures S1 and S2; STAR Methods). Model performance and robustness are evaluated using cross-validation and bootstrapping. The resulting coefficients are interpreted as quantitative contributions of genes to cell-cell interaction intensity, enabling both gene-level and pathway-level ranking of molecular activity associated with LIPSTIC signal.

Gloss outperforms other regularized linear regression methods

We assessed the performance of Gloss on four publicly available datasets for mouse *in vivo* cellular interactomes. This included one LIPSTICv1+scRNA-seq dataset profiling different immune cells from the myeloid lineage interacting with T cells in a mouse tumor model[14] and three uLIPSTIC+scRNA-seq datasets (Figure S1): immune cells interacting with the intestinal epithelium in healthy mice, immune cells interacting with CD8 T cells in systemic response across organs to lymphocytic choriomeningitis virus (LCMV) infection, and immune cells interacting with CD8 T cells in the mediastinal lymph node in LCMV infection.[13] For the LIPSTICv1+scRNA-seq dataset, we performed separate analysis in three largest cell subpopulations exhibiting sufficient variation of the LIPSTIC signal, enabling robust modeling: monocytes and macrophages (Mo/MF), conventional dendritic cells (DCs) of type 2 (cDC2), and mature immunoregulatory type 2 DCs (mRegDC2) (Figures 2A–2C). We also performed separate analysis in various cell subpopulations in the three uLIPSTIC+scRNA-seq datasets (Figures 2D–2F). We tested Gloss with two curated sets of pathways, MSigDB Hallmark pathways and KEGG pathways.[24–26] In this analysis, we compared Gloss with other regularized regression models that did not use additional pathway information, such as ridge, lasso, and elastic net regressions. We used R^2 as a metric of performance. Absolute R^2 varied across datasets, likely due to a combination of biological, technical, and statistical factors. We, therefore, focused on within-dataset comparisons across methods. In all cases, we observed that Gloss had better or similar predictive power compared with other regression methods (Figure 2). Thus, curated biological pathway information about genes improves predictive performance of the LIPSTIC signal in the LIPSTIC+scRNA-seq data. Furthermore, Gloss outperforms non-linear methods such as multilayer perceptron regression and random forest regression, while having a much faster runtime (Figure S3, STAR Methods).

Gloss enables association of pathway and gene activity with cell-cell interactions

To show that Gloss modeling can be used to associate gene and pathway activity with cell interaction intensity, we explored the results of Gloss for LIPSTICv1+scRNA-seq data for monocytes and macrophages interacting with $CD4^+$ T cells within a tumor[14]

CellPress
OPEN ACCESS

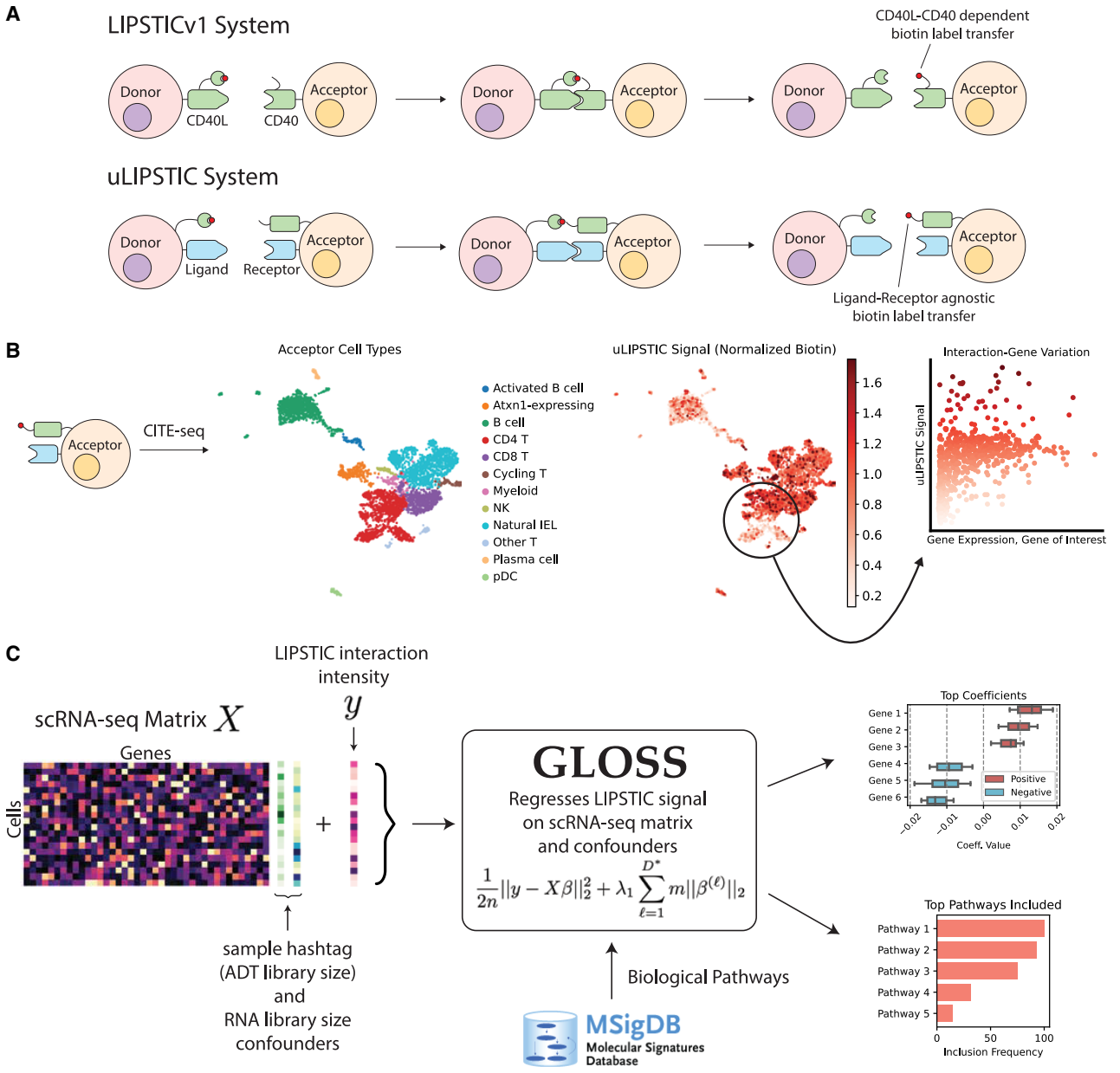


Figure 1. An overview of Gloss, a new computational method for modeling LIPSTIC+scRNA-seq data

(A) Schematic of LIPSTIC (LIPSTICv1) and uLIPSTIC. In LIPSTICv1 (top), label donor cells are engineered to transfer a quantifiable biotin label to their direct physically interacting partner cells via the interaction between CD40L on label donor cells and CD40 on label acceptor cells. See the LIPSTIC publication for more details.[12] uLIPSTIC (bottom) generalizes LIPSTIC beyond a specific CD40-CD40L receptor-ligand pairing and captures physical cell-cell interactions driven by any receptor and ligand. See the uLIPSTIC publication for more details.[13]

(B) An example of uLIPSTIC+scRNA-seq data for immune cells interacting with epithelial cells in the intestine.[13] (Left) The uLIPSTIC biotin signal is captured using a barcoded ADT sequencing along with the transcriptomic readout as in CITE-seq.[15,16] (Middle) uniform manifold approximation and projection (UMAP) visualization of the scRNA-seq data and cell type annotations. (Right) UMAP with the uLIPSTIC signal for each cell representing cumulative intensity of interaction of that cell with the intestinal epithelial cells (label donor cells in this experiment) during the course of the experiment, and a scatterplot illustrating the relationship between uLIPSTIC signal and gene activity. LIPSTIC+scRNA-seq data can be analyzed in two different ways: to describe the interacting and non-interacting cell types and to study the association of molecular activity (measured with scRNA-seq) with the uLIPSTIC interaction intensity within a cell type that exhibits a range of interaction intensities. In this work, we focus on the second type of analysis.

(C) Schematic of group lasso on scRNA-seq (Gloss). Gloss takes as input an scRNA-seq gene expression matrix and a vector of LIPSTIC values of interaction intensities for all cells. It also takes as input the values of confounding factors such as RNA library size (total read count) and ADT library size. It requires a curated list of biological pathways, i.e., (potentially overlapping) groups of genes involved in related functions. Gloss then regresses the LIPSTIC interaction intensity on scRNA-seq (and the confounding factors) using overlapping group lasso where the pathways serve as groups and uses cross-validation to estimate the regression parameters. The inferred coefficients of the model are then used to rank genes and pathways and estimate the robustness using bootstrapping.

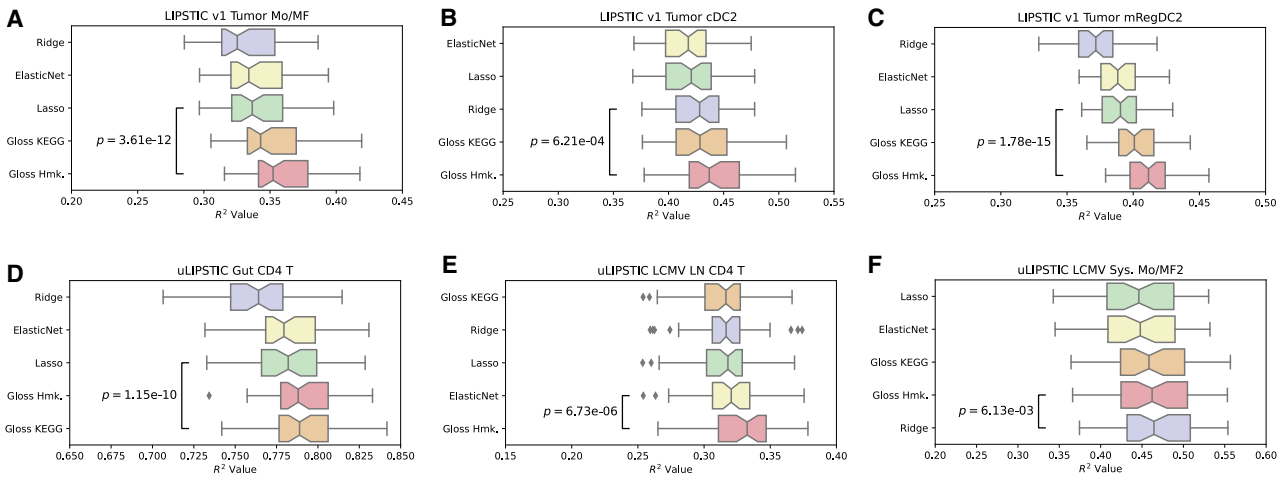


Figure 2. Gloss outperforms other regularized linear regression models

(A–C) The results of modeling the LIPSTIC signal in the LIPSTICv1+scRNA-seq data[14] for different cell subpopulations from myeloid lineage interacting with T cells in a mouse model of cancer. (A) Mo/MF, monocytes and macrophages; (B) cDC2, conventional dendritic cells of type 2; and (C) mRegDC2, mature immunoregulatory dendritic cells of type 2.

(D–F) Same results as in (A)–(C), but for the uLIPSTIC+scRNA-seq data[13] for (D) CD4 T cells interacting with intestinal epithelium, (E) for CD4 T cells interacting with CD8 T cells in the lymph node during lymphocytic choriomeningitis viral (LCMV) infection, and (F) for a subset of Mo/MF2 interacting with CD8 T cells in systemic LCMV response across organs. Hyperparameters for all methods were tuned using nested cross-validation (see STAR Methods). The boxplots show the distribution of test data R^2 values, over 50 splits of the data into train and test subsets. Boxplots: center line, median; notches, 95% confidence intervals around the median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range. Gloss was used with one of the two lists of pathways, MSigDB Hallmark (Hmk) or KEGG pathways.

p value from a paired Wilcoxon signed-rank test comparing the best performing of the two versions of the Gloss method with the best method other than Gloss (as ranked by median R^2).

(Figure 3). We ran Gloss twice with two different sets of pathways, MSigDB Hallmark and KEGG. Via bootstrapping, we identified top Hallmark and KEGG pathways most strongly contributing to the models (Figures 3A and 3B; STAR Methods). LIPSTICv1 specifically measures the strength of the CD40/CD40L-mediated cell interactions triggered during the immune response, which leads to DC activation and increased priming capability.[12,27,28] As expected, we observed top hits related to immune activation, such as "interferon gamma response," "inflammatory response," "TNF-α signaling via NF-κB," and "Il2-Stat5 signaling" in Hallmark and "cytokine-cytokine receptor interaction," "primary immunodeficiency," and "NF-κB signaling pathway" in KEGG (Figures 3A and 3B). Furthermore, top ranking of pathways "cytokine-cytokine receptor interaction," "neuroactive ligand-receptor interaction," and "cell adhesion molecules" suggested that Gloss captured other molecules beyond CD40/CD40L that may participate in the cell-cell interaction processes. Genes can be ranked by their contribution to the predictive power of the Gloss models, with bootstrapping used to estimate significance of the contribution. Furthermore, a gene contribution can be measured either overall, by summing all the coefficients for that gene (across all pathways it participates), or specifically with respect to a certain pathway (see STAR Methods). For example, as expected, we observed Cd40 among the top genes ranked by Gloss (Figures 3C and 3D). We also observed other genes expressing cytokines and cell surface molecules likely involved in cell-cell interactions, as well as genes associated with immune activation and T cell stimulation, such as Cxcl16, Cxcl9, Cd274, B2m, Ccl5, Slamf7, and Vcam1. Interestingly, some genes were ranked highly for a particular pathway but not overall, and vice versa. Indeed, in Gloss, each gene has a single-gene group and may appear in multiple pathway groups. The overall rank sums contributions across all copies (gene + pathway memberships), whereas a pathway-specific rank reflects contribution of a gene conditional on that pathway. For example, Cd274, also known as PD-L1, a target of clinically approved cancer immunotherapies, was among the most strongly associated as part of the interferon gamma response, but not overall in the Gloss model.

Gloss analysis also identified molecular pathways associated with cell-cell interactions in other LIPSTIC+scRNA-seq data and cell populations and demonstrated reproducibility across biological replicates (Figures S4 and S5). Furthermore, we also observed that cell interaction patterns learned by Gloss are generalizable to scRNA-seq data without LIPSTIC in matched experimental conditions and can yield interesting predictions (Figure S6). We additionally demonstrated in a proof-of-concept analysis that Gloss can be used beyond LIPSTIC+scRNA-seq to model protein expression signal in CITE-seq data and outperforms baseline linear regression methods that do not use pathway information (Figure S7).

In sum, Gloss modeling enables an efficient and interpretable way for association of pathway and gene activity with cell-cell interactions that can help formulate hypotheses for future studies (see below the examples of detailed Gloss analyses and interpretations) and in certain contexts could be generalizable to multi-modal data analysis beyond LIPSTIC+scRNA-seq.

CellPress
OPEN ACCESS

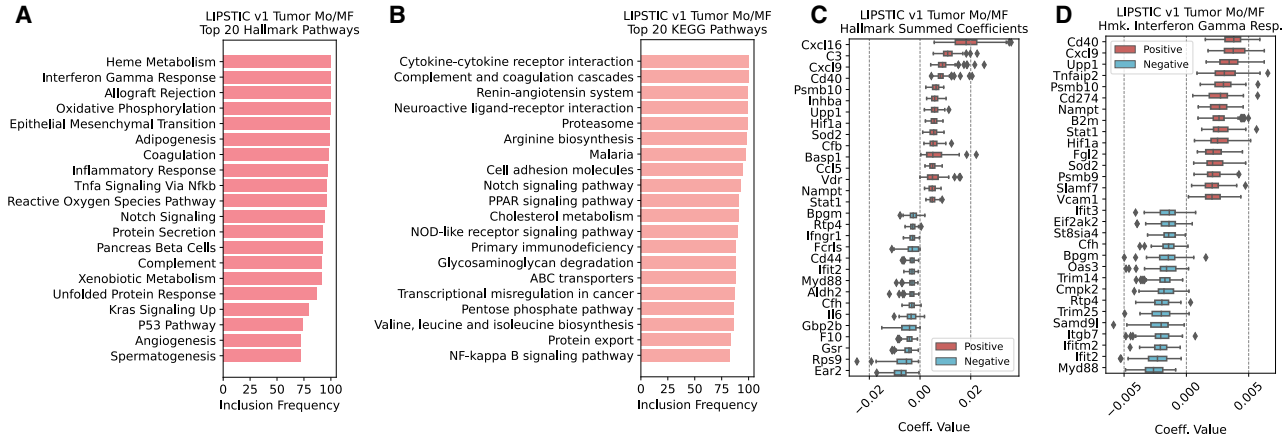


Figure 3. Gloss enables association of pathway and gene activity with cell-cell interactions

Gloss results for LIPSTICv1+scRNA-seq data for Mo/MF interacting with T cells in the tumor are used here as an example (see Figures 2A and S1A).

(A) After bootstrapping 100 times (retraining for subsamples with replacement from train data), the top 20 most frequently included pathways for Gloss using Hallmark pathways are shown. Frequency of inclusion with a non-zero coefficient among the 100 models, *x* axis.

(B) Same as (A), but for the frequency of KEGG pathways in Gloss with KEGG pathways.

(C) Top and bottom 15 gene coefficients from the analysis using Gloss with Hallmark pathways. The value is the sum of all coefficients for a gene, its single-gene group, and across pathways (see STAR Methods). Boxplots show the distribution over bootstrapping. Ranking of genes by median value.

(D) Top and bottom 15 gene coefficients specifically for the interferon gamma response pathway, from the analysis using Gloss with Hallmark pathways.

Boxplots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range.

Gloss results are enriched with curated ground-truth annotations

We wanted to explore in an unbiased and systematic manner how Gloss output compares with known biology and how Gloss compares with other regression and baseline methods. For each LIPSTIC+scRNA-seq experiment, we compiled a list of genes that are likely associated with cell interactions and used these genes as ground truth to assess Gloss performance and compare it with other methods (ridge regression and correlations).

Specifically, for the LIPSTICv1 experiment capturing cell-cell interactions driven by CD40 on LIPSTIC-labeled acceptor cells, we assumed that the genes whose protein products directly interact with CD40 or have a functional association with CD40 in the cell should be associated with the LIPSTIC signal. We thus used a curated mouse NicheNet signaling network[8] to define the high-confidence subnetwork of direct interacting partners of CD40 (see STAR Methods). Importantly, the NicheNet-derived subnetwork used for evaluation was generated using a distinct analytical framework and data sources focused on ligand-target and signaling relationships. While NicheNet integrates several large-scale resources that partially overlap with the databases underlying the MSigDB Hallmark and KEGG pathways used in Gloss (e.g., PathwayCommons and OmniPath), its network-based construction captures complementary information distinct from co-regulated or functionally grouped gene sets. Consistent with this, we found only modest overlaps between the NicheNet-derived gene sets and the pathway collections used in Gloss (data not shown), suggesting that these gene sets are useful for independent evaluation of Gloss results.

For *n* top-ranking genes outputted by Gloss (with either Hallmark or KEGG pathways), we did a Fisher's exact test for overlap with the NicheNet CD40 subnetwork and found significant enrichment of overlap, suggesting that Gloss results recovered the ground truth (Figure 4A). We also compared Gloss with ridge regression, and with a baseline method where genes were ranked by correlation with normalized LIPSTIC signal (using two different types of normalization, division-based and Gloss-based normalization; see STAR Methods), and observed in both cases that Gloss had higher overlap with the CD40 network than other methods (Figure 4A). Then, we repeated the same analysis for a different source of the ground-truth subnetwork, the cell type-specific CD40 GIANT subnetworks obtained from HumanBase,[29] and observed that Gloss again outperformed other methods (Figure 4B). Last, we also tested the enrichment of Gloss with a CD40-upregulated signature identified in human germinal center B cells by Basso et al.,[30] where we also observed a significant enrichment for Gloss (Figure S8A).

Next, we wanted to perform similar analysis for uLIPSTIC+scRNA-seq data. For CD4 T cells in the intestine, the gene Itgae, encoding for an integrin alpha E known to be involved in cell adhesion, was originally described as one of the strongest associates with the uLIPSTIC signal and was validated experimentally.[13] Therefore, we extracted a high-confidence NicheNet subnetwork of direct interacting partners of Itgae and used it in a similar analysis as above for this dataset (Figure 4C). We observed significant enrichment of overlap with Gloss top-ranking genes, and better or similar enrichment for Gloss compared with other approaches. Since uLIPSTIC captures interactions driven by any receptor-ligand pairs, we also used a curated list of genes encoding receptors and ligands involved in direct physical cell-cell contacts from CellChatDB.[31] In a similar analysis as above, Gloss showed significant enrichment and better performance compared with other approaches

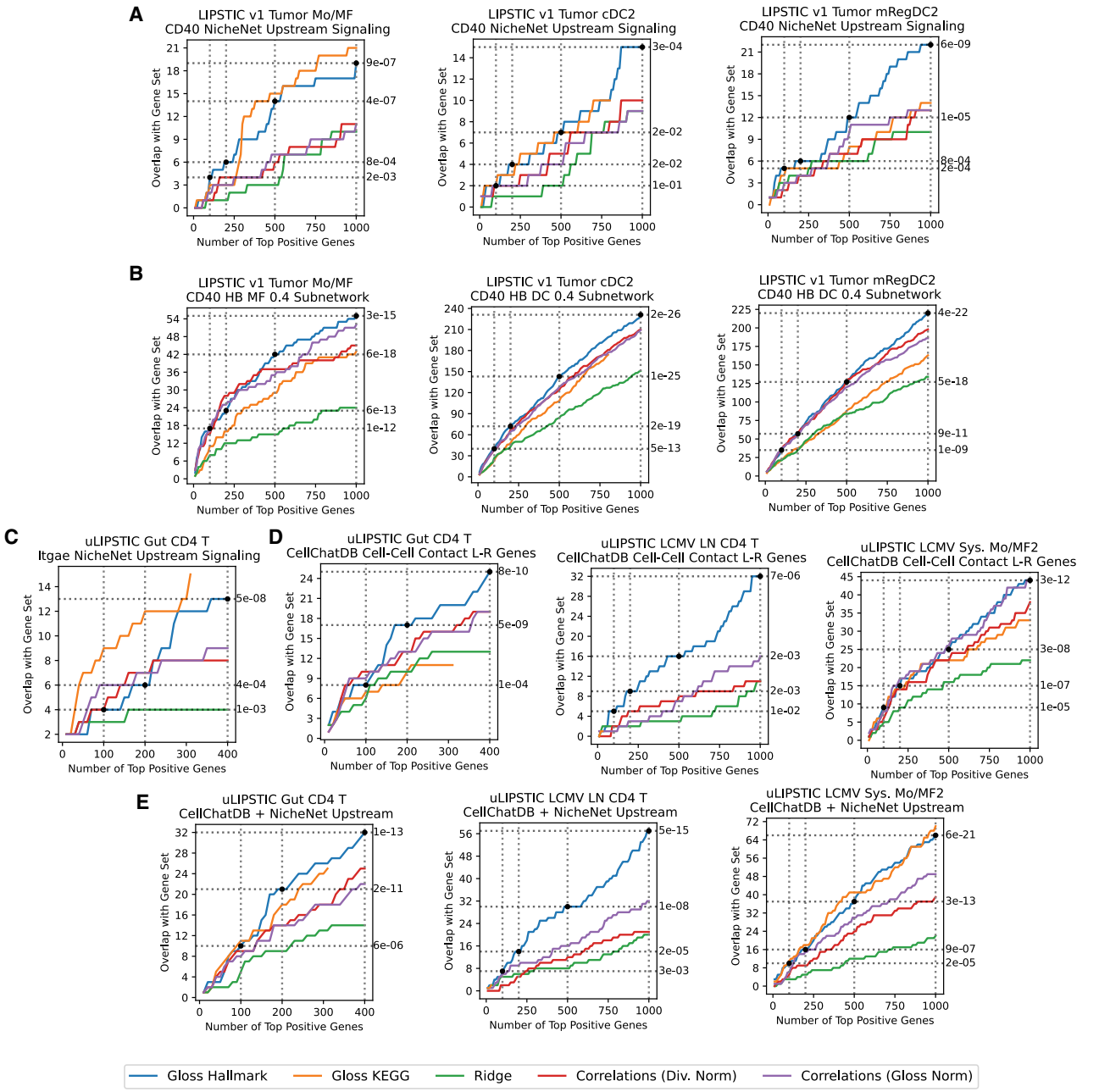


Figure 4. Genes prioritized by Gloss are enriched with curated ground-truth annotations

For each LIPSTIC+scRNA-seq experiment, we compiled a list of genes that are likely associated with cell interactions and used these genes as ground truth to assess Gloss performance and compare it with other methods.

(A) For the LIPSTICv1 experiment, we used the CD40 subnetwork of the mouse NicheNet signaling network[8] as the ground truth to compare with the Gloss results. For n top-ranking genes (x axis) in results of Gloss for Mo/MF (left), for cDC2 (center), and for mRegDC2 (right) interacting with T cells in the tumor, we did a Fisher's exact test for overlap with the CD40 NicheNet subnetwork (y axis) and found significant enrichment (p values shown for $n = 100, 200, 500, 1,000$ for Gloss Hallmark). The results were compared with ridge regression, and with a ranking of genes by correlation with normalized LIPSTIC signal (both division and Gloss normalized, see STAR Methods).

(B) Same as (A), but for the cell type-specific CD40 subnetworks obtained from HumanBase.[29] Macrophage-specific subnetwork for the Mo/MF dataset (left) and DC-specific subnetwork for the cDC2 and mRegDC2 datasets (center and right).

(C) Same as (A), but for the uLIPSTIC data for CD4 T cells interacting with intestinal epithelium. Here the ground truth is a NicheNet subnetwork associated with Itgae, as Itgae was validated experimentally.[13]

(D) Same as (C), but the ground truth is a curated set of genes encoding receptors and ligands involved in cell-cell contacts from CellChatDB.[31] The results are shown for three uLIPSTIC datasets (see Figure 2B).

(legend continued on next page)

for the three uLIPSTIC+scRNA-seq datasets (Figure 4D). Finally, we combined the previous two approaches based on NicheNet and CellChatDB to form ground-truth annotations. We formed a high-confidence NicheNet subnetwork directly upstream of the receptor and ligand genes involved in direct physical cell-cell contacts, as obtained from CellChatDB, and observed similar results for this set of genes (Figure 4E). We also observed significant enrichments of this latter ground-truth dataset in the results for the LIPSTICv1 data (Figures S8B and S8C). The enrichment results were qualitatively robust with respect to parameters used to generate NicheNet- or CellChatDB-derived subnetworks (Figure S9).

Thus, genes prioritized by Gloss were enriched with curated ground-truth annotations, and Gloss outperformed other methods with respect to the level of such enrichment.

Gloss performance on simulated perturbations of LIPSTIC+scRNA-seq data

Because of the scarcity of ground-truth molecular information about cell-cell interactions, we next wanted to extend the above analysis using computational simulations.

We started with the LIPSTICv1+scRNA-seq dataset. Assuming the CD40 subnetwork should be enriched in gene ranking from Gloss for LIPSTICv1+scRNA-seq data (see Figure 4A), we introduced *in silico* perturbations in the data to further probe the strength and significance of this relationship (Figure 5A).

The goal of this simulation was to generate a dataset with even stronger association between the LIPSTIC cell interaction intensity signal and genes in the CD40 subnetwork (see STAR Methods). Specifically, for each single cell in the data, we multiplied expression of genes involved in the CD40 subnetwork by a certain factor (above or below 1, causing increase or decrease of expression) and simultaneously multiplied the LIPSTIC signal by the same factor. The same factor was used for all genes and for the LIPSTIC signals in a cell, but different factors were chosen randomly across cells. To make for more realistic simulations, the perturbations were applied to raw read count data before normalization. For the resulting simulated data, we performed the same analysis as in Figure 4A, for 20 different simulations for each of the three LIPSTICv1+scRNA-seq data subsets. We calculated the enrichment of overlap of Gloss top-ranking genes with the CD40 subnetwork and compared it with the baseline approach, i.e., the ranking of genes by correlation with the normalized LIPSTIC signal (Figure 5A). In all cases, we observed that Gloss significantly outperformed the baseline correlation-based method.

As expected, in these simulations the enrichment for Gloss top-ranking genes was larger than in the original unperturbed data (Figure 5A), because in these simulations, a stronger relationship between the subnetwork and the LIPSTIC signal was "planted" into the data. We used an identical approach for the same HumanBase-derived ground-truth annotations as in Figure 4B, which demonstrated similarly significant results (Figure 5B). As a control, we performed simulations where perturbations were introduced to the same number of random genes with matched overall level of expression but unrelated to CD40. In these control simulations, the enrichment of the CD40 subnetwork was significantly lower as expected (Figure S10A), and the enrichment of the perturbed random gene set was largely not different when comparing Gloss and correlation-based gene rankings (Figure S10C). As an additional control, we also performed the simulations with the same gene perturbations as in Figure 5A, but without perturbing the LIPSTIC signal. In this case, the relationship between the CD40 gene subnetwork and the LIPSTIC signal was weakened, and indeed, we observed lower enrichment in Gloss top-ranking genes (Figure S10B). This analysis confirmed that the CD40 subnetwork that we constructed indeed contained genes associated with the cell-cell interactions measured by LIPSTIC and suggested that our simulation approach accurately captured this relationship and thus was appropriate for the assessment of Gloss performance.

We then used a similar simulation approach for the uLIPSTIC+scRNA-seq data, using the same ground-truth annotations as in Figures 4C–4E and again observed that Gloss outperformed the correlation-based baseline (Figures 5C and 5D). The simulation results were similar and significant when relaxing the stringency parameter in the definition of the NicheNet subnetwork (Figures S11A and S11B) or when planting a non-linear relationship between the gene expression and the LIPSTIC signal into the simulations (Figure S11C), confirming the robustness of the approach.

Thus, our simulation-based approach further confirmed that Gloss results were biologically meaningful and Gloss was better than the baseline correlation-based analysis for associating gene expression with the LIPSTIC cell-cell interaction intensity.

Gloss identifies molecular features of enhanced cell type-specific antigen presentation to T cells in the tumor during anti-Ctla4 immunotherapy treatment

Having demonstrated that Gloss effectively models LIPSTIC+scRNA-seq data and identifies genes associated with cell-cell interactions, we next sought to showcase how Gloss enables deeper reanalysis of previously published datasets. Specifically, we aimed to uncover molecular features of cell interactions that were not identified in the original studies, which lacked dedicated computational methods for analyzing such data.

We first focused on the LIPSTICv1+scRNA-seq dataset profiling myeloid cells in the tumor microenvironment interacting with T cells in the B16 melanoma mouse model.[14] In this dataset, the LIPSTIC signal reflects the cumulative intensity of interactions mediated by CD40-CD40L pairing, specifically between CD40 expressed on myeloid cells and CD40L constitutively expressed on $CD4^+$ T cells, while interactions may also be influenced by

(E) Same as (D), but the ground-truth gene set consists of the NicheNet subnetwork of genes directly interacting with genes encoding receptors and ligands involved in cell-cell contacts from CellChatDB.[31]

In the uLIPSTIC gut CD4 T cell analysis with the Gloss Hallmark pathways (C–E), enrichment p values are shown for $N = 100,200,400$. Gloss results are clipped to show overlap for only positive coefficients (see STAR Methods).

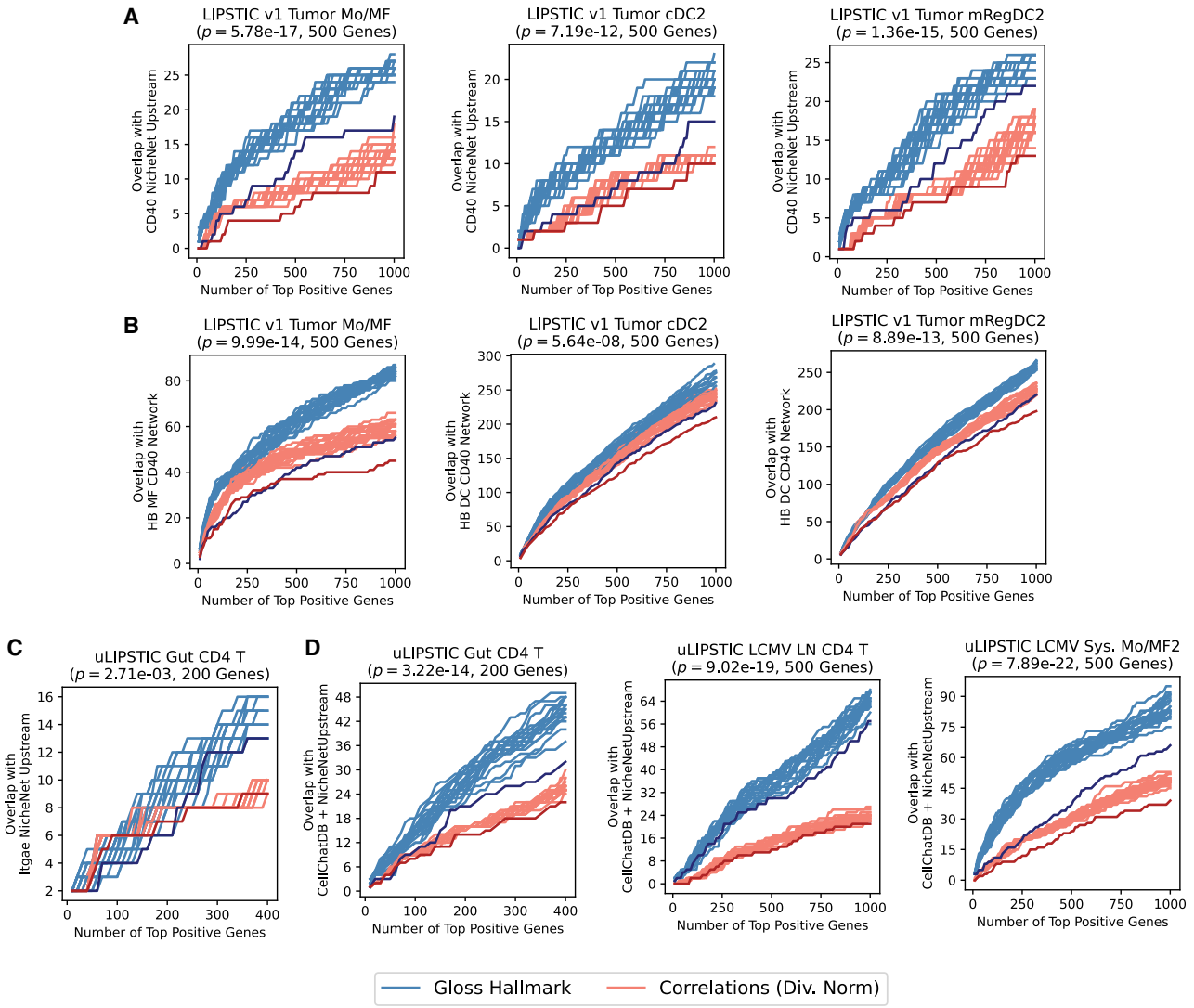


Figure 5. Gloss performance on simulated perturbations of LIPSTIC+scRNA-seq data

(A) Assuming the CD40 subnetwork should be enriched in gene ranking from Gloss for LIPSTICv1+scRNA-seq data (see Figure 4A), we introduced *in silico* perturbations in the data to further probe the strength and significance of this relationship. The goal of this simulation was to generate an array of datasets with even stronger association between the LIPSTIC cell interaction intensity signal and genes in the CD40 subnetwork. In each simulation, for each cell c, a random factor f_c between $2^{-0.5}$ and $2^{0.5}$ was chosen for c, then the raw read counts of all the genes in the CD40 subnetwork were multiplied by f_c in the cell c, and the LIPSTICv1 signal in the cell c was also multiplied by f_c (see STAR Methods for details). Then the resulting simulated data were analyzed in a standard way (as in Figure 4), and for 20 different simulations, the enrichment of overlap of top-ranking genes with the CD40 subnetwork was compared with gene ranking by correlation with the normalized LIPSTIC signal for the three LIPSTICv1+scRNA-seq cell subpopulations. Top, p value from a paired t test comparing the overlaps for the top 500 genes across simulations between Gloss and correlation-based gene ranking. Darker lines indicate enrichment for original, unperturbed data.
(B) Same analysis as in (A), but for the cell type-specific CD40 HumanBase subnetworks, as in Figure 4B.
(C and D) Gloss performance on simulated perturbations of the uLIPSTIC+scRNA-seq data. (C) Same analysis as in (A), but for the NicheNet Itgae subnetwork as in Figure 4C. (D) Same analysis as in (A), but for a NicheNet + CellChatDB subnetwork as in Figure 4E.
In (C) and (D), for the uLIPSTIC gut data, the range is limited to the top 400 genes, and the p value is from a paired t test comparing the overlaps for the top 200 genes, as in the unperturbed analysis in Figure 4 for these data.

other cell surface receptors and ligands. Using functional annotations based on marker genes and literature-curated cell type-specific gene signatures, the original study identified several major myeloid subpopulations, including Mo/MF, two types of conventional DCs (cDC1 and cDC2), and two subsets of regulatory DCs (mRegDC1 and mRegDC2). We have already presented the analysis of these data using Gloss (Figures 2A–2C, 3, 4A, 4B, 5A, 5B, S1A, S2B, S2C, S3A–S3C, S5, S10, and S11C), focusing on the three largest myeloid subpopulations with sufficient variation in the LIPSTIC signal: Mo/MF, cDC2, and mRegDC2. We now turned our attention to an additional aspect of this dataset: the experimental design included two mouse cohorts, one treated

CellPress
OPEN ACCESS

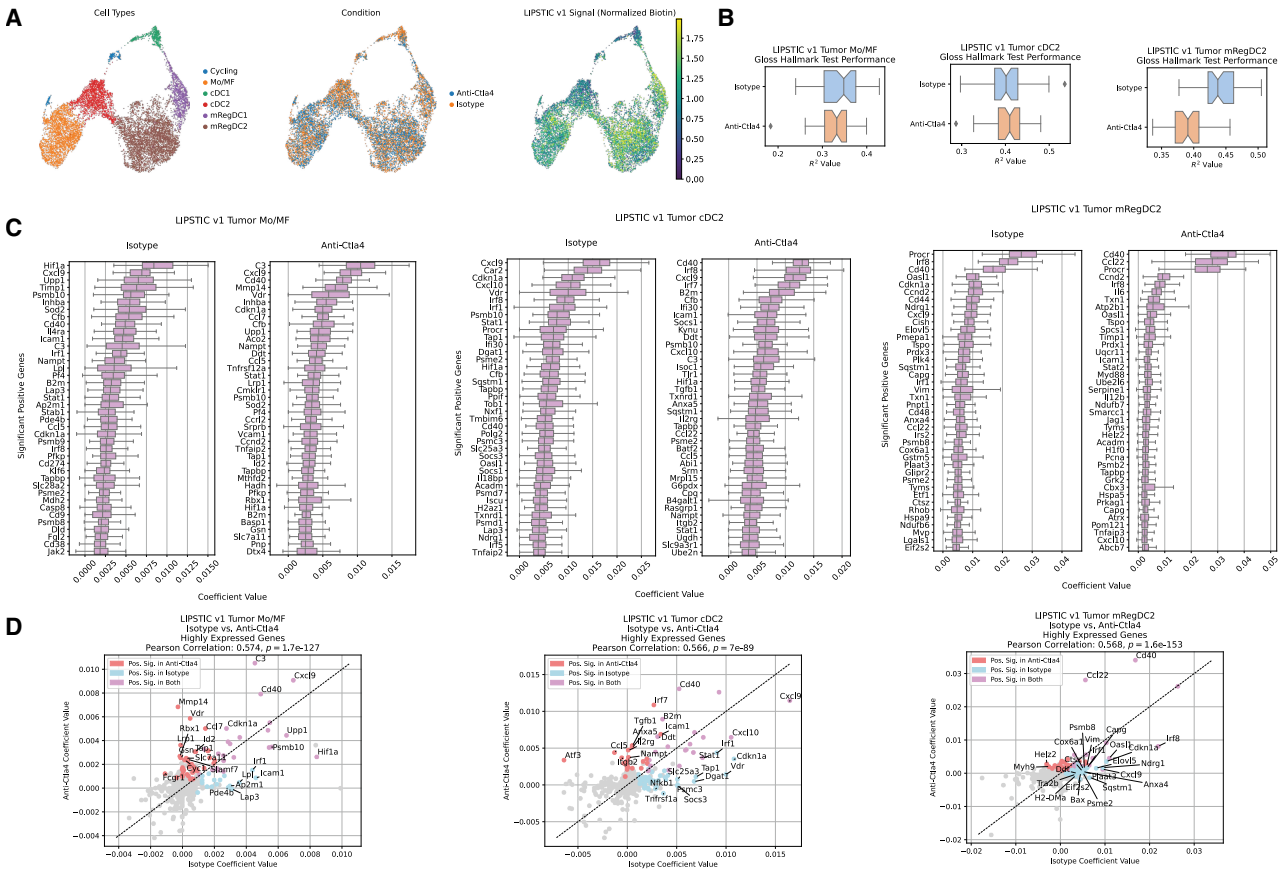


Figure 6. Gloss identifies molecular features of enhanced cell type-specific antigen presentation to T cells in the tumor during anti-Ctla4 immunotherapy treatment

(A) An overview of the LIPSTICv1+scRNA-seq data for the myeloid cells in the tumor and their interactions with T cells.[14] UMAPs: cell type annotations (left), treatment status (anti-Ctla4 immunotherapy or isotype control treatment, center), and normalized LIPSTIC signal representing the cumulative intensity of $CD40^{-}$CD40L-mediated interactions with T cells (right).

(B) Gloss performance on held-out test data (similar to Figures 2A–2C). The results for the three cell populations, Mo/MF, cDC2, and mRegDC2 cells; separately for the anti-Ctla4 treatment; and control.

(C) Top positive significant gene coefficients (see STAR Methods) in the Gloss models for different cell subsets (similar to Figure 3B).

(D) Scatterplot of highly expressed significant gene coefficients from the Gloss models for cells from the anti-Ctla4 and control mice for Mo/MF, cDC2, and mRegDC2 cells. Coefficients were chosen for all genes in each cell type that had at least a total count of 3,000 UMI across the two compared conditions.

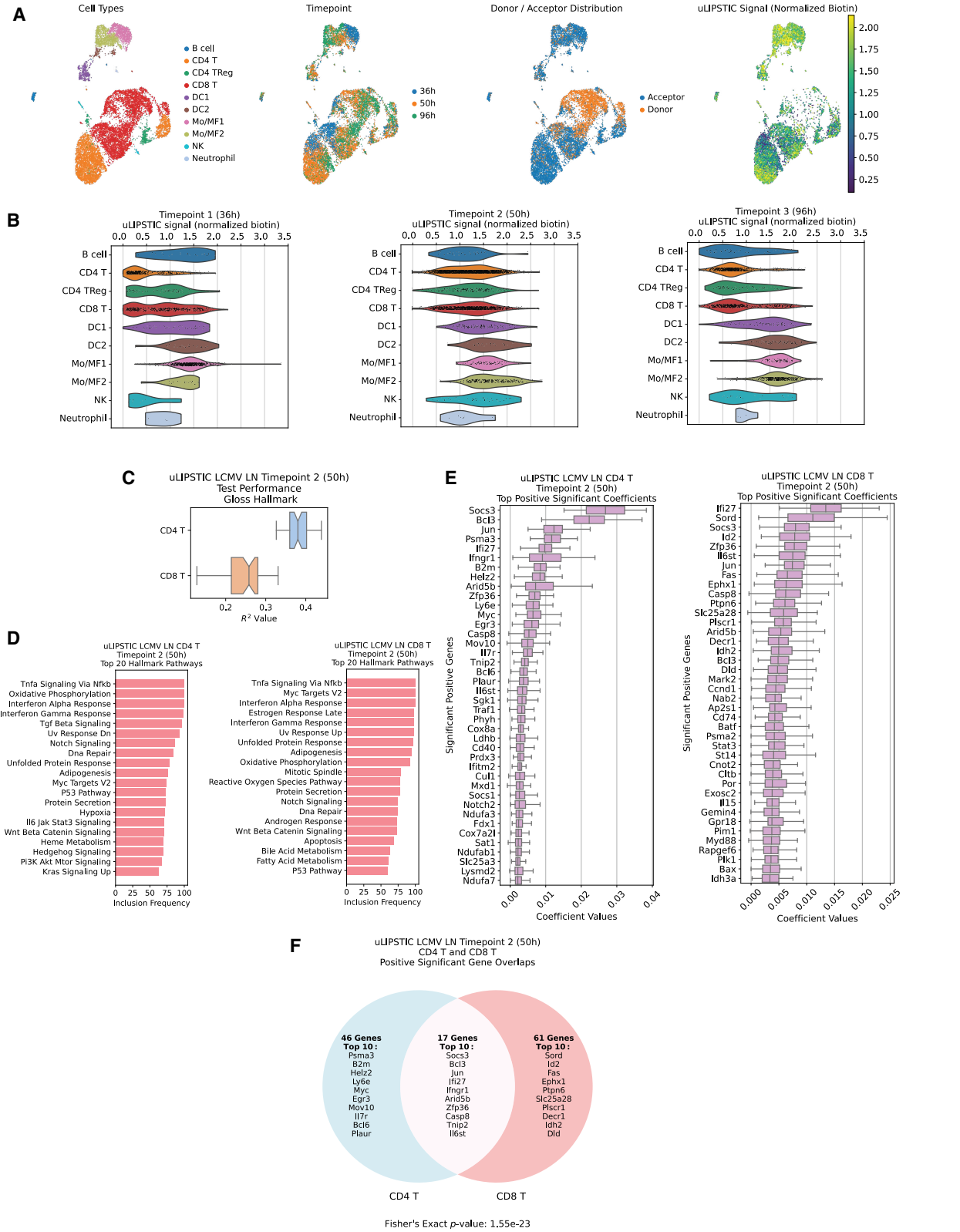
Boxplots: center line, median; notches (panel B), 95% confidence intervals around the median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range.

with anti-Ctla4 checkpoint blockade immunotherapy and the other receiving a control isotype treatment (Figure 6A). We aimed to explore how Gloss can uncover molecular features of cell-cell interactions that are specifically associated with immunotherapy treatment.

We applied Gloss separately to the Mo/MF, cDC2, and mRegDC2 subsets under the anti-Ctla4 treatment and control conditions and compared the results (Figures 6B–6D). Gloss models were generalizable to held-out test data (Figure 7B). We then focused on the key differences between treatment and control conditions, as well as differences across cell types (Figures 6C and 6D). The top significant gene coefficients were largely consistent and showed strong correlations between conditions, indicating that Gloss analysis is reproducible across datasets.

The original publication reported that anti-Ctla4 treatment promotes antigen presentation to T cells.[14] Indeed, the study reported an increased LIPSTIC signal following anti-Ctla4 treatment across multiple myeloid subpopulations including DCs and Mo/MF cells, both in the LIPSTIC+scRNA-seq data and through validation by flow cytometry. However, at the molecular level, the original analysis focused only on the CD40 target gene signature rather than individual genes, and a significant increase in this signature was detected only in the cDC2 population, but not in other cell types. This suggests that the study was underpowered to identify specific molecular associates of cell-cell interactions and antigen presentation.

In our analysis using Gloss, we observed a clear and significant increase in the Cd40 gene coefficient across all three major myeloid subpopulations under anti-Ctla4 treatment compared to



(legend on next page)

control, not only in cDC2 but also in Mo/MF and mRegDC2 cells (Figures 6C and 6D). This provides direct evidence, at the level of activity of individual genes, that anti-Ctla4 treatment enhances CD40-CD40L-mediated interactions, thereby promoting antigen presentation to T cells by diverse myeloid cell populations. Moreover, Gloss analysis revealed molecular features of enhanced antigen presentation that were specific to certain myeloid subpopulations. For example, in Mo/MF cells, but not in other cell types, the matrix metalloproteinase gene Mmp14 showed the largest increase in the anti-Ctla4 condition, consistent with its known role in regulating the extracellular matrix and facilitating cell-cell interactions, particularly between macrophages and T cells.[32,33] Additional cell surface molecules, including Ccl7 and Vdr, were also specifically enriched in the Mo/MF Gloss models. In mRegDC2, a recently defined population of DCs whose functional properties remain incompletely understood,[34,35] the chemokine Ccl22 exhibited a particularly significant increase in its Gloss coefficient under anti-Ctla4 treatment, consistent with its established role in mediating interactions with T cells.[36,37] Multiple other genes encoding cell surface or signaling proteins or transcription factors, such as Icam1, Itgb2, Ccl5, C3, Tgfb1, B2m, Lrp1, Id2, and Irf7, some of them known for their role in mediating cell-cell interactions, were enriched in a cell type-specific manner in the Gloss models for the anti-Ctla4 condition (Figures 6C and 6D). Together, these findings demonstrate that Gloss analysis of LIPSTIC+scRNA-seq data can effectively prioritize candidate genes associated with cell type-specific direct physical interactions between myeloid cells and T cells, providing additional biological insights beyond those reported in the original publication.

In summary, Gloss analysis revealed molecular features of enhanced, cell type-specific antigen presentation by myeloid cells to T cells within the tumor microenvironment during anti-Ctla4 checkpoint blockade treatment. These findings highlight and enable future studies of previously underappreciated cellular and molecular mechanisms of action for checkpoint blockade immunotherapies, extending beyond their direct effects on T cell priming and activation.

Gloss uncovers molecular features of direct physical T cell-T cell interactions in viral infection

We next applied Gloss to the uLIPSTIC+scRNA-seq dataset profiling immune cell interactions during the early response to LCMV infection.[13] In this experiment, antigen-specific CD8 T cells (P14 TCR transgenic) were adoptively transferred into recipient mice and served as uLIPSTIC label donor cells, enabling measurement of their direct physical interactions with other immune cell populations across multiple organs. Following LCMV infection, the data were collected at three distinct time points—36, 50, and 96 h post-infection—capturing the dynamic evolution of immune cell interactions during the early antiviral response (Figure 7A). We have previously presented a broad analysis of this dataset using Gloss (Figures 2E, 2F, 3D, 3E, 5D, S1C, S1D, S2E, S2F, S3E, S3F, S5B, S5C, S11A, and S11B). Here, we focus on a more detailed analysis of the mediastinal lymph node, the key secondary lymphoid organ where T cell priming occurs during early antiviral response. Specifically, we explored how Gloss can uncover additional properties of immune cell interactions not previously described in the original study.

The initial analysis in the uLIPSTIC publication focused primarily on antigen-presenting myeloid cells.[13] While interactions with DCs were detected as early as 36 h post-infection and peaked at 50 h, DCs accounted for only a small fraction of the CD8 T cell interactome. Instead, the majority of interactions occurred with monocytic lineage cells rather than with DCs, and these interactions were shown to be antigen dependent. This revealed a previously underappreciated role for monocytic cells as dominant interaction partners of CD8 T cells during the early antiviral response, suggesting their direct involvement in antigen presentation *in vivo*.

In our reanalysis, we recapitulated these trends for myeloid antigen-presenting cells (Figures 7A and 7B). Strikingly, however, we also observed a pronounced enrichment of the LIPSTIC signal at the 50 h time point within CD4 T cells and endogenous CD8 T cells, prompting further investigation using Gloss. Importantly, for this analysis, we focused specifically on endogenous CD8 T cells by excluding P14 donor cells, which also served as uLIPSTIC label donors.

We trained separate Gloss models for CD4 and CD8 T cells at 50 h post-infection, and both models demonstrated strong generalization to held-out test data (Figure 7C), supporting the robustness and interpretability of the models. We next examined the top enriched pathways and gene coefficients to gain insight into the molecular features associated with these T cell interactions. Gloss analysis revealed that many of the same pathways and genes were associated with the uLIPSTIC interaction signal in both CD4 and CD8 T cells (Figures 7D–7F). The most enriched pathways included "TNF-α signaling via NF-κB," "interferon-α response," and "interferon-γ response," all indicative of immune activation and inflammatory signaling (Figure 7D).

Similarly, several of the top gene coefficients were shared between the two models. These included suppressor of cytokine signaling 3 (Socs3), a negative regulator of cytokine signaling

Figure 7. Gloss uncovers molecular features of direct physical T cell-T cell interactions in viral infection

(A) Overview of the uLIPSTIC+scRNA-seq data for immune cells interacting with antigen-specific CD8 T cells (donors of the uLIPSTIC label, also profiled) in the lymph node during early response to LCMV infection (profiled at 36, 50, and 96 h upon infection).[13] UMAPs: cell type annotations (left), time point after infection (center left), LIPSTIC label donor (antigen-specific adoptively transferred CD8 T cells) or acceptor status (center right), and normalized uLIPSTIC signal (right).

(B) Violin plots showing distributions of the normalized LIPSTIC signal for non-donor cells across cell types and time points. In what follows, Gloss analysis for CD8 T cells excludes adoptively transferred label donor cells and includes only endogenous cells that were LIPSTIC label acceptors.

(C) Gloss performance on held-out test data (similar to Figure 2E) for CD4 T and CD8 T cells at 50 h after infection.

(D) Top Gloss pathways for CD4 T (left) and CD8 T cells (right) at 50 h after infection. Similar to Figure 3A.

(E) Top Gloss positive significant gene coefficients for CD4 T (left) and CD8 T cells (right) at 50 h after infection. Similar to Figure 6C.

(F) Overlap of top genes from CD4 and CD8 T cell Gloss models at 50 h after infection.

Boxplots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range.

that fine-tunes T cell activation through the JAK/STAT pathway; Ifngr1, the receptor for interferon gamma, a key cytokine driving T cell activation and effector functions; Bcl3, a transcriptional co-regulator involved in T cell survival and inflammatory responses; Ifi27, an interferon-stimulated gene implicated in antiviral defense; Jun, a component of the AP-1 transcription factor complex that regulates T cell activation and differentiation; and Il6st (gp130), a common signal transducer for several interleukin family cytokines, important for T cell differentiation and survival.[38–45] These results indicate that both endogenous CD4 and CD8 T cells directly engage in physical interactions with antigen-specific CD8 T cells responding to infection, relying at least in part on shared activation-associated molecular mechanisms. In contrast, Gloss analysis of the Mo/MF1 and Mo/MF2 subpopulations, the cells that were the primary focus of the original study,[13] revealed distinct molecular features underlying their interactions with activated antiviral CD8 T cells. Among the strongest coefficients in the Mo/MF1 and Mo/MF2 models, but not observed in T cell models, were Cxcl9 and Cxcl10, chemokines expressed by myeloid cells recruiting and positioning activated $Cxcr3^{+}$ CD8 T cells via the Cxcl9/10-Cxcr3 axis[36,46–49] (Figure S12). Conversely, top features of the T cell Gloss models were absent from the myeloid models. This suggests the existence of specialized interaction programs that may be broadly shared among T cells but not employed by myeloid cells. In addition to these shared T cell programs of interactions with activated antiviral CD8 T cells, Gloss identified cell type-specific gene signatures. For example, in CD4 T cells the gene Bcl6, a master regulator of CD4 T follicular helper (Tfh) cell differentiation,[50] was prioritized, suggesting that CD4 T cells en route to becoming Tfh cells may be involved in interactions with CD8 T cells in this system.

Overall, our Gloss analysis of the uLIPSTIC+scRNA-seq data uncovers both shared and cell type-specific molecular programs underlying direct interactions between subsets of T cells during the early antiviral response. These findings highlight previously underappreciated modes of T cell cooperation and suggest new avenues for exploring how such interactions shape immune responses.

DISCUSSION

We introduced Gloss, a predictive modeling framework for analyzing LIPSTIC+scRNA-seq data to uncover molecular features of direct cell-cell interactions. By leveraging group lasso regression with curated pathway information, Gloss outperforms correlation-based and standard regression models, offering both improved predictive accuracy and biological interpretability.

Our analyses demonstrate that LIPSTIC+scRNA-seq technologies, including LIPSTICv1 and uLIPSTIC, reliably capture molecular associates of physical cell-cell interactions. Using Gloss, we identified distinct molecular programs of antigen presentation by myeloid cells during anti-Ctla4 cancer immunotherapy, suggesting additional mechanisms through which checkpoint blockade therapies modulate immune responses. Gloss also revealed direct T cell-T cell interactions during viral infection, highlighting shared molecular programs between CD4 and CD8 T cells underlying their direct interactions with activated CD8 T cells, but distinct from those employed by myeloid cells.

More broadly, Gloss illustrates how computational modeling using gene expression and pathway activity can provide a unified framework for studying cellular communication across modalities. Extending pathway-regularized predictive models to spatial transcriptomics or ligand-receptor co-expression analyses could enable systematic quantification and comparison of molecular determinants of intercellular interactions. Likewise, applying Gloss to CITE-seq highlights its broader potential for modeling signaling-related modalities directly from transcriptomes, in line with previous related approaches.[51–55] Similar extensions could model interaction-strength proxies in other data types, such as the α-mixing factor in PIC-seq, which reflects inferred contact intensity between sequenced, paired physically interacting cells (PICs).[56]

Our analysis also highlights important considerations for normalization and modeling confounders. Because LIPSTIC signals and transcriptomes may share global activity shifts, the inclusion of total RNA counts or other global covariates as confounders can suppress biologically meaningful variation if not carefully examined. Future work should systematically evaluate alternative normalization schemes and feature-weighting strategies to distinguish between global activation effects and gene-specific regulatory programs.

Finally, while we demonstrated Gloss using MSigDB Hallmark and KEGG pathways, the framework is flexible and can incorporate any curated or data-driven gene sets, including cell type-specific regulatory networks or inferred ligand-receptor modules. Extensions of Gloss could incorporate network-based priors or non-linear modeling, although such approaches would require larger datasets and additional computational development.

Overall, Gloss provides a generalizable and interpretable framework for identifying gene- and pathway-level determinants of cellular interactions. Its combination of predictive modeling and biological interpretability offers a foundation for future multimodal studies aimed at mechanistic understanding of intercellular communication.

Limitations of the study

While Gloss provides a robust and interpretable framework for modeling molecular activity underlying cell-cell interactions, several limitations remain. First, LIPSTIC+scRNA-seq datasets are still scarce and vary in experimental design and signal dynamic range, which may limit generalization across biological systems. Second, Gloss relies on transcriptomic features and therefore cannot capture post-transcriptional, spatial, or biophysical factors that also influence cellular interactions. Third, pathway-based regularization depends on existing annotations, which may be incomplete or biased toward well-studied processes. Finally, while we demonstrate proof-of-concept applications beyond LIPSTIC (e.g., CITE-seq), broader benchmarking across diverse multimodal datasets will be required to fully assess generalizability and optimal parameterization of Gloss.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Yuri Pritykin (pritykin@princeton.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data: This paper analyzes existing, publicly available data. The LIPSTICv1+scRNA-seq data[14] are available at GEO using accession number GSE275471. The uLIPSTIC+scRNA-seq data[13] are available at GEO using accession number GSE253000 and at https://github.com/pritykinlab/ulipstic-analysis. Additional single-cell genomics data from Ishizuka et al.[57] are available at GEO using accession number GSE110746 and from 10× Genomics on their website.[58]
- Code: The Gloss software package is available at https://github.com/pritykinlab/Gloss. All code used to produce the analysis presented in the paper is available at https://github.com/pritykinlab/gloss-paper-analysis. The package, along with the analysis code, has been archived on Zenodo at https://doi.org/10.5281/zenodo.17956349.
- Additional information: Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank Aleksey Chudnovskiy and Tiago B.R. Castro for providing the processed LIPSTICv1 data and for helpful discussions. This study was partially funded by NIH/NIAID grant DP2AI171161, Ludwig Institute for Cancer Research and Weill Cancer Hub East (Y.P.). G.D.V. is an HHMI investigator and is supported by NIH grant R01AI173086.

AUTHOR CONTRIBUTIONS

Conceptualization, T.A. and Y.P.; methodology, T.A., S.K.W., and Y.P.; investigation, T.A.; software, T.A. and S.K.W.; writing – original draft, T.A. and Y.P.; writing – review & editing, T.A., S.K.W., G.D.V., and Y.P.; funding acquisition, G.D.V. and Y.P.; supervision, G.D.V. and Y.P.

DECLARATION OF INTERESTS

G.D.V. is an adviser for and owns stock in the Vaccine Company, Inc.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2026.101301.

Received: June 19, 2025
Revised: October 28, 2025
Accepted: January 6, 2026
Published: February 10, 2026

REFERENCES

1. Armingol, E., Baghdassarian, H.M., and Lewis, N.E. (2024). The diversification of methods for studying cell–cell interactions and communication. Nat. Rev. Genet. *25*, 381–400.
2. Dimitrov, D., Türei, D., Garrido-Rodriguez, M., Burmedi, P.L., Nagai, J.S., Boys, C., Ramirez Flores, R.O., Kim, H., Szalai, B., Costa, I.G., et al. (2022). Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. Nat. Commun. *13*, 3224.
3. Liu, Z., Sun, D., and Wang, C. (2022). Evaluation of cell-cell interaction methods by integrating single-cell RNA sequencing data with spatial information. Genome Biol. *23*, 218.
4. Almet, A.A., Cang, Z., Jin, S., and Nie, Q. (2021). The landscape of cell–cell communication through single-cell transcriptomics. Curr. Opin. Syst. Biol. *26*, 12–23.
5. 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. Nat. Protoc. *15*, 1484–1506.
6. Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nat. Commun. *12*, 1088.
7. Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. Nat. Methods *17*, 159–162.
8. Sang-aram, C., Browaeys, R., Seurinck, R., and Saeys, Y. (2025). Unraveling cell–cell communication with NicheNet by inferring active ligands from transcriptomics data. Nat. Protoc. *20*, 1439–1467.
9. Baysoy, A., Bai, Z., Satija, R., and Fan, R. (2023). The technological landscape and applications of single-cell multi-omics. Nat. Rev. Mol. Cell Biol. *24*, 695–713.
10. Palla, G., Fischer, D.S., Regev, A., and Theis, F.J. (2022). Spatial components of molecular tissue biology. Nat. Biotechnol. *40*, 308–318.
11. Wilk, A.J., Shalek, A.K., Holmes, S., and Blish, C.A. (2024). Comparative analysis of cell–cell communication at single-cell resolution. Nat. Biotechnol. *42*, 470–483.
12. Pasqual, G., Chudnovskiy, A., Tas, J.M.J., Agudelo, M., Schweitzer, L.D., Cui, A., Hacohen, N., and Victora, G.D. (2018). Monitoring T cell–dendritic cell interactions in vivo by intercellular enzymatic labelling. Nature *553*, 496–500.
13. Nakandakari-Higa, S., Walker, S., Canesso, M.C.C., van der Heide, V., Chudnovskiy, A., Kim, D.Y., Jacobsen, J.T., Parsa, R., Bilanovic, J., Parigi, S.M., et al. (2024). Universal recording of immune cell interactions in vivo. Nature *627*, 399–406.
14. Chudnovskiy, A., Castro, T.B.R., Nakandakari-Higa, S., Cui, A., Lin, C.H., Sade-Feldman, M., Phillips, B.K., Pae, J., Mesin, L., Bortolatto, J., et al. (2024). Proximity-dependent labeling identifies dendritic cells that drive the tumor-specific CD4+ T cell response. Sci. Immunol. *9*, eadq8843.
15. Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017).

Simultaneous epitope and transcriptome measurement in single cells. Nat. Methods *14*, 865–868.

16. Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B.Z., Mauck, W.M., Smibert, P., and Satija, R. (2018). Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biol. *19*, 224.
17. Heumos, L., Schaar, A.C., Lance, C., Litinetskaya, A., Drost, F., Zappia, L., Lücken, M.D., Strobl, D.C., Henao, J., Curion, F., et al. (2023). Best practices for single-cell analysis across modalities. Nat. Rev. Genet. *24*, 550–572.
18. Cao, J., Cusanovich, D.A., Ramani, V., Aghamirzaie, D., Pliner, H.A., Hill, A.J., Daza, R.M., McFaline-Figueroa, J.L., Packer, J.S., Christiansen, L., et al. (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science *361*, 1380–1385.
19. Rang, F.J., de Luca, K.L., de Vries, S.S., Valdes-Quezada, C., Boele, E., Nguyen, P.D., Guerreiro, I., Sato, Y., Kimura, H., Bakkers, J., and Kind, J. (2022). Single-cell profiling of transcriptome and histone modifications with EpiDamID. Mol. Cell *82*, 1956–1970.e14.
20. Ma, S., Zhang, B., LaFave, L.M., Earl, A.S., Chiang, Z., Hu, Y., Ding, J., Brack, A., Kartha, V.K., Tay, T., et al. (2020). Chromatin potential identified by shared single-cell profiling of RNA and chromatin. Cell *183*, 1103–1116.e20.
21. Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. Cell *167*, 1853–1866.e17.
22. Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., et al. (2016). A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. Cell *167*, 1867–1882.e21.
23. Jacob, L., Obozinski, G., and Vert, J.P. (2009). Group Lasso with overlap and graph Lasso. In Proceedings of the 26th Annual International Conference on Machine Learning. ICML '09 (Association for Computing Machinery), pp. 433–440.
24. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.
25. Castanza, A.S., Recla, J.M., Eby, D., Thorvaldsdóttir, H., Bult, C.J., and Mesirov, J.P. (2023). Extending support for mouse data in the Molecular Signatures Database (MSigDB). Nat. Methods *20*, 1619–1620.
26. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark gene set collection. Cell Syst. *1*, 417–425.
27. Elgueta, R., Benson, M.J., De Vries, V.C., Wasiuk, A., Guo, Y., and Noelle, R.J. (2009). Molecular mechanism and function of CD40/CD40L engagement in the immune system. Immunol. Rev. *229*, 152–172.
28. Cella, M., Scheidegger, D., Palmer-Lehmann, K., Lane, P., Lanzavecchia, A., and Alber, G. (1996). Ligation of CD40 on dendritic cells triggers production of high levels of interleukin-12 and enhances T cell stimulatory capacity: T–T help via APC activation. J. Exp. Med. *184*, 747–752.
29. Greene, C.S., Krishnan, A., Wong, A.K., Ricciotti, E., Zelaya, R.A., Himmelstein, D.S., Zhang, R., Hartmann, B.M., Zaslavsky, E., Sealfon, S.C., et al. (2015). Understanding multicellular function and disease with human tissue-specific networks. Nat. Genet. *47*, 569–576.
30. Basso, K., Klein, U., Niu, H., Stolovitzky, G.A., Tu, Y., Califano, A., Cattoretti, G., and Dalla-Favera, R. (2004). Tracking CD40 signaling during germinal center development. Blood *104*, 4088–4096.
31. Jin, S., Plikus, M.V., and Nie, Q. (2025). CellChat for systematic analysis of cell–cell communication from single-cell transcriptomics. Nat. Protoc. *20*, 180–219.
32. Klose, A., Zigrino, P., and Mauch, C. (2013). Monocyte/Macrophage MMP-14 modulates cell infiltration and t-cell attraction in contact dermatitis but not in murine wound healing. Am. J. Pathol. *182*, 755–764.
33. Kessenbrock, K., Plaks, V., and Werb, Z. (2010). Matrix metalloproteinases: regulators of the tumor microenvironment. Cell *141*, 52–67.
34. Maier, B., Leader, A.M., Chen, S.T., Tung, N., Chang, C., LeBerichel, J., Chudnovskiy, A., Maskey, S., Walker, L., Finnigan, J.P., et al. (2020). A conserved dendritic-cell regulatory program limits antitumour immunity. Nature *580*, 257–262.
35. Li, J., Zhou, J., Huang, H., Jiang, J., Zhang, T., and Ni, C. (2023). Mature dendritic cells enriched in immunoregulatory molecules (mregDCs): a novel population in the tumour microenvironment and immunotherapy target. Clin. Transl. Med. *13*, e1199.
36. Kohli, K., Pillarisetty, V.G., and Kim, T.S. (2022). Key chemokines direct migration of immune cells in solid tumors. Cancer Gene Ther. *29*, 10–21.
37. Piseddu, I., Röhrle, N., Knott, M.M.L., Moder, S., Eiber, S., Schnell, K., Vetter, V., Meyer, B., Layritz, P., Kühnemuth, B., et al. (2020). Constitutive expression of CCL22 is mediated by T cell-derived GM-CSF. J. Immunol. *205*, 2056–2065.
38. Palmer, D.C., and Restifo, N.P. (2009). Suppressors of cytokine signaling (SOCS) in T cell differentiation, maturation, and function. Trends Immunol. *30*, 592–602.
39. Sercan, Ö., Stoycheva, D., Hämmerling, G.J., Arnold, B., and Schüler, T. (2010). IFN-γ receptor signaling regulates memory CD8+ T cell differentiation. J. Immunol. *184*, 2855–2862.
40. Siegel, J.P. (1988). Effects of interferon-γ on the activation of human T lymphocytes. Cell. Immunol. *111*, 461–472.
41. Liu, H., Zeng, L., Pan, M., Huang, L., Li, H., Liu, M., Niu, X., Zhang, C., and Wang, H. (2023). Bcl-3 regulates T cell function through energy metabolism. BMC Immunol. *24*, 35.
42. Tang, W., Wang, H., Claudio, E., Tassi, I., Ha, H.I., Saret, S., and Siebenlist, U. (2014). The oncoprotein and transcriptional regulator Bcl-3 governs plasticity and pathogenicity of autoimmune T cells. Immunity *41*, 555–566.
43. Shojaei, M., and McLean, A.S. (2025). Interferon-stimulated gene IFI27 as a multifaceted candidate target in precision medicine. Trends Immunol. *46*, 219–228.
44. Papavassiliou, A.G., and Musti, A.M. (2020). The multifaceted output of c-Jun biological activity: focus at the junction of CD8 T cell activation and exhaustion. Cells *9*, 2470.
45. Nishihara, M., Ogura, H., Ueda, N., Tsuruoka, M., Kitabayashi, C., Tsuji, F., Aono, H., Ishihara, K., Huseby, E., Betz, U.A.K., et al. (2007). IL-6–gp130–STAT3 in T cells directs the development of IL-17+ Th with a minimum effect on that of Treg in the steady state. Int. Immunol. *19*, 695–702.
46. Tokunaga, R., Zhang, W., Naseem, M., Puccini, A., Berger, M.D., Soni, S., McSkane, M., Baba, H., and Lenz, H.J. (2018). CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation–a target for novel cancer therapy. Cancer Treat Rev. *63*, 40–47.
47. House, I.G., Savas, P., Lai, J., Chen, A.X.Y., Oliver, A.J., Teo, Z.L., Todd, K.L., Henderson, M.A., Giuffrida, L., Petley, E.V., et al. (2020). Macrophage-derived CXCL9 and CXCL10 are required for antitumor immune responses following immune checkpoint blockade. Clin. Cancer Res. *26*, 487–504.
48. Colvin, R.A., Campanella, G.S.V., Sun, J., and Luster, A.D. (2004). Intracellular domains of CXCR3 that mediate CXCL9, CXCL10, and CXCL11 function. J. Biol. Chem. *279*, 30219–30227.
49. Marcovecchio, P.M., Thomas, G., and Salek-Ardakani, S. (2021). CXCL9-expressing tumor-associated macrophages: new players in the fight against cancer. J. Immunother. Cancer *9*, e002045.
50. Nurieva, R.I., Chung, Y., Martinez, G.J., Yang, X.O., Tanaka, S., Matskevitch, T.D., Wang, Y.H., and Dong, C. (2009). Bcl6 mediates the development of T follicular helper cells. Science *325*, 1001–1005.

51. Ntranos, V., Yi, L., Melsted, P., and Pachter, L. (2019). A discriminative learning approach to differential expression analysis for single-cell RNA-seq. Nat. Methods *16*, 163–166.

52. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. *20*, 296.

53. Skinnider, M.A., Squair, J.W., Kathe, C., Anderson, M.A., Gautier, M., Matson, K.J.E., Milano, M., Hutson, T.H., Barraud, Q., Phillips, A.A., et al. (2021). Cell type prioritization in single-cell data. Nat. Biotechnol. *39*, 30–34.

54. Huang, M., Wang, J., Torre, E., Dueck, H., Shaffer, S., Bonasio, R., Murray, J.I., Raj, A., Li, M., and Zhang, N.R. (2018). SAVER: gene expression recovery for single-cell RNA sequencing. Nat. Methods *15*, 539–542.

55. Zhou, Z., Ye, C., Wang, J., and Zhang, N.R. (2020). Surface protein imputation from single cell transcriptomes by deep neural networks. Nat. Commun. *11*, 651.

56. Giladi, A., Cohen, M., Medaglia, C., Baran, Y., Li, B., Zada, M., Bost, P., Blecher-Gonen, R., Salame, T.M., Mayer, J.U., et al. (2020). Dissecting cellular crosstalk by sequencing physically interacting cells. Nat. Biotechnol. *38*, 629–637.

57. Ishizuka, J.J., Manguso, R.T., Cheruiyot, C.K., Bi, K., Panda, A., Iracheta-Vellve, A., Miller, B.C., Du, P.P., Yates, K.B., Dubrot, J., et al. (2019). Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade. Nature *565*, 43–48.

58. 10× Genomics (2018). 10k PBMCs from a healthy donor – gene expression with a panel of TotalSeq™-B antibodies. https://www.10xgenomics.com/datasets/10-k-pbm-cs-from-a-healthy-donor-gene-expression-and-cell-surface-protein-3-standard-3-0-0.

59. Bertrand, Q., Klopfenstein, Q., Bannier, P.A., Gidel, G., and Massias, M. (2022). Beyond L1: faster and better sparse models with skglm. Proceedings of the 36th International Conference on Neural Information Processing Systems, 38950–38965. 2823.

60. Moufad, B., Bannier, P.A., Bertrand, Q., Klopfenstein, Q., and Massias, M. (2025). skglm: improving scikit-learn for regularized generalized linear models. J. Mach. Learn. Res. *26*, 1–6.

61. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15.

62. Fang, Z., Liu, X., and Peltz, G. (2023). GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. Bioinformatics *39*, btac757.

63. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

64. Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., et al. (2013). API design for machine learning software: experiences from the scikit-learn project. In ECML PKDD Workshop: Languages for Data Mining and Machine Learning, pp. 108–122.

65. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272.

66. 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. Nat. Methods *16*, 1289–1296.

67. Massias, M., Gramfort, A., and Salmon, J. (2018). Celer: a fast solver for the Lasso with dual extrapolation. Proceedings of the 35th International Conference on Machine Learning *80*, 3321–3330.

68. Massias, M., Vaiter, S., Gramfort, A., and Salmon, J. (2020). Dual extrapolation for sparse GLMs. J. Mach. Learn. Res. *21*, 1–33.

69. Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., and Peterson, H. (2023). g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). Nucleic Acids Res. *51*, W207–W212.

70. Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D., Connor, R., Funk, K., Kelly, C., Kim, S., et al. (2021). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. *50*, D20–D26.

71. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233.

72. Virshup, I., Bredikhin, D., Heumos, L., Palla, G., Sturm, G., Gayoso, A., Kats, I., and Koutrouli, M. (2023). The scverse project provides a computational ecosystem for single-cell omics data analysis. Nat. Biotechnol. *41*, 604–606.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
uLIPSTIC datasets	Nakandakari-Higa et al.[13]	GEO: GSE253000, https://github.com/pritykinlab/ulipstic-analysis
LIPSTIC dataset	Chudnovskiy et al.[14]	GEO: GSE275471
10× CITE-seq dataset	10× Genomics	https://www.10xgenomics.com/datasets/10-k-pbm-cs-from-a-healthy-donor-gene-expression-and-cell-surface-protein-3-standard-3-0-0
Ishizuka2019 dataset	Ishizuka et al.[57]	GEO: GSE110746
MSigDB Hallmark gene sets	Liberzon et al.[26] Castanza et al.[25]	https://www.gsea-msigdb.org/gsea/msigdb
NicheNet v2	Browaeys et al.[7] Sang-aram et al.[8]	https://doi.org/10.5281/zenodo.7074291
HumanBase GIANT networks	Greene et al.[29]	https://humanbase.io/download
GC B Cells CD40 Signature	Basso et al.[30]	https://ashpublications.org/blood/article/104/13/4088/18960/Tracking-CD40-signaling-during-germinal-center
Software and algorithms		
Gloss	This paper	https://github.com/pritykinlab/Gloss
skglm	Bertrand et al.[59] Moufad et al.[60]	https://contrib.scikit-learn.org/skglm/
scanpy	Wolf et al.[61]	https://github.com/scverse/scanpy
GSEApy	Fang et al.[62]	https://github.com/zqfang/GSEApy
scikit-learn	Pedregosa et al.[63] Buitinck et al.[64]	https://scikit-learn.org/stable/
SciPy	Virtanen et al.[65]	https://scipy.org
Harmony and Harmonypy	Korsunsky et al.[66]	https://github.com/immunogenomics/harmony https://github.com/slowkow/harmonypy
CellChat	Jin et al.[31]	https://github.com/jinworks/CellChat

METHOD DETAILS

The regression model

Given the LIPSTIC interaction intensity and scRNA-seq transcriptomic features of a sample of cells of a specific cell type, we would like to determine the most highly associated and predictive genes for the LIPSTIC signal. Considering a set of pathways, we also would like to know which of these pathways are implicated in this relationship. Gloss posits this as a regression problem.

Data inputs

Gloss takes as input the following data.

1. An scRNA-seq matrix $X^S \in \mathbb{N}^{N\times G}$ as raw, unnormalized UMI counts, where N is the number of cells and G is the number of genes.
2. Each cell's LIPSTIC interaction intensity $u \in \mathbb{N}^{N\times 1}$ as ADT-seq read count.
3. A set of gene groups D with d groups, which are assumed to represent biological pathways, intersected with the genes that are included in X^S. These gene groups can be overlapping or non-overlapping, and they do not have to fully encompass all the genes in X^S.
4. (optional) Each cell's scRNA-seq library size $L \in \mathbb{N}^{N\times 1}$.

5. (optional) Each cell's mouse-of-origin sample hashtag library size $H \in \mathbb{N}^{N\times 1}$. This is the number of raw sample hashtag reads for the mouse each cell is assigned to.

Gloss assumes that the input scRNA-seq matrix has been annotated to specific cell types, and that cells and genes with insufficient signal have been filtered out from the data (as is standard in scRNA-seq data analysis).

Transformation of inputs

Gloss first normalizes the raw inputs X^S, L, H, y across all the cells (see Data preprocessing and normalization), which transforms $X^S \to X^{S'} \in \mathbb{R}^{N\times G}, L \to L' \in \mathbb{R}^{N\times 1}, H \to H' \in \mathbb{R}^{N\times 1}$, and $u \to u' \in \mathbb{R}^{N\times 1}$. We then augment the pathway set D by adding every gene as a distinct single-gene group, resulting in an augmented set D^* with $d^* = d + G$ groups.

Then, for each group of genes including the individual gene groups $\ell_1, \ell_2, \cdots, \ell_{d^*} \in D^*$, and using $|$ to denote concatenation, we construct

$$X^{S''} := X^{S'\ell_1} | X^{S'\ell_2} | \cdots | X^{S'\ell_{d^*}}, \quad \text{(Equation 1)}$$

where $X^{S'\ell_i}$ indicates the subset of $X^{S'}$ whose genes are included in the gene set ℓ_i. This effectively constructs $X^{S''}$ as an augmented version of $X^{S'}$, where genes that are in multiple groups have additional duplicated copies, and the presence of all its copies are indexed by the groups in D^*. That is, every gene has its own single-gene group, and has an additional copy for every pathway in D that it belongs to. We then concatenate our confounder variables L and H to $X^{S'}$ to get our final regression matrix X which we will regress y on:

$$X := X^{S''} | L' | H'. \quad \text{(Equation 2)}$$

We set $y = u'$, and we then subset X, y to observations of a specific cell type when running Gloss.

Confounding factors

We include the sample hashtag as a confounding factor because it is measured along with the LIPSTIC signal as part of the ADT sequencing library, and the two exhibit a significant technical positive correlation unrelated to biological interaction (Figure S2). In correlation-based analyses of the LIPSTIC signal,[13] this relationship motivates normalization of the LIPSTIC signal by dividing it by the sample hashtag. The regressive framework in Gloss allows us to include this factor in the regression feature matrix X directly, without the need to normalize the LIPSTIC signal by that factor. We also include RNA library size as a confounder to help prevent genes that are highly correlated with library size from being selected as interaction-associated genes, if there is an association with library size and the LIPSTIC signal (Figure S2).

Regression objective

Gloss regresses cell type-specific y on X to derive β such that $y = X\beta$. To accomplish this, Gloss minimizes the following objective function:

$$\widehat{\beta} = \underset{\beta}{\operatorname{argmin}} \frac{1}{2n} ||y - X\beta||_2^2 + \lambda_1 \sum_{\ell=1}^{D^*} m ||\beta^{(\ell)}||_2 \quad \text{(Equation 3)}$$

where

$$m := \begin{cases} 1 & \text{if } p_\ell = 1 \ \& \ \ell \text{ is a confounder} \\ \lambda_2 & \text{if } p_\ell = 1 \ \& \ \ell \text{ is a gene} \\ \sqrt{p_\ell} & \text{if } p_\ell > 1 \end{cases} \quad \text{(Equation 4)}$$

where (ℓ) denotes group membership, p_ℓ denotes the size of the group, and λ_1 and λ_2 are tunable hyperparameters. The objective is a modified group lasso model, due to the presence of the unsquared $||\cdot||_2$ norm over groups provided as input into the model. The overlapping group lasso[23] allows the duplication of features present in multiple groups, and then the standard group lasso penalty is applied over this duplicated feature matrix. In this formulation, we have included all genes, including those not originally in the pathway groups. Hence each gene, including the ones that are in the pathways already, has its own additional group, to make the penalty more fair across all profiled genes.

Modifications from standard group lasso

The custom group weighting penalty m is unique to Gloss. Standard group lasso scales penalties by $\sqrt{p_\ell}$. However, the dominance of single-gene groups in our feature set renders this scaling suboptimal. Gloss retains $\sqrt{p_\ell}$ for multi-gene pathways, but introduces a tunable weight $\lambda_2 \geq 1$ to explicitly penalize single-gene groups. Thus Gloss has two tunable hyperparameters λ_1, λ_2, whereas standard group lasso has only one. This increases flexibility in application to scRNA-seq data analysis.

Solving the objective for β

The Gloss objective is solved using block coordinate descent over the groups D^*, according to the weights specified by m and the hyperparameters λ_1 and λ_2. Gloss uses an efficient implementation using Anderson acceleration for coordinate descent using the Celer[67,68] solver, as specified in the package skglm.[59,60]

Interpretation of β and group inclusion

The vector of coefficients β has values for confounding variables, a copy of all the genes, and an additional copy of each gene for each time it appears in a pathway group. Summing coefficients across all instances of a gene (single-gene features plus pathway memberships) yields a cumulative gene-level contribution. The values of coefficients for different copies of a gene across different pathways allow for additional interpretation. Additionally, as the penalty $||\cdot||_2$ over each group has a discontinuity at 0, some groups are zeroed out entirely, similar to how the $||\cdot||_1$ norm implies the same for individual features. The subset of groups that are included with non-zero coefficients are considered important to the model.

Baseline regression models

The baseline linear regression models of ridge, elastic net, and lasso, and the baseline nonlinear models of multilayer perceptron (MLP) and random forest were all applied on the normalized scRNA-seq matrix appended with the two normalized confounders, without any gene duplications. These were all taken from their implementations in scikit-learn.[63]

Cross-validation and hyperparameter fine-tuning

Nested cross-validation (an inner loop for tuning hyperparameters and an outer loop for testing generalization) were used to tune Gloss and all other models tested.

All models tested used a consistent 5-fold outer loop, but differed in evaluation and selection of hyperparameters in their inner loops. This is shown in Table S1. Note that in the uLIPSTIC LCMV CD4 T cell subset analysis, a restrictive λ_1 with $\lambda_2 = 1$ was picked for Gloss with KEGG, preventing any pathways from being selected and thus effectively reducing Gloss KEGG to a lasso model.

Pathways and enrichment gene sets

Acquisition

The pathways and gene sets used in Gloss analysis and benchmarking were collected as described in the key resources table and Table S1.

Filtering and processing

These datasets were filtered and processed as follows:

NicheNet upstream signaling subnetworks. This network is both directional (source-target gene-gene edges) and multiedge (the same two genes can have multiple edges between them). We first considered only edges whose targets were of interest (listed below), so genesets consisted of genes upstream of these genes of interest and the targets of interest themselves. To leverage the mulitiedge quality of the network, high-confidence networks were constructed with specific thresholds t of a minimum number of edges going from a source gene to the target gene of interest (each edge indicates evidence from a distinct source): For CD40, we set the threshold to $t = 2$. For Itgae, due to relatively limited information on Itgae, we kept the threshold at $t = 1$. For the combination with CellChatDB, we first filtered NicheNet to the subnetwork with gene targets only in the CellChatDB geneset, and then filtered it to three thresholds of $t = 5$, $t = 6$, and $t = 7$. The threshold of 7 is used in the main geneset, and 5 and 6 are used for additional analysis.

Humanbase Networks. These networks provide tissue-specific undirected gene association probabilities. The subnetwork was constructed for both the macrophage and dendritic cell networks, restricting to edges connected to CD40. These were each further filtered using probability threshold 0.4 for both the macrophage network and the dendritic cell network. As these genes are human genes, the g:Profiler[69] ortholog mapping tool was used to map each of these genesets to mouse orthologs.

Basso et al. *signature.* The human genes in the set of upregulated CD40 genes in the signature were mapped to mouse orthologs using g:Profiler.[69] Genes that were not mapped to any orthologs were manually examined and mapped to an ortholog if available using NCBI.[70]

Additional signatures. The NicheNet CD40 $t = 1$ and HumanBase probability threshold of 0.2, 0.6, and 0.8 signatures for both macrophages and dendritic cells were also included as supplementary signatures.

Pathways

Table S1 provides a description of parameters of the pathways used in Gloss.

Curated gene sets

Table S1 provides a description of the parameters of the genesets used for Gloss evaluation analysis.

Data preprocessing and normalization

The LIPSTICv1 and uLIPSTIC data, along with cell type annotations, were taken from the respective publications.[13,14] The LIPSTICv1 data[14] was filtered to match the cell and gene filtering strategy used in the uLIPSTIC datasets - see Data and Code Availability and the original uLIPSTIC publication.[13] The CD8 effector T cell and CD8 T cell type subsets in the uLIPSTIC LCMV datasets were merged to

one CD8 T cell subset. To prepare data for regression, Gloss first removes genes with zero counts, and then normalizes scRNA-seq gene expression using log-library size normalization over each cell. Then the expression of each gene is z-scored to have 0 mean and unit variance across the entire profiled dataset. Both the sample hashtag and RNA library size confounders are log-normalized, using a pseudocount equal to the 0.1th percentile, and then z-scored across the dataset. The LIPSTIC interaction signal is log normalized, using a pseudocount of the 5th percentile, z-scored, and then offset so the minimum value is zero.

In the LCMV datasets, the CD8 T cell donors were profiled in addition to the acceptors; a P14 flag was used to determine the donor cells.[13] For these datasets, the donors were removed after the above normalization, before subsetting to specific cell types and then using as input to Gloss's feature transformation and regression.

LIPSTIC data

The data subsets used for Gloss regression, as well as their original main datasets and their sizes (after preprocessing), are described in Table S1. Scanpy[61] was used to process and handle this data. Note that in all cases, the number of genes (features in the Gloss model) are larger than the number of cells (observations in Gloss), further justifying the use of a regularized linear regression model.

The CD4 T cells in the uLIPSTIC gut dataset were annotated to include a small number of Treg cells, while in the two LCMV datasets, Treg cells have a separate annotation than the rest of the CD4 T cells. See Figure S1 for further details about the datasets used in our analysis. All UMAP plots with the normalized LIPSTIC signal are clipped at maximum 98% and minimum 2%.

Table S1 shows the number of cells in each condition or timepoint-specific subset.

Data simulations

Geneset-based simulation of LIPSTIC signal

We perturbed both the raw counts of the selected genes and the LIPSTIC signal in the LIPSTICv1 and uLIPSTIC data to simulate geneset-based dependencies with the LIPSTIC signal. To accomplish this, for each cell c, a random factor $r \in [-0.5, 0.5]$ was uniformly drawn, and the raw UMI count for each gene in the geneset was multiplied by $f_c = 2^r$, and the LIPSTIC signal was also multiplied by $f_c = 2^r$. Resulting values were rounded to the nearest integer. This simulation process was run 20 times to get 20 independent simulations.

Simulations in this fashion were done using the CD40-based genesets in NicheNet and Humanbase for the LIPSTICv1 data, as well as the NicheNet Itgae geneset for the uLIPSTIC CD4 Gut data and combined CellChatDB + NicheNet genesets for each of the uLIPSTIC datasets.

This transformation was applied over each whole dataset before applying any normalization before inputting into Gloss. The only exception was the LCMV datasets, where donors are also profiled; in this case, the transformation was applied only to the profiled acceptor cells.

Control perturbation simulations

The NicheNet CD40 geneset was used to create two types of control simulations in LIPSTICv1+scRNA-seq data. First, we repeated the simulation by applying the same per-cell transformation to a random gene set matched in size (and matched by expression level) to the CD40 genes present in the data, resampling a new random set for each of the 20 runs. The LIPSTIC signal transformation was the same as in the main simulation. These random genes were chosen to match the numbers of genes in the CD40 geneset that were in the gene transcript count bins of $[0, 10]$, $[10, 10^2]$, $[10^2, 10^3]$, $[10^3, 10^4]$, and $[10^4, 10^5]$.

The second control simulation perturbed the scRNA-seq matrix for the CD40-related genes but left the LIPSTIC signal unchanged.

Alternative nonlinear simulation of LIPSTIC signal

An alternative nonlinear version of the perturbation was used where for each cell c, a random factor $r \in [-0.5, 0.5]$ was uniformly drawn, and the raw UMI count for each gene in the geneset was multiplied by $f_c = 2^r$, but the LIPSTIC signal was taken to the power of $1 + r$. We rounded both of these computations to the nearest integer. This simulation process was run 20 times to get 20 independent simulations. This was tested only using the CD40 NicheNet Upstream gene set for the LIPSTICv1 data (Figure S11C).

Prediction of LIPSTIC signal in non-LIPSTIC data

We used our cell type-specific Gloss models trained on the LIPSTICv1+scRNA-seq data (measuring interactions with CD4 T cells) for myeloid cells in B16 melanoma mouse model and applied them to myeloid cells in an independent scRNA-seq dataset Ishizuka2019[57] profiling immune cells in B16 melanoma mouse model under ADAR1 knockout and control conditions. After filtering, log library size with z-scoring was used to normalize the data, which was then clustered using the Leiden[71] clustering algorithm to reannotate the data. Clusters were annotated using marker genes from the original publication. This was done using scanpy.[61]

To match populations for prediction, Harmony[66] was used to integrate the Ishizuka et al.[57] data with the Chudonvskiy et al. LIPSTICv1[14] data, which we used to match different populations to models trained on the LIPSTICv1 data. The monocytes were matched to the Mo/MF model, the Migratory DCs to the mRegDC2 model, and the cDC2s to the cDC2 model.

For prediction over each cell in the Ishizuka2019 dataset, the gene expression coefficients and intercept for each cell type specific model were averaged across the 100 bootstraps, and then the shared genes in the Ishizuka2019 dataset were used as input to this averaged model. The coefficients assigned to RNA and ADT library size were not used, which allows normalized LIPSTIC signal to be predicted.

The predictions revealed a consistent increase in inferred myeloid–T cell interaction intensity in the ADAR1 knockout condition, in line with the known role of ADAR1 as post-transcriptional silencer of immune activation (Figure S6). Specifically, Ishizuka et al. showed that loss of ADAR1 function improves anti-PD-1 immunotherapy response,[57] and our results suggest a specific underlying mechanism, via increased intensity of interactions between T cells and antigen-presenting myeloid cells. This proof-of-concept analysis demonstrates that, although broad cross-dataset generalization is not expected, Gloss can in certain cases transfer learned interaction patterns to independent non-LIPSTIC data and yield interesting predictions.

CITE-seq analysis

We demonstrate as a proof of concept that Gloss can be applied to CITE-seq data to provide interpretable predictions of protein abundance from transcriptomes. To accomplish this, 10k human PBMC CITE-seq dataset was acquired from the 10× Genomics Website.[58] After filtering, log library size with *z*-scoring was used to normalize the data, which was then clustered using the Leiden[71] clustering algorithm. Clusters were annotated as T cells or other cell types based on expression of CD3D, CD3E, and CD3G. This was done using scanpy.[61]

Then we restricted the analysis to the T cell compartment. We modeled three key proteins, CD4, CD8A, and PD-1, which mark helper CD4 T cells, cytotoxic CD8 T cells, and activated or exhausted T cells (Figure S7). Over all the annotated T cells, the CD4, CD8A, and PD-1 CITE-seq signals were independently predicted using the same workflow used to predict LIPSTIC signal using Gloss and Hallmark (human) pathways, except that the CITE-seq library size was directly used in lieu of the sample hashtag. Note: during CITE-seq preprocessing, Pdcd1 was filtered out from the scRNA-seq analysis because of low UMI counts, so it was not used in Gloss modeling.

Gloss significantly outperformed baseline methods (lasso, ridge, elastic net) that rely solely on individual gene features, demonstrating the benefit of incorporating pathway-level information through group lasso regularization. Gloss identified expected dominant transcripts, such as CD4 for CD4 protein and CD8A and CD8B for CD8 protein, as well as cell type-specific markers such as CD40LG for CD4 T cells and CTSW for CD8 T cells. It also highlighted transcripts associated with T cell activation and exhaustion, including TOX, TIGIT, and GZMK. Together, these results indicate that Gloss effectively integrates pathway information to enhance predictive power and interpretability, even in modalities beyond LIPSTIC+scRNA-seq.

Software and compatibility

Gloss is built on top of the GroupLasso estimator of skglm,[59,60] and is designed as a class built on top of scikit-learn[63,64] base regression estimator. Gloss expects data to be provided as input in the form of AnnData and is compatible with the scanpy[61] and scverse[72] ecosystem, as well as scikit-learn infrastructure.

QUANTIFICATION AND STATISTICAL ANALYSIS

Computation of enrichments

Enrichments were calculated using the GSEApy implementation of Enrichr local enrichment,[62] which applies the Fisher's exact test to assess overlap between a query gene set and the annotation gene set of interest, using all genes in the scRNA-seq matrix as the background.

Bootstrapping coefficients for gloss and ridge regression

To estimate confidence intervals, we performed bootstrapping. For this, the optimal parameters from the optimization using the entire dataset were used. We generated 100 bootstrap resamples (sampling with replacement), applied Gloss regression, computed gene coefficient confidence intervals and pathway inclusion frequencies. Gene coefficients were then ranked by their median value calculated over the bootstrap samples.

Significance for positive bootstrapped coefficients were determined by testing whether the 95% confidence interval computed using top and bottom 2.5% percentiles of a coefficient was entirely above zero; if this was the case, the coefficient was determined to be positive and significant.

Top coefficients in enrichment plots

Enrichment plots showing overlap with curated or simulated ground truth gene sets (e.g., Figures 4 and 5) were based on median-ranked Gloss coefficients from the bootstrap analysis. Only genes with positive median coefficients were included. The uLIPSTIC gut CD4 T cell data yielded fewer such genes, and thus *x* axes in the corresponding plots were limited to the top 400 positive genes.

Performance and runtime

Model performance was evaluated using 50 random 80/20 train-test splits for each dataset. After parameter tuning, each model was trained and tested across these splits. Runtime analysis measured the duration of model fitting (model.fit()) on the training data. For benchmarking, we used scikit-learn implementations of Ridge, MLP, and RandomForest, and the faster skglm implementations of Lasso and ElasticNet.[59,60]

In Figures 2 and S7, comparisons between the top non-Gloss method and top Gloss method (determined by median R^2 value) were done using Wilcoxon signed-rank tests.

In Figure S6, comparisons of the predicted signal between the ADAR1-knockout and control conditions were done using Mann-Whitney 2-sided U-tests.

Overlap between top gloss coefficients

In Figure 7F, the Venn diagram shows the overlap of the top 10 positive significant genes for CD4 T and CD8 T cells at 50 h post-infection, ranked by their median coefficients (in CD4 T cells). The significance of the overlap was assessed using Fisher's exact test, with the list of all genes from the uLIPSTIC LCMV LN analysis used as background.

LIPSTIC and gene expression correlation analyses

Correlation-based analysis was used as a baseline for comparison with Gloss. Spearman correlations were calculated between normalized gene expression and sample-hashtag normalized LIPSTIC signal. In most comparisons, for normalization the raw LIPSTIC read counts were divided by the sample-hashtag library size for each cell.

Alternatively, normalization using Gloss was applied. Gloss can be used to normalize the LIPSTIC signal by regressing it on confounding variables only (i.e., not including gene expression features), which reduces the regression model to a standard lasso regression. This enables normalization either within each cell type or across the full dataset. In some visualizations and correlation analyses (e.g., Figures 4 and S2) we used this Gloss confounder-normalized LIPSTIC signal.

Correlations, Venn diagram overlap, method performance statistical comparisons, and prediction statistical comparisons were all done using the SciPy package.[65]

Cell Reports Methods, Volume *6*

Supplemental information

Predictive modeling of molecular activity underlying physical cell-cell interactions

Tamjeed Azad, Sarah K. Walker, Gabriel D. Victora, and Yuri Pritykin

Document S1: Supplementary Figures S1–S12

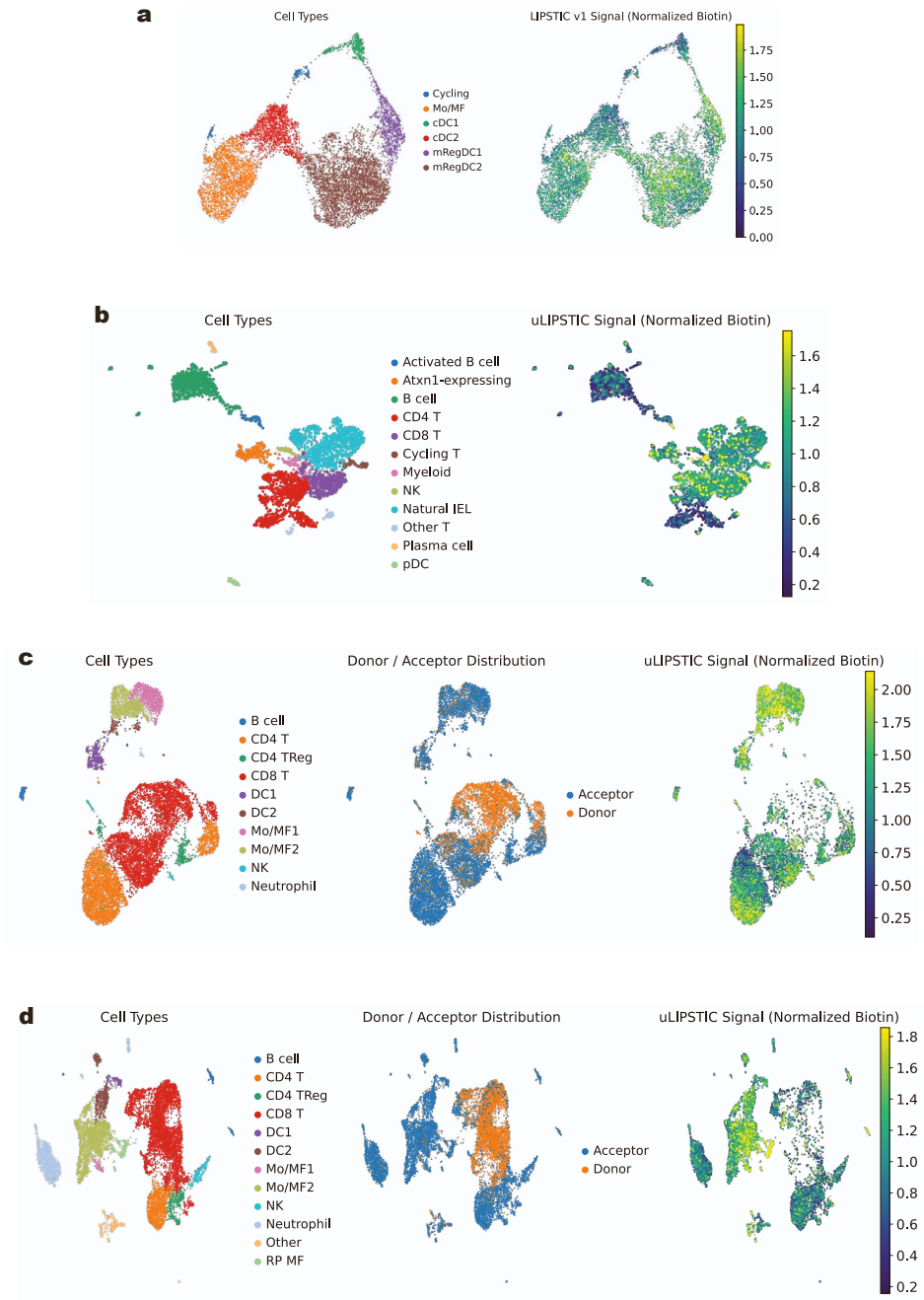


Figure S1. UMAP visualizations of the datasets used for Gloss analysis, related to STAR Methods.
(a) The LIPSTICv1 data showing myeloid cells interacting with CD4 T cells in the tumor microenvironment (Chudnovskiy et al. 2024 [S1]).
(**b**) The uLIPSTIC data profiling immune cells interacting with the mouse small intestine epithelium (gut) (Nakandari-Higa et al. 2024 [S2]).
(**c-d**) Two uLIPSTIC datasets taken in the context of LCMV from the lymph node (LN) (c) and pooled across the sites of systemic (sys.) infection, i.e. the lung, liver, and spleen (d), interacting with CD8 T cells (Nakandari-Higa et al. 2024 [S2]). For all datasets, the LIPSTIC signal was clipped at min 2% and max 98%.

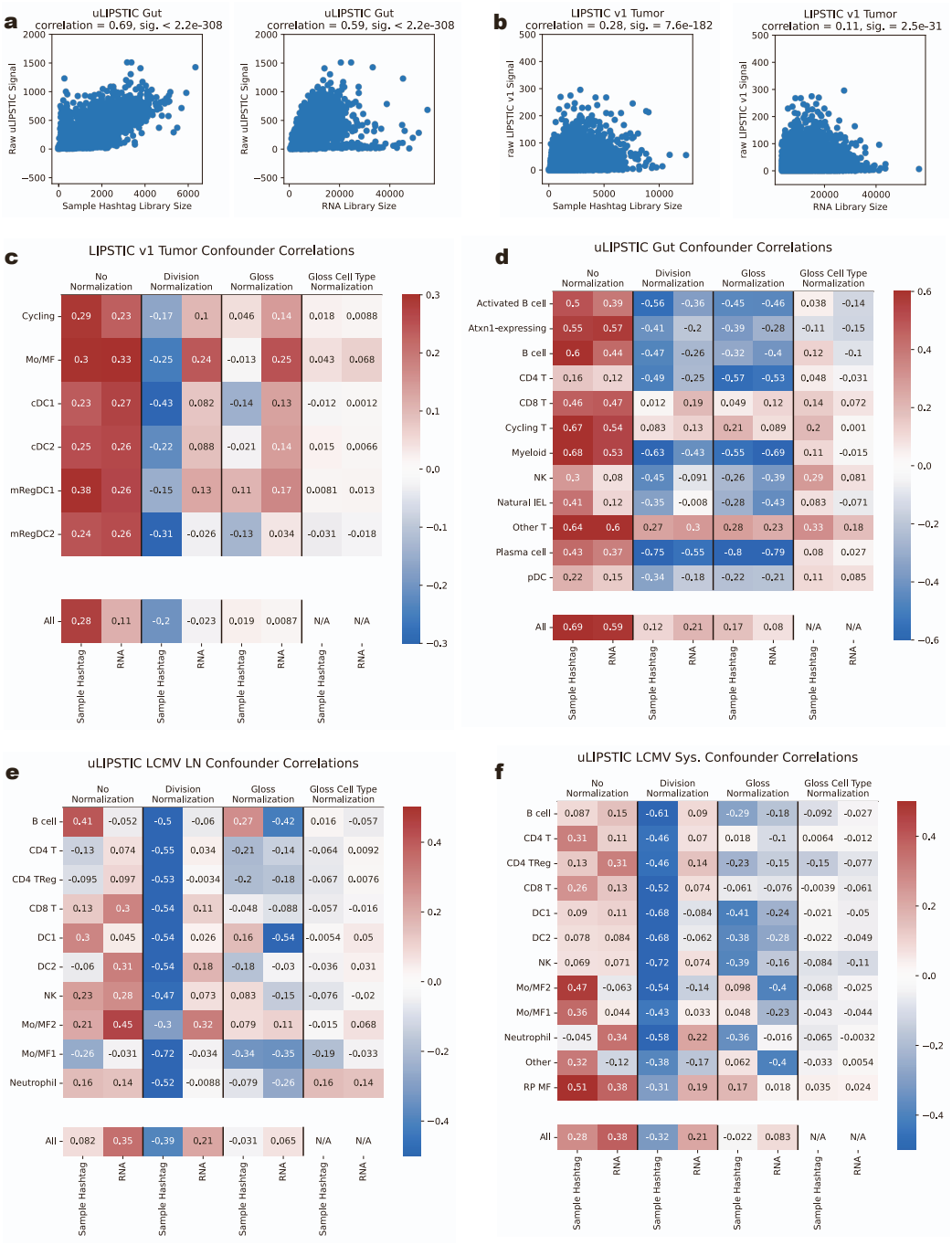


Figure S2. Correlations of the LIPSTIC signal with confounders, related to STAR Methods.
(a-b) Scatter plot of the LIPSTIC signal and the sample hashtag sequencing library size and the RNA sequencing library size for all cells in the uLIPSTIC gut data (a) and the LIPSTICv1 tumor data (b). Spearman correlation and significance p-value. For visualization purposes, the y-axis values were clipped at 2000 in (a) and at 500 in (b).
(c-f) Spearman correlations of sample hashtag library size and RNA library size for different LIPSTIC signal normalizations shown across cell types for the four datasets. See Methods for details about different normalizations.

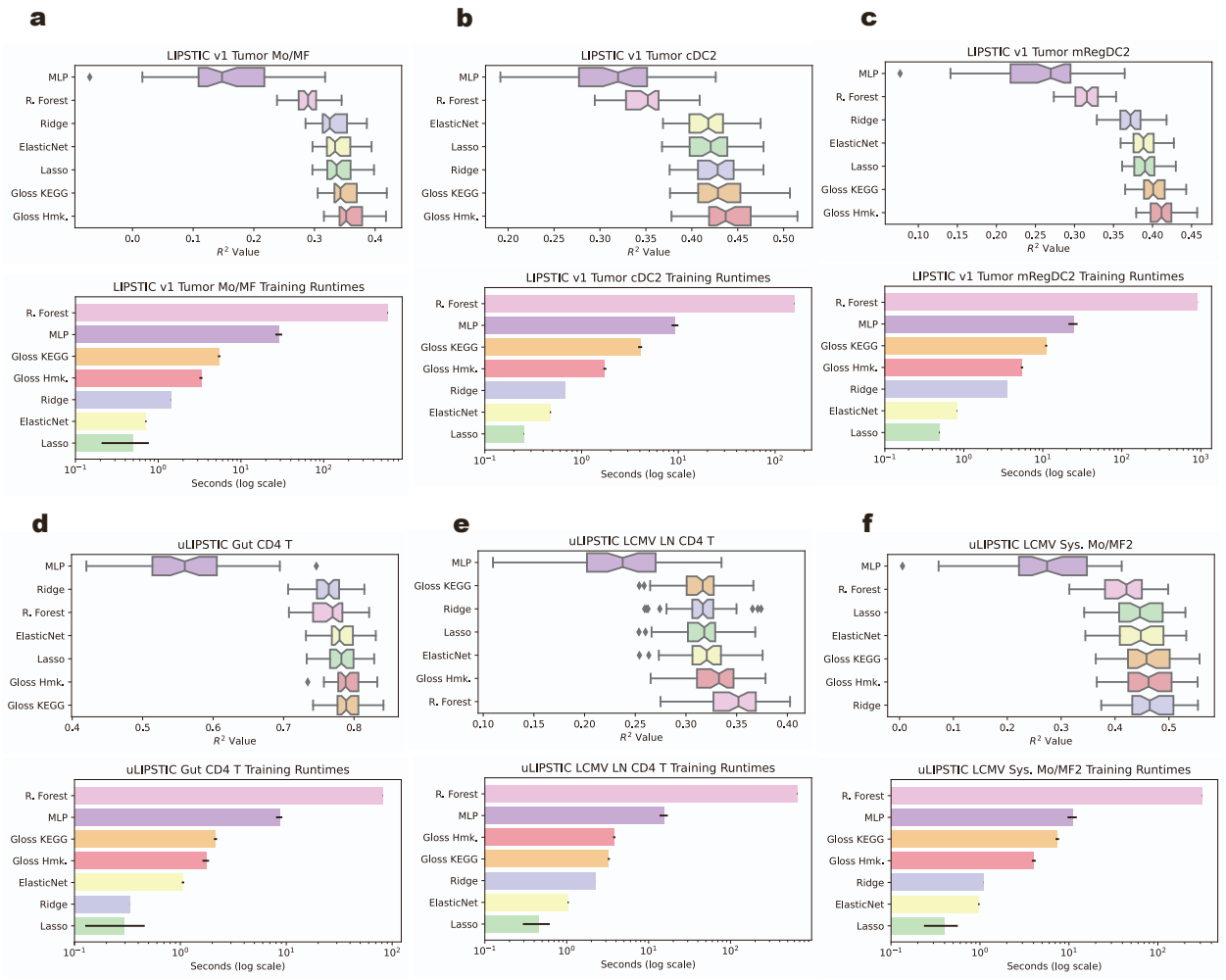


Figure S3. Gloss outperforms other nonlinear regression models, related to Figure 2. The same evaluations as in Fig. 2, but also showing results for MLP and random forest models. See Methods for details. In addition, barplots below each R^2 evaluation panel show the mean time (in seconds, log scale) taken to fit over the training data, when evaluating over the 50 train-test splits. Error bars represent estimated 95% confidence intervals.

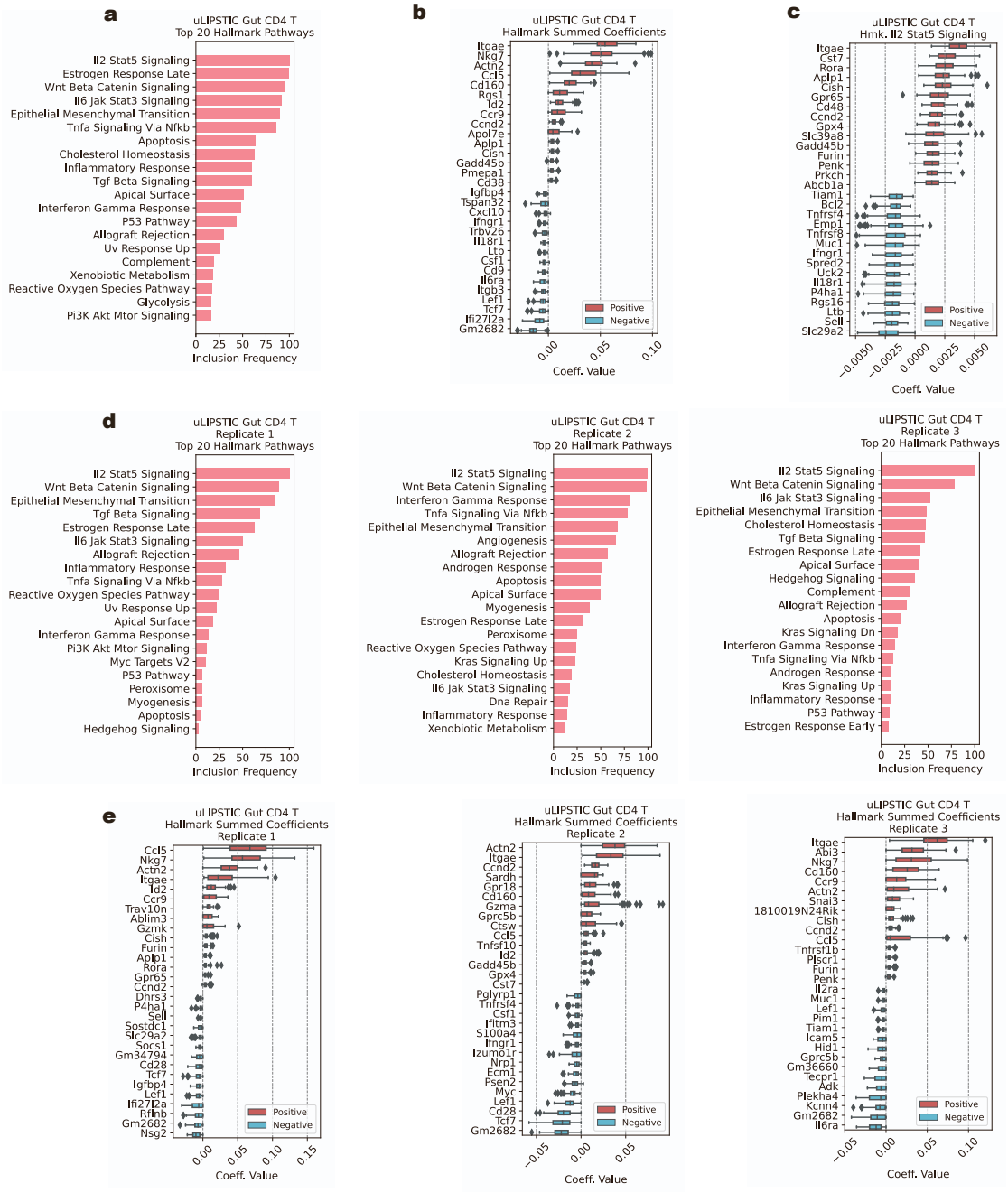


Figure S4. Top pathways and gene coefficients for the uLIPSTIC Gut CD4 T cells, with replicates, related to Figure 3.

(**a-c**) Similar to Fig. 3, but showing the results for the uLIPSTIC Gut CD4 T subset. (**d-e**) Highlighting reproducibility of results across mouse replicates for (d) top pathways and (e) gene coefficients.

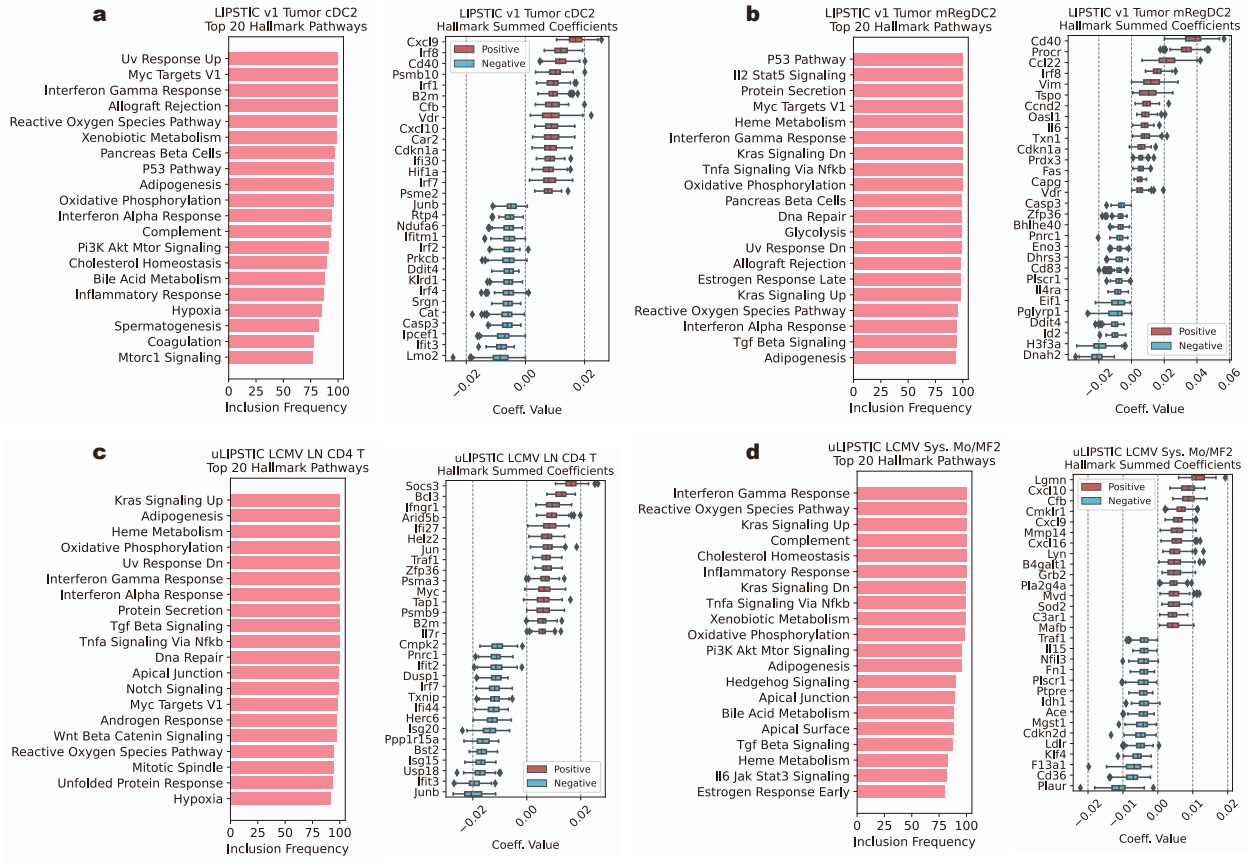


Figure S5. Top pathways and gene coefficients for the LIPSTIC data, related to Figure 3.
(a-b) Similar to Fig. 3, but showing the results for the LIPSTICv1 cDC2 subset (a) and the LIPSTICv1 mRegDC2 subset (b).
(c-d) Similar to Fig. 3, but showing the results for the CD4 T cell subset in the uLIPSTIC LCMV LN data (c), and for the Mo/MF2 subset in the uLIPSTIC LCMV Sys. data (d).

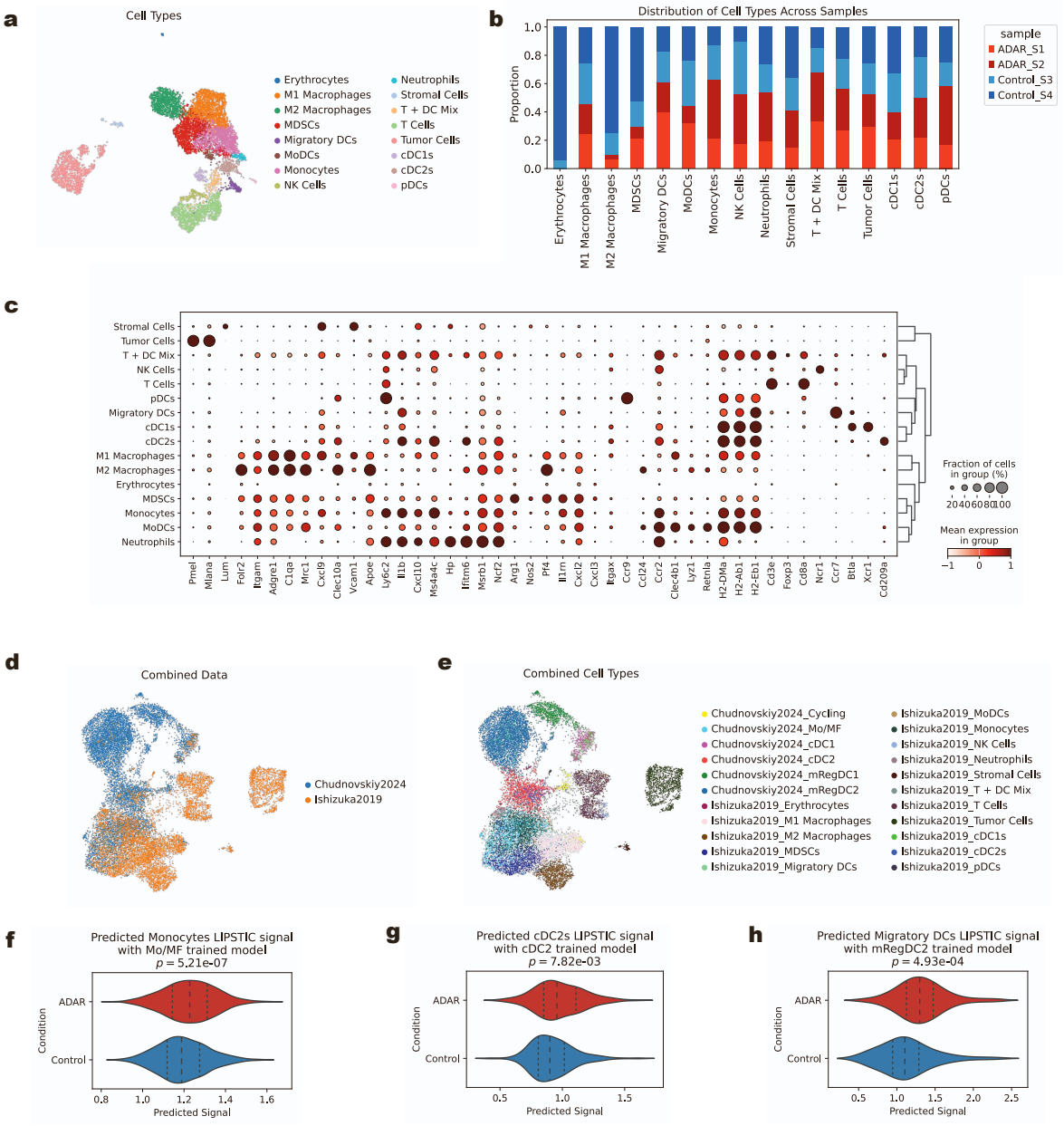


Figure S6. Transferring Gloss predictions to matched non-LIPSTIC data, related to Figure 3.
(a) Reannotation of the scRNA-seq data from Ishizuka et al., 2019 [S3], which matches the experimental context of the LIPSTICv1+scRNA-seq dataset (Chudnovskiy et al. 2024 [S1]) (measuring interactions of CD4 T cells with immune cells in the B16 melanoma mouse model, analyzed in Fig. 2a-c, 3, 4a,b, 6). The Ishizuka2019 data profiles the effect of ADAR1 knockout on immune cells in tumors.
(b) Sample composition across annotated cell types in ADAR1-knockout and control conditions.
(c) Gene expression dot plot of canonical marker genes across annotated cell type subsets.
(d-e) UMAP of harmonized integrated scRNA-seq data analyzed in panels **a-c** and scRNA-seq component of the LIPSTICv1+scRNA-seq data, colored by (d) dataset, (e) cell type annotations.
(f-h) Violin plots of predicted LIPSTIC signal obtained by applying Gloss models trained on matched cell types from the LIPSTICv1+scRNA-seq data to the Ishizuka2019 dataset, stratified by condition. Predictions in (f) monocytes, (g) cDC2 cells, (h) migratory DCs. The predicted interaction intensity values were significantly higher for ADAR1 knockout than for control (top, p-values, two-sided Mann–Whitney U), in line with the known role of ADAR1 as post-transcriptional silencer of immune activation.

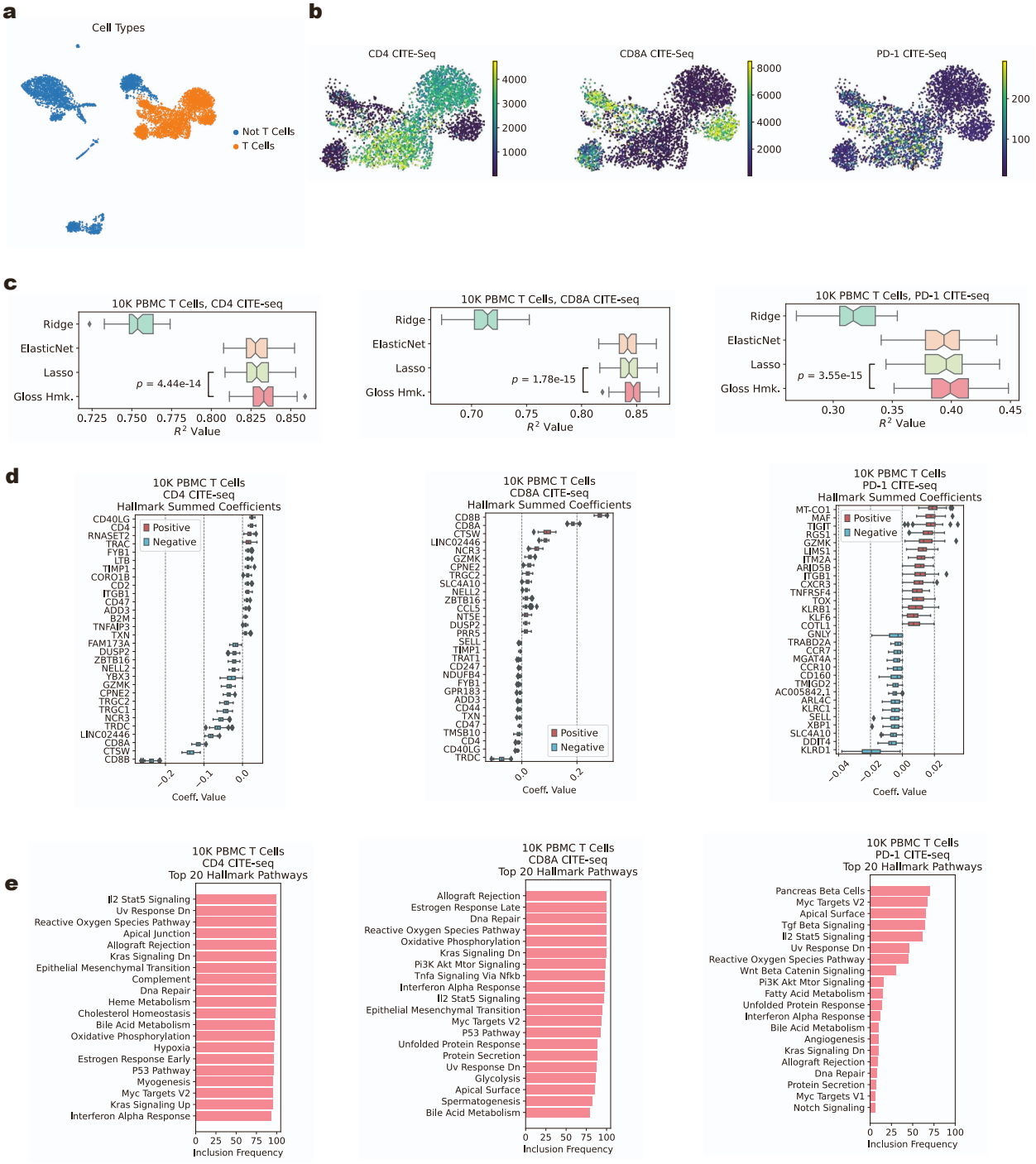


Figure S7. Applying Gloss to predict CITE-seq signal in human PBMCs, related to Figure 3.
(a) UMAP of annotated T cells from the 10x Genomics PBMC dataset (10x Genomics [S4]).
(b) UMAPs showing CITE-seq signal intensities for CD4, CD8A, and PD-1 within the T cell compartment.
(c) Predictive performance of Gloss compared with baseline linear regression methods (Lasso, Ridge, and Elastic Net) for CD4, CD8A, and PD-1 protein expression (similar to Fig. 2). Gloss outperformed Lasso, the 2nd best method (paired Wilcoxon signed-rank test).
(d) Top Gloss gene coefficients contributing to the prediction of CD4, CD8A, and PD-1.
(e) Top Gloss pathways most strongly associated with each target protein.

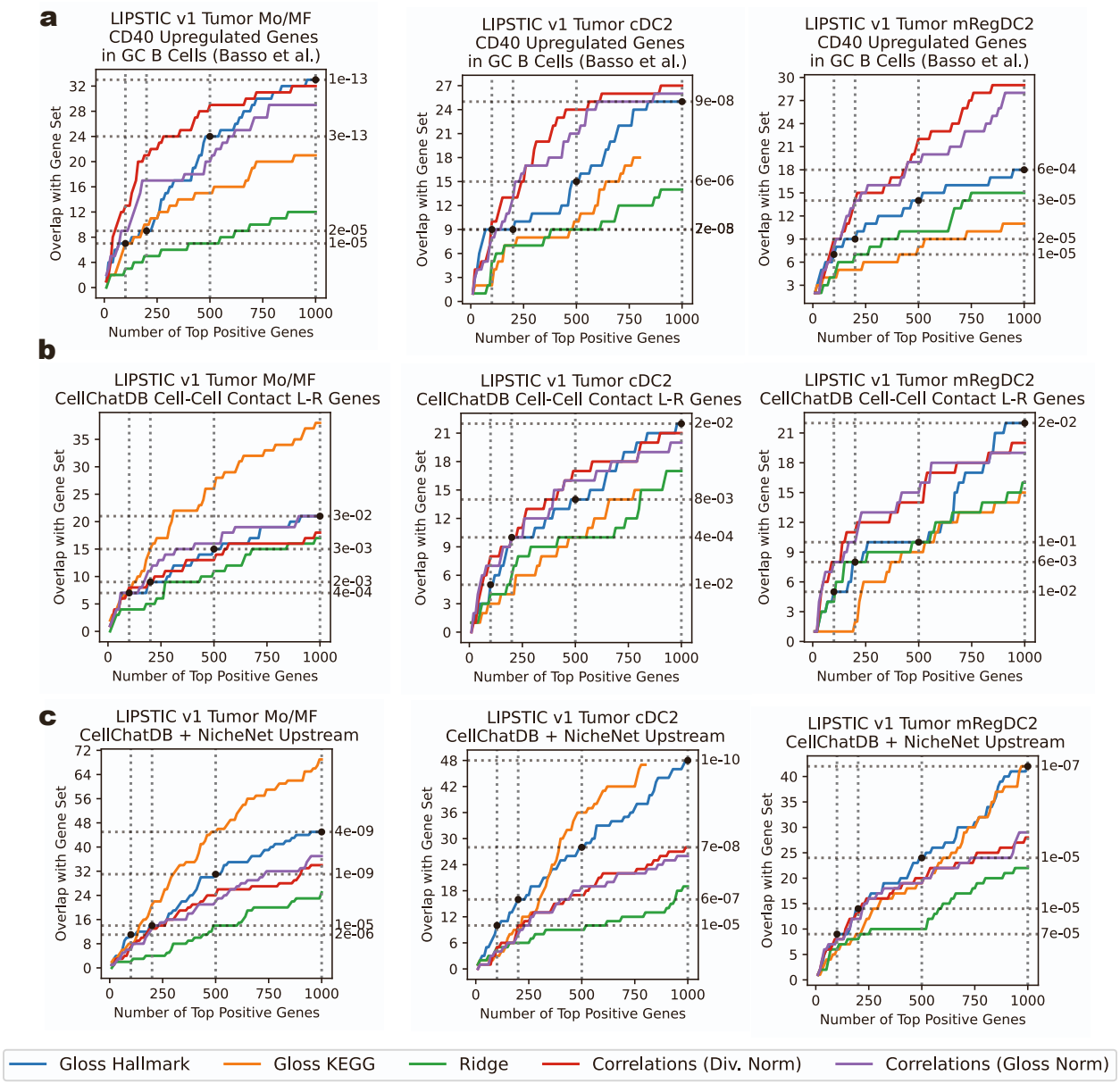


Figure S8. Additional ground truth enrichment analysis, related to Figure 4.

Results of analysis similar to Fig. 4, but for additional curated ground truth annotations tested for the Mo/MF, cDC2, and mRegDC2 cells in the LIPSTICv1 tumor dataset.

(a) Enrichment with a signature of CD40 upregulated genes in germinal center B cells from Basso et al. 2004 [S5].

(b) Enrichment of the gene set tested in Fig. 4d, but here on LIPSTICv1 instead of uLIPSTIC data.

(c) Enrichment of the gene set tested in Fig. 4e, but here on LIPSTICv1 instead of uLIPSTIC data.

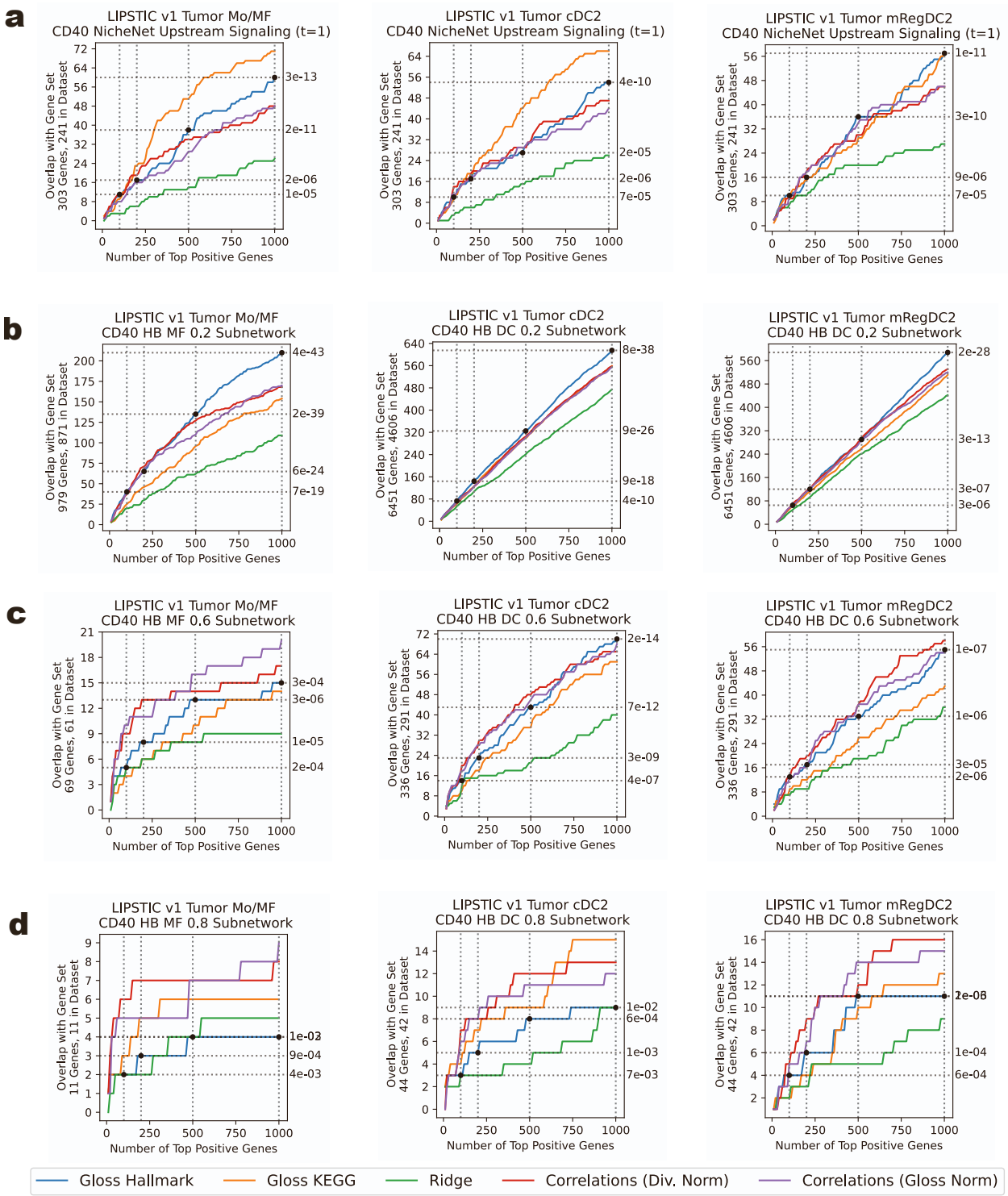


Figure S9. Gene set enrichment analysis of Gloss results with varying parameters, related to Figure 4. Results of analyses analogous to Fig. 4, testing additional parameter cutoffs for curated ground-truth annotations in Mo/MF, cDC2, and mRegDC2 cell populations from the LIPSTICv1+scRNA-seq tumor dataset.

(a) Enrichment using the CD40 NicheNet gene set with $t = 1$, without high-confidence filtering (see Methods).

(b-d) Enrichment using Humanbase GIANT network gene sets with probability cutoffs ranging from 0.2 to 0.8: (b) 0.2, (c) 0.6, and (d) 0.8.

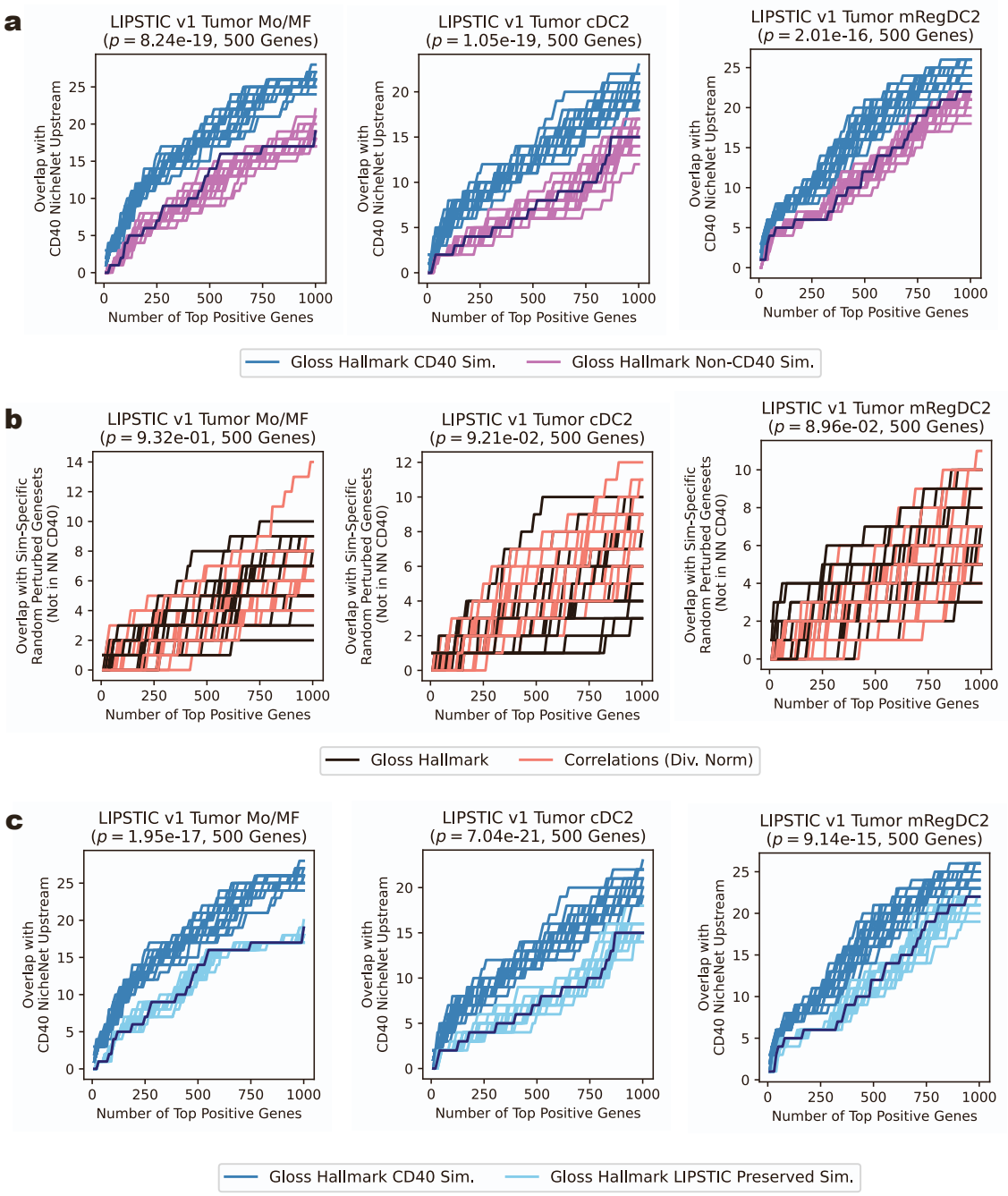


Figure S10. Comparison of Gloss results for the CD40 NicheNet subnetwork-based simulation and two control simulations, related to Figure 5.

(a) The Gloss results for the CD40 NicheNet subnetwork simulations from Fig. 5a are compared with Gloss trained on 20 other simulations where random genes not in the CD40 NicheNet subnetwork but with matched expression are perturbed (see Methods). Top, p-value from a Welch's unpaired t-test comparing the overlaps for top 500 genes from Gloss across the main and control simulations with the CD40 NicheNet subnetwork. Darker blue lines indicate enrichment for original, unperturbed data.

(b) For the control simulations from panel (a) where random gene sets were perturbed, the overlap with this random gene set (instread of the CD40 subnetwork) was compared between the Gloss results and correlation-based analysis.

(c) Similar to panel (a), but for comparison of the Gloss results for the CD40 NicheNet subnetwork simulations from Fig. 5a to the control simulations where the gene perturbations were the same but the LIPSTIC signal was unperturbed (see Methods).

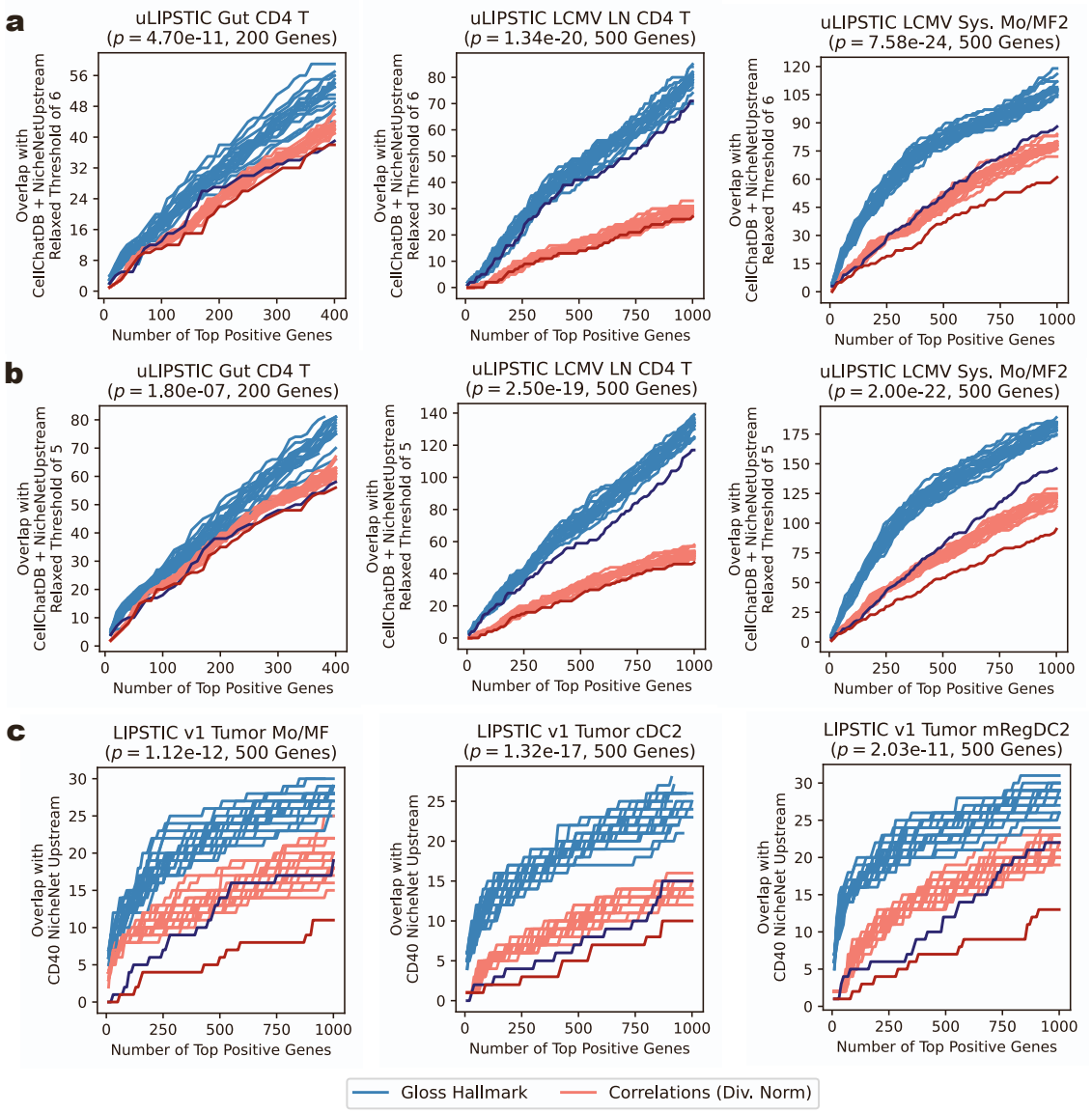


Figure S11. Robustness of the simulation analysis, related to Figure 5.

(a-b) Results of analysis similar to Fig. 5d, but for different, more relaxed parameters in the definition of a NicheNet subnetwork (see Methods). While in the main analysis presented in Fig. 5d, the edge confidence threshold of $t = 7$ was used for selecting the high-confidence edges in the NicheNet network, here the threshold was (a) $t = 6$ or (b) $t = 5$. Top, p-value from a paired t-test comparing the overlaps for top indicated number of genes between the results for Gloss and correlation-based analysis. Darker lines indicate enrichment for original, unperturbed data.

(c) Results for simulations similar to those presented in Fig. 5a, but with a non-linear relationship between the LIPSTIC signal and gene expression. Here, for each cell c, a random parameter α_c from $[-0.5, 0.5]$ was chosen, and then the gene expression read count was multiplied by 2^{α_c}, and the LIPSTIC signal read count s was transformed to $s^{1+\alpha_c}$ (instead of multiplying by 2^{α_c} like in Fig. 5a). The results are shown for Mo/MF, cDC2, and mRegDC2 cells in the LIPSTICv1 tumor dataset.

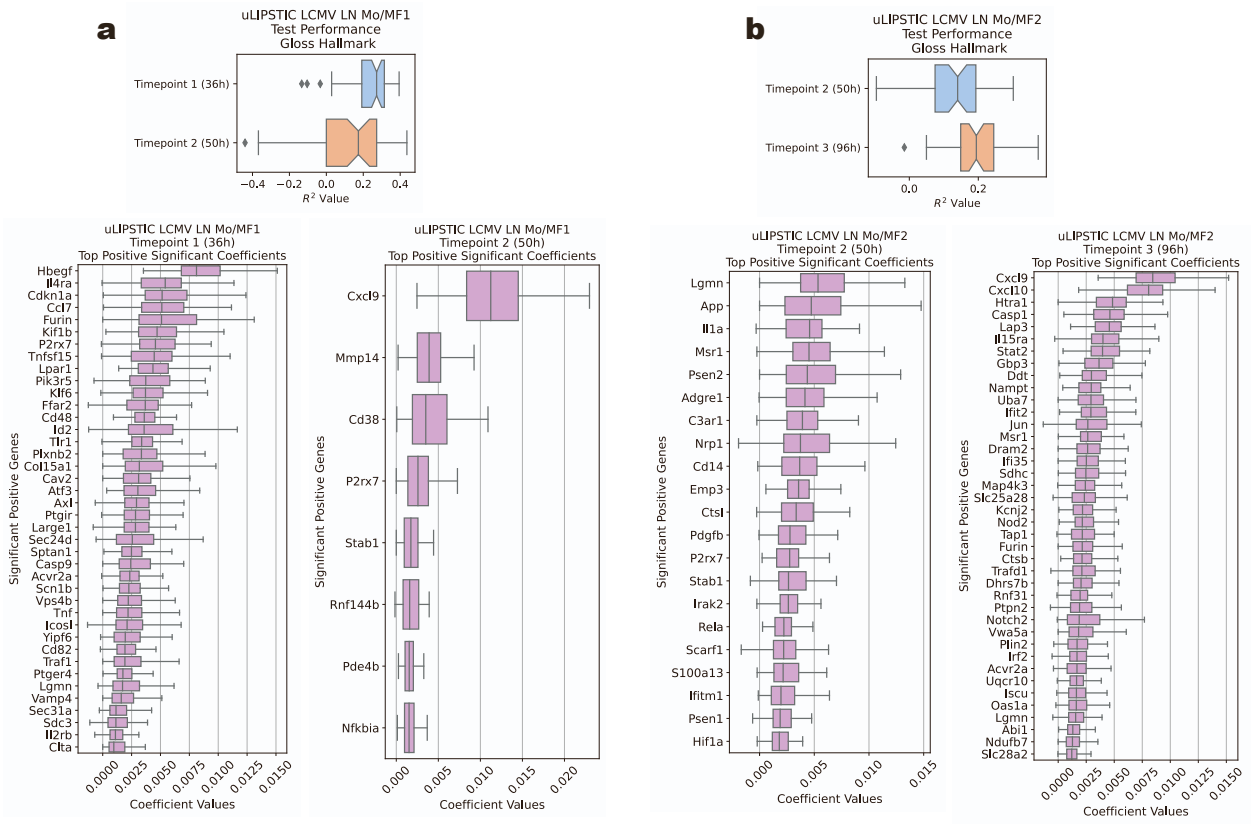


Figure S12. Gloss analysis of the uLIPSTIC+scRNA-seq data for the two subpopulations of monocytes and macrophages in LCMV infection, related to Figure 7.

Gloss performance for the **(a)** Mo/MF1 and **(b)** Mo/MF2 subpopulations of monocytes/macrophages as defined in the original publication (Nakandakari-Higa et al. 2024 [S2]). Top: results for held out test data for the two timepoints, 36h and 50h after initiation of the infection. Bottom: Top positive significant gene coefficients from the Gloss models.

Supplementary References

S1. Chudnovskiy, A., Castro, T.B.R., Nakandakari-Higa, S., Cui, A., Lin, C.H., Sade-Feldman, M., Phillips, B.K., Pae, J., Mesin, L., Bortolatto, J., Schweitzer, L.D., Pasqual, G., Lu, L.F., Hacohen, N., and Victora, G.D. (2024). Proximity-dependent labeling identifies dendritic cells that drive the tumor-specific CD4+ T cell response. Science Immunology *9*, eadq8843.

S2. Nakandakari-Higa, S., Walker, S., Canesso, M.C.C., van der Heide, V., Chudnovskiy, A., Kim, D.Y., Jacobsen, J.T., Parsa, R., Bilanovic, J., Parigi, S.M., Fiedorczuk, K., Fuchs, E., Bilate, A.M., Pasqual, G., Mucida, D., Kamphorst, A.O., Pritykin, Y., and Victora, G.D. (2024). Universal recording of immune cell interactions in vivo. Nature *627*, 399–406.

S3. Ishizuka, J.J., Manguso, R.T., Cheruiyot, C.K., Bi, K., Panda, A., Iracheta-Vellve, A., Miller, B.C., Du, P.P., Yates, K.B., Dubrot, J., Buchumenski, I., Comstock, D.E., Brown, F.D., Ayer, A., Kohnle, I.C., Pope, H.W., Zimmer, M.D., Sen, D.R., Lane-Reticker, S.K., Robitschek, E.J., Griffin, G.K., Collins, N.B., Long, A.H., Doench, J.G., Kozono, D., Levanon, E.Y., and Haining, W.N. (2019). Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade. Nature *565*, 43–48.

S4. 10x Genomics (2018). 10k PBMCs from a healthy donor — gene expression with a panel of TotalSeq$^{\text{TM}}$-B antibodies. https://www.10xgenomics.com/datasets/10-k-pbm-cs-from-a-healthy-donor-gene-expression-and-cell-surface-protein-3-standard-3-0-0. Accessed: 2025-10-21.

S5. Basso, K., Klein, U., Niu, H., Stolovitzky, G.A., Tu, Y., Califano, A., Cattoretti, G., and Dalla-Favera, R. (2004). Tracking CD40 signaling during germinal center development. Blood *104*, 4088–4096.

S6. Castanza, A.S., Recla, J.M., Eby, D., Thorvaldsdóttir, H., Bult, C.J., and Mesirov, J.P. (2023). Extending support for mouse data in the molecular signatures database (MSigDB). Nature Methods *20*, 1619–1620.

S7. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark gene set collection. Cell Systems *1*, 417–425.

S8. Cokelaer, T., Pultz, D., Harder, L.M., Serra-Musach, J., and Saez-Rodriguez, J. (2013). BioServices: a common Python package to access biological Web Services programmatically. Bioinformatics *29*, 3241–3242.

S9. Sang-aram, C., Browaeys, R., Seurinck, R., and Saeys, Y. (2025). Unraveling cell–cell communication with NicheNet by inferring active ligands from transcriptomics data. Nature Protocols *20*, 1439–1467.

S10. Jin, S., Plikus, M.V., and Nie, Q. (2025). CellChat for systematic analysis of cell–cell communication from single-cell transcriptomics. Nature Protocols *20*, 180–219.

S11. Greene, C.S., Krishnan, A., Wong, A.K., Ricciotti, E., Zelaya, R.A., Himmelstein, D.S., Zhang, R., Hartmann, B.M., Zaslavsky, E., Sealfon, S.C., Chasman, D.I., FitzGerald, G.A., Dolinski, K., Grosser, T., and Troyanskaya, O.G. (2015). Understanding multicellular function and disease with human tissue-specific networks. Nature Genetics *47*, 569–576.