Cell Reports, Volume *25*

Supplemental Information

Analysis of Single-Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics

Manu P. Kumar, Jinyan Du, Georgia Lagoudas, Yang Jiao, Andrew Sawyer, Daryl C. Drummond, Douglas A. Lauffenburger, and Andreas Raue

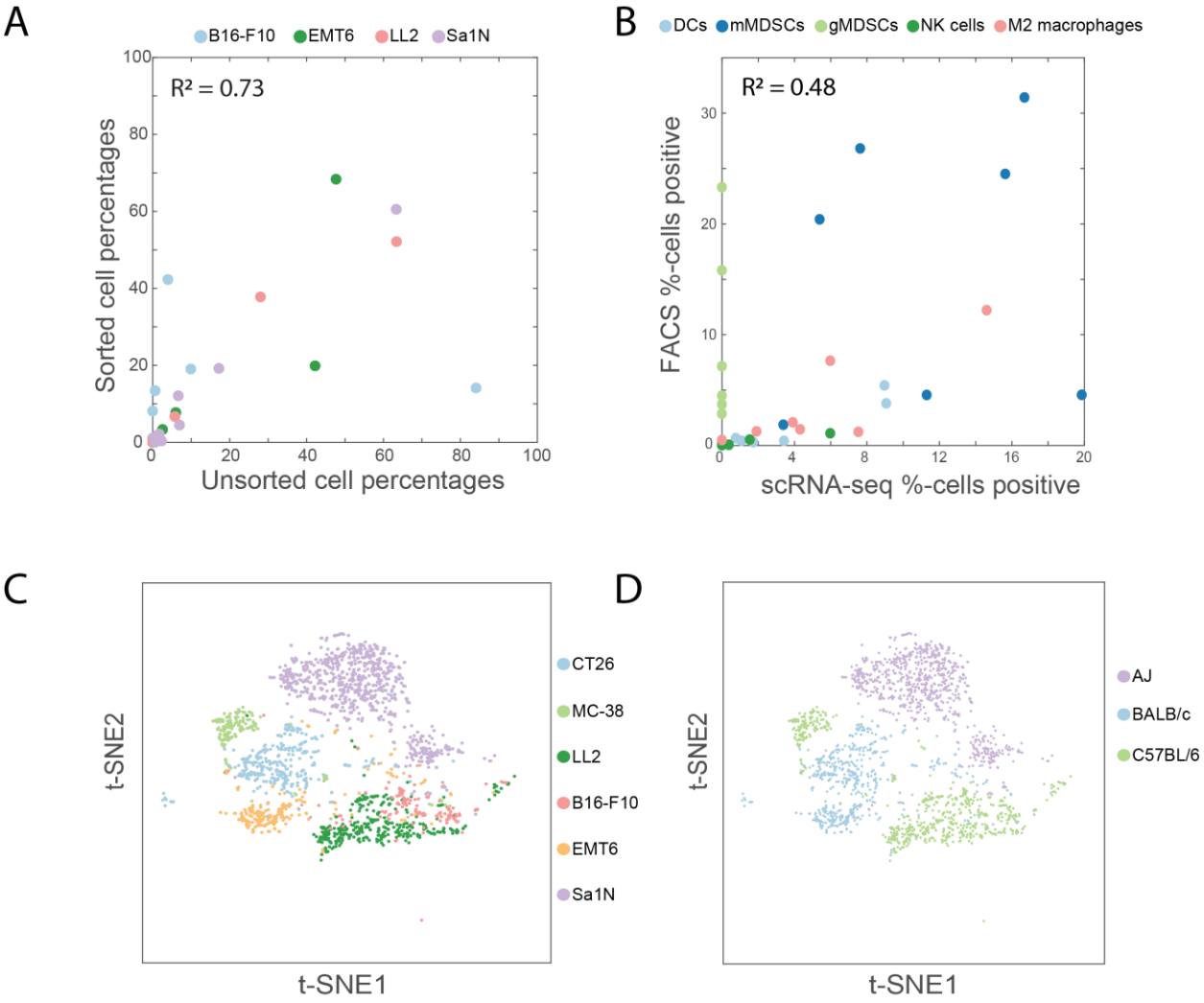
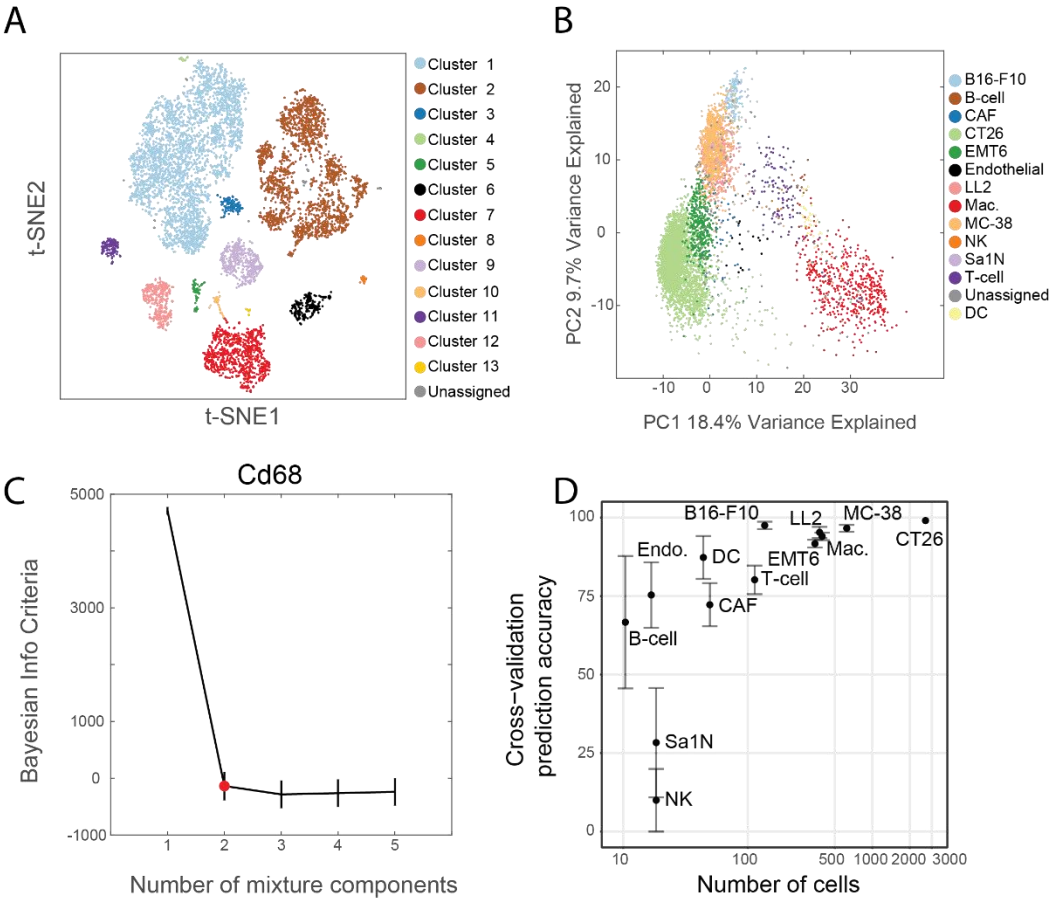


Figure S1. Single cell RNA sequencing of mouse syngeneic tumors and comparison to flow cytometry. Related to Figure 1.

(A) Percentage of cell types measured for samples that are enriched for CD45+ cells (sorted) or unenriched for any specific markers (unsorted). For most tumor models, enriching for CD45+ cells does not appreciably change cell type percentages. **(B)** Percentage of cell types identified in a tumor is consistent across tumor models when measured by either scRNA-seq (x-axis) or flow cytometry (y-axis). **(C, D)** t-SNE coordinates of predicted macrophages (**C)** Macrophages are colored by tumor model from which the cell originated. The two colon cancer models (MC-38 and CT26) cluster together **(D)** Macrophages are colored by the mouse strain. Macrophages also appear to cluster by mouse strain, except for cells from the MC-38 model (compare with Figure S1C) which cluster with the other colon cancer model (CT26).



E

Actual \ Predicted	B16-F10	B-cell	CAF	CT26	EMT6	Endo.	LL2	Mac.	MC-38	NK	Sa1N	T-cell	DC
B16-F10	166	0	0	0	0	0	0	0	0	0	0	0	0
B-cell	0	12	0	0	0	0	0	0	0	0	0	0	0
CAF	0	0	54	0	0	0	0	0	0	0	0	0	0
CT26	0	0	0	3326	0	0	0	1	0	0	0	0	0
EMT6	0	0	0	6	406	0	0	0	2	0	0	0	0
Endo.	0	0	0	0	0	16	0	0	0	0	0	0	0
LL2	0	0	0	0	0	0	454	5	0	0	0	0	0
Mac.	0	0	0	0	0	0	0	487	0	0	0	0	0
MC-38	0	0	0	2	0	0	0	1	770	0	0	0	0
NK	0	0	0	0	0	0	0	0	0	11	0	0	0
Sa1N	0	0	0	0	0	0	0	1	0	0	12	0	0
T-cell	0	0	0	0	0	0	0	8	0	0	0	114	0
DC	0	0	0	0	0	0	0	2	0	0	0	0	50

Figure S2. Classification of cell types based on scRNAseq data of syngeneic mouse tumor models. Related to Figure 1.

(A) DBSCAN clustering of t-SNE projected single-cell sequencing data used for identifying tumor specific cell type markers **(B)** Principal component analysis of the training data set. Each point represents an individual cell in the training data plotted along the first two principal components. PC1 appears to separate tumor from immune cells while PC2 separates tumor types **(C)** Illustrative example of the Gaussian mixture model selection approach for *Cd68*. BIC values were calculated using five-fold cross-validation for models containing one through five Gaussian components. The model with the fewest number of components within one standard error of the mean (error bars) was selected (indicated by red dot). **(D)** Cell type specific classification accuracy determined using 5-fold cross validation. Error bars are standard error of the mean. **(E)** Confusion matrix showing error rate for predicting cell type labels on the full training data set.

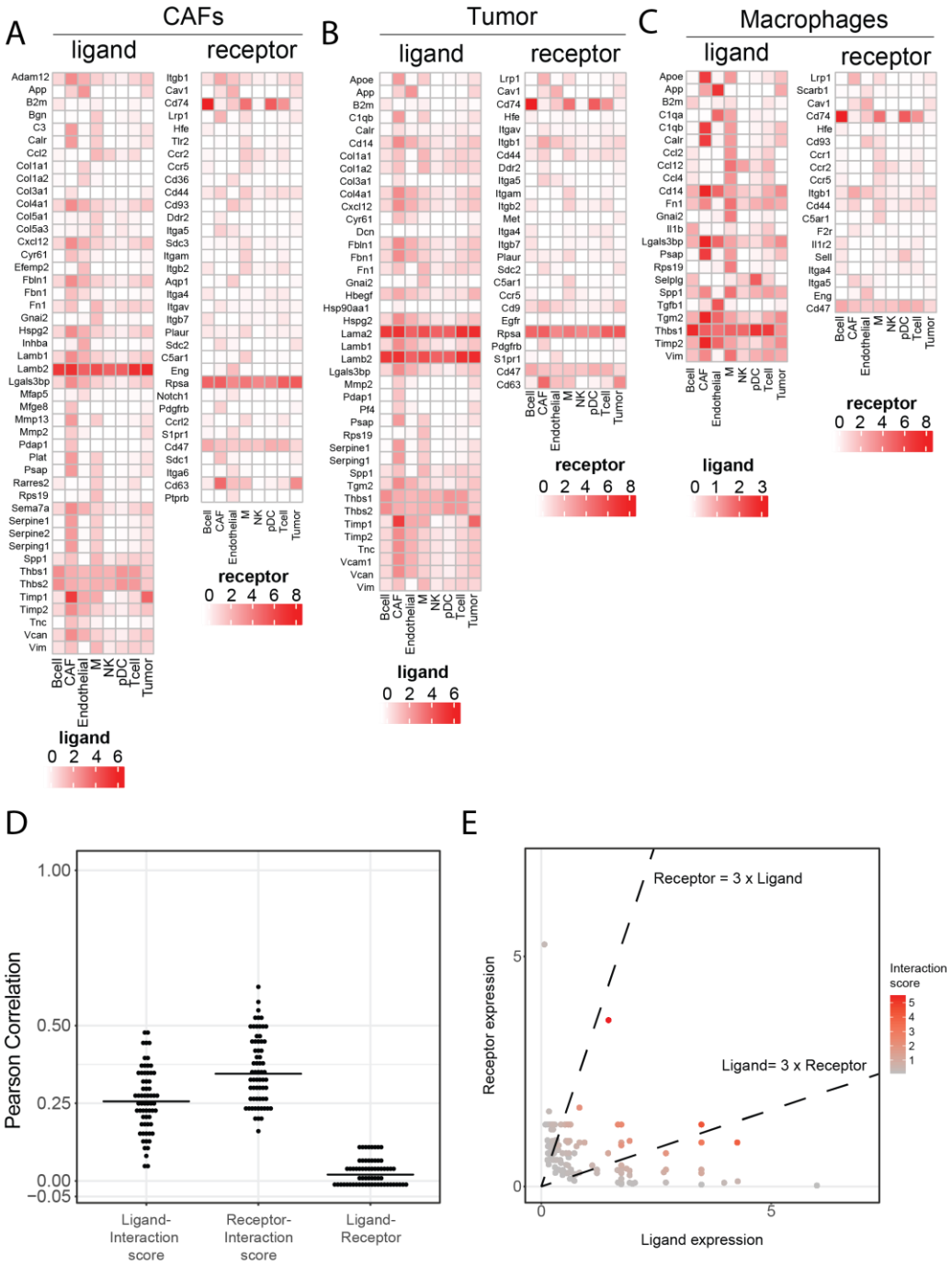


Figure S3. Characterizing known receptor ligand cell-cell interactions and their association with phenotypes of interest. Related to Figure 2.

(A, B, C) Heatmap shows either individual ligand or receptor expression for the interactions displayed in Figure 2. **(D)** Boxplot showing the correlation of ligand expression with interaction scores (left), receptor expression with interaction scores (middle) and receptor expression with ligand expression. Each dot represents a cell type pair and the Pearson correlation is computed across all receptor-ligand interactions (Table S6). Horizontal black lines indicate median correlation across all cell type pairs **(E)** Scatter plot showing receptor expression versus ligand expression for all Tumor-Tumor interactions in the syngeneic mouse models. Each point represents one receptor-ligand interaction and are colored by the average interaction score across all syngeneic models.

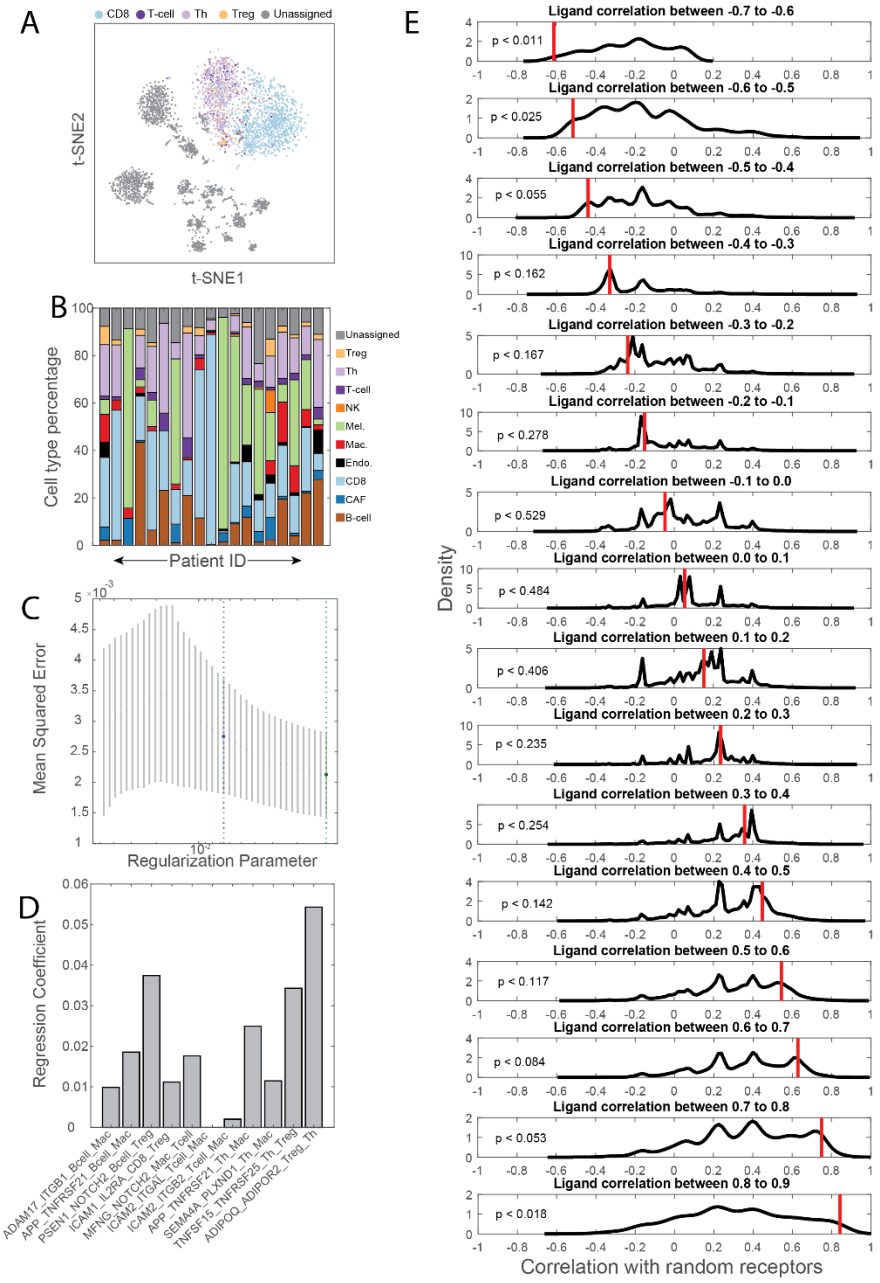


Figure S4. Analysis of interactions present in human metastatic melanoma. Related to Figure 4.
(A) t-SNE plot of cells from the Tirosh et al. dataset. Cells classified as T-cells were further classified into CD8+, Th, and Treg cells. **(B)** Percentages of cell types identified per tumor model. **(C)** Mean squared error (MSE) across a range of regularization parameters for predicting the percentage of Tregs. Error bars show the standard error of the mean. Green circles indicate the regularization parameter with the lowest MSE and blue circles indicate the sparsest model with MSE within one SE of the minimum MSE. **(D)** Regression coefficients for features selected by LASSO model trained for predicting the percentage of Tregs out of total T-cells **(E)** Distribution of spearman correlation coefficients measuring the correlation between the percentage of Tregs and interaction scores computed using randomized receptor-ligand pairs. Each panel shows the distribution of spearman correlation coefficients conditioned on the ligand correlation. Red lines indicate the median value of ligand correlation used for each subplot. p-values indicate the probability that the spearman correlation of a random ligand-receptor pair with Treg percentage is larger than the correlation of the ligand with Treg percentage.