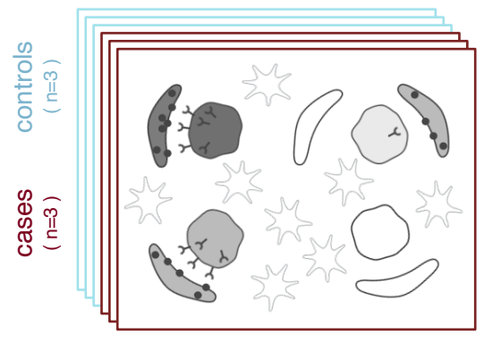
	ligrec_extra (non_curated) (#8350)	
	binary	complex
raw	6998	1352
deduplicate	6882	1351
split and create pair all-vs-all		8616
deduplicate		3272
check PPI (# 80.237)		1262
merge	7344	

curated (#5141)	
binary	complex
4384	757
4300	757
	4966
	2185
	852
4651	

Final (deduplicated)	6941

Supplemental Figure 1. Community ligand-receptor database

a Calculate general statistics



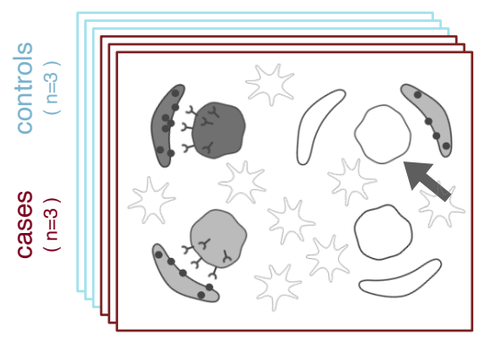
cell type	# cells	# active cells	expression	normalised expression
	5	3	4, 5, 7	2.5, 3.1, 4.4
	4	3	1, 5, 6	0.8, 4.2, 5.0
	8	-	-	
	total: 17			

b Set thresholds for cell types and LRP of interest

- minimum number of cells in a cell type
  e.g. > 3 => both cell types pass the threshold

- minimum number of active cells in a cell type
  e.g. > 1 => both cell types pass the threshold

- minimum normalised expression in a cell
  e.g. > 1.5 => one cell doesn't pass the threshold

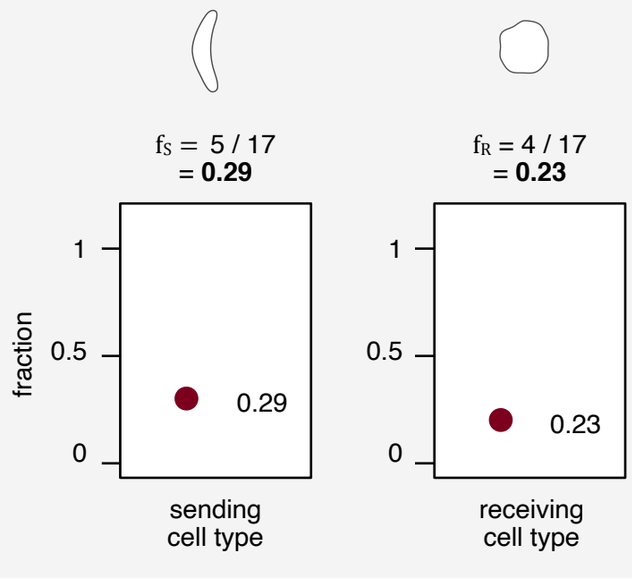


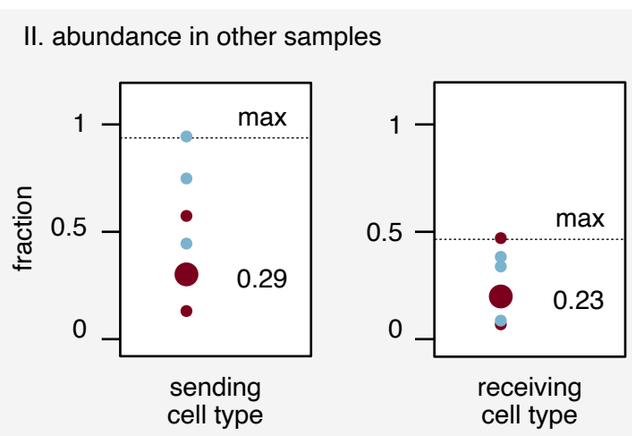
cell type	# cells	# active cells	expression	normalised expression
	5	3	4, 5, 7	2.5, 3.1, 4.4
	4	**2**	**5, 6**	**4.2, 5.0**
	8	-	-	
	total: 17			

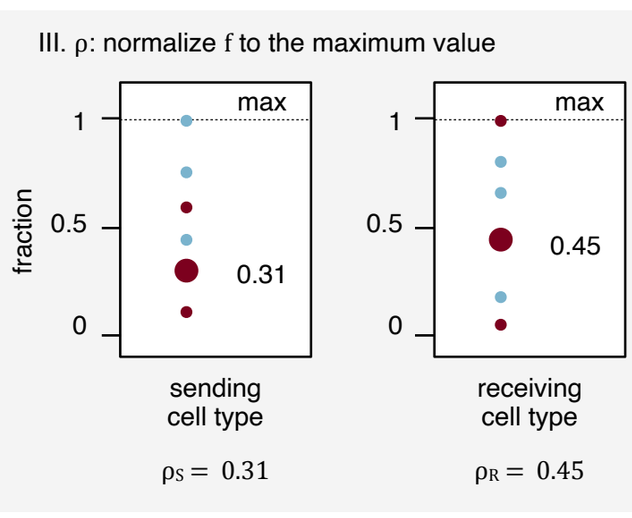
c Calculate individual components

ρ (relative abundance of cell type)

I. abundance (f) of sending and receiving cell types in sample of interest

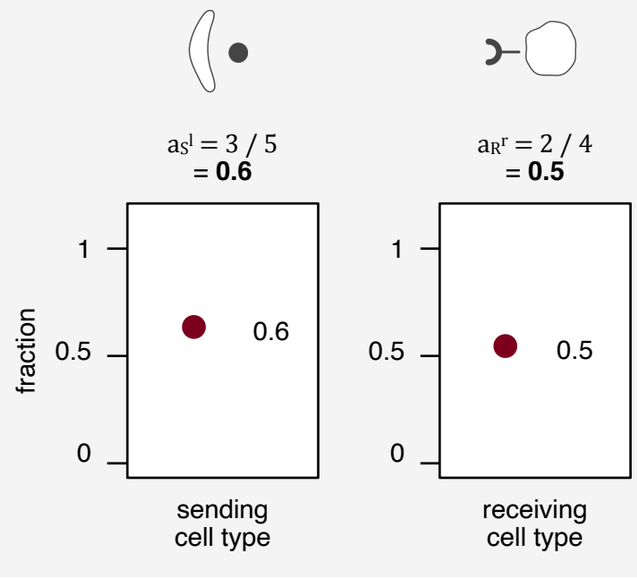


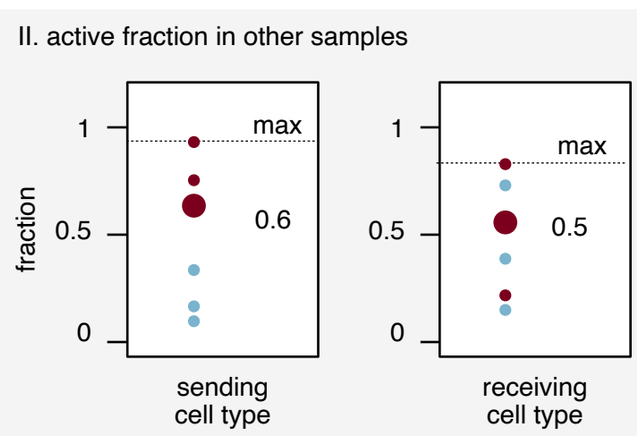


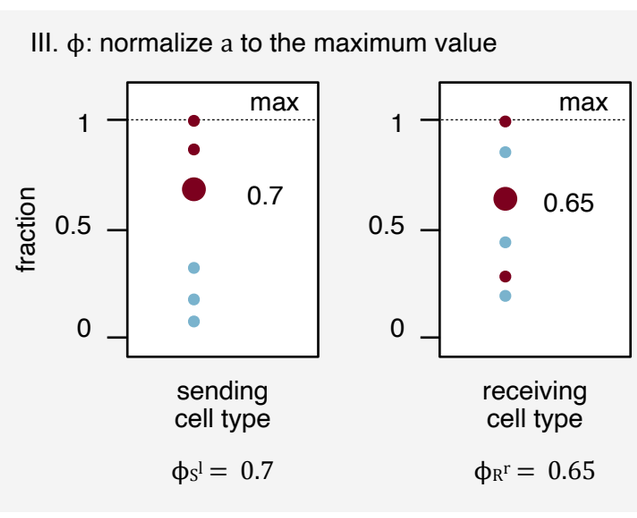


ϕ (relative active fraction)

I. active fraction (a) in sample of interest

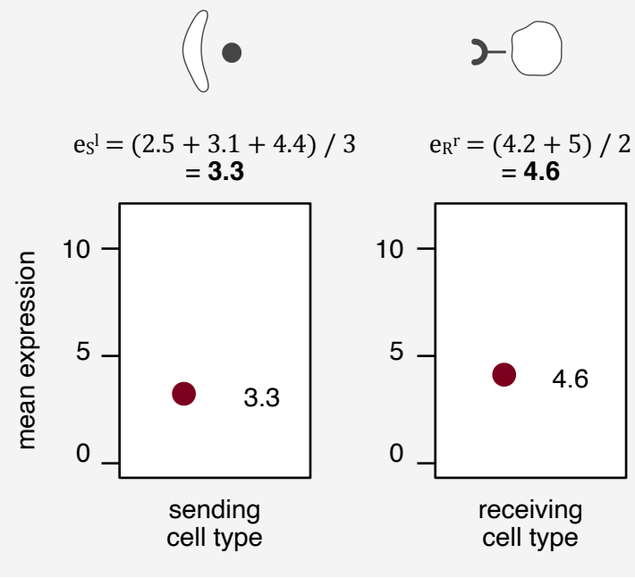


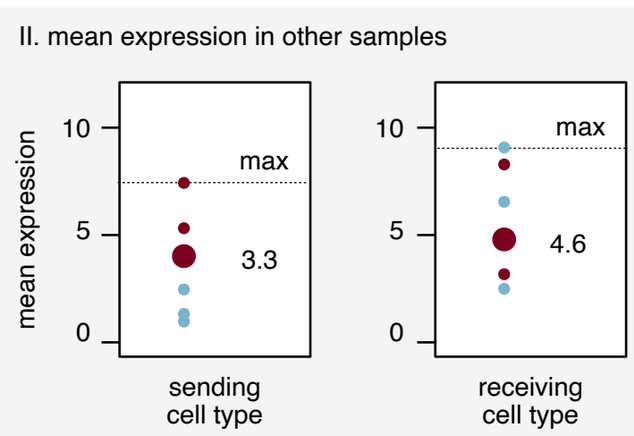


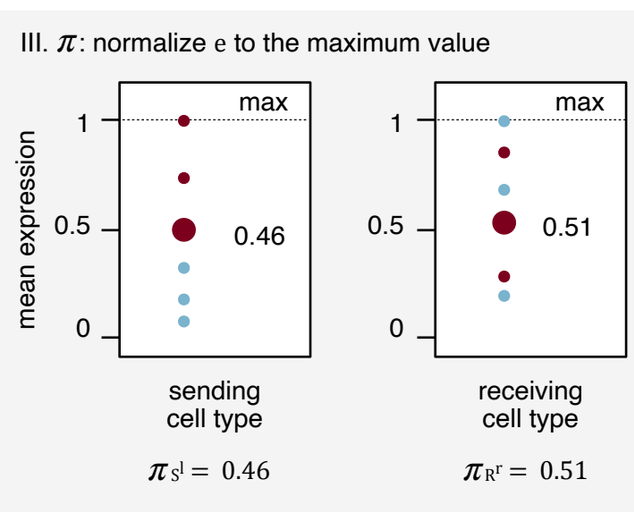


π (relative expression in the active fraction)

I. mean expression (e) in the active fraction in sample of interest





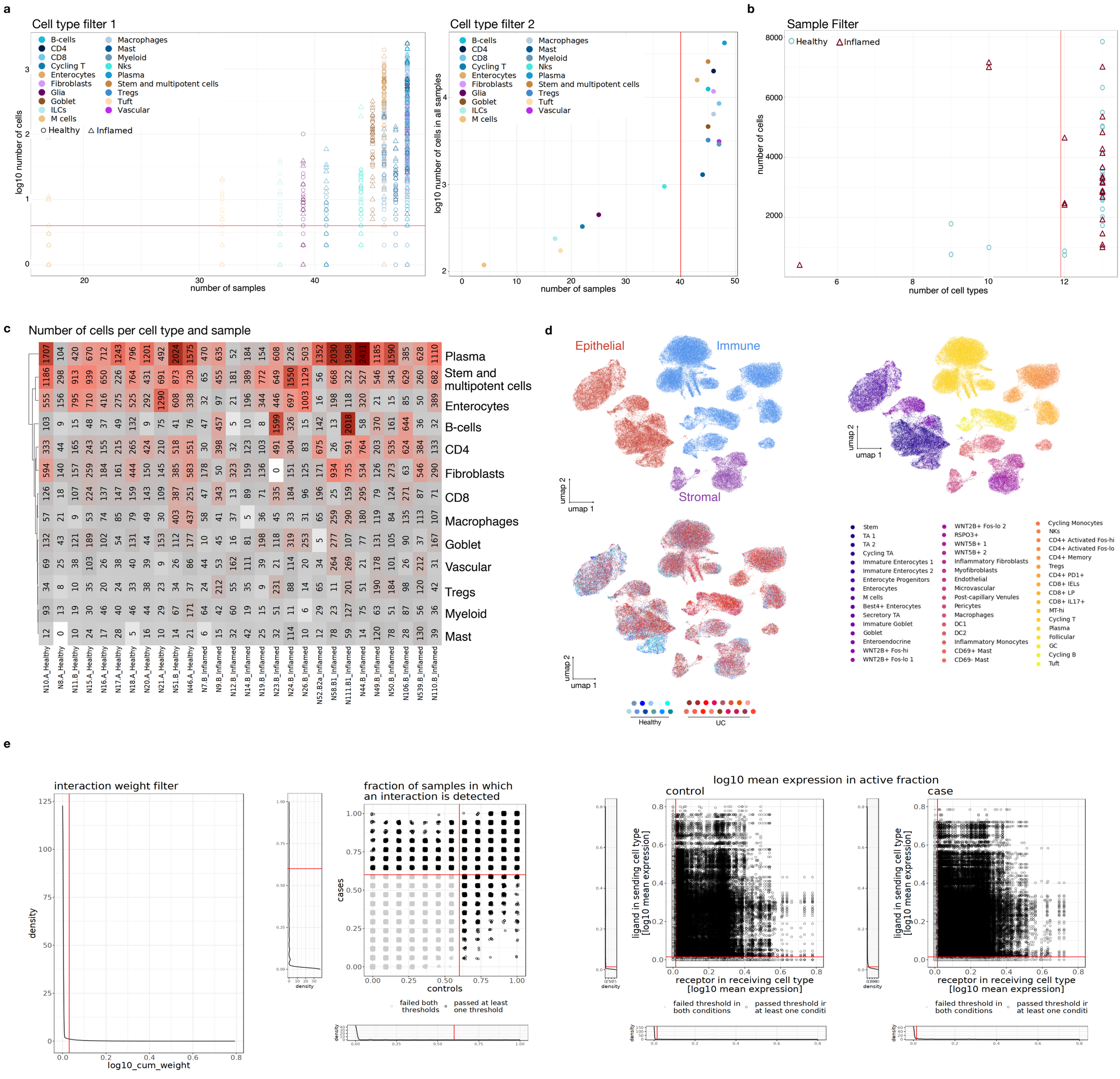


d Calculate the weight

$$w_{(S,R)_i}^{(l,r)_k} = \left(\rho_{S_i} \times \rho_{R_i}\right) \times \left(\phi_{S_i}^{l_k} \times \phi_{R_i}^{r_k}\right) \times \left(\pi_{S_i}^{l_k} \times \pi_{R_i}^{r_k}\right)$$

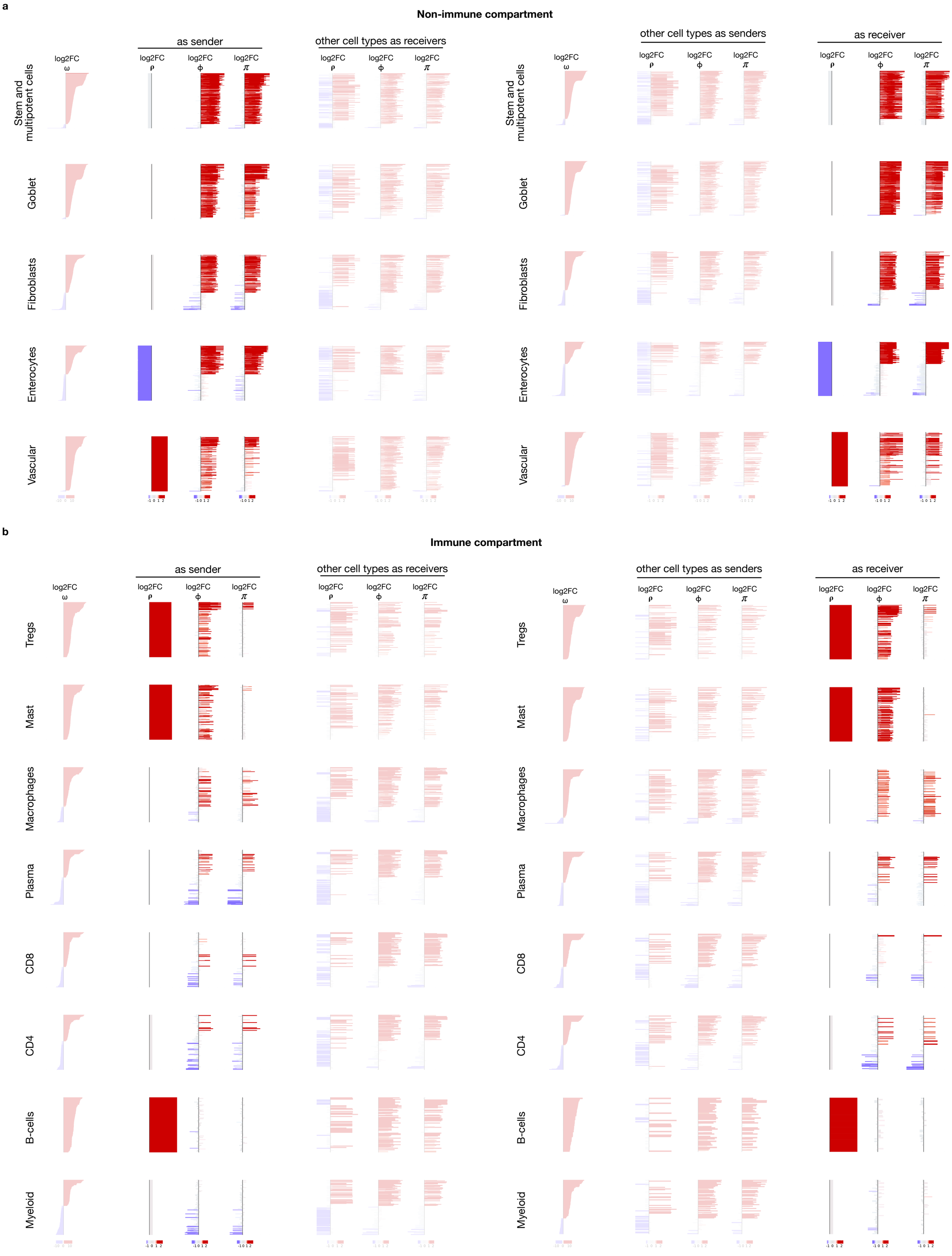
$$w = \left(0.31 \times 0.45\right) \times \left(0.7 \times 0.65\right) \times \left(0.46 \times 0.51\right) = 0.015$$

Supplemental Figure 2. Formula to calculate individual interaction weight between a sending cell type of interest expressing a ligand of interest and a receiving cell type of interest expressing a receptor of interest. a, Step 1: calculate general statistics for each cell type in each sample: number of cells, number of active cells (i.e. cell expressing the ligand or receptor of interest respectively). **b**, Step 2: set and apply thresholds: i) minimum number of cells in a cell type, ii) minimum number of active cells in a cell type, iii) minimum normalised expression in a cell. **c**, Step 3: calculate the relative cell type abundance, relative active fraction, and relative mean expression in the active fraction for the sending and the receiving cell types. **d**, Step 4: multiply all six value to get the interaction weight.

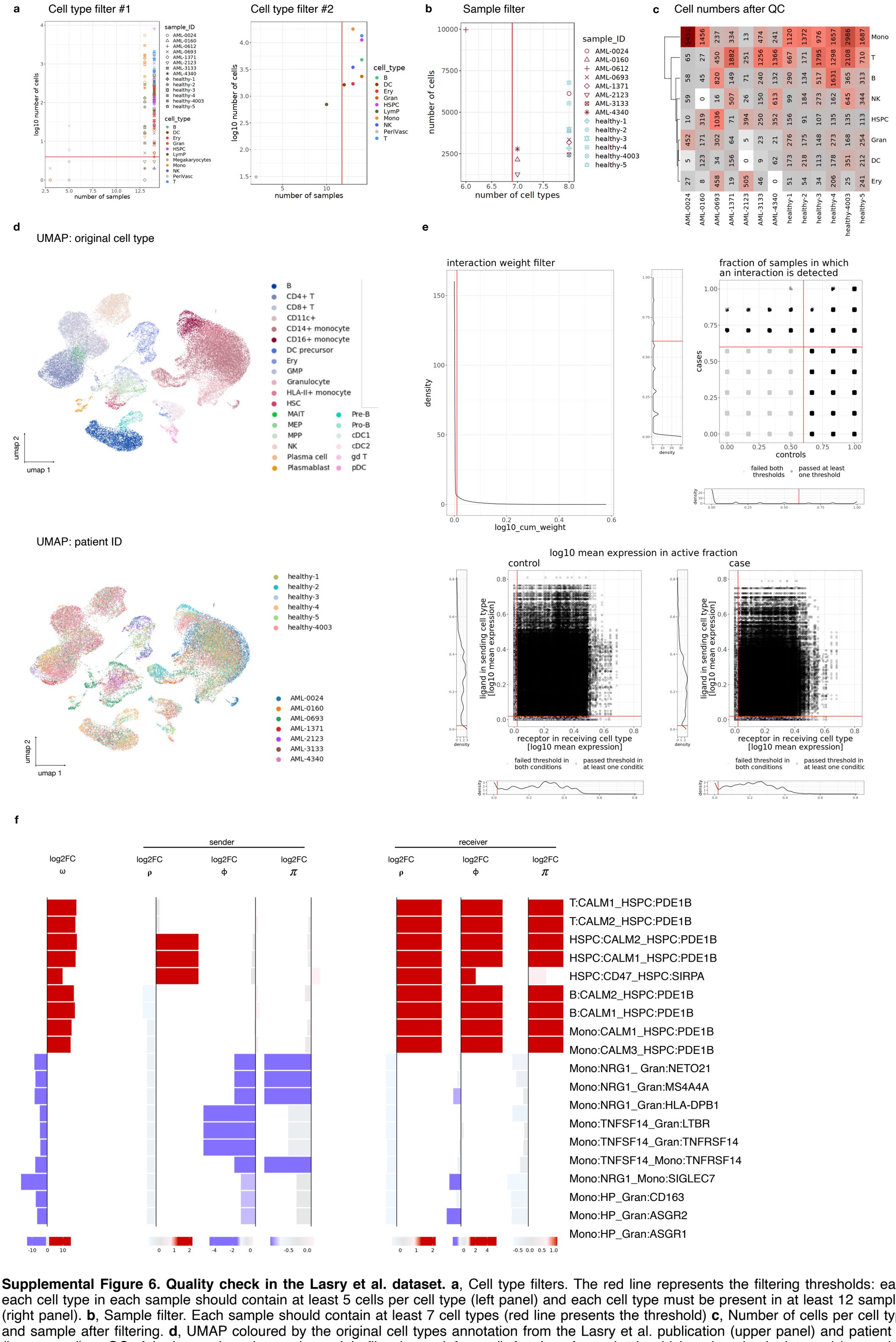


Supplemental Figure 3. Quality check in the Smillie et al. dataset. a, Cell type filters. The red line represents the filtering thresholds: each cell type in each sample should contain at least 5 cells per cell type (left panel) and each cell type must be present in at least 40 samples (right panel). **b,** Sample filter. Each sample should contain at least 12 cell types (red line presents the threshold). **c,** Number of cells per cell type and sample after filtering. **d,** UMAP visualization of the dataset after batch correction colored by tissue type (top left), original cell types (top right) and samples (bottom left). **e,** QC of the interactions: interaction weight filter (far left panel), fraction of samples in which an iteration is detected (middle left panel), mean expression of the ligands and the receptors in the active fraction of the corresponding cell types in the control cohort (middle right panel) and the case cohort (far right panel).

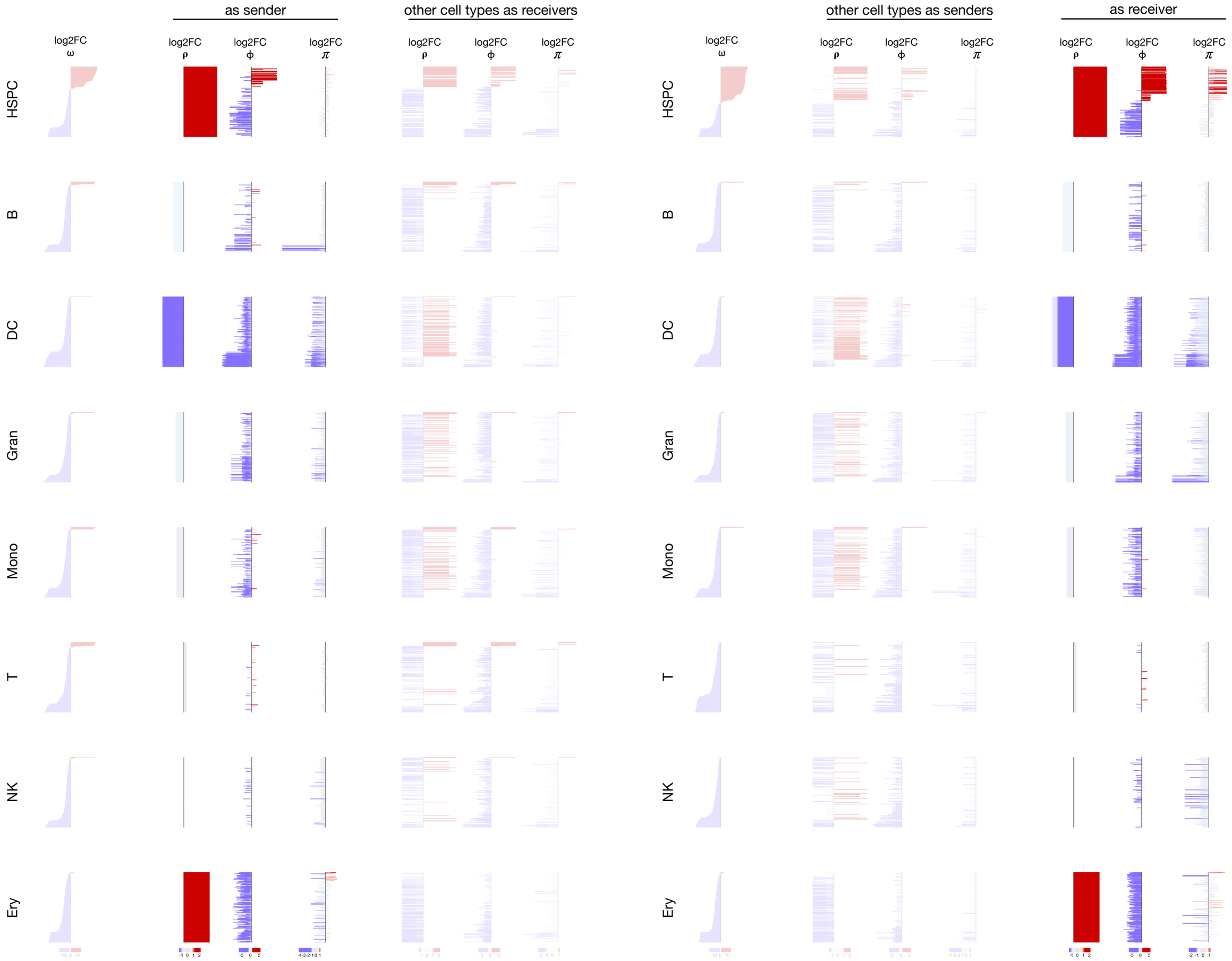
Supplemental Figure 4. Interaction analysis of Smillie et al. data a, Number of significant (p-val < 0.05) interactions per cell type, split depending on if the cell type acts as a sender, or a receiver and if the interactions are up- or downregulated. The colore indicates whether the counterpart cell types belong to the non-immune compartment (gold) or to the immune compartment (blue). **b,** Forest plot showcasing the individual components of significantly up- and downregulated interactions with *CD44* as a receiver. **c,** Bar plots showcase the distribution of the genes OSMR and IL6ST over all Fibroblasts. P-values were calculated using t-tests (p-val. <0.05), with an asterisk marking significance. The error bars were calculated with the python function sem() for the standard error of the mean. Bonferroni correction was performed. **d,** Forest plot showing the individual interactions between VEGFA on Stem and multipotent cells (left side) and FLT1/KDR on Vascular cells(right side), split by components. **e,** Dot-plot of *VEGFA, EPAS1, HRE complex and PHD family* expression in stem and multipoint cells and *FLT1, KDR, MAP3K and NOS3* expression in vascular cells. The color represents mean gene expression for a cell type per sample. Dot-size represents the fraction of the respective cell type in the cell-specific compartment. Statistics were done by performing t-tests. Significant findings (p-val. <0.05) are labeled with an asterisk.



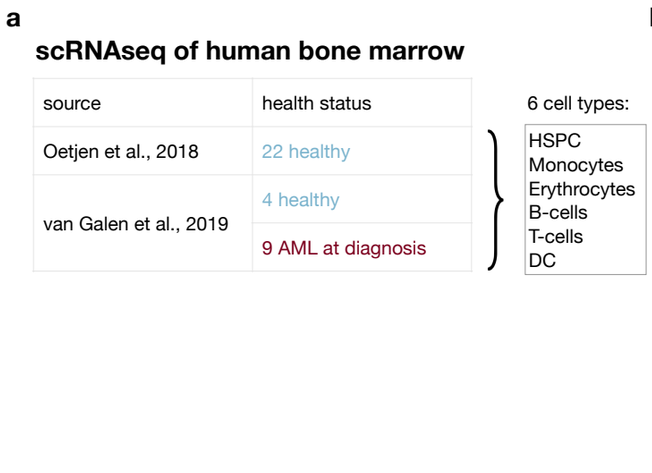
Supplemental Figure 5. Differential interactions in Smillie et al. data split by components and cell types The figure shows all significantly up- and downregulated interactions for the non-immune (**a)** and immune (**b)** cell types. For each cell type, two forest plots are plotted: a plot where the cell type acts as a sender (left part) or as a receiver (right part). Each forest plot contains log2 fold change representation for the individual components ρ, ϕ and π , as well as the total interaction weight w.

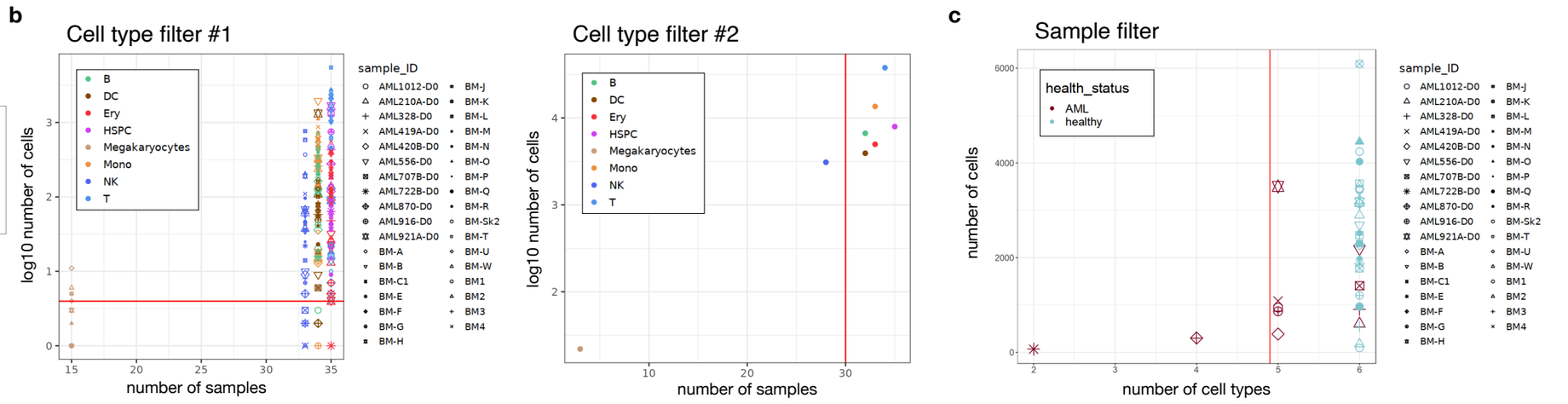


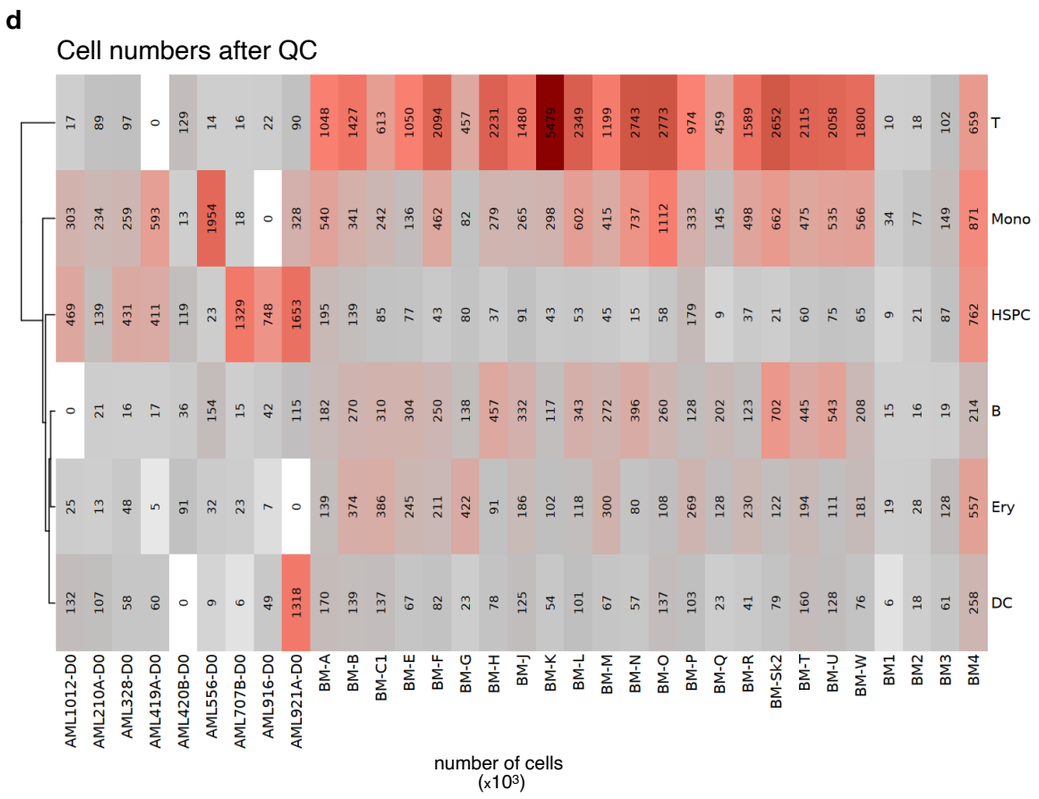
Supplemental Figure 6. Quality check in the Lasry et al. dataset. a, Cell type filters. The red line represents the filtering thresholds: each each cell type in each sample should contain at least 5 cells per cell type (left panel) and each cell type must be present in at least 12 samples (right panel). **b**, Sample filter. Each sample should contain at least 7 cell types (red line presents the threshold) **c**, Number of cells per cell type and sample after filtering. **d**, UMAP coloured by the original cell types annotation from the Lasry et al. publication (upper panel) and patient ID (lower panel). **e**, QC of the interactions: interaction weight filter (upper left panel), fraction of samples in which an iteration is detected (upper right panel), mean expression of the ligands and the receptors in the active fraction of the corresponding cell types in the control cohort (lower left panel) and the case cohort (lower right panel). **f**, Forest plot showcasing the individual components of top up-and down-regulated interactions.

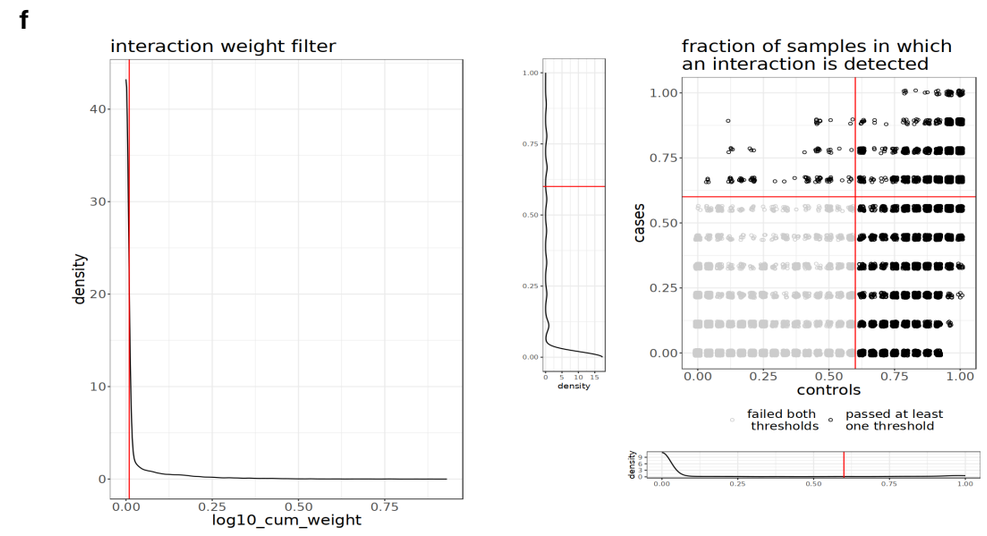


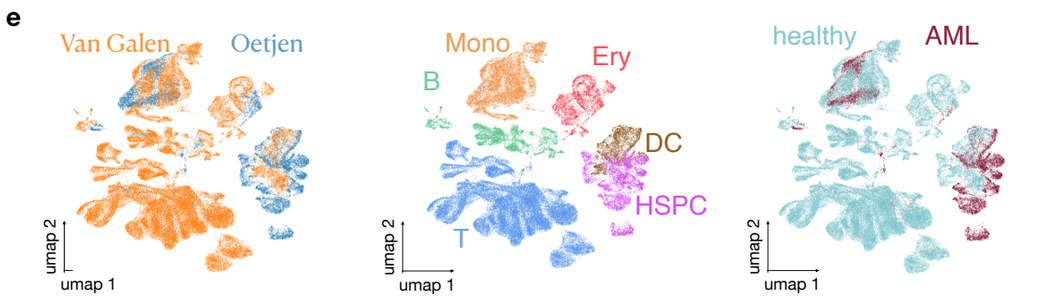
Supplemental Figure 7. Interactions in Lasry et al. data split by components and cell types The figure shows all significantly up- and downregulated interactions for all cell types. For each cell type, two forest plots are plotted: a plot where the cell type acts as a sender (left part) or as a receiver (right part). Each forest plot contains log2 fold change representation for the individual components $\boldsymbol{\rho}$, ϕ and π , as well as the total interaction weight w.

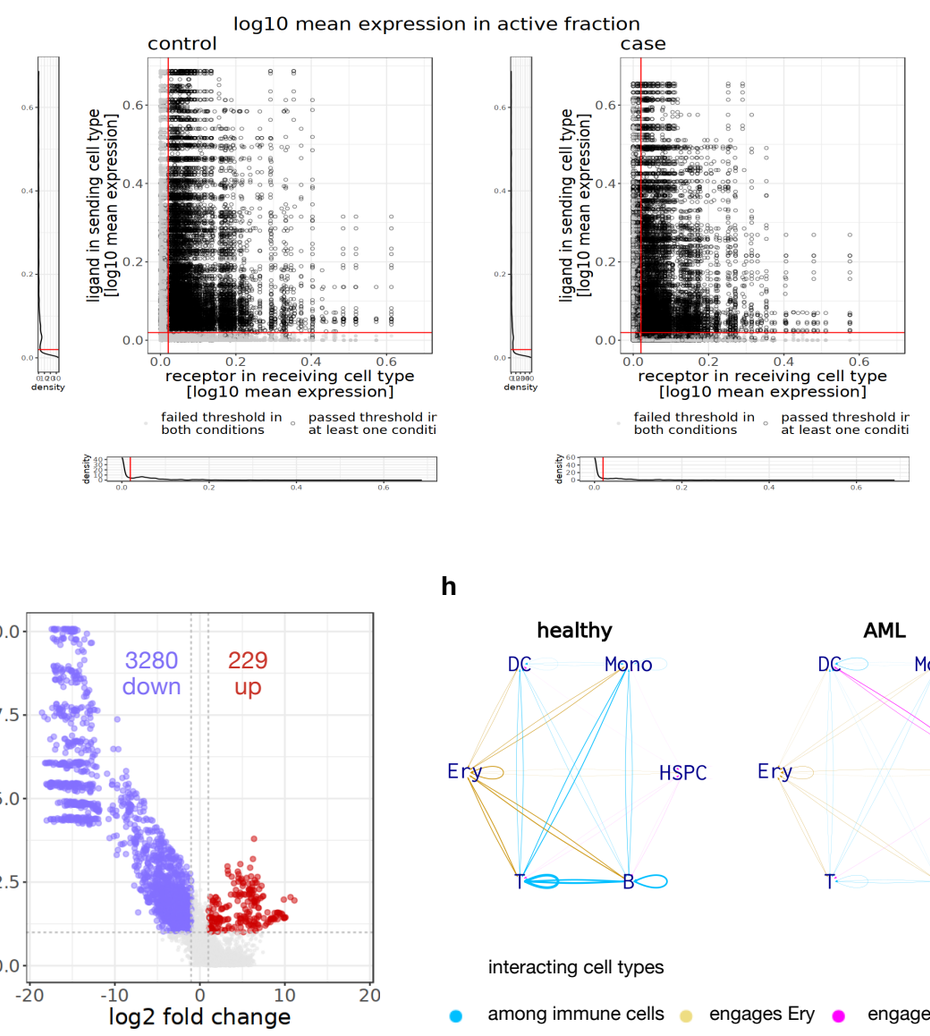


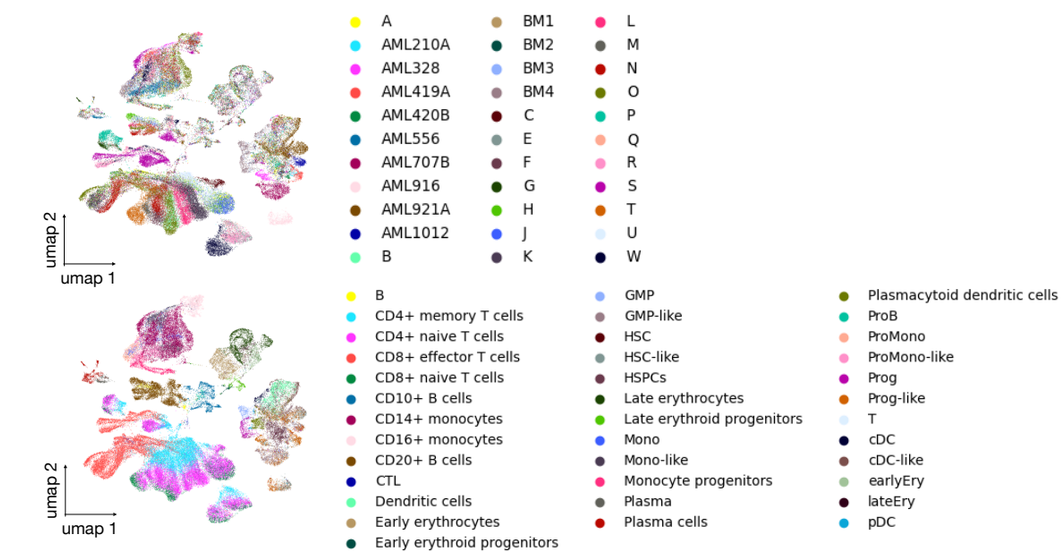






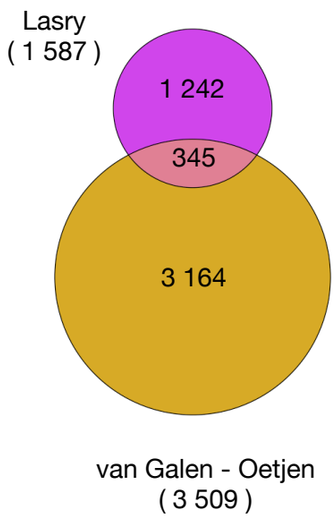




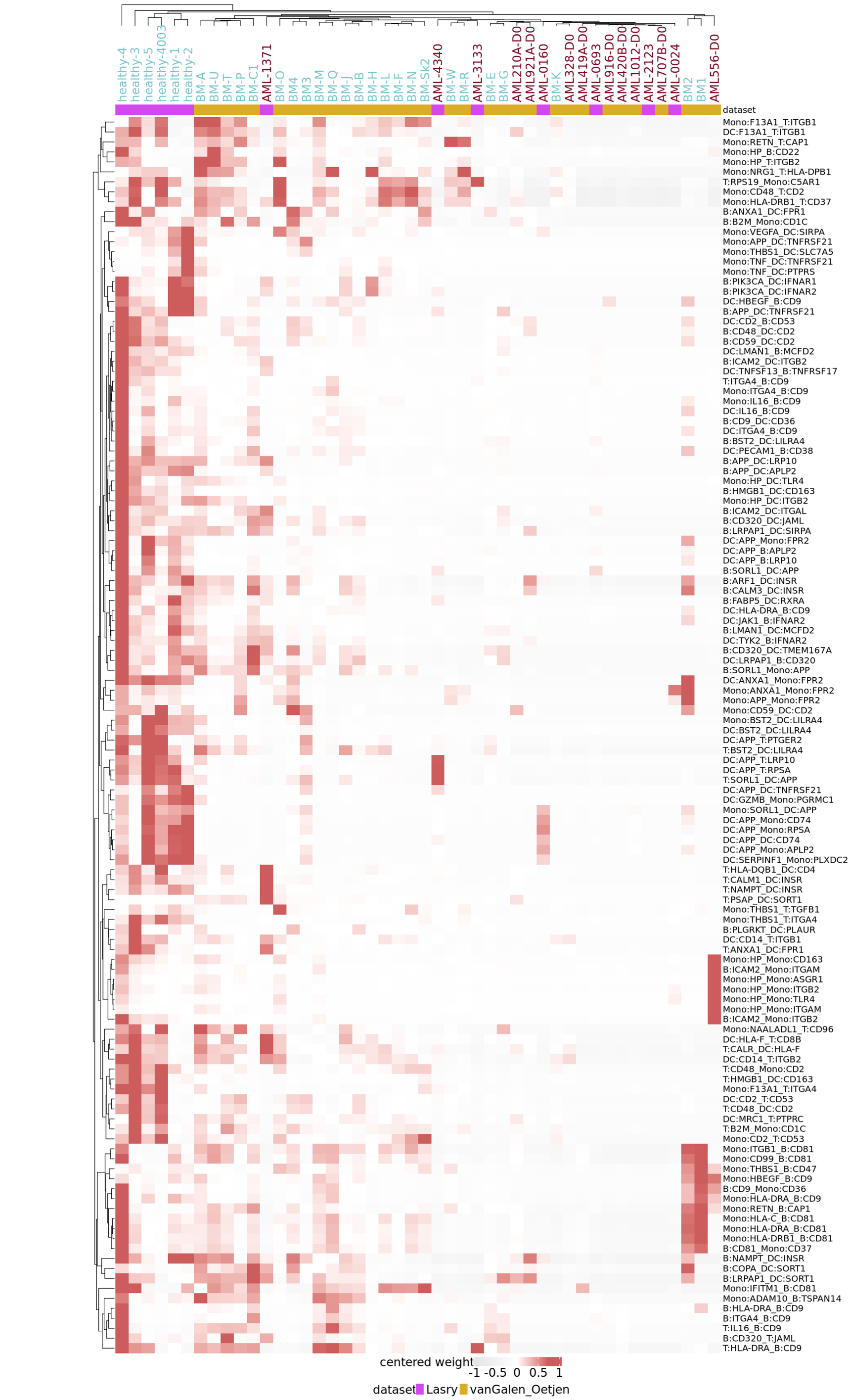


Supplemental Figure 8. Quality check in the integrated van Galen et al. - Oetjen et al. dataset. **a**, Overview of both scRNAseq datasets used for human bone marrow analysis. The table summarizes the source of the datasets, health status of the samples, and the 6 merged cell types. **b**, Cell type filters. The red line represents the filtering thresholds: each cell type in each sample should contain at least 5 cells per cell type (left panel) and each cell type must be present in at least 30 samples (right panel). **c**, Sample filter. Each sample should contain at least 5 cell types (red line presents the threshold). **d**, Number of cells per cell type and sample after filtering. **e**, UMAP visualization of the dataset after batch correction colored by dataset (top left panel), cell type (top center panel), health status (top right panel), samples (middle panel) and original cell types (bottom panel). **f**, QC of the interactions: interaction weight filter (upper left panel), fraction of samples in which an iteration is detected (upper right panel), mean expression of the ligands and the receptors in the active fraction of the corresponding cell types in the control cohort (lower left panel) and the case cohort (lower right panel). **g**, Volcano plot of differential interactions (AML vs healthy). **h**, Network graph of differential interactions of healthy (left panel) and AML (right panel) cohort. One line represents cumulative weight of all differential interactions between the two connected cell types.

a



b

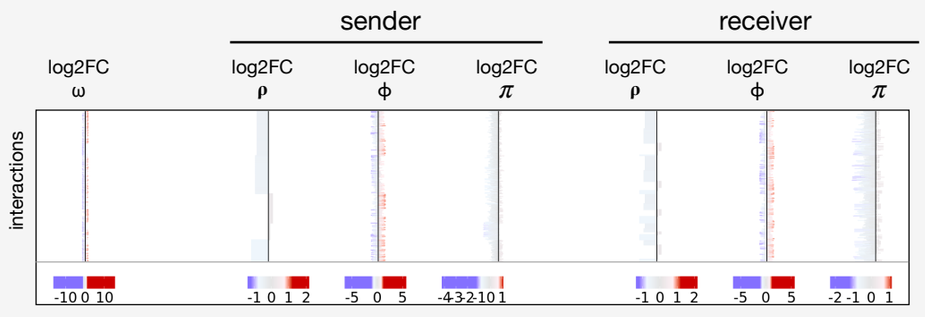


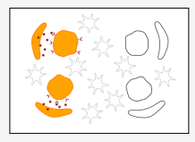
Supplemental Figure 9. Overlap between differential interactions of the Lasry (granulocytes and NKs excluded) and the integrated van Galen et al. - Oetjen et al. datasets. **a**, Euler diagram of the differential interactions of Lasry et al and integrated van Galen et al. - Oetjen et al. **b**, Heatmap of the 124 overlapping differential interactions among the immune cell types.

$$w_{(S,R)_i}^{(l,r)_k} = \underbrace{\left(\rho_{S_i} \times \rho_{R_i}\right) \times \left(\phi_{S_i}^{l_k} \times \phi_{R_i}^{r_k}\right) \times \left(\pi_{S_i}^{l_k} \times \pi_{R_i}^{r_k}\right)}_{\text{6 components}}$$

stable overall communication

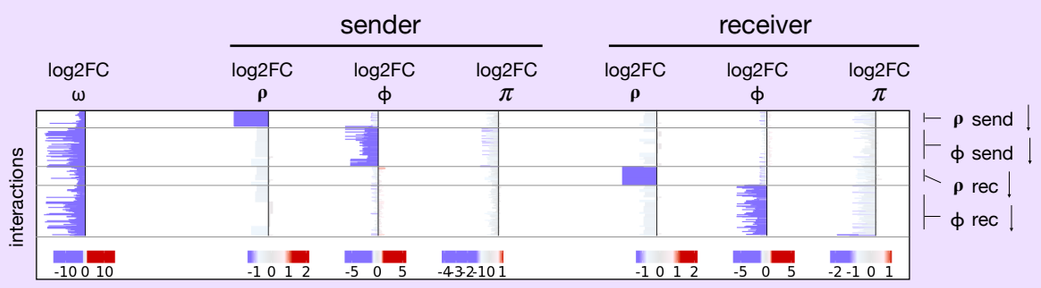
1. none of the components change ⟶ **no change**

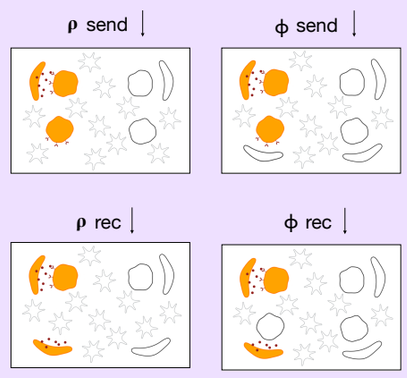




increase / decrease in communication

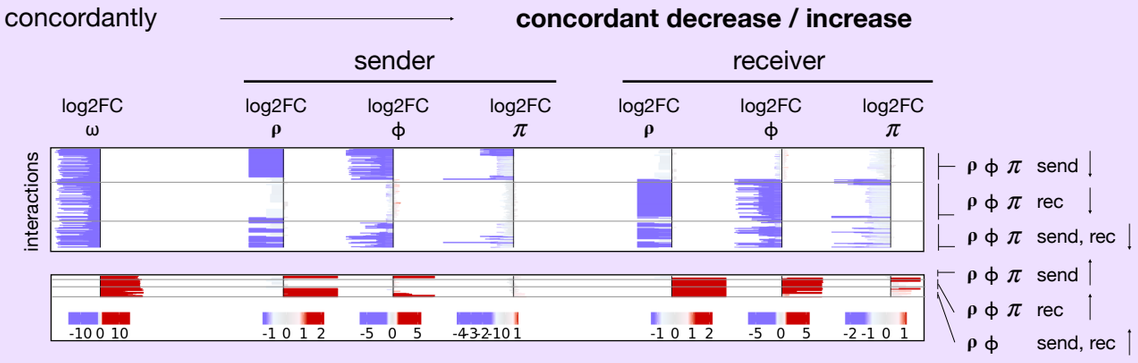
2. one component changes ⟶ **simple decrease / increase**

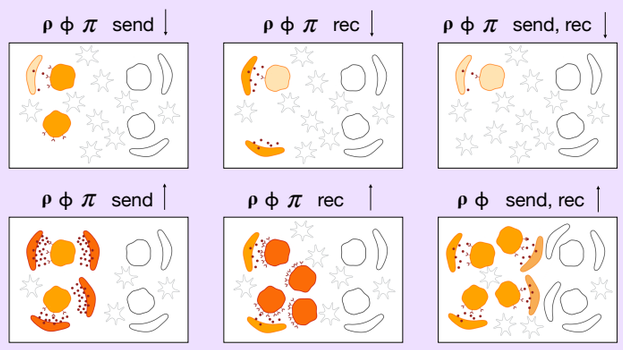




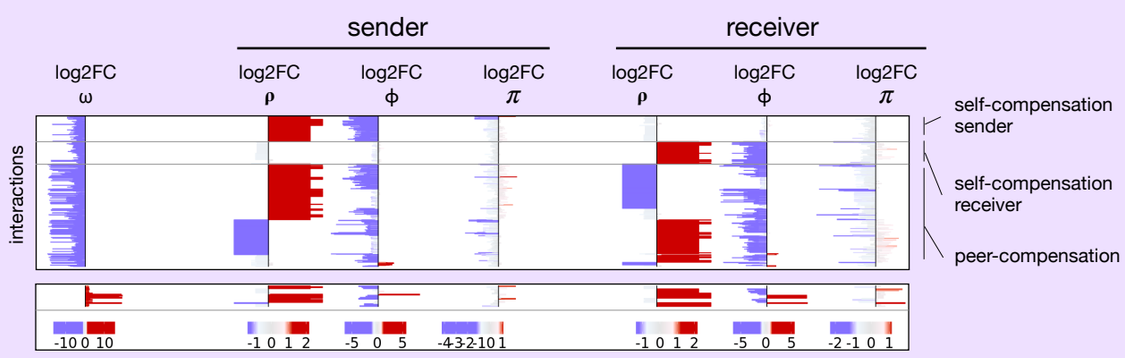
3. several components change

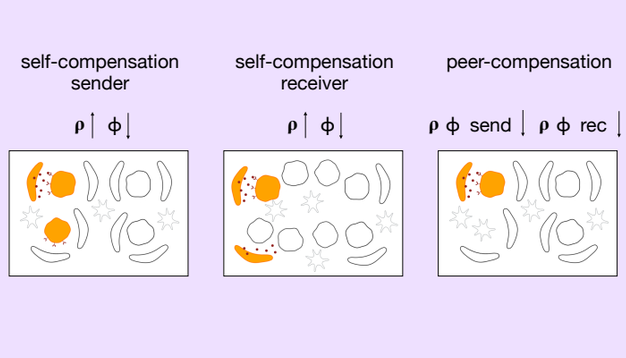
concordantly ⟶ **concordant decrease / increase**





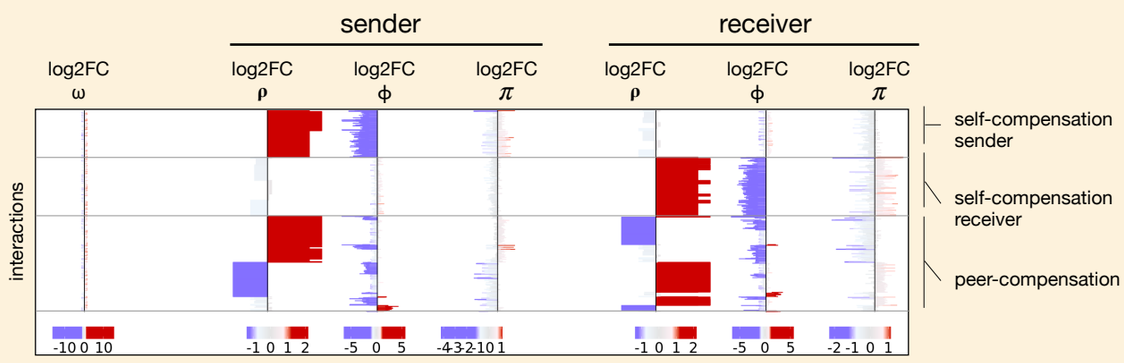
disconcordantly ⟶ **insufficient compensation**

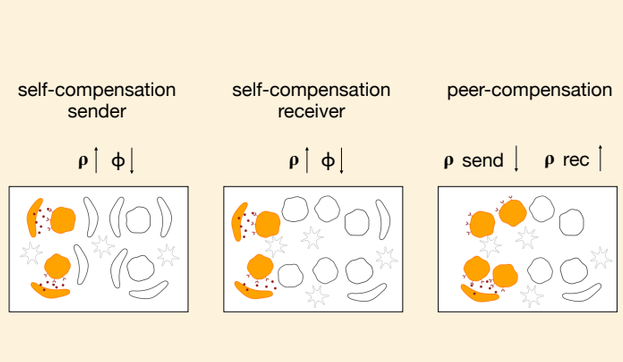




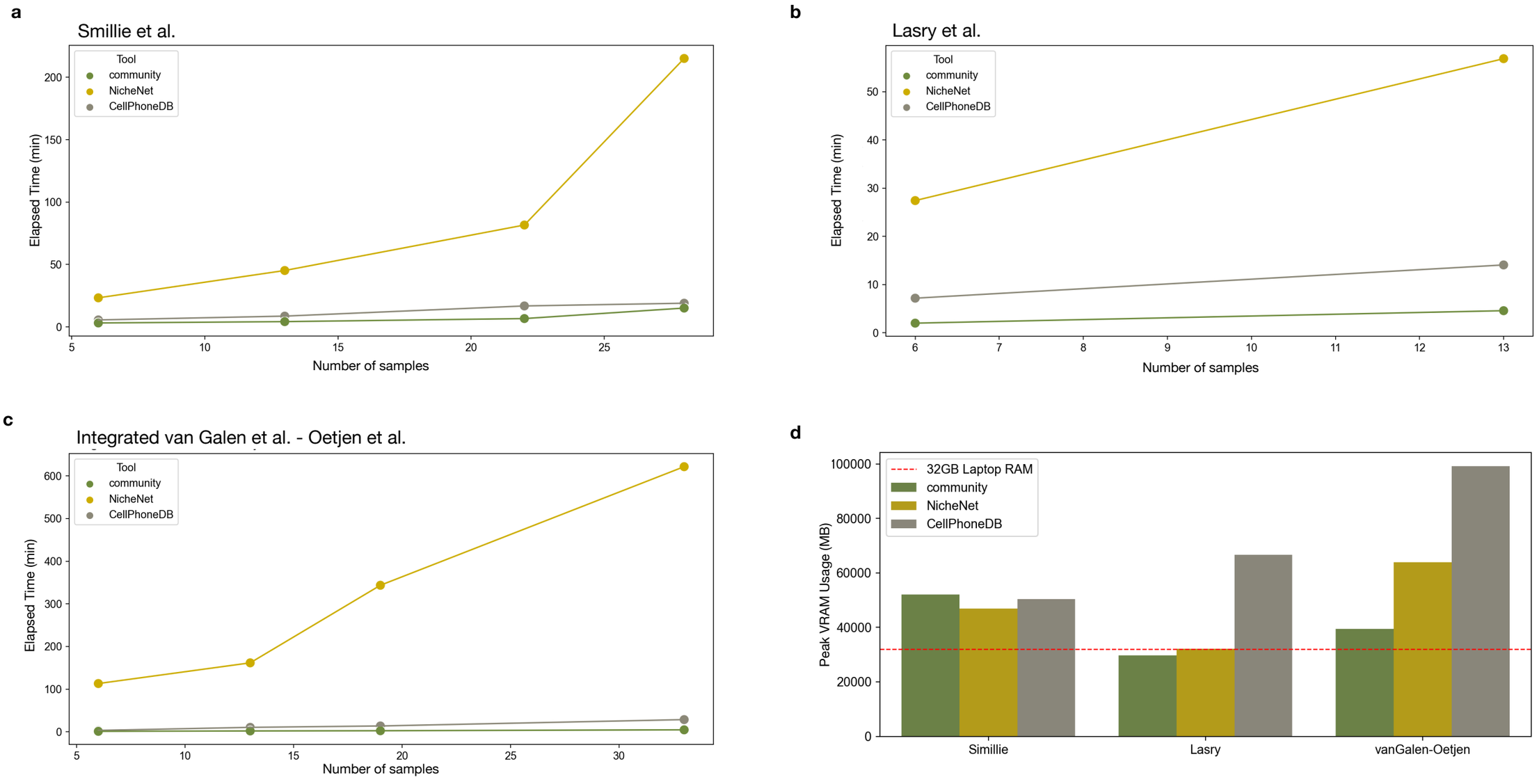
stable overall communication

sufficient compensation





Supplemental Figure 10. Interplay of individual components shapes the resulting communication shift



Supplemental Figure 11. Resource usage. a-c,Benchmarking of elapsed time for the *community*, NicheNet and CellPhoneDB tools across the three datasets with with increasing number of samples: Smillie et al. (**a**), Larry et al. (**b**), and integrated van Galen et al. - Oetjen et al. (**c**) **d**, Comparison of peak RAM usage among the tools and across three full datasets: 22 samples for Smillie et al., 13 samples for Lasry et al., and 33 samples for the integrated van Galen-Oetjen datasets. The dashed red line indicates the capacity of 32 GB, a specification often seen in laptops used for data-intensive tasks.