MultiNicheNet parameter interpretation guidelines

Parameter	Consequences of being more lenient	Consequences of being more stringent	Notes and recommendations
Sample filtering: ***min_cells = 10*** For each cell type, the considered samples are those samples with nr. of cells >= *min_cells* parameter	You can be more lenient by decreasing this value. This will keep more samples per cell type in the analysis, and thus possibly more cell types as well. However, DE analysis-based results and prioritization criteria may be less trustworthy if based on several samples with a low nr of cells. Based on the mock analysis, the potential disadvantage of being lenient is likely to be limited (if not too extreme).	You can be more stringent by increasing this value. This can lead to a loss of cell types relevant to the condition of interest. For example, cell types that are more abundant in the condition of interest can be left out from analysis if not sufficiently present in the steady-state condition. Based on the mock analysis, being more stringent will probably not lead to improved prioritization of interactions between abundant cell types.	For datasets with several lowly abundant cell types of interest, we recommend using *min_cells = 5*. We explicitly **recommend against using *min_cells < 5* and *min_cells > 50*.**
Gene filtering - sample proportion: ***min_sample_prop = 0.50*** For each cell type, we consider genes expressed if they are expressed in at least *min_sample_prop* fraction of samples in the smallest condition.	You can be more lenient by decreasing this value. Based on the mock analysis, the influence of decreasing this parameter is limited.	You can be more stringent by increasing this value. Based on the mock analysis, the influence of increasing this parameter is limited.	We recommend using the default value.
Gene filtering - cell proportion: ***fraction_cutoff = 0.05*** Genes are considered to be expressed in a sample if they have non-zero expression values in *fraction_cutoff* fraction of cells of that cell type.	You can be more lenient by decreasing this value. This will lead to a higher probability of discovering weakly-expressed interactions. The potential downside is that condition-specific strongly-expressed interactions may be more hidden among all the weakly-expressed interactions while exploring the top predictions.	You can be more stringent by increasing this value. This will lead to risking the loss of condition-specific but weakly expressed interactions. Based on the mock analysis, the potential benefit of better prioritizing strongly expressed interactions is likely to be limited.	We recommend using the default value, unless you are specifically interested in prioritizing (very) weakly expressed interactions. We explicitly **recommend against using *fraction_cutoff > 0.10***.

Thresholds before the ligand activity analysis: define the geneset_oi per cell type: ****logFC_threshold = 0.50*** ****p_val_adj = FALSE*** ****p_val_threshold = 0.05*** These thresholds are used to define the up- and downregulated genes per cell type/contrast combination. These genesets are then used as input for the ligand activity predictions.	You can be more lenient by decreasing the logFC_threshold and keeping *p_val_adj = FALSE*. This will increase the size of the genesets of interest. As a result, the results of the ligand activity analysis can change.	You can be more stringent by increasing the *logFC_threshold*, and changing *p_val_adj = TRUE*. This will decrease the size of the genesets of interest. As a result, the results of the ligand activity analysis can change.	We recommend inspecting the nr. of DE genes for all cell types based on the default thresholds and adapting accordingly. **The ratio geneset_oi/background should ideally be between 0.005 and 0.1**. If the smallest group >= 20 samples, we typically recommend using *p_val_adj = TRUE*. If the biological difference between the conditions is very large, we typically recommend increasing the *logFC_threshold* and/or using *p_val_adj = TRUE*.
The nr. of considered top-predicted prior target genes of each ligand: ***top_n_target = 250*** For target gene inference during visualizations and construction of the intercellular regulatory network, we only consider the *top_n_target* genes with the highest regulatory potential scores as targets of a ligand.	You can be more lenient by increasing this value. This will consider more genes as potential target genes of a ligand. However, some considered target genes could be supported by less prior knowledge.	You can be more stringent by decreasing this value. This will consider fewer genes as potential target genes of a ligand. But, the considered target genes are supported by more prior knowledge.	We recommend users to test other settings in case they would be interested in exploring fewer, but more confident target genes, or vice versa.

Parameter	Consequences of different parameter settings	Notes and recommendations
ligand_activity_down	*ligand_activity_down = TRUE: For prioritization based on ligand activity: consider the maximum of up- and downregulation* Both LR pairs with high upregulatory and downregulatory activity will be strongly prioritized. *ligand_activity_down = FALSE: consider only upregulated activity* Only LR pairs with high upregulatory activity will be strongly prioritized. Because ligand activity is only one criterion, LR pairs with high downregulatory activity may still be prioritized if they score high on the other criteria.	The setting "*ligand_activity_down = FALSE" is more complete, but* the setting "*ligand_activity_down = FALSE" results in more interpretable results. In any case, we would rather suggest users validate LR pairs with high upregulatory activity (= target gene enrichment in the condition of interest) compared to high downregulatory activity (= target gene enrichment in the other condition(s)).*
scenario	*scenario = "regular": The setting "regular" is strongly recommended and gives each criterion equal weight.* *scenario = "lower_DE": It halves the weight for DE criteria, and doubles the weight for ligand activity.* This setting will emphasize interactions with higher ligand activity compared to strong differential expression. *scenario = "no_frac_LR_expr": It does not consider the criterion that represents the fraction of samples with sufficient expression of the L-R pair.* This setting will result in prioritized interactions that may be strongly differentially expressed and active in general, but not expressed in all samples.	The setting "lower_DE" is recommended in case you hypothesize that the differential CCC patterns in your data are less likely to be driven by DE (eg in cases of differential migration into a niche) no_frac_LR_expr: if you have a limited nr of samples in your conditions (< 10) and you don't want to penalize interactions if they are not sufficiently expressed in some samples.