Input files

RaCInG needs at least four input files to function. Two of these files define the quantifications, and two others define the compatibility matrices. A fifth optional file can be provided: the sign file for ligand receptor interactions. In this document, we will provide the structure of the csv input files.

Cell-type compatibility with ligand type

	Ligand 1 name	**Ligand 2 name**	**Ligand 3 name**	…
Cell type 1	0	1	1	…
Cell type 2	1	0	0	…
Cell type 3	0	1	0	…
⋮	⋮	⋮	⋮	

Notes:

- Values should be 1 (secretion possible) or 0 (secretion impossible).

Cell-type compatibility with receptor type

	Receptor 1	**Receptor 2**	**Receptor 3**	…
Cell type 1	0	1	1	…
Cell type 2	1	0	0	…
Cell type 3	0	1	0	…
⋮	⋮	⋮	⋮	

Notes:

- Cell type identifiers should be the same as in the previous csv-file.
- Values should be 1 (secretion possible) or 0 (secretion impossible).

Cell-type quantification

	Cell type 1	**Cell type 2**	**Cell type 3**	…
Patient 1	0.143527	0.2314527	0.135624	…
Patient 2	0.1678542	0.3214521	0.0953214	…
Patient 3	0.1137845	0.264512	0.0512389	…
⋮	⋮	⋮	⋮	

Notes:

- Patient rows should sum up to one. Code ensure that even if this is not the case, the final matrix after reading it in does have this property. Input values are then transformed such that rows sum up to one, and relative cell type value compared with the row sum remains the same.
- Cell type identifiers should be the same as in the previous csv-files.

Interaction weights

	Lig_Rec 1	**Lig_Rec 2**	**Lig_Rec 3**	…

Patient 1	5.3975	3.4678	10.3457	…
Patient 2	0.7913	5.1379	1.3974	…
Patient 3	2.39456	1.37945	0.12345	…
⋮	⋮	⋮	⋮	

Notes:

- Column identifiers should consist of a ligand identifier, followed by an underscore and a receptor identifier. For example, “ITGB2_ICAM2”.
- Patient identifiers should be the same as in the previous csv-file.
- Ligand and receptor identifiers should be the same as in the previous csv-files.

Sign interactions (optional)

	Receptor 1	**Receptor 2**	**Receptor 3**	…
Ligand 1	0	-1	1	…
Ligand 2	1	0	-1	…
Ligand 3	0	-1	0	…
⋮	⋮	⋮	⋮	

Notes:

- Ligand identifiers should be the same as in the previous csv-files.
- Receptor identifiers should be the same as in the previous csv-files.
- Values should be 1 (promoting), -1 (inhibiting) or 0 (unknown/nonexistent).

Property .csv files

	Property 1	**Property 2**	**Property 3**	…
Patient 1	0.5348	2.13458	1.48632	…
Patient 2	1.02365	7.26589	0.314578	…
Patient 3	2.31546	10.1245	1.23564	…
⋮	⋮	⋮	⋮	

Notes:

- Patient identifiers are the same as in the previous files.
- Properties consist of cell types separated by underscores.
- Only one type of property per file (e.g. only wedges).

Feature .txt files (triangles)

```
Triangle_type

Weight_type,number_patients,number_cells,number_graphs,average_deg

Patient_number,number_cells,average_deg

['cell1' 'cell2' 'cell3' …]

Count,total_number_feature_average,total_number_feature_std

Composition – Average:
```

```
0,0,0,number_this_subfeature_average
0,0,1,number_this_subfeature_average
#Etc until all sub features are done.
Composition – Std:
0,0,0,number_this_subfeature_std
0,0,1,number_this_subfeature_std
#Etc until all sub features are done.
Patient_number,number_cells,average_deg
['cell1' 'cell2' 'cell3' …]
Count,total_number_feature_average,total_number_feature_std
Composition – Average:
0,0,0,number_this_subfeature_average
0,0,1,number_this_subfeature_average
#Etc until all sub features are done.
Composition – Std:
0,0,0,number_this_subfeature_std
0,0,1,number_this_subfeature_std
#Etc until all sub features are done.
#Etc until all patients are done.
```

Notes:

- Number_cells is the number of cells per generated graph
- The count line gives the total count of the feature (not subdivided into specific cell types).
- In the composition lines the numbers correspond to the cell types in the cell-type array.
- This is the structure used for only wedges, trust triangles and cycle triangles.
- An enter at the end of the file is required.
- Filename: $cancer_$feature_$average(_norm).txt

Feature .txt files (limiting direct communication)

```
Weight_type,cancer_type,number_patients
Patient_number
['cell1' 'cell2' 'cell3' …]
0,0,limiting_number_this_subfeature
0,1,limiting_number_this_subfeature
#Etc until all sub features are done.
```

```
Patient_number
['cell1' 'cell2' 'cell3' …]
0,0,limiting_number_this_subfeature
0,1,limiting_number_this_subfeature
#Etc until all sub features are done.
#Etc until all patients are done.
```

Notes:

- Information about average degree and number of cells is not needed, because we consider limiting values.
- There is no separation between average and standard deviation, because there one deterministic limiting value for each feature.
- This is the structure used for only direct communication.
- An enter at the end of the file is required.
- Filename: $cancer_D(_norm).txt