SUPPLEMENTARY 1

General information

PhysiSpatial is a workflow which combines different packages and analysis in Python and R to generate all the files needed to have your spatial transcriptomics data into a simulation in PhysiCell/PhysiBoSS. Moreover, it generates the personalized boolean models using the Fumia Model of Cancer. In future updates we will implement other Boolean Models to be personalized as well as more parameters to customize your simulations.

Requisites

- .h5 file with the RNA expression information
- Folder called **''spatial''** with the **images** (png), the **position list** (csv),and the **scalefactors** (json), all usually provided in the database.
- A folder called **''PROFILE-master''** with the tool PROFILE which you can download here: https://github.com/ArnauMontagud/PROFILE_v2

How to Run

Run the shell script ***PhysiSpatial_setup.sh***

Scripts Explanation

clustering_ESTIMATE.py

With this script we do the analysis with scanpy of the spatial transcriptomics data (you can customize the parameters, explained in the script). It generates as output the file: ''input_Ensembl2Entrez.csv'' and a first clustering process using Leiden. This script changes the Ensemble notation to Entrez which is the one used by the R package ESTIMATE.

Ensembl2Entrez.R

The purpose of this script is to change the Ensembl notation to Entrez which is a requirement of some packages we use. This script is used at some points with little changes.

ESTIMATE_3kmeans.R

This script uses the package ESTIMATE to generate 4 scores (immune score, stromal score, ESTIMATE score and tumor purity). With those scores we perform a clustering using kmeans algorithm. In this first version we decided to cluster into 3 clusters but in further updates we will improve this but Silhouette, Elbow and Gap statistic methods for cluster optimization agree with the number of clusters between 2-4 (Fig 1S).

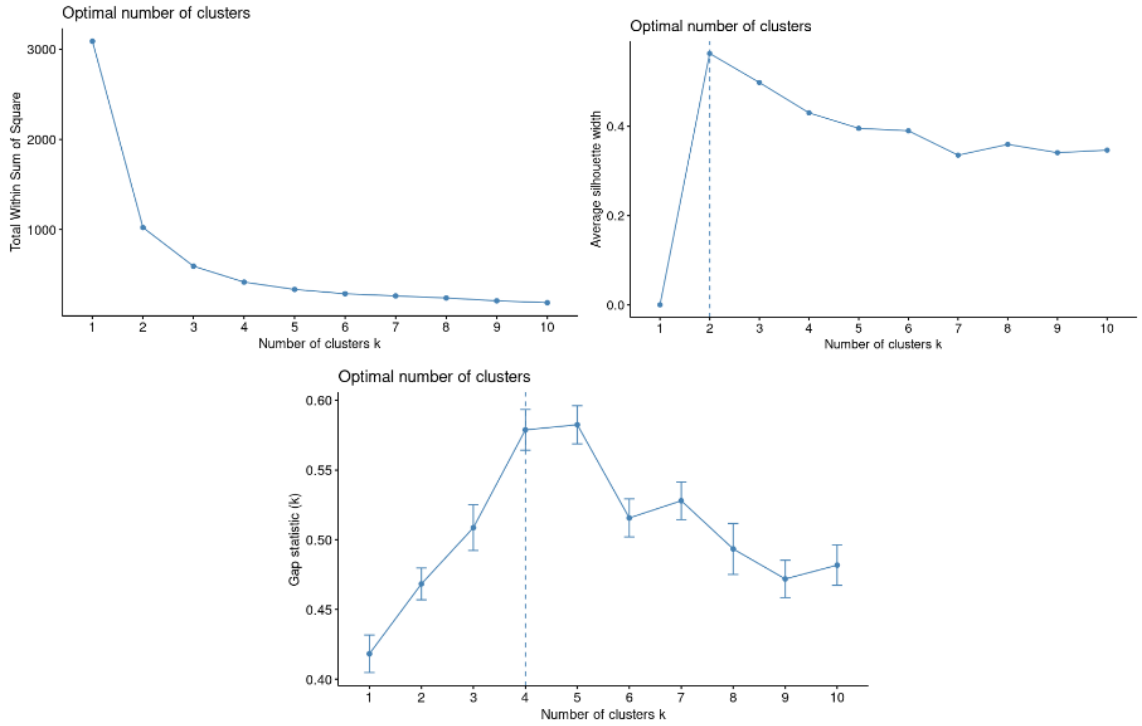

Fig 1S. Analysis of cluster number with different methods. The first one is Elbow method, second one is Silhouette method and the last one is Gap Statistic.

Each cancer subpopulation has a different mean of the scores and they seem to have different characteristics as we can see at Fig 2S.

MEANS

Group.1	StromalScore	ImmuneScore	ESTIMATEScore	TumorPurity
1	595.4042	1583.12378	2178.5280	0.6007893
2	-550.9280	57.68825	-493.2397	0.8597852
3	167.6065	685.62188	853.2284	0.7439146

STD

Group.1	StromalScore	ImmuneScore	ESTIMATEScore	TumorPurity
1	296.9639	446.2921	439.9321	0.05206096
2	264.6059	279.7907	429.1417	0.03129059
3	284.4708	253.1789	372.0215	0.03643835

Fig 2S. Mean and standard deviation for the ESTIMATE scores for the cells population for the 3 cancer clusters. See how the 3 clusters have different signatures of the scores and the tumor being from the more ‘’tumor pure’’ to the less Cluster2, Cluster 3 and Cluster 1.

input_preprocess_PROFILE.py

This script generates the output to perform the personalization of the Boolean models and another file which will be used to generate the final csv with the simulation information.

set_up_physi.py

To understand what this script does, we need to introduce the scanpy and the PhysiCell formats of position. The spatial information in the spatial transcriptomic data is saved in two ways (Fig 2S)

[12]:

	x_coord	y_coord	col	row
0	6080	3722	16	0
1	17782	15632	102	50
2	9763	4445	43	3
3	6447	17734	19	59
4	16716	7079	94	14
...	...	...	...	...
3975	9587	17507	42	58
3976	7947	17976	30	60
3977	7549	14413	27	45
3978	9439	21068	41	73
3979	10852	5399	51	7

3980 rows × 4 columns

Fig 2S. Spatial information saved in the scRNA ST object. In the two first columns we have the pixel position and in the last two the position is saved as the grid position number.

The main objective of this script is to change the positions to the PhysiCell format because they do not work in the same way (Fig 3S). Moreover, we add to the simulation extra margins.

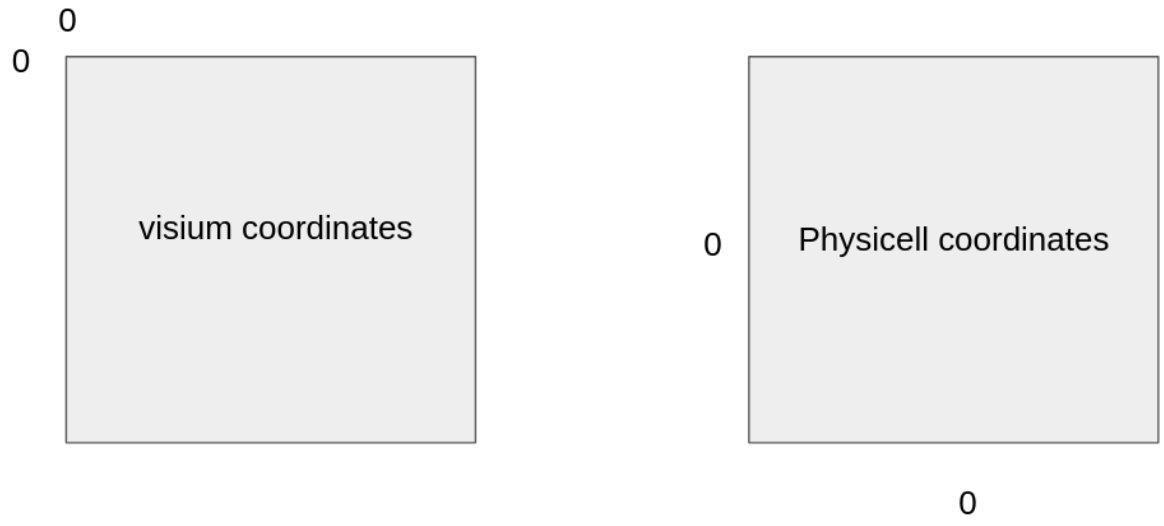

Fig 3S. Scheme of Visium and PhysiCell coordinates formats.

Fumia_Visium_Profile_onlyRNA.R

This script is an adaptation of a script provided by the PROFILE tool. It performs a previous processing of the data generating the files needed to create the personalized boolean models. For more information see [1].

[1] Arnau Montagud, Jonas Béal, Luis Tobalina, Pauline Traynard, Vigneshwari Subramanian, Bence Szalai, Róbert Alföldi, László Puskás, Alfonso Valencia, Emmanuel Barillot, Julio Saez-Rodriguez, Laurence Calzone (2022) Patient-specific Boolean models of signalling networks guide personalised treatments eLife 11:e72626