Nov 25, 2025

# QUICHE simulation design

DOI
https://dx.doi.org/10.17504/protocols.io.6qpvrwjk2lmk/v1

Jolene Ranek[1], Mike Angelo[1]

[1]Department of Pathology, Stanford University, Stanford CA, USA

Jolene Ranek





**DOI:** https://dx.doi.org/10.17504/protocols.io.6qpvrwjk2lmk/v1

**Protocol Citation:** Jolene Ranek, Mike Angelo 2025. QUICHE simulation design. **protocols.io** https://dx.doi.org/10.17504/protocols.io.6qpvrwjk2lmk/v1

**Manuscript citation:**
Ranek, J. S., Greenwald, N. F., Goldston, M., Fullaway, C. C., Sowers, C., Kong, A., Mouron, S., Quintela-Fandino, M., West, R. B., and Angelo, M. QUICHE reveals structural definitions of anti-tumor responses in triple negative breast cancer. *bioRxiv*. 2025

**Protocol status:** Working
**We use this protocol and it's working**

**Created:** November 25, 2025

**Last Modified:** November 25, 2025

**Protocol Integer ID:** 233510

**Keywords:** quiche simulation design, quiche simulation design this protocol, triple negative breast cancer, simulation design for the manuscript, simulation design, automated computational, quiche

## Abstract

This protocol describes the simulation design for the manuscript "The automated computational workflow QUICHE reveals structural definitions of anti-tumor responses in triple negative breast cancer" by Ranek *et al*.

## *In silico* simulation design

1 To validate our approach and benchmark spatial enrichment methods on their ability to identify condition-specific differences in the spatial organization of cell types from individual patient samples, we simulated multi-patient cohorts with different spatial topologies (e.g., unstructured, structured), where the ground truth organization was known. Here, each multi-patient cohort consisted of 20 patient samples evenly split across two conditions. For each patient sample, we fixed the number of cells and cell types to be the same, and then varied the spatial organization of cells according to each patient's condition label (e.g., condition A consisted of immune cells enriched in the cancer core).

## Unstructured simulation

2 We generated simulated datasets with unstructured spatial topologies as follows. For each patient sample, we randomly sampled the spatial coordinates of 5000 cells across five cell types (cell types A, B, C, D, E) from a uniform distribution within a $2048 \times 2048$ pixel domain. Each image was then discretized into a $g \times g$ grid. To define ground truth condition-specific niches, we selected one grid from each image and enforced cell types to co-occur within that region (condition 1: cell types A, C, and E; condition 2: cell types B and D). The grid size $g$ was used to determine the niche prevalence (e.g., $5 \times 5$ grid indicates that a niche occupies $1/25$ or $4\%$ of a sample). The output of this approach is a ground truth cell-to-cell type and cell-to-niche assignments. To evaluate how well methods can identify condition-specific differences, we varied both the size of a niche in a patient sample (grid size = 3, 4 , 5, 6, 7, 8, 9, 10) and its prevalence across patient samples (20, 40, 60, 80, 100% of samples). Method performance was assessed over five random trials for each parameter configuration.

## Structured simulation

3 For a more realistic simulation design, we leveraged multiplexed proteomic imaging data from the Spain TNBC cohort to define cell spatial coordinates and tissue domains as follows.

3.1 **Tumor selection:** To generate simulated datasets with structural spatial topologies, we randomly selected tumors with compartmentalized tumor-immune spatial architectures using the mixing score defined in Ref. [1] as,

$$m = \frac{\sum_{i=1}^{n_r} \sum_{j \in \mathcal{N}_i} T_j}{\sum_{i=1}^{n_r} \sum_{j \in \mathcal{N}_i} R_j + \sum_{i=1}^{n_t} \sum_{j \in \mathcal{N}_i} T_j}.$$

The mixing score quantifies the degree of mixing between reference (tumor) and target (immune) cells within a fixed spatial radius, $s$, by computing the ratio of heterotypic (cancer-immune) to homotypic interactions (cancer-cancer, immune-immune). Here, $n_r$ and $n_t$ represent the number of reference and target cells, respectively. The indicator functions $T_j$ and $R_j$ denote whether a target or reference cell type is within the neighborhood $\mathcal{N}$of cell $i$. Specifically, $T_j = 1$ if cell $j$ is a target cell type within the neighborhood $\mathcal{N}_i$, and $T_j = 0$ otherwise. Similarly, $R_j = 1$ if cell $j$ is a reference cell type within the neighborhood $\mathcal{N}_i$, and $R_j = 0$ otherwise. The neighborhood $\mathcal{N}_i$ is defined as $\{j : \|(u_i, v_i) - (u_j, v_j)\| \leq s\}$, where $(u_i, v_i)$ and $(u_j, v_j)$ are the spatial coordinates of cells $i$ and $j$, respectively. Tumors were considered to have compartmentalized spatial architectures if they had more homotypic than heterotypic interactions.

3.2 **Tumor compartments:** To define structural regions in each tumor image, we followed the approach outlined previously [2]. Briefly, tumor regions (cancer core, cancer border, stroma) were defined by taking the union of the E-cadherin signal with the cancer cell segmentation masks. Tumor compartment masks were then thresholded (0.0015), binarized, and eroded by 100 pixels to delineate cancer cells within the core vs border; all of the remaining cells were defined as stroma. Image processing was performed using the scikit-image v0.19.3 package in Python.

3.3 **Differential spatial enrichment:** We simulated condition-specific cell type enrichment within structured spatial topologies (condition 1: immune cells within the cancer core, condition 2: immune cells within the cancer border). For each patient sample, we randomly sampled cells from tumor regions (condition 1: cells within the cancer core; condition 2: cells within the cancer border) using a Gaussian distribution with a specified radius (radius = 50, 100, 250, 500). Next, we defined ground truth niches by reclassifying a subset of cells within the Gaussian radius as immune cells, thereby creating localized clusters of immune infiltration. Finally, we varied the (1) niche size (radius = 100, 250, 500 pixels), (2) niche sparsity (5, 10, 25% of immune cells), and (3) niche prevalence across patient samples (20, 40, 60, 80, 100% samples). Method performance was evaluated across five random trials for each parameter configuration.

## Expression

4 We used Splatter [3] to simulate expression data in our *in silico* data. Simulation parameters were estimated from a single-cell RNA sequencing dataset [4]. Next, the estimated parameters (mean_rate = 0.0173, mean_shape = 0.54, lib_loc = 12.6, lib_scale = 0.423, out_prob = 0.000342, out_fac_loc = 0.1, out_fac_scale = 0.4, bcv = 0.1, bcv_df = 90.2, dropout = None) were used in the Splatter groups function (python wrapper scprep SplatSimulate v1.2.3 of splatter v1.18.2) to simulate expression data (p = 500) for each cell type (e.g., cancer core, cancer border, stroma, immune).

## References

5 [1] Keren, L., Bosse, M., Marquez, D., et al. A structured tumor-immune microenvironment in triple-negative breast cancer revealed by multiplexed ion beam imaging. Cell 174(6):1373–1387.e19, 2018.

[2] Greenwald, N.F., Nederlof, I., Sowers, C., et al. Temporal and spatial composition of the tumor microenvironment predicts response to immune checkpoint inhibition. bioRxiv 2025.01.26.634557, 2025.

[3] Zappia, L., Phipson, B., & Oshlack, A. Splatter: simulation of single-cell RNA sequencing data. Genome Biol. 18(1):174, 2017.

[4] Loh, K.M., Chen, A., Koh, P.W., et al. Mapping the pairwise choices leading from pluripotency to human bone, heart, and other mesoderm cell types. Cell 166(2):451–467, 2016.