Supporting Information for “SPF: A Spatial and Functional Data Analytic Approach to cell Imaging data”

Thao Vu1, Julia Wrobel1, Benjamin G. Bitler 2,3, Erin L. Schenk4, Kimberly R. Jordan 5, Debashis Ghosh1*

1 Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA
2 University of Colorado Comprehensive Cancer Center, Aurora, CO 80045, USA
3 Department of OB/GYN, Division of Reproductive Sciences, The University of Colorado, Aurora, CO 80045, USA
4 Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA
5 Department of Immunology and Microbiology, University of Colorado School of Medicine, Aurora, CO 80045, USA

*debashis.ghosh@cuanschutz.edu

S.1 NSCLC dataset

S.1.1 Model output by fitting mcf

Family: Cox PH
Link function: identity

Formula:
\[ \text{stime} \sim \text{stage\_group} + \text{Age} + \text{total\_cell} + \text{ti(t\_int, func\_cov, by = l\_int, bs = c("cr", "cr"), k = c(k, k), mc = c(FALSE, TRUE))} \]

Parametric Terms:
Approximate significance of smooth terms:

<table>
<thead>
<tr>
<th>edf</th>
<th>Ref.df</th>
<th>Chi.sq</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ti(t_int,func_cov):l_int</td>
<td>7.164</td>
<td>9.430</td>
<td>8.871</td>
</tr>
</tbody>
</table>

S.1.2 Exploratory analysis of MHCII in tumor vs. stromal cells

![Graphs showing distribution of MHCII expression](image)

**Fig S1.** Distribution of MHCII expression in tumor (turquoise) vs. stromal (red) cells in (A) A representative MHCII$^{hi}$ sample. (B) A representative MHCII$^{lo}$ sample.

S.1.3 Model output by fitting Moran’s I correlation in MHCII

Family: Cox PH

Link function: identity

Formula:

```r
stime ~ stage_group + Age + total_cell + ti(t_int, func_cov, 
by = l_int, bs = c("cr", "cr"), k = c(k, k), mc = c(FALSE, 
TRUE))
```
Parametric Terms:

<table>
<thead>
<tr>
<th>Term</th>
<th>df</th>
<th>Chi.sq</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>stage_group</td>
<td>3</td>
<td>6.058</td>
<td>0.108807</td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td>13.326</td>
<td>0.000262</td>
</tr>
<tr>
<td>total_cell</td>
<td>1</td>
<td>1.264</td>
<td>0.260872</td>
</tr>
</tbody>
</table>

Approximate significance of smooth terms:

<table>
<thead>
<tr>
<th>Term</th>
<th>edf</th>
<th>Ref.df</th>
<th>Chi.sq</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ti(t_int,func_cov):l_int</td>
<td>2.349</td>
<td>2.584</td>
<td>0.449</td>
<td>0.873</td>
</tr>
</tbody>
</table>
S.2 TNBC dataset

S.2.1 Representative TNBC images

Fig S2. Representative TNBC tissue samples. Top Row: Pixel-level images of (A) Patient 1 and (B) Patient 2; color-coded from cell segmentation process to be associated with cell-level data. Bottom Row: Corresponding cell-level images for (C) Patient 1 and (D) Patient 2. Each color represents a cell classification group. Dot size is proportional to cell size.
S.2.2 Exploratory analysis of P53 in tumor vs. immune cells

Fig S3. P53 expression in TNBC. Distribution of P53 expression in tumor (turquoise) vs. immune cells (red) in: (A) A representative P53 positive sample. (B) A representative P53 negative sample. (C) Corresponding Moran’s I correlation using P53 expression in P53 positive (red) and P53 negative (black) samples. Moran’s I values above 0 indicate a direct relationship in P53 expression between tumor and immune cells. Negative Moran’s I values suggest an inverse association in P53 expression between tumor and immune cells.

S.3 Ovarian cancer dataset

Tissue microarray (TMA) slides of 132 ovarian cancer patients were stained with antibodies specific for CD3, CD4, CD8, CD19, CD68, cytokeratin, Ki67, pStat, and IER3. The slides were imaged using Vectra 3.0 microscope (Akoya Biosystems) and then segmented to define cells within the tumor and stroma regions as well as to assign cell phenotype utilizing inForm software. More details can be found in Jordan et al. [1]. Fig. S4 visualized images of four representative patients with turquoise and red points denoting tumor and stromal cells, respectively. Within the cohort, we excluded 18 patients from the analysis due to missingness of clinical information. Similar to the applications described in the main text, we explored the impact of spatial distribution of cells in the TME on patient survival outcome. Fig. S5 (A) captured the interactions between tumor and stromal cells as a function of cell distance for all patients in the dataset while Fig. S5 (B) depicted the Moran’s I correlation between tumor and stromal cells within the TME with respect to Ki67 marker expression. Mcf and Moran’s I curves were separately included in the model as functional covariates with survival outcome displayed in Fig. S5 (C). The estimated functional surface shown Fig. S6 (A) had the values decreased from red to blue, in
correspondence with a decrease in hazard of mortality, while holding the scalar predictors fixed. If there were more stromal cells surrounding tumor cells at any given distance $r$, which resulted in the larger mcf values, the hazard of death increased. On the contrary, if tumor and stromal cells were clustered together in their compartments, which resulted in smaller mcf values, the estimated survival improved. In similar fashion, Fig. S6 (B) displayed the estimated functional surface $\hat{F}$ after fitting the model using the Moran’s I correlation curve in Ki67 expression as functional covariates. Specifically, positive correlation in Ki67 intensity between tumor and stromal cells in the neighborhood of less than 150\(\mu m\) caused an increase in the risk of mortality. However, if the positive association continued past 175\(\mu m\), the hazard of mortality decreased.

Fig S4. Representative images in the ovarian cancer dataset. Example images of four representative patients (A) Patient 1, (B) Patient 51, (C) Patient 75, and (D) Patient 103, with red and turquoise points denoting stromal and tumor cells, respectively.
Fig S5. Ovarian cancer dataset. (A) Mcf curves for all patients. Note that mcf values below 1 indicate strong clustering of cells of same type, while values above 1 suggest higher level of mixing in of the two cell types. (B) Moran’s I correlation between tumor and stromal cells across subjects using Ki67 marker expression. Moran’s I values above 0 indicate a direct relationship in Ki67 expression between tumor and stromal cells. Negative Moran’s I values suggest an inverse association in Ki67 expression between tumor and stromal cells.

Fig S6. Estimated surfaces from AFCM using ovarian cancer dataset. (A) Estimated surface from AFCM using mcf curves as functional covariates, with values of $\hat{F}$ decreasing from positive (red) to negative (blue). (B) Estimated surface from AFCM using Moran’s I correlation in Ki67 expression between tumor and stromal cells, with values of $\hat{F}$ decreasing from positive (red) to negative (blue). Positive $\hat{F}$ corresponds to increased risk of mortality while negative $\hat{F}$ associates with reduced hazard of death.
References