ASA SMI

Packages

Some packages that we need to load for analyses like `tidyverse` and `spatialTIME` available from CRAN.

```
#make vector of packages to install and load
packages = c("tidyverse", "spatialTIME", "here")
#check if packages are installed and if not, install
install.packages(setdiff(packages, rownames(installed.packages())))
t = lapply(packages, require, character.only = T)
```

Importing Data

The clinical data, per-cell spatial data, and sample summary data are in the `data/` subdirectory

```
#read clinical csv
clinical_data <- read.csv(here::here("data/deidentified_clinical.csv"))
head(clinical_data)
```

```
##   age    race  sex deidentified_sample deidentified_id status
## 1  66 unknown Male     TMA3_[12,Q].tif             198      A
## 2  60 unknown Male     TMA3_[10,I].tif             127      A
## 3  55 unknown Male      TMA3_[8,L].tif              38      A
## 4  59 unknown Male      TMA3_[7,N].tif             122      B
## 5  57 unknown Male      TMA3_[5,B].tif             992      B
## 6  52 unknown Male      TMA3_[9,D].tif             690      B
```

```
#read summary csv
sample_summary <- read.csv(here::here("data/deidentified_summary.csv"), sep="")
head(sample_summary)
```

```
##   deidentified_id deidentified_sample Total.Cells
## 1             198     TMA3_[12,Q].tif        1472
## 2             272     TMA3_[13,S].tif        1095
## 3             127     TMA3_[10,I].tif         332
## 4             122      TMA3_[7,N].tif        1240
## 5             122      TMA3_[7,R].tif        1028
## 6             371      TMA3_[7,U].tif        2286
##   FOXP3..Opal.620..Positive.Cells CD3..Opal.570..Positive.Cells
## 1                               6                            37
## 2                               3                            19
## 3                               0                            16
```

```
## 4                                0                                5
## 5                                3                               90
## 6                               10                               36
##   CD8..Opal.520..Positive.Cells PD1..Opal.650..Positive.Cells
## 1                            29                             0
## 2                             7                             0
## 3                            18                             0
## 4                            17                             0
## 5                            86                             0
## 6                            18                             0
##   PDL1..Opal.540..Positive.Cells CD3..FOXP3..Cells CD3..CD8..Cells
## 1                              0                 6               15
## 2                              0                 1                1
## 3                              0                 0                7
## 4                              0                 0                2
## 5                              0                 3               32
## 6                              0                 5                9
##   CD3..CD8..FOXP3..Cells CD3..PD1..Cells CD3..PD.L1..Cells CD8..PD1..Cells
## 1                      1               0                 0               0
## 2                      0               0                 0               0
## 3                      0               0                 0               0
## 4                      0               0                 0               0
## 5                      0               0                 0               0
## 6                      0               0                 0               0
##   CD3..CD8..PD.L1..Cells X..FOXP3..Opal.620..Positive.Cells
## 1                      0                          0.407609
## 2                      0                          0.273973
## 3                      0                          0.000000
## 4                      0                          0.000000
## 5                      0                          0.291829
## 6                      0                          0.437445
##   X..CD3..Opal.570..Positive.Cells X..CD8..Opal.520..Positive.Cells
## 1                         2.513587                         1.970109
## 2                         1.735160                         0.639269
## 3                         4.819277                         5.421687
## 4                         0.403226                         1.370968
## 5                         8.754864                         8.365759
## 6                         1.574803                         0.787402
##   X..PD1..Opal.650..Positive.Cells X..PDL1..Opal.540..Positive.Cells
## 1                                0                                 0
## 2                                0                                 0
## 3                                0                                 0
## 4                                0                                 0
## 5                                0                                 0
## 6                                0                                 0
##   X..CD3..FOXP3..Positive.Cells X..CD3..CD8..Positive.Cells
## 1                      0.407609                    1.019022
## 2                      0.091324                    0.091324
## 3                      0.000000                    2.108434
## 4                      0.000000                    0.161290
```

```
## 5                      0.291829                    3.112840
## 6                      0.218723                    0.393701
##   X..CD3..CD8..FOXP3..Positive.Cells X..CD3..PD1..Positive.Cells
## 1                           0.067935                           0
## 2                           0.000000                           0
## 3                           0.000000                           0
## 4                           0.000000                           0
## 5                           0.000000                           0
## 6                           0.000000                           0
##   X..CD3..PD.L1..Positive.Cells X..CD8..PD1..Positive.Cells
## 1                             0                           0
## 2                             0                           0
## 3                             0                           0
## 4                             0                           0
## 5                             0                           0
## 6                             0                           0
##   X..CD3..CD8..PD.L1..Positive.Cells Area.Analyzed..µm.. Area.Analyzed..mm..
## 1                                  0           297716.84          0.29771684
## 2                                  0           202001.39          0.20200139
## 3                                  0            95751.26          0.09575126
## 4                                  0           268881.72          0.26888172
## 5                                  0           242663.67          0.24266367
## 6                                  0           348478.28          0.34847828
```

```
#read spatial list
load(here::here("data/deidentified_example.RData"))
names(example_spatial_small)
```

```
##  [1] "TMA3_[13,S].tif" "TMA3_[8,L].tif"  "TMA3_[2,J].tif"  "TMA3_[2,N].tif"
##  [5] "TMA3_[2,K].tif"  "TMA3_[8,T].tif"  "TMA3_[5,B].tif"  "TMA3_[7,R].tif"
##  [9] "TMA3_[7,N].tif"  "TMA3_[10,I].tif" "TMA3_[7,U].tif"  "TMA3_[12,Q].tif"
## [13] "TMA3_[2,A].tif"  "TMA3_[9,D].tif"  "TMA3_[6,E].tif"
```

Create Multiplex ImmunoFlourescent (mif) Object

spatialTIME functions use a custom `mif` object which can be created using the `create_mif()` function. The `mif` object has 6 slots storing the:

- Clinical data which must contain:
  - a column whose column name matches one in the sample dataset.
- Sample summary data including counts/percentages of each positive cells for each single or combination of markers and total number of cells for each core. In order to use `create_mif()` function this table must contain:
  - a column whose column name matches one in the clinical dataset
  - a column whose column name matches one in the spatial list.

- Spatial list (1 per each core):
  - This object should be a list object where each element of the list corresponds to a core and each element should be a
    $n \times p$ dataframe (
    n = number of cells) containing:
    1. a column name that matches a column name for the sample file (for merging and potential downstream analysis linking to clinical variables),
    2. `XMin`, `XMax`, `YMin`, and `YMax` which defines the area that a cell occupies which is eventually used to assign a location for each cell with the x-position being the mean of the `XMin` and `XMax` and y-position being the mean of the `YMin` and `YMax`
    3. a set of columns that indicate whether a cell is positive to one or multiple markers.
- patient id:
  - a column name used to merge clinical and summary data
- sample_id:
  - a column name used to merge the spatial and summary data
- derived:
  - where all of the plots and spatial clustering measures are stored

Creating MIF object

```
# Make sure the variable types are the same for deidentified_id and
# deidentified_sample in their corresponding datasets
x <- spatialTIME::create_mif(clinical_data = clinical_data %>%
                               mutate(deidentified_id = as.character(deidentified_id)
),
                             sample_data = sample_summary %>%
                               mutate(deidentified_id = as.character(deidentified_id)
),
                             spatial_list = example_spatial_small,
                             patient_id = "deidentified_id", sample_id = "deidentifie
d_sample")

x #prints a summary of how many patients, samples, and spatial files are present
```

```
## 14 patients spanning 15 samples and 15 spatial data frames were found
```

Visualizing TMAs

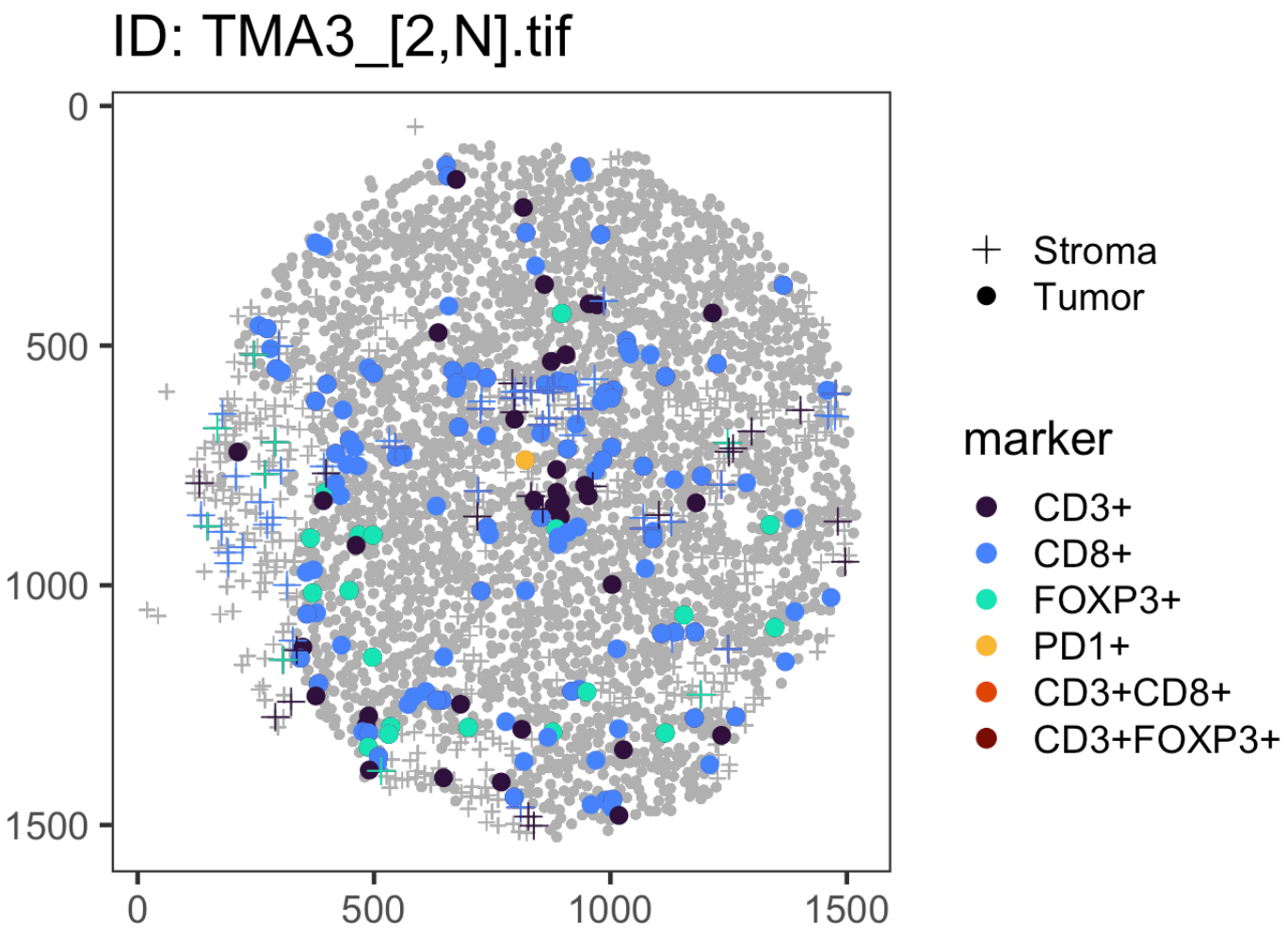
A plot of for each sample in the spatial list is created and assigned to the `derived` slot in the mif object. If an argument to `filename` is provided then a PDF with one figure per page is created.

The order of the phenotypes provided matters. When phenotypes are derived by multiple markers, the last provided phenotype will overlap previous phenotypes - so put the marker combinations before the individual markers.

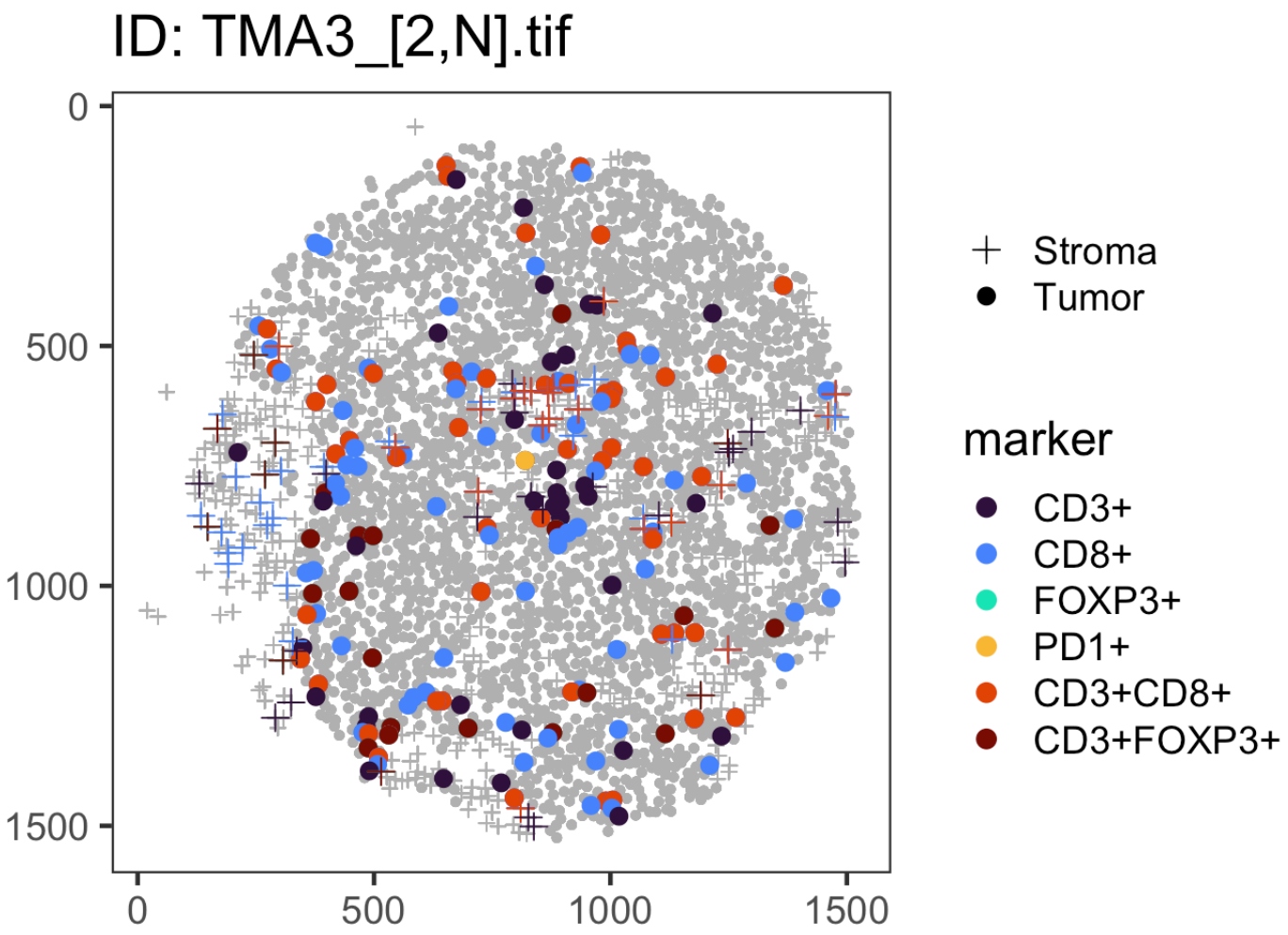
```
names_one <- c("CD3..CD8.","CD3..FOXP3.","CD3..Opal.570..Positive",
               "CD8..Opal.520..Positive","FOXP3..Opal.620..Positive",
               "PDL1..Opal.540..Positive", "PD1..Opal.650..Positive")

names_two <- c("CD3..Opal.570..Positive","CD8..Opal.520..Positive",
               "FOXP3..Opal.620..Positive","PDL1..Opal.540..Positive",
               "PD1..Opal.650..Positive","CD3..CD8.","CD3..FOXP3.")

#add an element in the `derived` object position
x <- spatialTIME::plot_immunoflo(x, plot_title = "deidentified_sample",
                                 mnames = names_one, cell_type = "Classifier.Label")
x[["derived"]][["spatial_plots"]][[4]] +
  scale_color_manual(breaks = names_two,
                     values = viridis::turbo(7),
                     labels = names_two %>%
                       gsub("..Opal.*", "+", .) %>%
                       gsub("\\.\\.", "+", .) %>%
                       gsub("\\.", "+", .))
```



```
# name order matters
x <- spatialTIME::plot_immunoflo(x, plot_title = "deidentified_sample",
                                 mnames = names_two, cell_type = "Classifier.Label")
x[["derived"]][["spatial_plots"]][[4]] +
  scale_color_manual(breaks = names_two,
                     values = viridis::turbo(7),
                     labels = names_two %>%
                       gsub("..Opal.*", "+", .) %>%
                       gsub("\\.\\.", "+", .) %>%
                       gsub("\\.", "+", .))
```



Complete spatial randomness

Complete spatial randomness (CSR) is the assumption that the cells are randomly distributed across the region, with no evidence of clustering or repulsion, and that the cell intensity is constant across the entire region. When tissue cores are collected, rips or tears can create regions where it appears there are no cells (when this is not the case). This violates the CSR assumption and therefore theoretical estimates of CSR may not be accurate. Damage can occur to tissue cores due to how they are collected. This damage can lead to

rips and tears in the cores which results in regions where it appears that cells are not located and is not actually the case. Due to these violations of the CSR assumption, the theoretical estimate for CSR may not be accurate.

Univariate analysis

The spatialTIME package can compute 3 count-based spatial statistics: Ripley's K, Besag's L and Marcon's M. Ripley's K measures the average number of cells within a specified radius of a cell. Edge corrections for Ripley's K are used to account for points outside of the observed region and assume that the distribution of cells is the same inside and outside of the observed region. There are 3 main edge corrections available in spatialTIME - isotropic, translational or border. We recommend using either isotropic or translational edge corrections when there are a small number of cells. The package can also compute a distance-based measure: G, which is the proportion of cells whose distance to its nearest neighbor is less than a specified amount (`r`). Edge corrections for G apply different cell inclusion criteria and are either reduced sample and Hasnisch.

Univariate - count

The `ripleys_k` function reports a permuted and theoretical estimate of CSR, the observed value for the specified statistic, and the full permutation distribution of the statistic if `keep_perm_dis = TRUE` .

In this example, the number of permutations is 10, but this should be increased to closer to 100 for a more reliable estimate when doing a real analysis.

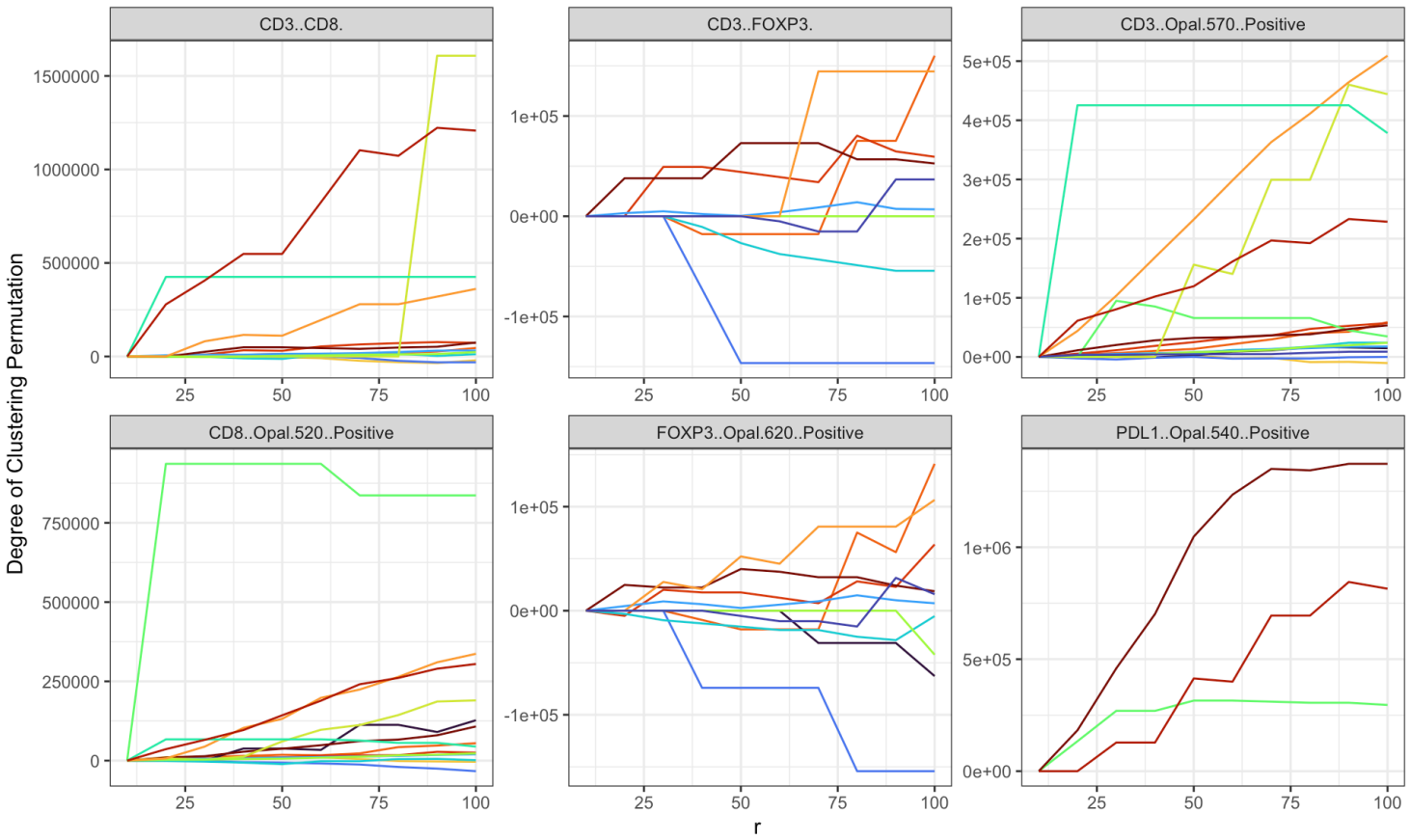
```
x <- ripleys_k(mif = x, mnames = names_two,
               num_permutations = 10, method = "K",
               edge_correction = 'translation', r = seq(0,100,10),
               keep_perm_dis = FALSE, workers = 1)

tail(x$derived$univariate_Count)
```

```
##      deidentified_sample                   Marker   r Theoretical CSR
## 1045      TMA3_[6,E].tif PDL1..Opal.540..Positive  50        7853.982
## 1046      TMA3_[6,E].tif PDL1..Opal.540..Positive  60       11309.734
## 1047      TMA3_[6,E].tif PDL1..Opal.540..Positive  70       15393.804
## 1048      TMA3_[6,E].tif PDL1..Opal.540..Positive  80       20106.193
## 1049      TMA3_[6,E].tif PDL1..Opal.540..Positive  90       25446.900
## 1050      TMA3_[6,E].tif PDL1..Opal.540..Positive 100       31415.927
##      Permuted K Observed K Degree of Clustering Permutation
## 1045   6192.018    1053674                           1047482
## 1046   9383.705    1244139                           1234756
## 1047  22213.934    1371986                           1349772
## 1048  28685.026    1371986                           1343301
## 1049  31967.088    1404456                           1372489
## 1050  31967.088    1404456                           1372489
##      Degree of Clustering Theoretical
## 1045                          1045820
## 1046                          1232830
## 1047                          1356592
## 1048                          1351879
## 1049                          1379009
## 1050                          1373040
```

```
values = viridis::turbo(length(unique(x$derived$univariate_Count$deidentified_sample)
))
names(values) = unique(x$derived$univariate_Count$deidentified_sample)

x$derived$univariate_Count  %>%
  filter(Marker != 'PD1..Opal.650..Positive') %>%
  ggplot(aes(x = r, y = `Degree of Clustering Permutation`)) +
  geom_line(aes(color = deidentified_sample), show.legend = FALSE) +
  facet_wrap(Marker~., scales = 'free') +
  theme_bw() +
  scale_color_manual(values = values)
```



Positive values for degree of cluster when using `method = 'K'` and `method = 'L'` indicates evidence of spatial clustering, while negative values correspond to spatial regularity.

On the other hand, if using `method = 'M'` then values > 1 correspond to clustering and values < 1 correspond to regularity. These values can also be interpreted as the the percent difference from spatial clustering, for example if $M = 0.5$ that means there is 50% less spatial clustering than expected under CSR.

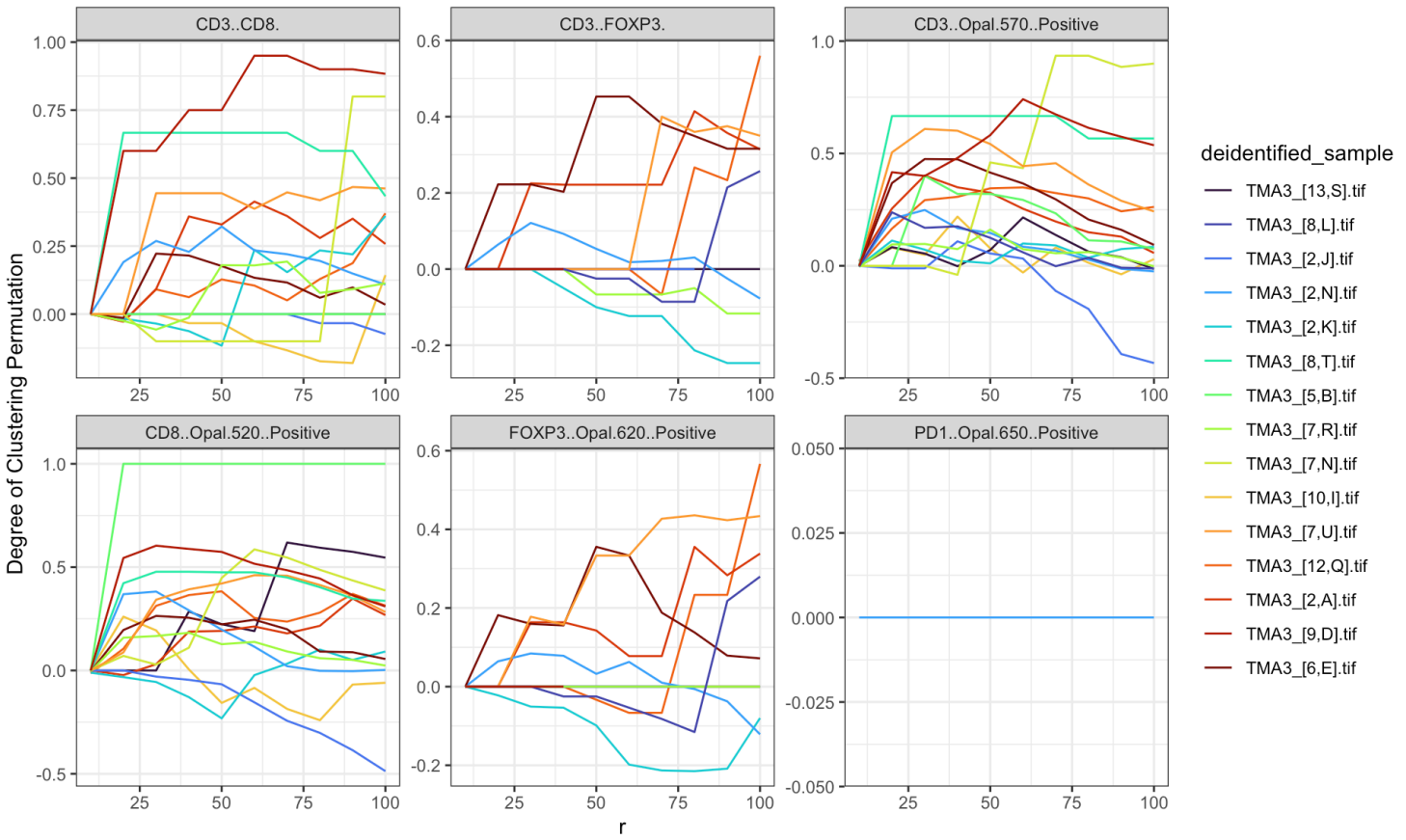
Univariate - nearest neighbor

```
x <- NN_G(mif = x, mnames = names_two, num_permutations = 10,
                edge_correction = 'rs', r = seq(0,100,10),
                keep_perm_dis = FALSE, workers = 1)

tail(x$derived$univariate_NN)
```

```
##      deidentified_sample                  Marker   r Theoretical CSR
## 1045      TMA3_[6,E].tif PDL1..Opal.540..Positive  50      0.05649350
## 1046      TMA3_[6,E].tif PDL1..Opal.540..Positive  60      0.08032866
## 1047      TMA3_[6,E].tif PDL1..Opal.540..Positive  70      0.10772238
## 1048      TMA3_[6,E].tif PDL1..Opal.540..Positive  80      0.13831817
## 1049      TMA3_[6,E].tif PDL1..Opal.540..Positive  90      0.17172707
## 1050      TMA3_[6,E].tif PDL1..Opal.540..Positive 100      0.20753591
##      Permuted CSR Observed Degree of Clustering Theoretical
## 1045    0.0850000        1                         0.9435065
## 1046    0.0975000        1                         0.9196713
## 1047    0.1116667        1                         0.8922776
## 1048    0.1116667        1                         0.8616818
## 1049    0.2068651        1                         0.8282729
## 1050    0.1932143        1                         0.7924641
##      Degree of Clustering Permutation
## 1045                        0.9150000
## 1046                        0.9025000
## 1047                        0.8883333
## 1048                        0.8883333
## 1049                        0.7931349
## 1050                        0.8067857
```

```
x$derived$univariate_NN  %>%
  filter(Marker != 'PDL1..Opal.540..Positive') %>%
  ggplot(aes(x = r, y = `Degree of Clustering Permutation`)) +
  geom_line(aes(color = deidentified_sample)) +
  facet_wrap(Marker~., scales = 'free') + theme_bw() +
  scale_color_manual(values = values)
```



The interpretation of the degree of clustering for G that values > 0 indicate spatial clustering of the cell types of interest, while values < 0 indicate dispersion of these cells.

Bivariate analysis

Bivariate - count

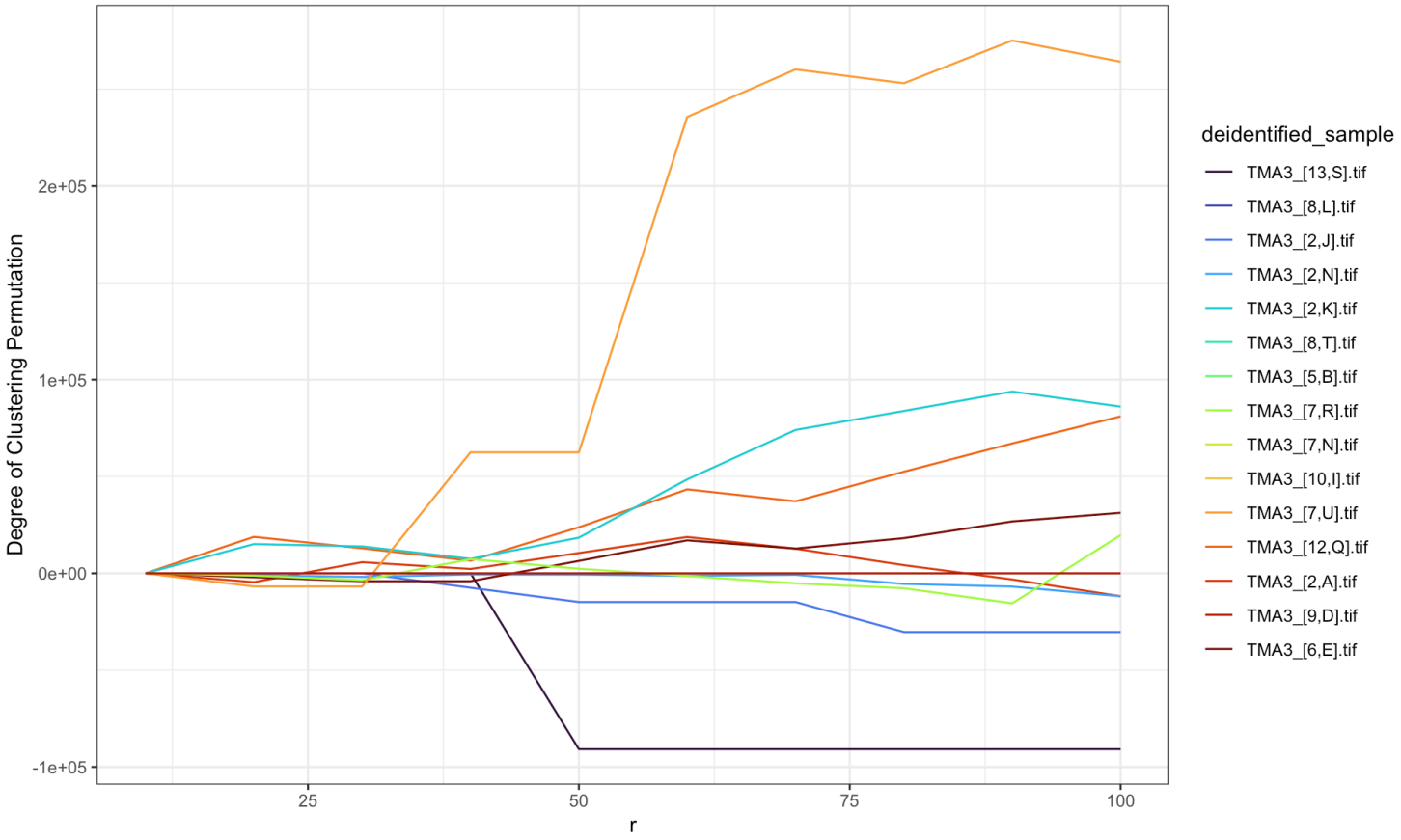
In bivariate analysis, we are interested in how many cells of Type 1 (Counted) are clustered in proximity to Type 2 (Anchor). Essentially the radius is centered around a cell of Type 2 and then the cells of Type 1 are counted.

```
x <- bi_ripleys_k(mif = x, mnames = c("CD3..CD8.", "CD3..FOXP3."),
                  num_permutations = 10, method = "K",
                  edge_correction = 'translation', r = seq(0,100,10),
                  keep_perm_dis = FALSE, workers = 1, exhaustive = TRUE)

head(x$derived$bivariate_Count)
```

```
##   deidentified_sample    anchor     counted  r Theoretical CSR Permuted K
## 1     TMA3_[13,S].tif CD3..CD8. CD3..FOXP3. 10        314.1593          0
## 2     TMA3_[13,S].tif CD3..CD8. CD3..FOXP3. 20       1256.6371          0
## 3     TMA3_[13,S].tif CD3..CD8. CD3..FOXP3. 30       2827.4334          0
## 4     TMA3_[13,S].tif CD3..CD8. CD3..FOXP3. 40       5026.5482          0
## 5     TMA3_[13,S].tif CD3..CD8. CD3..FOXP3. 50       7853.9816          0
## 6     TMA3_[13,S].tif CD3..CD8. CD3..FOXP3. 60      11309.7336          0
##   Observed K Degree of Clustering Permutation Degree of Clustering Theoretical
## 1          0                                0                       -314.1593
## 2          0                                0                      -1256.6371
## 3          0                                0                      -2827.4334
## 4          0                                0                      -5026.5482
## 5          0                                0                      -7853.9816
## 6          0                                0                     -11309.7336
```

```
x$derived$bivariate_Count  %>%
  filter(anchor == 'CD3..FOXP3.') %>%
  ggplot(aes(x = r, y = `Degree of Clustering Permutation`)) +
  geom_line(aes(color = deidentified_sample), show.legend = TRUE) +
  theme_bw() + scale_color_manual(values = values)
```



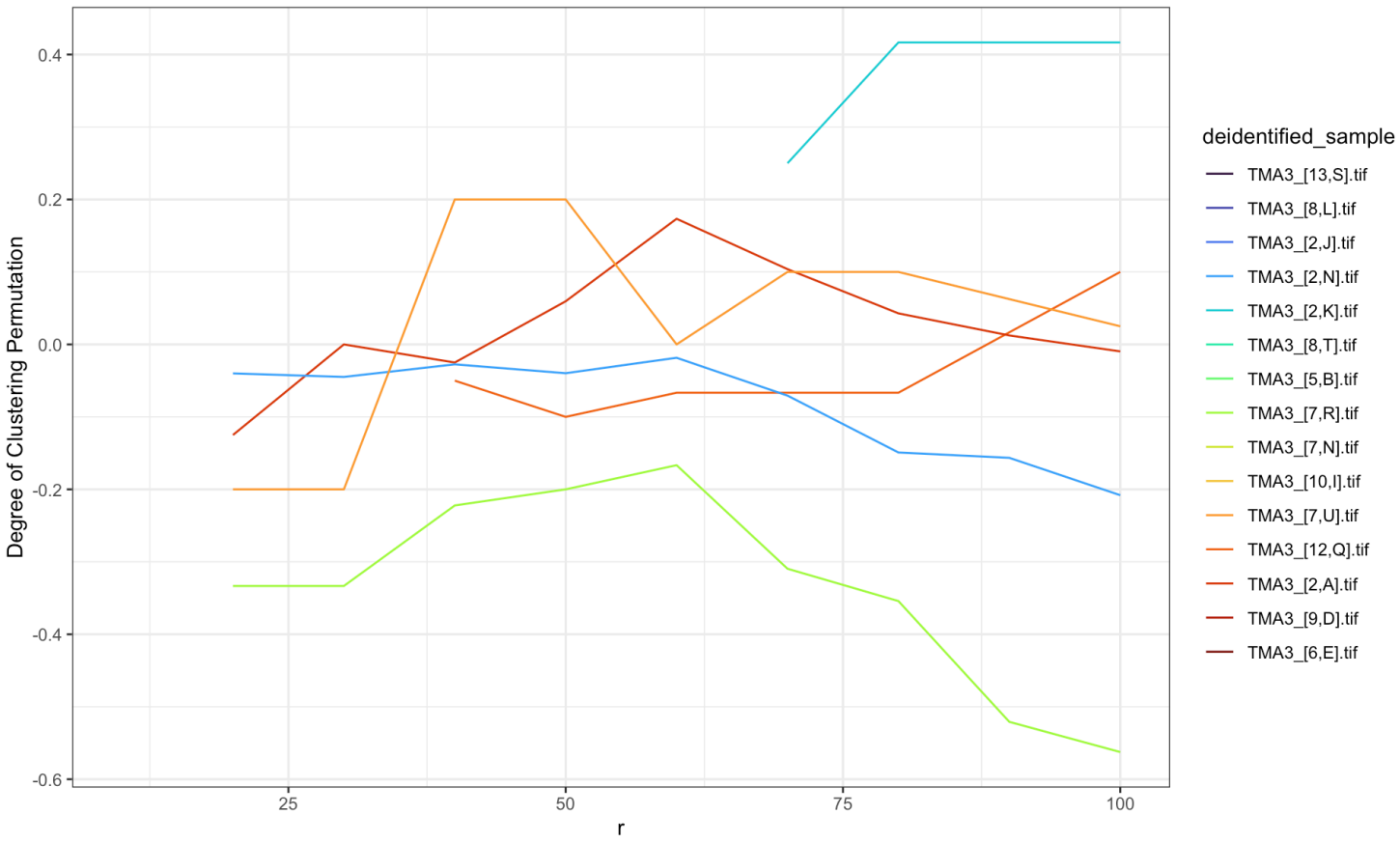
Bivariate - nearest neighbor

```
x <- bi_NN_G(mif = x, mnames = c("CD3..CD8.", "CD3..FOXP3."), num_permutations = 10,
               edge_correction = 'rs', r = seq(0,100,10),
               keep_perm_dis = FALSE, workers = 1, overwrite = TRUE)

head(x$derived$bivariate_NN)
```

```
##   deidentified_sample    anchor     counted  r Theoretical CSR Permuted G
## 1     TMA3_[13,S].tif CD3..CD8. CD3..FOXP3. 10     0.000371030          0
## 2     TMA3_[13,S].tif CD3..CD8. CD3..FOXP3. 20     0.001483294          0
## 3     TMA3_[13,S].tif CD3..CD8. CD3..FOXP3. 30     0.003334318          0
## 4     TMA3_[13,S].tif CD3..CD8. CD3..FOXP3. 40     0.005919988          0
## 5     TMA3_[13,S].tif CD3..CD8. CD3..FOXP3. 50     0.009234567          0
## 6     TMA3_[13,S].tif CD3..CD8. CD3..FOXP3. 60     0.013270714          0
##   Observed G Degree of Clustering Permutation Degree of Clustering Theoretical
## 1          0                               NaN                    -0.000371030
## 2          0                               NaN                    -0.001483294
## 3          0                               NaN                    -0.003334318
## 4          0                               NaN                    -0.005919988
## 5          0                               NaN                    -0.009234567
## 6          0                               NaN                    -0.013270714
```

```
x$derived$bivariate_NN  %>%
  filter(anchor == 'CD3..FOXP3.') %>%
  ggplot(aes(x = r, y = `Degree of Clustering Permutation`)) +
  geom_line(aes(color = deidentified_sample), show.legend = TRUE) +
  theme_bw() +  scale_color_manual(values = values)
```

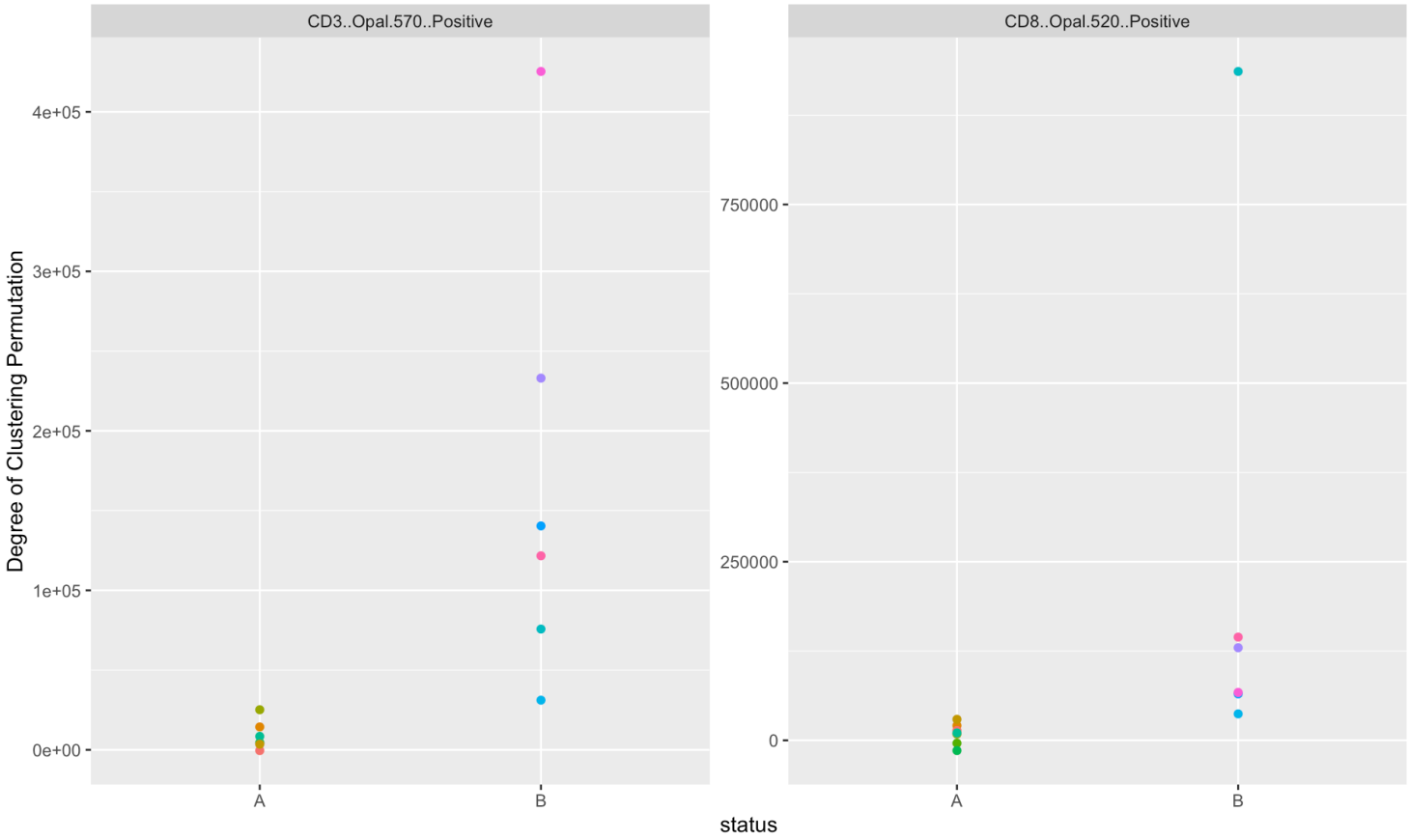


Additional analyses

```
x <- ripleys_k(mif = x, mnames = names_two, num_permutations = 10,
               edge_correction = 'translation', r = c(0,50),
               keep_perm_dis = FALSE, workers = 1, overwrite = TRUE)

inner_join(x$clinical, x$derived$univariate_Count) %>%
  filter(grepl('CD3..O|CD8..O', Marker)) %>%
  ggplot(aes(x = status, y = `Degree of Clustering Permutation`)) +
  geom_point(aes(color = deidentified_sample), show.legend = FALSE) +
  facet_wrap(Marker ~., scales = 'free')
```

```
## Joining, by = "deidentified_sample"
```


CD3..Opal.570..Positive
CD8..Opal.520..Positive
Degree of Clustering Permutation
4e+05
3e+05
2e+05
1e+05
0e+00
750000
500000
250000
0
A
B
status