

PVALB Exporatory Analysis

Table of contents

| | |
|---|-----------|
| Summary | 1 |
| PVALB Spot Calling | 2 |
| Choose threshold | 2 |
| PVALB+ spots and gene expression in V12F14-053_A1 | 3 |
| PVALB+ Spots (Not very informative) | 5 |
| PNN+ & PVALB+ Spots | 8 |
| Session Info | 10 |

Summary

Objective: The purpose of this report is to describe what would be a gene expression threshold to call PVALB+ spots ($\log_counts(PVALB) > 0$), and descriptive analysis summarizing proportion of PVALB+ and PVALB+/PNN+ spots across donors as well as spatial domains.

Summary:

- $\log_counts(PVALB) > 0$ seems to be a good threshold to call PVALB+ spots
- There seems to be an difference between the proportion of PNN+/PVALB+ spots across diagnosis groups (did not test for statistical significance).
- It would be challenging to do Pseudobulk analysis among PNN+/PVALB+ spots between diagnosis groups. I think it is possible to do it, but the power would be pretty low for this analysis.

PVALB Spot Calling

Choose threshold

Firstly, we used one sample (V12F14-053_A1) to decide/examine what a good threshold of PVALB gene expression would be.

The five point descriptive statistics for the log2counts of PVALB in this sample is listed below, and the distribution of PVALB is visualized in a histogram.

Overall, we can see that most of the spots doesn't have a PVALB expression. Less than a quarter of the spots express PVALB. Hence, we decide `logcounts(PVALB)>0` could be a good threshold to call PVALB+ spots.

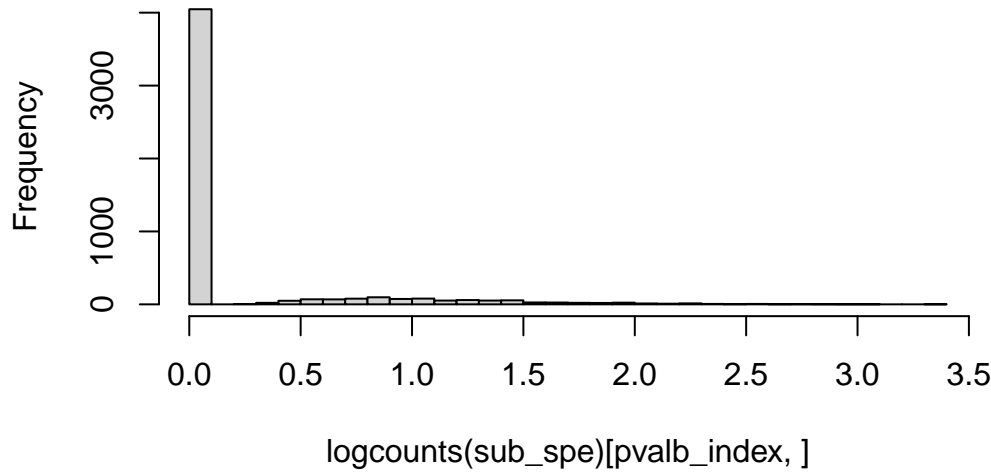
```
# five point summary statistics for the gene in this one sample
# V12F14-053_A1
logcounts(sub_spe)[pvalb_index, ] |> summary()
```

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--|--------|---------|--------|--------|---------|--------|
| | 0.0000 | 0.0000 | 0.0000 | 0.2064 | 0.0000 | 3.3038 |

```
#      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
# 0.0000  0.0000  0.0000  0.2064  0.0000  3.3038
```

```
hist(logcounts(sub_spe)[pvalb_index, ], breaks = 40)
```

Histogram of logcounts(sub_spe)[pvalb_index,]



```
# It seems there's a clear separation between 0 and non-0 expression. Hence, we will use the
```

NOTES: it is possible that in some spots, there's pvalb expression, just not being picked up/sequenced due to any technical reasons. This is out of the scope of this study.

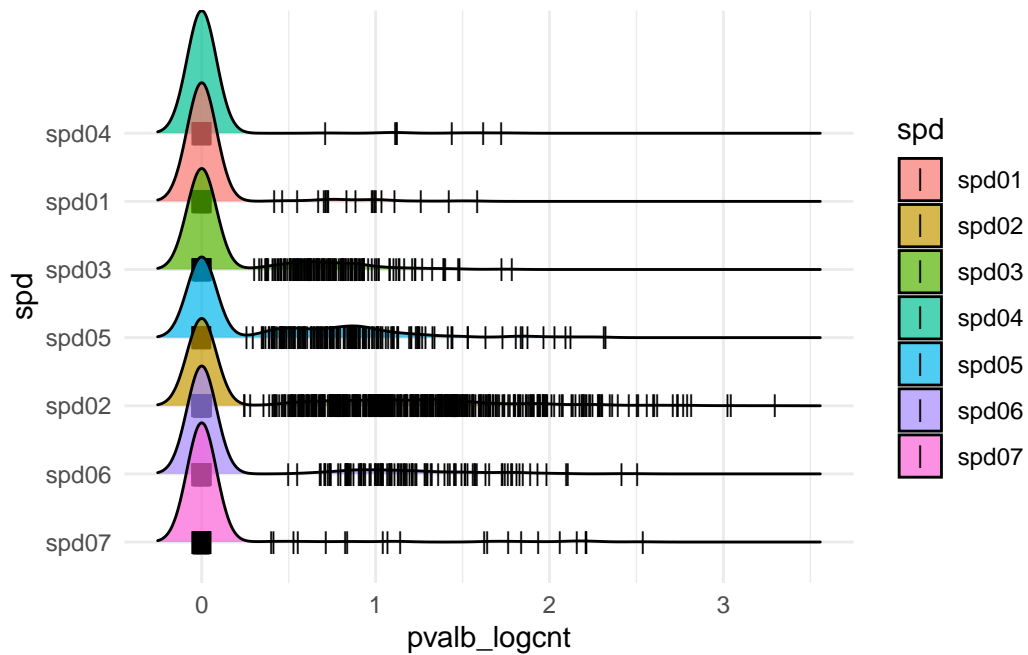
PVALB+ spots and gene expression in V12F14-053_A1

In this sample, around 2% of spots have PVALB expression.

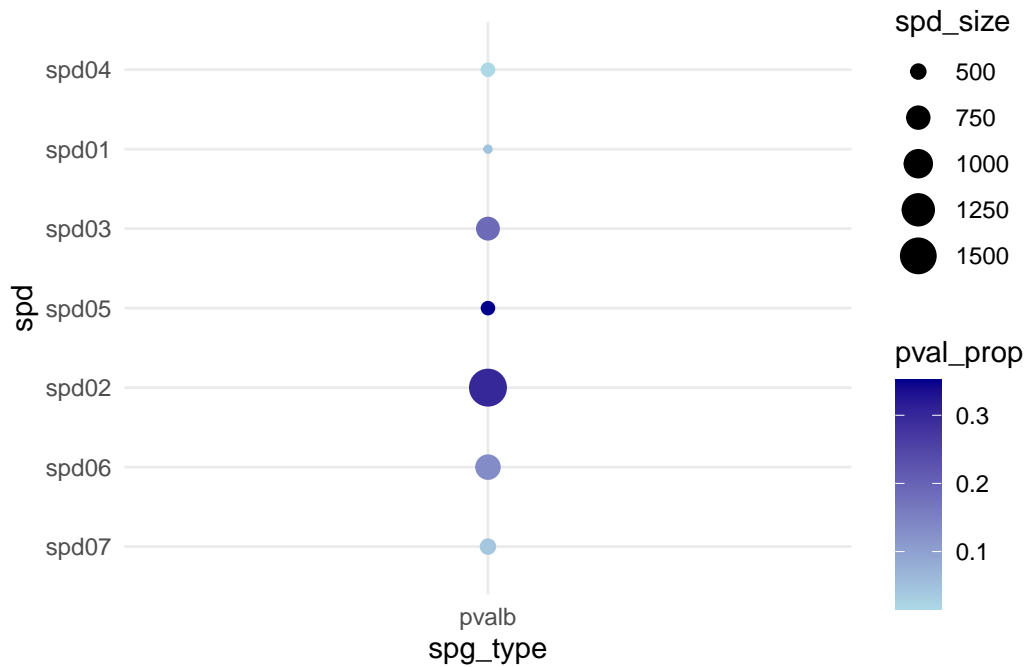
```
[1] 0.02858021
```

The following shows that the gene expression of PVALB across SPD07. We can see that there's most spots in SPD02 (Layer 3/4) express PVALB. Additionally, across all spatial domain, there's gap between 0 and express and hence, $\text{logcount}(\text{PVALB}) > 0$ seems to be a good threshold.

Picking joint bandwidth of 0.0845



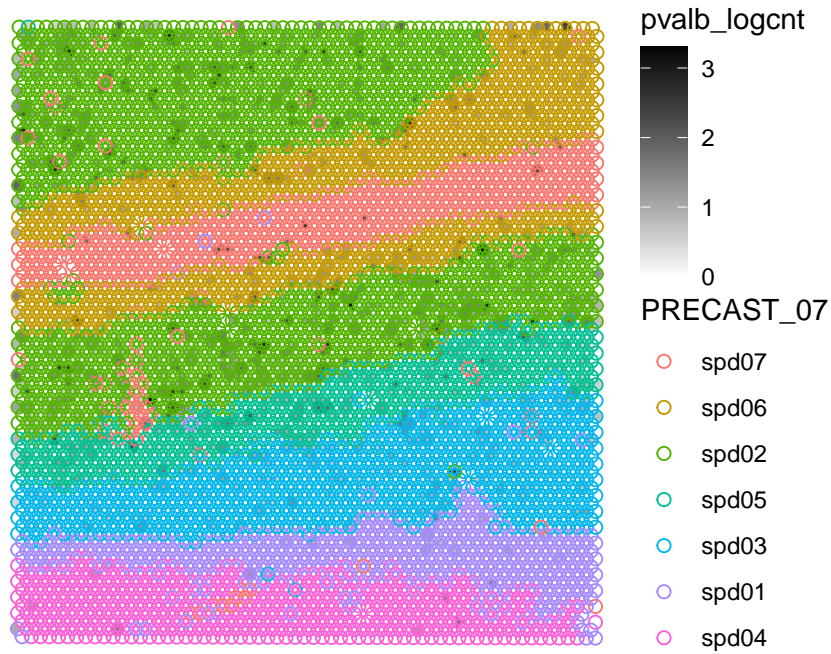
The following plot show the proportion (color) of PVALB+ spots across spatial domains, with and without accounts for individual donor.



Here is the spot plot showing the gene expression of PVALB (log2count) - also showing higher gene expression in SpD02 (L3/4)

Scale for fill is already present.

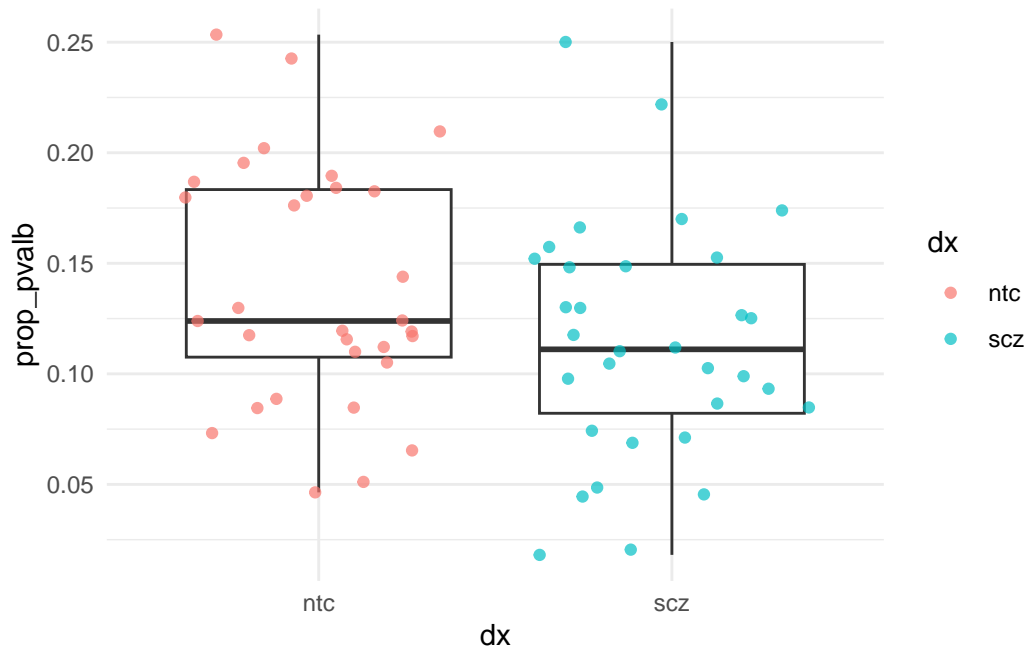
Adding another scale for fill, which will replace the existing scale.



PVALB+ Spots (Not very informative)

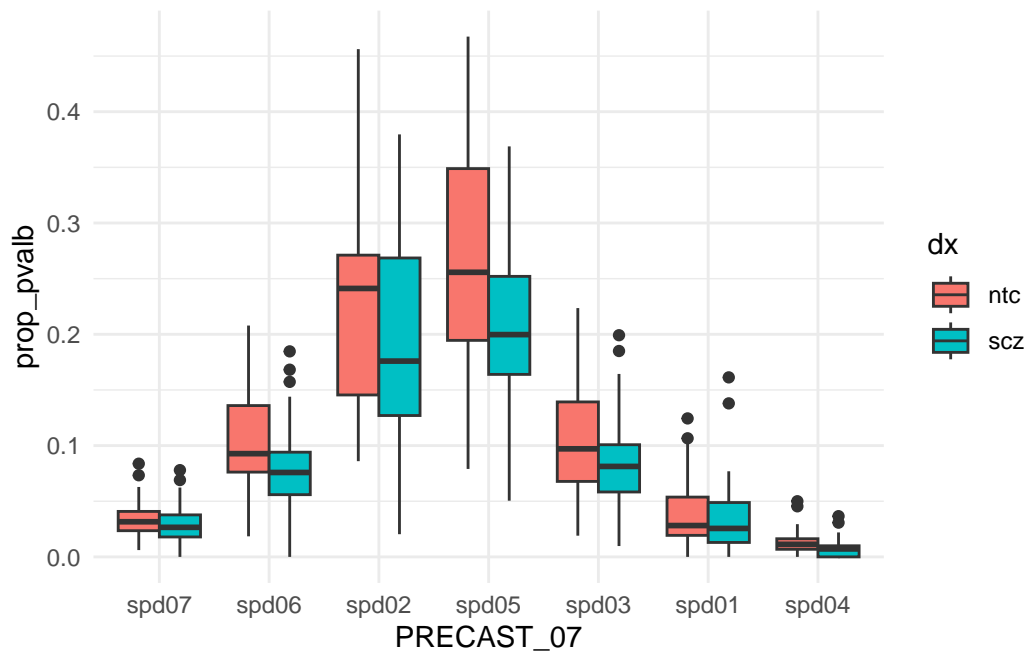
In this section, we conduct descriptive analysis and visualization among all donors/samples.

We first visualize the proportion of PVALB+ spots within each donor across diagnosis group. Qualitatively, it seems like the proportion of pvalb+ spots is slightly higher among NTC group than SCZ group.

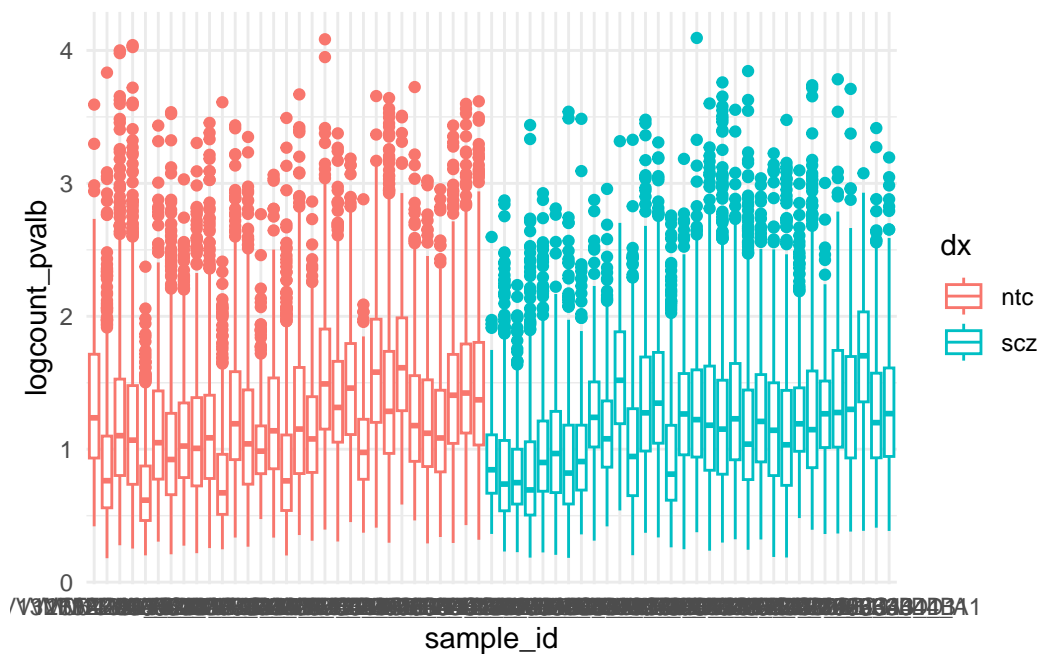


Here is the boxplot accounts for spatial domain. There's more PVALB+ spots in SpD02(L3/4) and SpD05(L5). It still seems like there's more pvalb+ spots among NTC group than SCZ group.

``summarise()`` has grouped output by 'sample_id'. You can override using the ``.groups`` argument.



Here it visualizes the distribution of gene expression, instead of the proportion of positive spots. (Probably not that useful though)

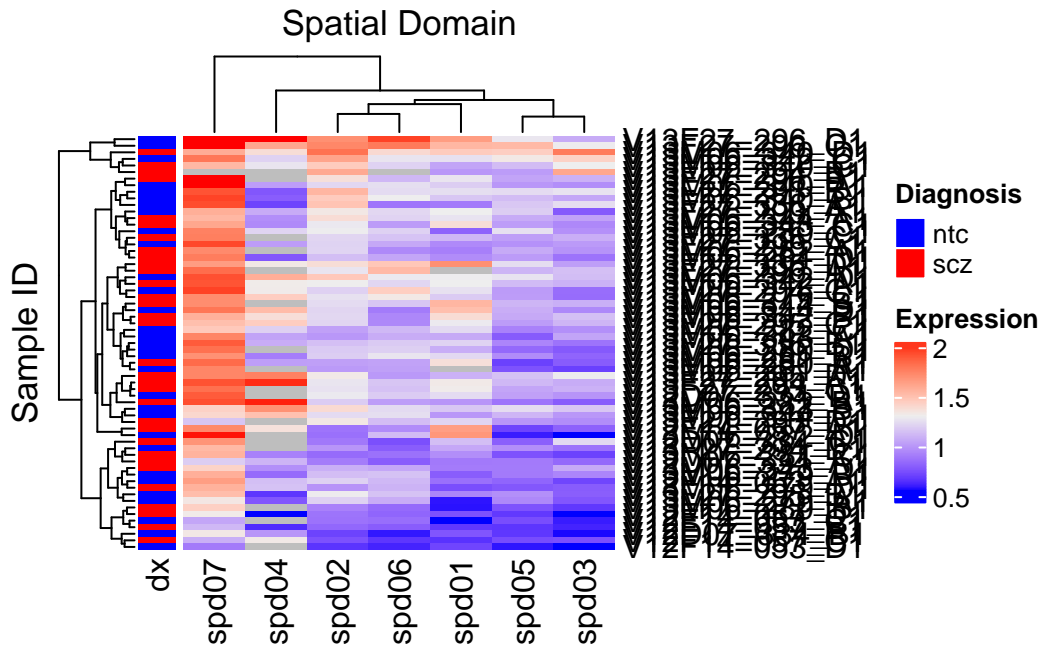


Heatmap showing median logcount(PVALB) among PVALB+ spots per spatial domain. I

don't think this is that informative.

NOTE: Here we visualize the median expression of PVALB+ spots per SpD and sample. This is different from visualize the mean/median PVALB expression of per SpD-sample_id

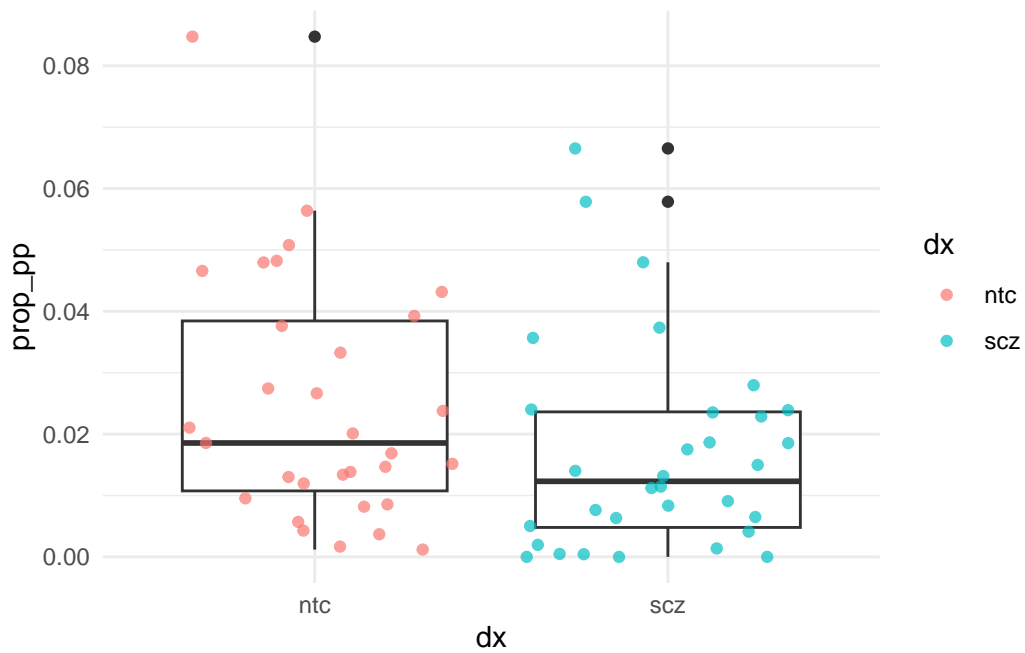
``summarise()`` has grouped output by 'sample_id'. You can override using the ``groups`` argument.



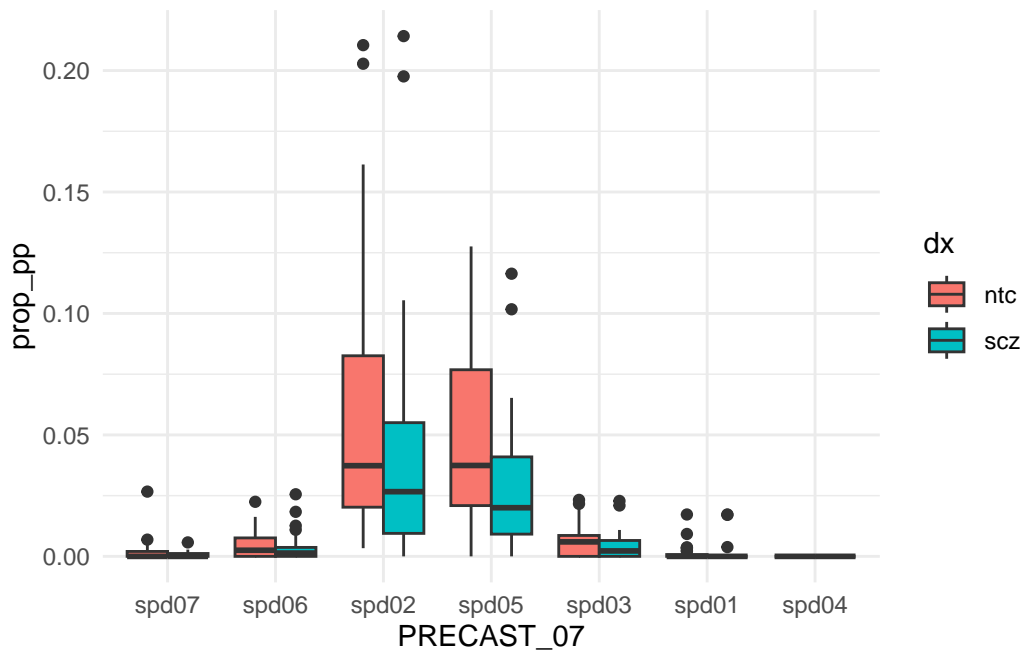
PNN+ & PVALB+ Spots

In this section, we aim to investigate the PNN+ and PVALB+ spots, ie. spots that have both PNN expression and PVALB expression.

Similarly, we first visualize the proportion of PNN+/PVALB+ spots accounts for domain or not.

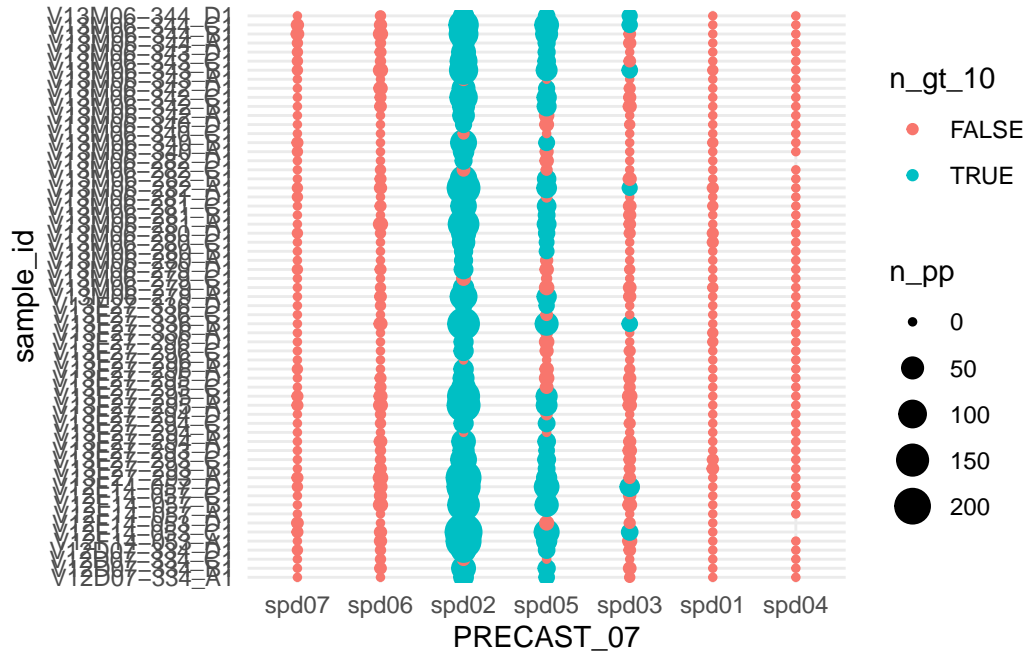


``summarise()`` has grouped output by 'sample_id'. You can override using the ``groups`` argument.



Overall, we can see that there's not a lot of PNN+/PVALB+ spots in most of the spatial domain, etc Layer 3/4 and L5.

``summarise()`` has grouped output by `'sample_id'`. You can override using the ``groups`` argument.



Session Info

```
- Session info -----
setting  value
version  R version 4.4.1 (2024-06-14)
os       macOS Sonoma 14.6.1
system   aarch64, darwin20
ui       X11
language (EN)
collate  en_US.UTF-8
ctype    en_US.UTF-8
tz       America/New_York
date     2024-10-30
pandoc   3.1.12.1 @ /opt/homebrew/bin/ (via rmarkdown)
```

```

- Packages -----
package      * version date (UTC) lib source
abind        1.4-8   2024-09-12 [1] CRAN (R 4.4.1)
beeswarm     0.4.0   2021-06-01 [1] CRAN (R 4.4.0)
Biobase      * 2.64.0  2024-04-30 [1] Bioconductor 3.19 (R 4.4.0)
BiocGenerics * 0.50.0  2024-04-30 [1] Bioconductor 3.19 (R 4.4.0)
Cairo        1.6-2   2023-11-28 [1] CRAN (R 4.4.0)
circlize     0.4.16  2024-02-20 [1] CRAN (R 4.4.0)
cli          3.6.3   2024-06-21 [1] CRAN (R 4.4.0)
clue         0.3-65  2023-09-23 [1] CRAN (R 4.4.0)
cluster      2.1.6   2023-12-01 [1] CRAN (R 4.4.1)
codetools    0.2-20  2024-03-31 [1] CRAN (R 4.4.1)
colorspace   2.1-1   2024-07-26 [1] CRAN (R 4.4.0)
ComplexHeatmap * 2.20.0  2024-04-30 [1] Bioconductor 3.19 (R 4.4.0)
crayon       1.5.3   2024-06-20 [1] CRAN (R 4.4.0)
DelayedArray 0.30.1  2024-05-30 [1] Bioconductor 3.19 (R 4.4.0)
digest       0.6.37  2024-08-19 [1] CRAN (R 4.4.1)
doParallel   1.0.17  2022-02-07 [1] CRAN (R 4.4.0)
dplyr        * 1.1.4   2023-11-17 [1] CRAN (R 4.4.0)
escheR       * 1.4.0   2024-05-30 [1] Bioconductor 3.19 (R 4.4.0)
evaluate     1.0.0   2024-09-17 [1] CRAN (R 4.4.1)
fanside      1.0.6   2023-12-08 [1] CRAN (R 4.4.0)
farver       2.1.2   2024-05-13 [1] CRAN (R 4.4.0)
fastmap      1.2.0   2024-05-15 [1] CRAN (R 4.4.0)
forcats      * 1.0.0   2023-01-29 [1] CRAN (R 4.4.0)
foreach      1.5.2   2022-02-02 [1] CRAN (R 4.4.0)
generics     0.1.3   2022-07-05 [1] CRAN (R 4.4.0)
GenomeInfoDb * 1.40.1  2024-06-16 [1] Bioconductor 3.19 (R 4.4.0)
GenomeInfoDbData 1.2.12  2024-08-05 [1] Bioconductor
GenomicRanges * 1.56.1  2024-06-16 [1] Bioconductor 3.19 (R 4.4.0)
GetoptLong   1.0.5   2020-12-15 [1] CRAN (R 4.4.0)
ggbeeswarm   * 0.7.2   2023-04-29 [1] CRAN (R 4.4.0)
ggplot2      * 3.5.1   2024-04-23 [1] CRAN (R 4.4.0)
ggridges     * 0.5.6   2024-01-23 [1] CRAN (R 4.4.0)
GlobalOptions 0.1.2   2020-06-10 [1] CRAN (R 4.4.0)
glue         1.8.0   2024-09-30 [1] CRAN (R 4.4.1)
gtable       0.3.5   2024-04-22 [1] CRAN (R 4.4.0)
here         * 1.0.1   2020-12-13 [1] CRAN (R 4.4.0)
hms          1.1.3   2023-03-21 [1] CRAN (R 4.4.0)
htmltools    0.5.8.1 2024-04-04 [1] CRAN (R 4.4.0)
httr         1.4.7   2023-08-15 [1] CRAN (R 4.4.0)
IRanges      * 2.38.1  2024-07-03 [1] Bioconductor 3.19 (R 4.4.1)
iterators    1.0.14  2022-02-05 [1] CRAN (R 4.4.0)

```

| | | | | | |
|----------------------|----------|------------|-----|-------------------|-----------|
| jsonlite | 1.8.9 | 2024-09-20 | [1] | CRAN | (R 4.4.1) |
| knitr | 1.48 | 2024-07-07 | [1] | CRAN | (R 4.4.0) |
| labeling | 0.4.3 | 2023-08-29 | [1] | CRAN | (R 4.4.0) |
| lattice | 0.22-6 | 2024-03-20 | [1] | CRAN | (R 4.4.1) |
| lifecycle | 1.0.4 | 2023-11-07 | [1] | CRAN | (R 4.4.0) |
| lubridate | * 1.9.3 | 2023-09-27 | [1] | CRAN | (R 4.4.0) |
| magick | 2.8.4 | 2024-07-14 | [1] | CRAN | (R 4.4.0) |
| magrittr | 2.0.3 | 2022-03-30 | [1] | CRAN | (R 4.4.0) |
| Matrix | 1.7-0 | 2024-04-26 | [1] | CRAN | (R 4.4.1) |
| MatrixGenerics | * 1.16.0 | 2024-04-30 | [1] | Bioconductor 3.19 | (R 4.4.0) |
| matrixStats | * 1.4.1 | 2024-09-08 | [1] | CRAN | (R 4.4.1) |
| munsell | 0.5.1 | 2024-04-01 | [1] | CRAN | (R 4.4.0) |
| pillar | 1.9.0 | 2023-03-22 | [1] | CRAN | (R 4.4.0) |
| pkgconfig | 2.0.3 | 2019-09-22 | [1] | CRAN | (R 4.4.0) |
| png | 0.1-8 | 2022-11-29 | [1] | CRAN | (R 4.4.0) |
| purrr | * 1.0.2 | 2023-08-10 | [1] | CRAN | (R 4.4.0) |
| R6 | 2.5.1 | 2021-08-19 | [1] | CRAN | (R 4.4.0) |
| RColorBrewer | 1.1-3 | 2022-04-03 | [1] | CRAN | (R 4.4.0) |
| Rcpp | 1.0.13 | 2024-07-17 | [1] | CRAN | (R 4.4.0) |
| readr | * 2.1.5 | 2024-01-10 | [1] | CRAN | (R 4.4.0) |
| rjson | 0.2.23 | 2024-09-16 | [1] | CRAN | (R 4.4.1) |
| rlang | 1.1.4 | 2024-06-04 | [1] | CRAN | (R 4.4.0) |
| rmarkdown | 2.28 | 2024-08-17 | [1] | CRAN | (R 4.4.0) |
| rprojroot | 2.0.4 | 2023-11-05 | [1] | CRAN | (R 4.4.0) |
| S4Arrays | 1.4.1 | 2024-05-30 | [1] | Bioconductor 3.19 | (R 4.4.0) |
| S4Vectors | * 0.42.1 | 2024-07-03 | [1] | Bioconductor 3.19 | (R 4.4.1) |
| scales | 1.3.0 | 2023-11-28 | [1] | CRAN | (R 4.4.0) |
| sessioninfo | * 1.2.2 | 2021-12-06 | [1] | CRAN | (R 4.4.0) |
| shape | 1.4.6.1 | 2024-02-23 | [1] | CRAN | (R 4.4.0) |
| SingleCellExperiment | * 1.26.0 | 2024-04-30 | [1] | Bioconductor 3.19 | (R 4.4.0) |
| SparseArray | 1.4.8 | 2024-05-30 | [1] | Bioconductor 3.19 | (R 4.4.0) |
| SpatialExperiment | * 1.14.0 | 2024-04-30 | [1] | Bioconductor 3.19 | (R 4.4.0) |
| stringi | 1.8.4 | 2024-05-06 | [1] | CRAN | (R 4.4.0) |
| stringr | * 1.5.1 | 2023-11-14 | [1] | CRAN | (R 4.4.0) |
| SummarizedExperiment | * 1.34.0 | 2024-04-30 | [1] | Bioconductor 3.19 | (R 4.4.0) |
| tibble | * 3.2.1 | 2023-03-20 | [1] | CRAN | (R 4.4.0) |
| tidyr | * 1.3.1 | 2024-01-24 | [1] | CRAN | (R 4.4.0) |
| tidyselect | 1.2.1 | 2024-03-11 | [1] | CRAN | (R 4.4.0) |
| tidyverse | * 2.0.0 | 2023-02-22 | [1] | CRAN | (R 4.4.0) |
| timechange | 0.3.0 | 2024-01-18 | [1] | CRAN | (R 4.4.0) |
| tzdb | 0.4.0 | 2023-05-12 | [1] | CRAN | (R 4.4.0) |
| UCSC.utils | 1.0.0 | 2024-05-06 | [1] | Bioconductor 3.19 | (R 4.4.0) |
| utf8 | 1.2.4 | 2023-10-22 | [1] | CRAN | (R 4.4.0) |

| | | | | |
|-------------|--------|------------|-----|-----------------------------|
| vctrs | 0.6.5 | 2023-12-01 | [1] | CRAN (R 4.4.0) |
| vipor | 0.4.7 | 2023-12-18 | [1] | CRAN (R 4.4.0) |
| viridisLite | 0.4.2 | 2023-05-02 | [1] | CRAN (R 4.4.0) |
| withr | 3.0.1 | 2024-07-31 | [1] | CRAN (R 4.4.0) |
| xfun | 0.47 | 2024-08-17 | [1] | CRAN (R 4.4.0) |
| XVector | 0.44.0 | 2024-04-30 | [1] | Bioconductor 3.19 (R 4.4.0) |
| yaml | 2.3.10 | 2024-07-26 | [1] | CRAN (R 4.4.0) |
| zlibbioc | 1.50.0 | 2024-04-30 | [1] | Bioconductor 3.19 (R 4.4.0) |

[1] /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/library
