

V(D)J-seq based plots for Gearty et al. (2021)

Friederike Dündar | Weill Cornell Medicine

2021

Contents

Clonotype frequencies	2
Counting clonotypes	3
Verbose clonotype number statements	4
Per mouse	5
Focus on unique = non-shared) ones	8
Fractions	9
Counting cells	9
Pancreas cells	10
Tcf1-hi cells	11
Tcf1-lo cells	12
How many Tcf.hi make it to the pancreas?	12
Alluvial plots	15

SessionInfo 16

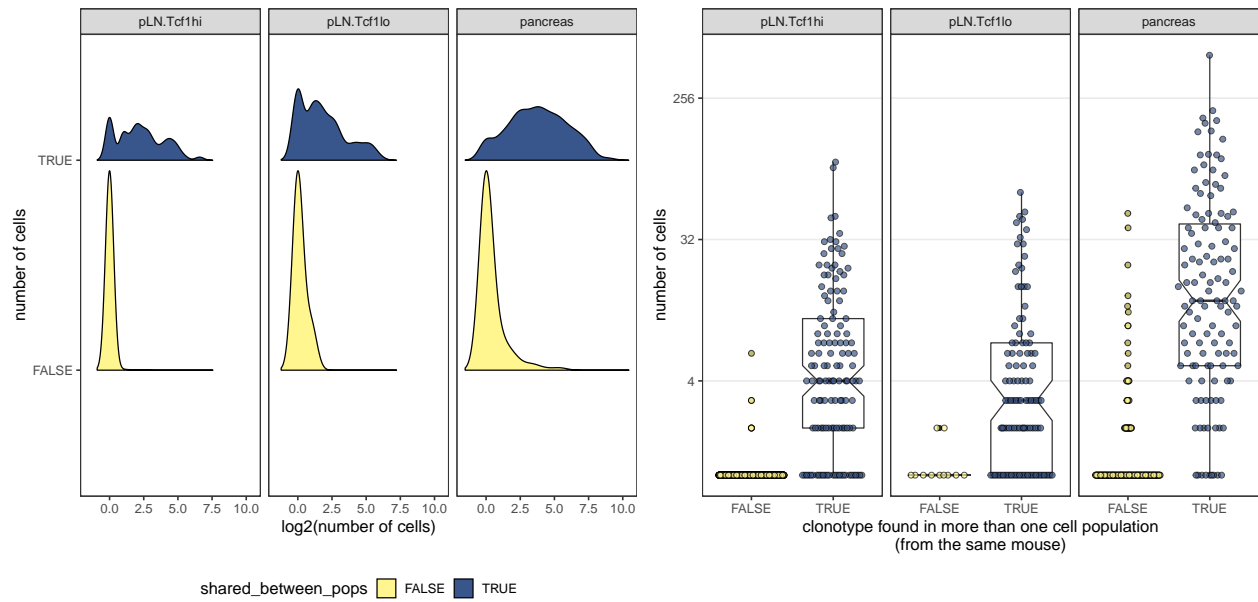
The plots here are based on the results of a V(D)J-seq experiment where CD8 T cells from pancreatic lymphnodes (pLN) and the pancreas (panc) were obtained and their TCR gene sequences were profiled with 10X Genomics' V(D)J sequencing strategy.

Cached data here: ~/Library/Caches/BiocFileCache/7ad848b1a31b_5wj8h9qssai2u0ya2sre2tsudahvvoni.rds

Clonotype frequencies

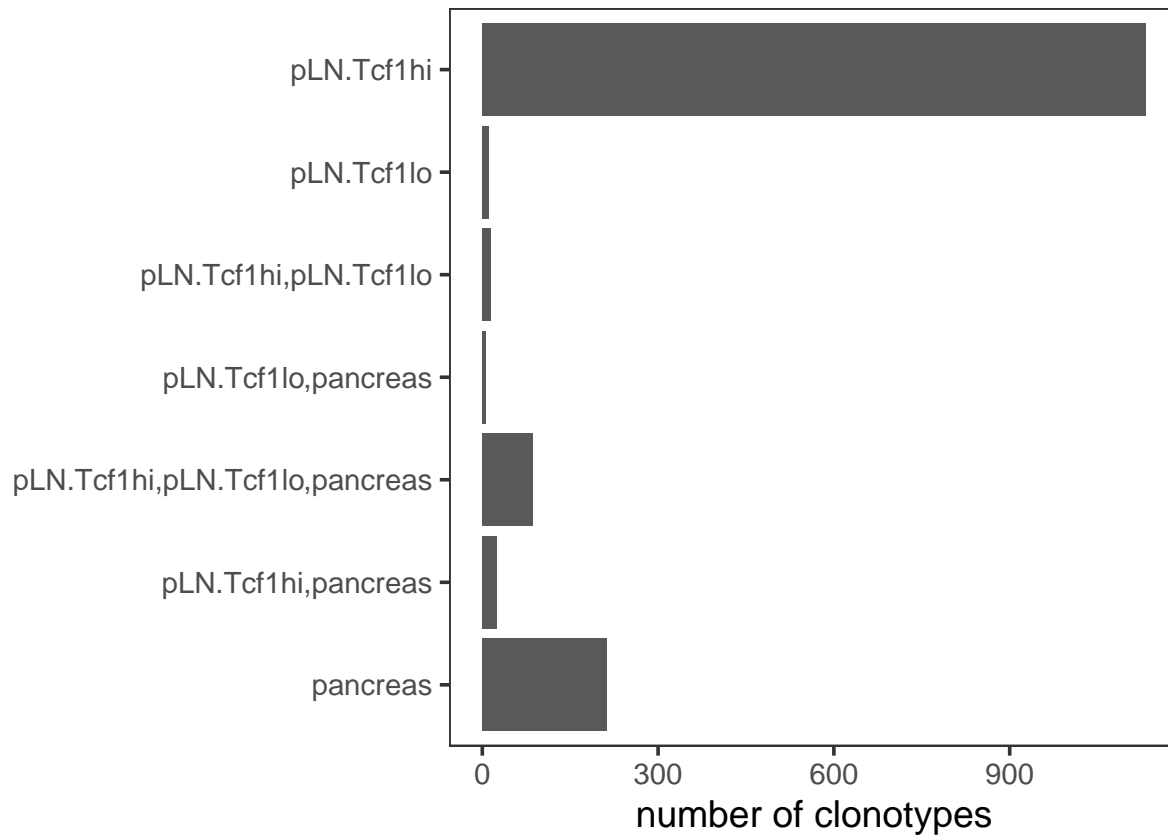
Number of cells per clonotype per population

Split by whether a clonotype is unique or found across multiple populations



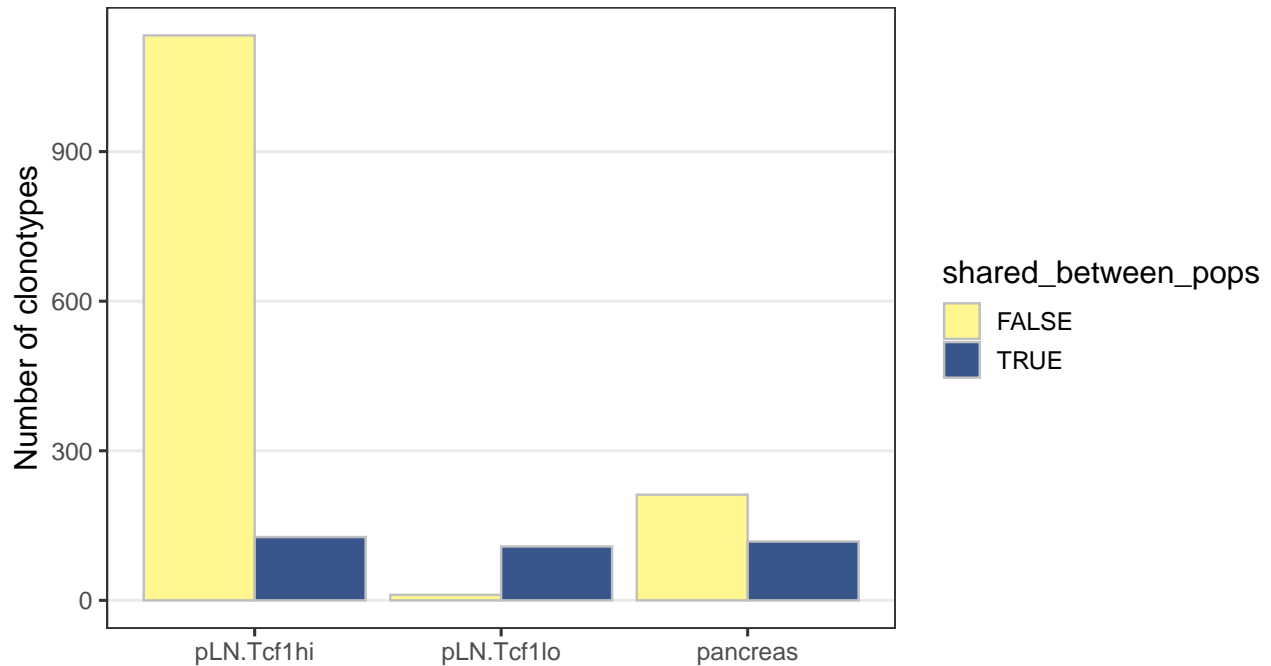
These plots demonstrate that clonotypes that are found in just one population (“FALSE”; yellow) tend to be found in single cells (N=1) whereas clonotypes that are found across multiple populations (“TRUE”; darkblue) are usually found in multiple cells per population.

Counting clonotypes

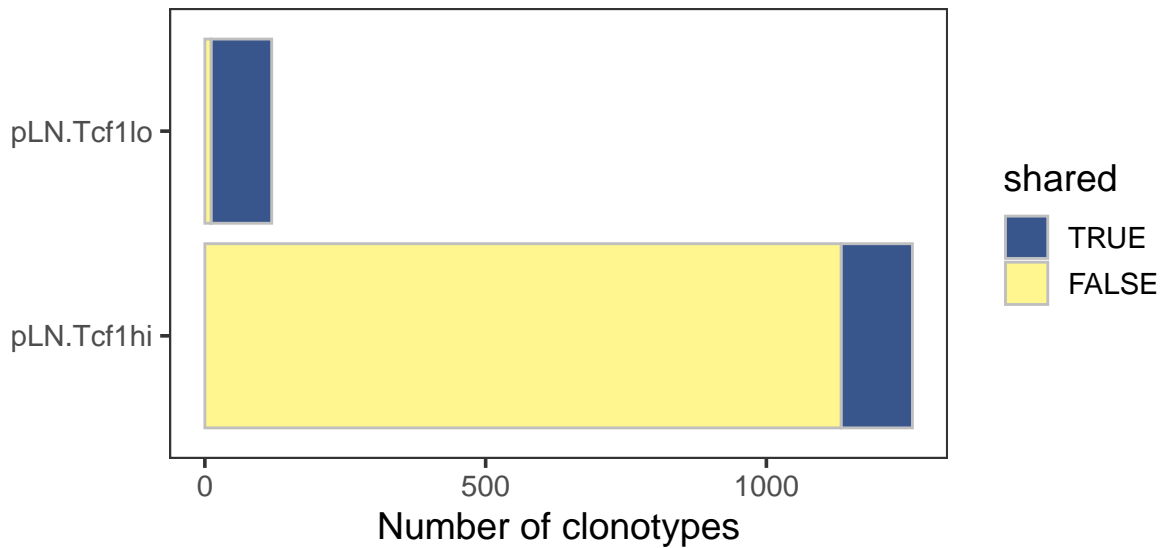


Here are the numbers:

##	allPops	N
## 1:	pLN.Tcf1hi,pLN.Tcf1lo,pancreas	87
## 2:	pLN.Tcf1hi,pancreas	25
## 3:	pLN.Tcf1lo,pancreas	6
## 4:	pancreas	212
## 5:	pLN.Tcf1hi	1133
## 6:	pLN.Tcf1lo	11
## 7:	pLN.Tcf1hi,pLN.Tcf1lo	15



And here's the same figure, just stacked:



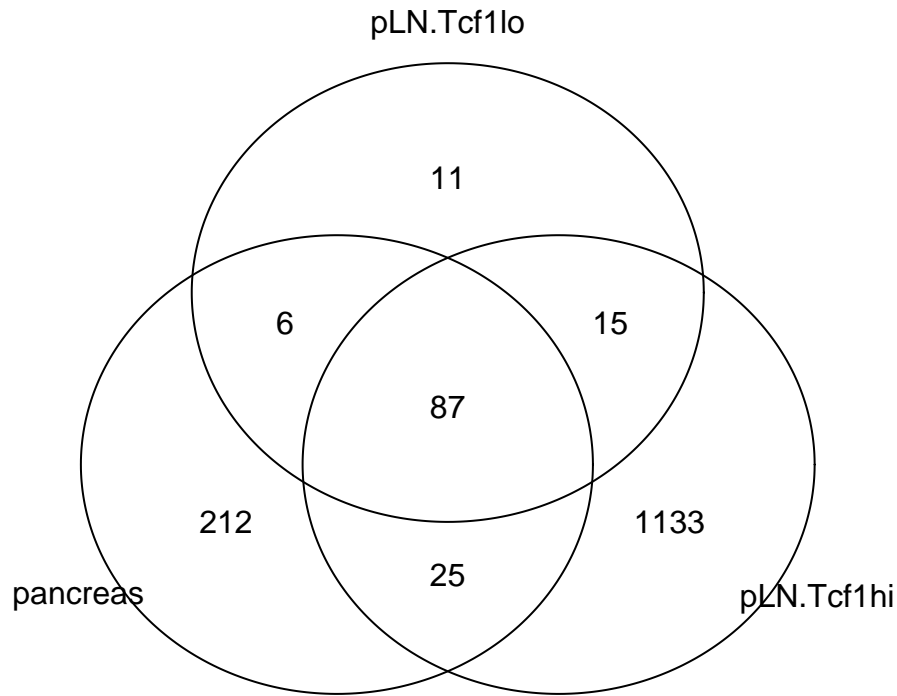
The plots above don't show the counts at the individual mouse level, which may help simplify things. The actual numbers underlying the plots above:

```
##      Tcf.status FALSE TRUE
## 1: pLN.Tcf1hi   1133  127
## 2: pLN.Tcf1lo    11  108
## 3:  pancreas    212  118
```

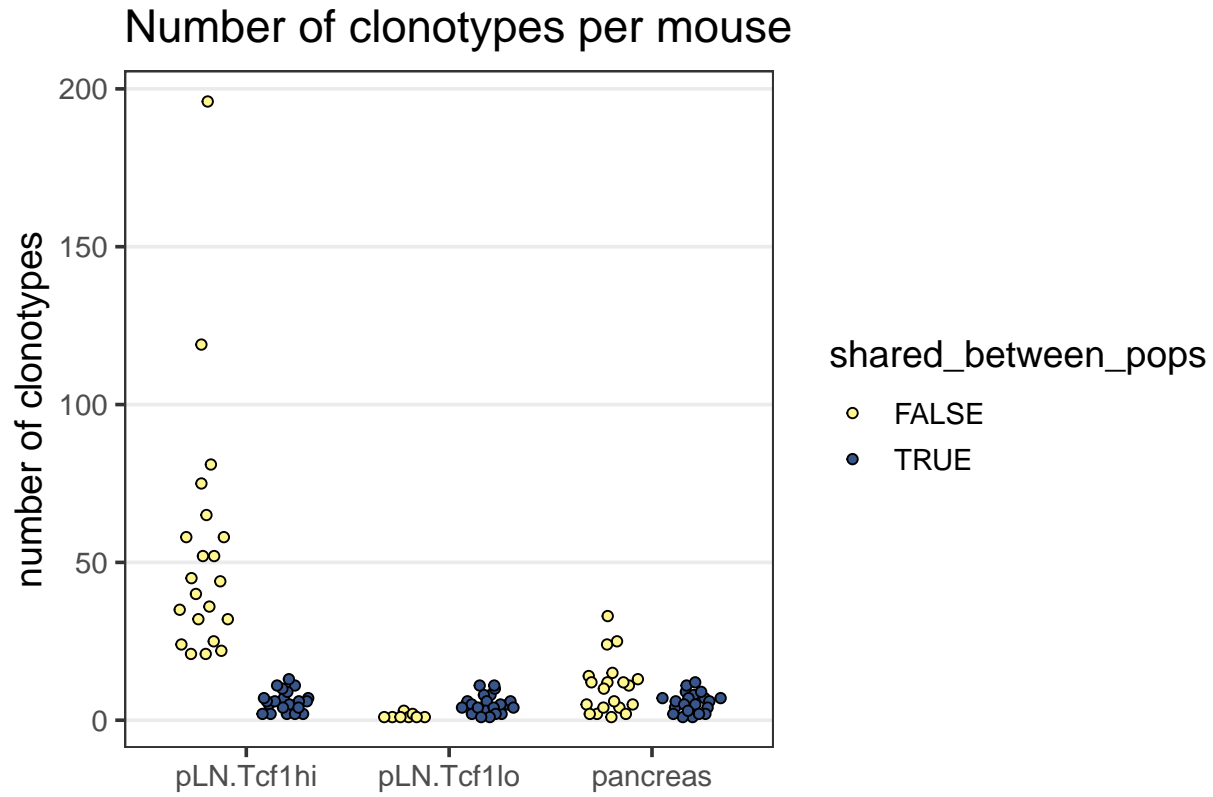
Verbose clonotype number statements

- There are 1489 clonotypes **across all cells** of which 133 are found in more than one population (“shared”).
- There are 330 clonotypes in the **pancreas** of which 118 are shared with at least one pLN population.

- Of these 330 shared pancreas clonotypes 112 are found in the Tcf1hi population (and 93 in the Tcf1-lo population)
- There are 1260 in the **pLN.Tcf1hi population** of which 127 are shared with other populations.
 - Of these 127 shared clonotypes, `length(unique(tcrCells[Tcf.status == "pLN.Tcf1hi" & new_cid %in% shared_across_pops & new_cid %in% tcfHI_and_panc]$new_cid))` are found in the pancreas, too.



Per mouse To bring the individual mouse information back in, we can use scatter plots again:



##	Mouse	Tcf.status	FALSE	TRUE
##	1: m.2024	pLN.Tcf1hi	58	7
##	2: m.2024	pLN.Tcf1lo	1	5
##	3: m.2024	pancreas	4	7
##	4: m.2412	pLN.Tcf1hi	58	9
##	5: m.2412	pLN.Tcf1lo	NA	8
##	6: m.2412	pancreas	12	8
##	7: m.2413	pLN.Tcf1hi	44	11
##	8: m.2413	pLN.Tcf1lo	NA	10
##	9: m.2413	pancreas	24	11
##	10: m.2414	pLN.Tcf1hi	32	5
##	11: m.2414	pLN.Tcf1lo	3	6
##	12: m.2414	pancreas	11	6
##	13: m.2808	pLN.Tcf1hi	22	4
##	14: m.2808	pLN.Tcf1lo	NA	1
##	15: m.2808	pancreas	33	5
##	16: m.2810	pLN.Tcf1hi	52	10
##	17: m.2810	pLN.Tcf1lo	NA	8
##	18: m.2810	pancreas	14	9
##	19: m.2811	pLN.Tcf1hi	40	7
##	20: m.2811	pLN.Tcf1lo	NA	6
##	21: m.2811	pancreas	15	7
##	22: m.2812	pLN.Tcf1hi	36	6
##	23: m.2812	pLN.Tcf1lo	NA	4
##	24: m.2812	pancreas	10	6
##	25: m.2985	pLN.Tcf1hi	45	2
##	26: m.2985	pLN.Tcf1lo	NA	3
##	27: m.2985	pancreas	4	3
##	28: m.2989	pLN.Tcf1hi	21	7

```

## 29: m.2989 pLN.Tcf1lo    NA    4
## 30: m.2989 pancreas      12    7
## 31: m.3758 pLN.Tcf1hi    25    2
## 32: m.3758 pLN.Tcf1lo    NA    2
## 33: m.3758 pancreas      25    2
## 34: m.3759 pLN.Tcf1hi    32    5
## 35: m.3759 pLN.Tcf1lo    NA    4
## 36: m.3759 pancreas       5    5
## 37: m.3761 pLN.Tcf1hi    35    6
## 38: m.3761 pLN.Tcf1lo     1    5
## 39: m.3761 pancreas      12    7
## 40: m.3762 pLN.Tcf1hi   196    2
## 41: m.3762 pLN.Tcf1lo     1    2
## 42: m.3762 pancreas       1    1
## 43: m.3766 pLN.Tcf1hi   119   13
## 44: m.3766 pLN.Tcf1lo     1   11
## 45: m.3766 pancreas      13   12
## 46: m.3770 pLN.Tcf1hi    21    2
## 47: m.3770 pLN.Tcf1lo    NA    1
## 48: m.3770 pancreas       2    2
## 49: m.3771 pLN.Tcf1hi    24    4
## 50: m.3771 pLN.Tcf1lo     2    4
## 51: m.3771 pancreas       2    4
## 52: m.3773 pLN.Tcf1hi    52    2
## 53: m.3773 pLN.Tcf1lo     1    2
## 54: m.3773 pancreas       5    1
## 55: m.3774 pLN.Tcf1hi    65   11
## 56: m.3774 pLN.Tcf1lo    NA   11
## 57: m.3774 pancreas       6    9
## 58: m.3778 pLN.Tcf1hi    75    6
## 59: m.3778 pLN.Tcf1lo    NA    6
## 60: m.3778 pancreas      NA    2
## 61: m.3780 pLN.Tcf1hi    81    6
## 62: m.3780 pLN.Tcf1lo     1    5
## 63: m.3780 pancreas       2    4
##      Mouse Tcf.status FALSE TRUE

```

And here are the mean values:

```

##      Tcf.status    FALSE    TRUE
## 1: pLN.Tcf1hi 53.95238 6.047619
## 2: pLN.Tcf1lo  1.37500 5.142857
## 3: pancreas 10.60000 5.619048

```

I.e., for pLN.Tcf1hi there are, on average, 54 clonotypes per mouse that are not shared (=unique), whereas we detect only 1 to 2 unique clonotypes per mouse in the pLN.Tcf1lo population.

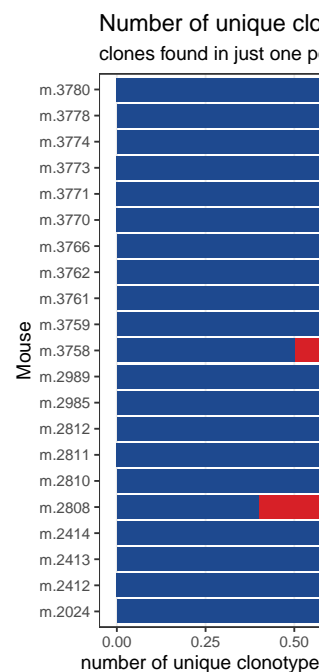
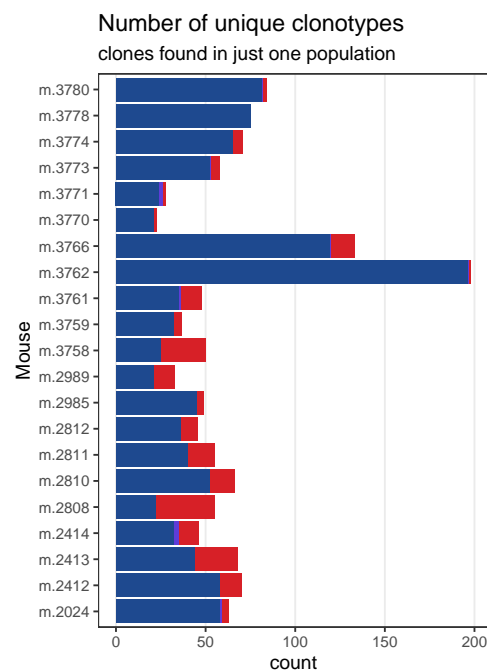
These are the values if we do not split by shared or not-shared:

```

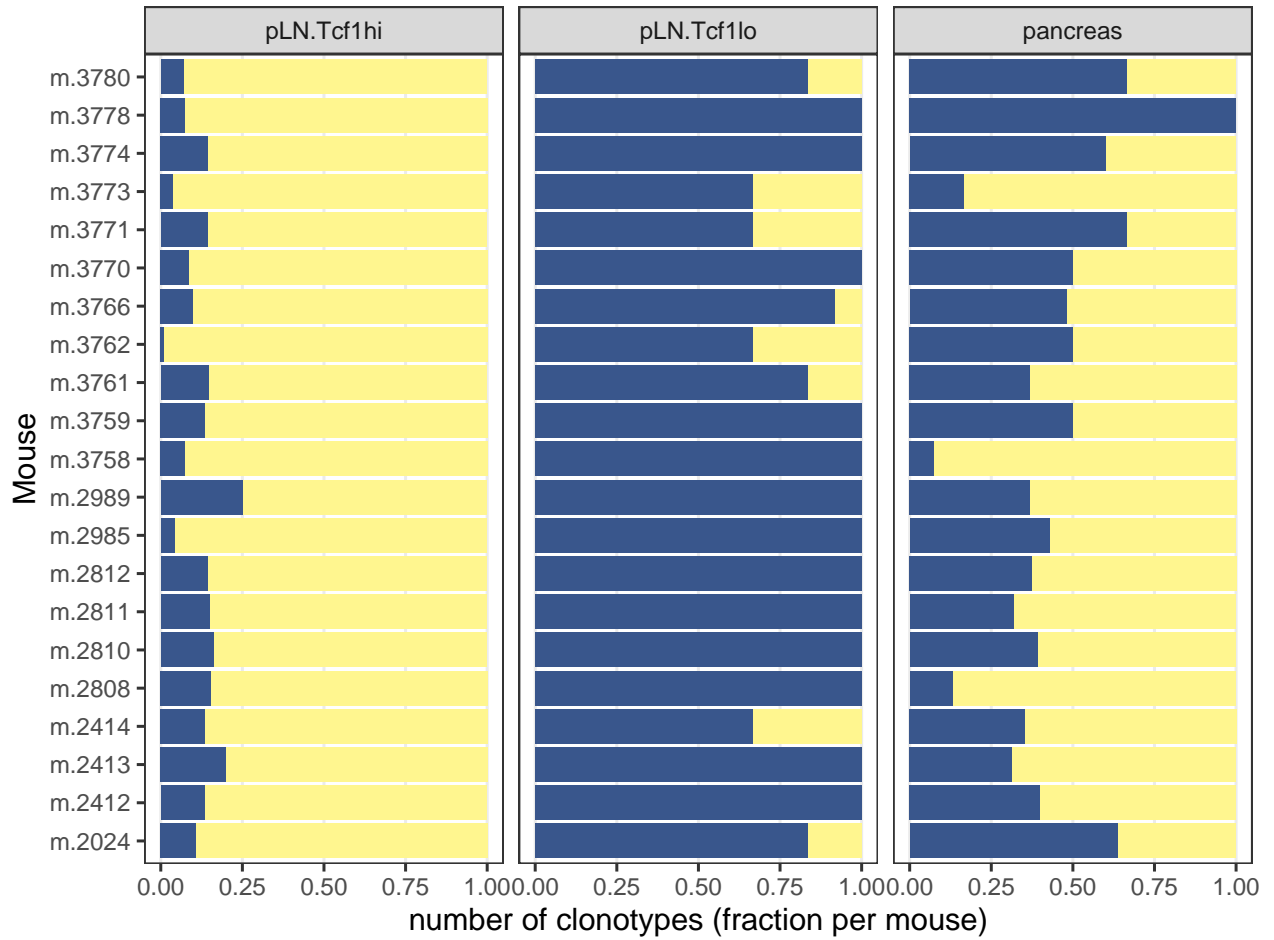
##      Tcf.status      V1
## 1: pancreas 15.714286
## 2: pLN.Tcf1hi 60.000000
## 3: pLN.Tcf1lo  5.666667

```

Focus on unique = non-shared) ones



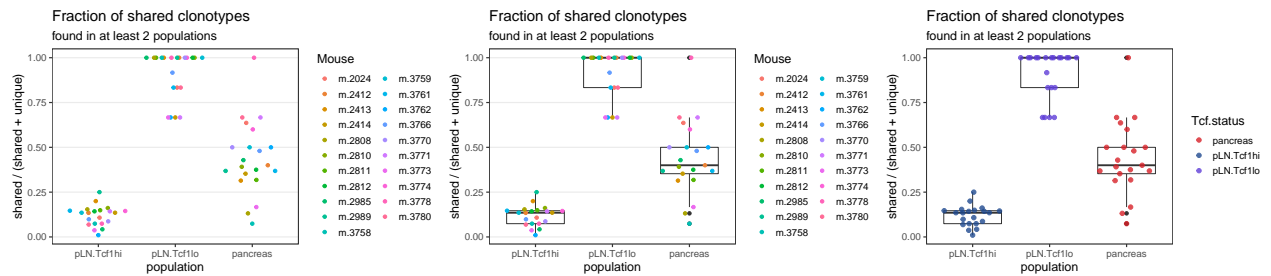
Shared vs. unique clonotypes



Fractions

shared_between_pops FALSE TRUE

The following plots show the same data, just with small tweaks each iteration:



```
## Tcf.status V1
## 1: pancreas 0.4000000
## 2: pLN.Tcf1hi 0.1351351
## 3: pLN.Tcf1lo 1.0000000
```

Counting cells

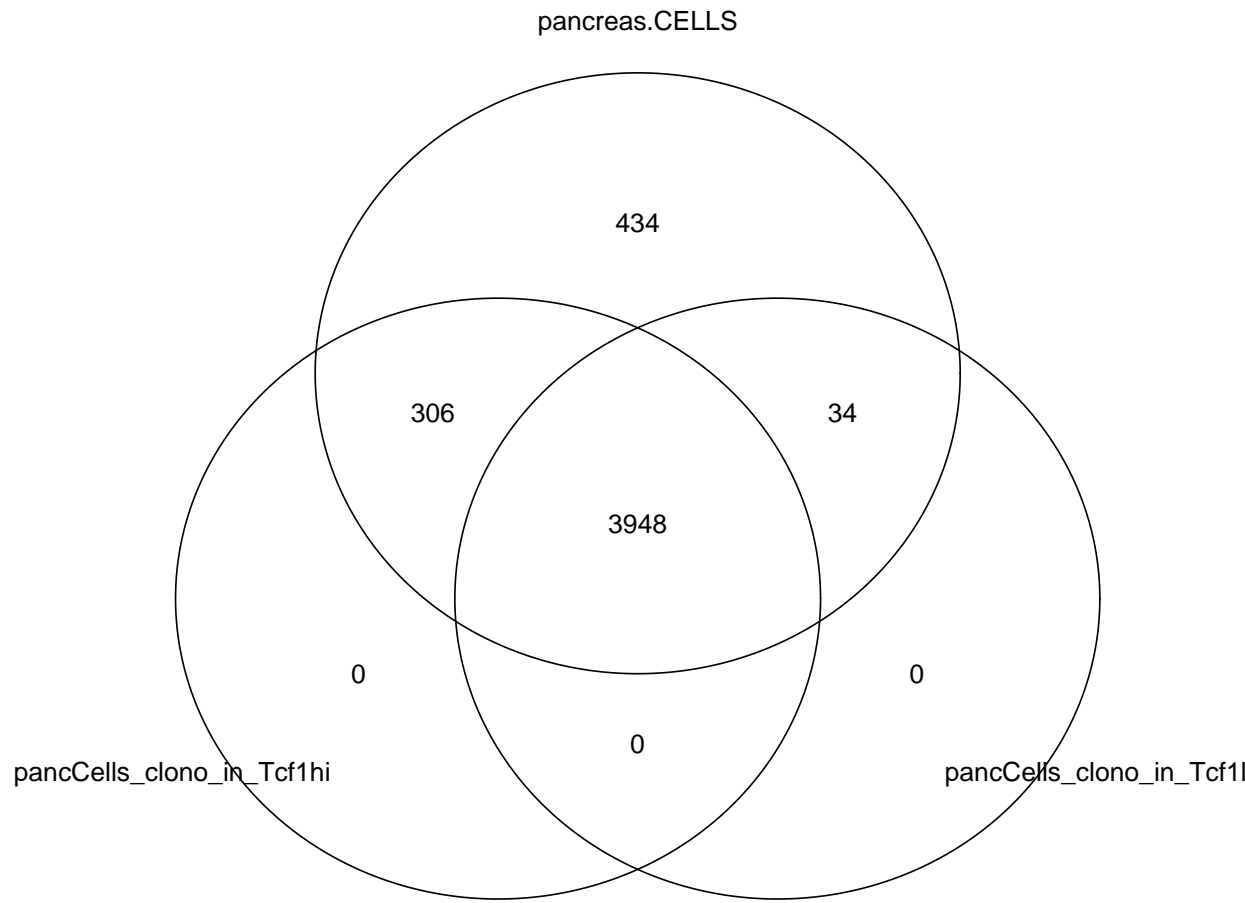
```
## Tcf.status N
## 1: pancreas 4722
```

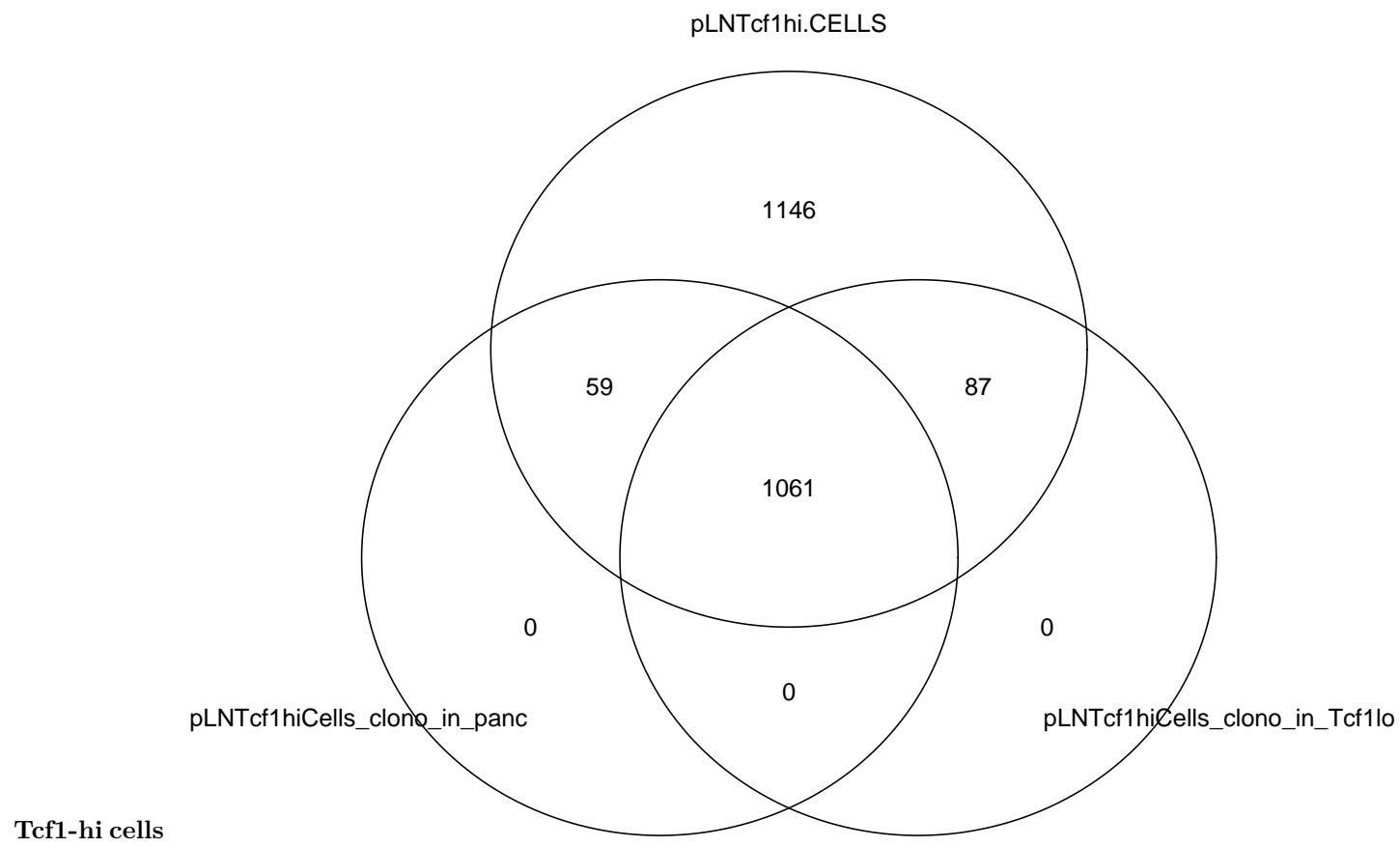
```
## 2: pLN.Tcf1hi 2353
## 3: pLN.Tcf1lo 821
```

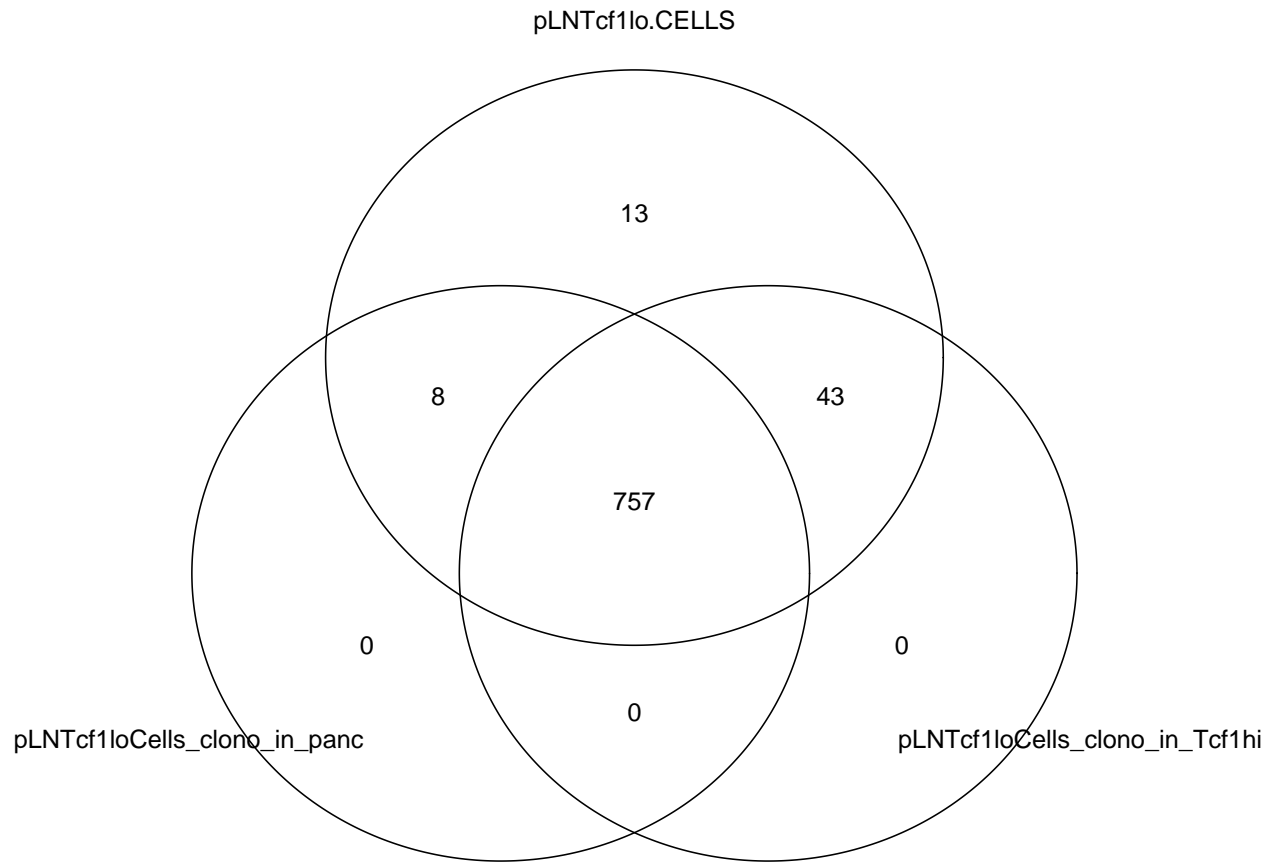
There are 4722 cells in total from the pancreas. 4288 (91%) cells of these have a clonotype that's either found in Tcf1.hi and/or Tcf1.lo in the pLN ("shared"). `nrow(tcrCells[Tissue == "panc" & new_cid %in% shared_across_pops & new_cid %in% tcfHI_and_panc])` (89%) cells have a clonotype that's shared with pLN.Tcf1.hi (and possibly *also* in Tcf1.lo)

In the pLN.Tcf1.hi-population, we have 1120 (of 2353 total) cells that have a clonotype that's found in the pancreas, too.

The following Venn diagrams depict the **number of cells** from each population that have clonotypes that are either exclusive to that population or found in any of the other populations, too.



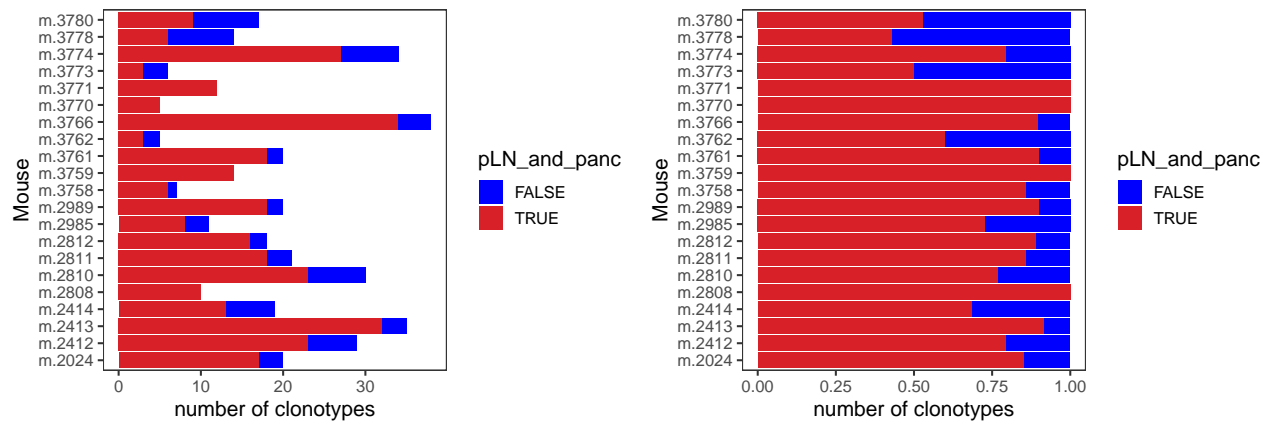




Tcf1-lo cells

How many Tcf.hi make it to the pancreas? Here, I extract all clonotypes that aren't unique; this includes cases where a clonotype might only be found in pLN-Tcf.lo + pancreas:

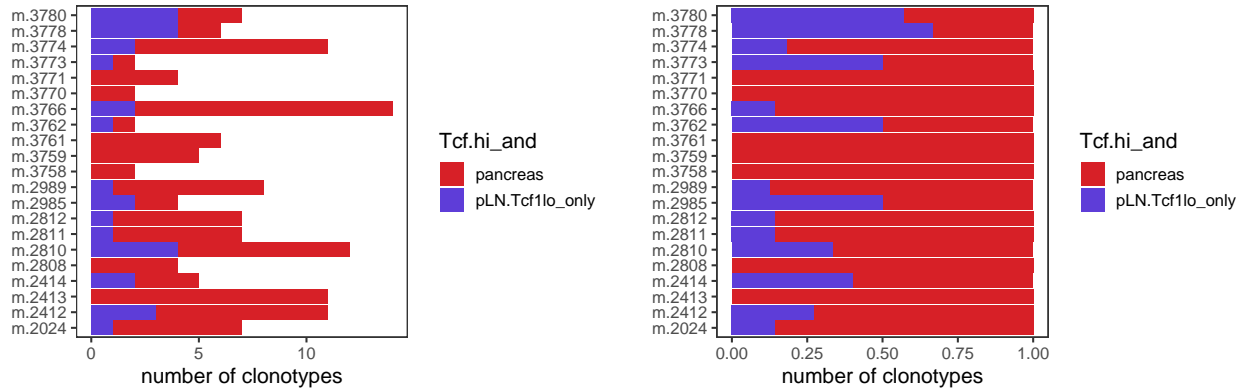
Non-unique clonotypes



And here we focus on clonotypes where at least one cell is found in the pLN.Tcf1hi population, more directly addressing the question of how many Tcf1-hi-cells make it to the pancreas.

How often do Tcf1^{hi} cells that we find in multiple populations end up in the pancreas?

Non-unique clonotypes with min. 1 cell in the pLN-Tcf1hi population



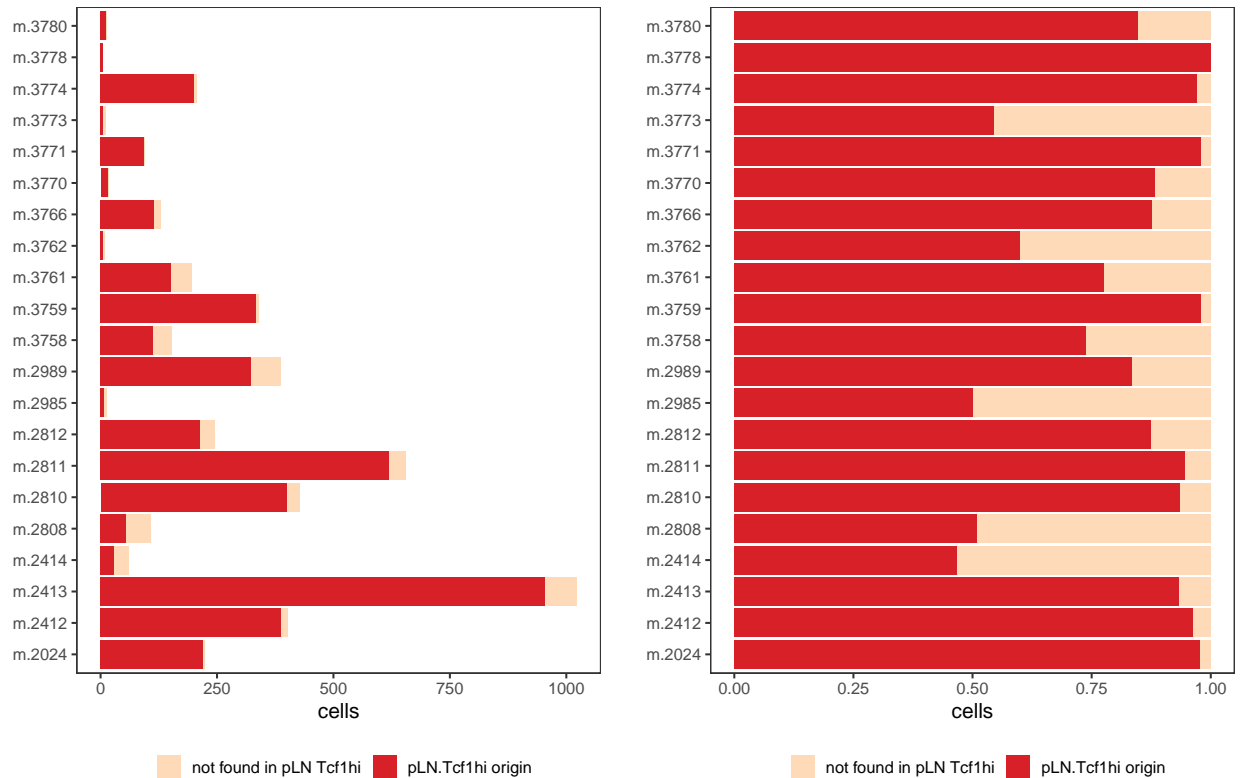
This shows that there is some variability between mice, but in most cases, we see well above 50% of the clonotypes ending up in the pancreas.

We can try to translate that to cell numbers (instead of clonotype numbers) to get a better sense of how many cells likely had their origin in pLN1-Tcf1hi. That also illustrates the point of the expansion of those clonotypes that we actually do find shared. I.e. many of the unique clonotypes in the pancreas are also not expanded.

How many *cells* in the pancreas represent clonotypes that we also see in pLN.Tcf1hi?

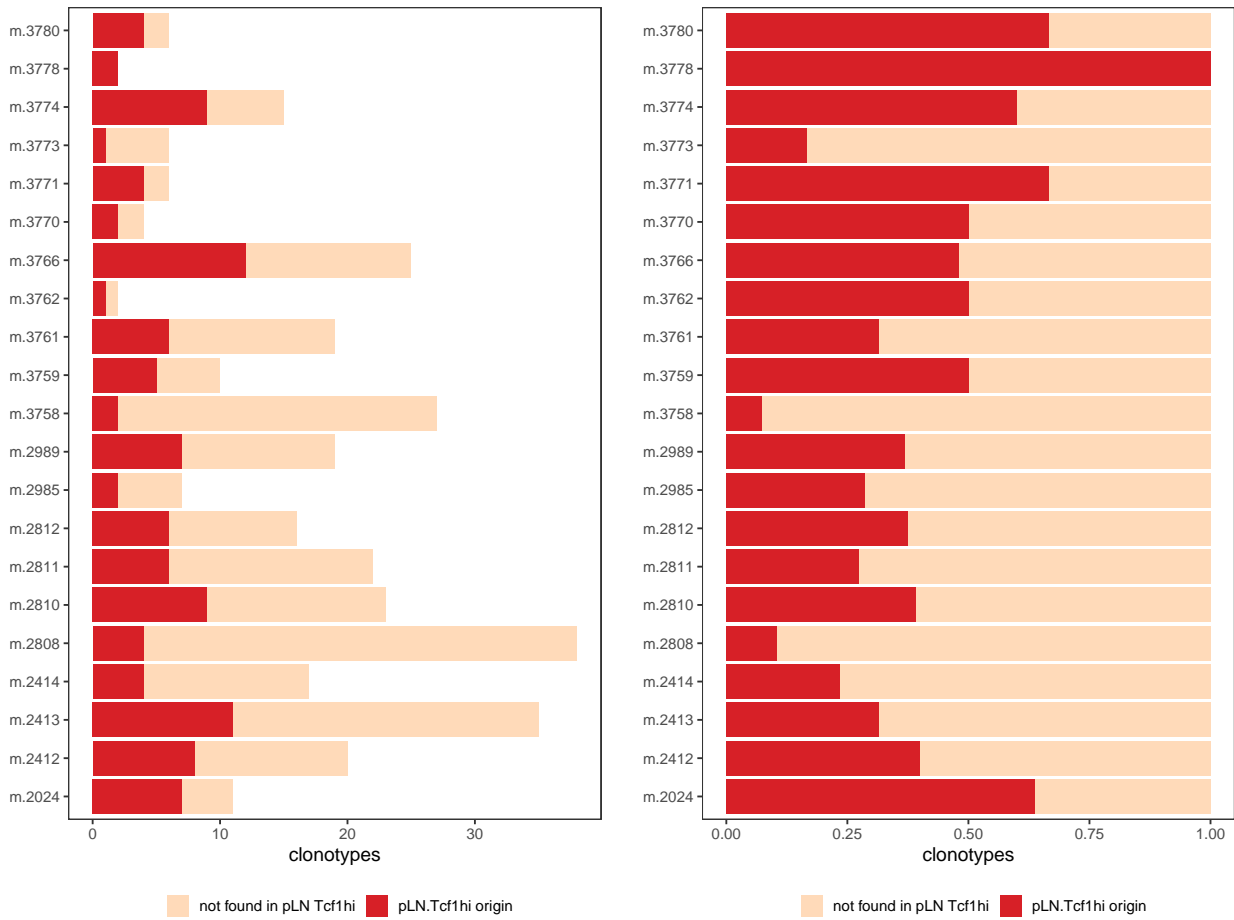
```
##
## not found in pLN Tcf1hi      pLN.Tcf1hi origin
##                               468          4254
```

Where do cells from the pancreas originate from?



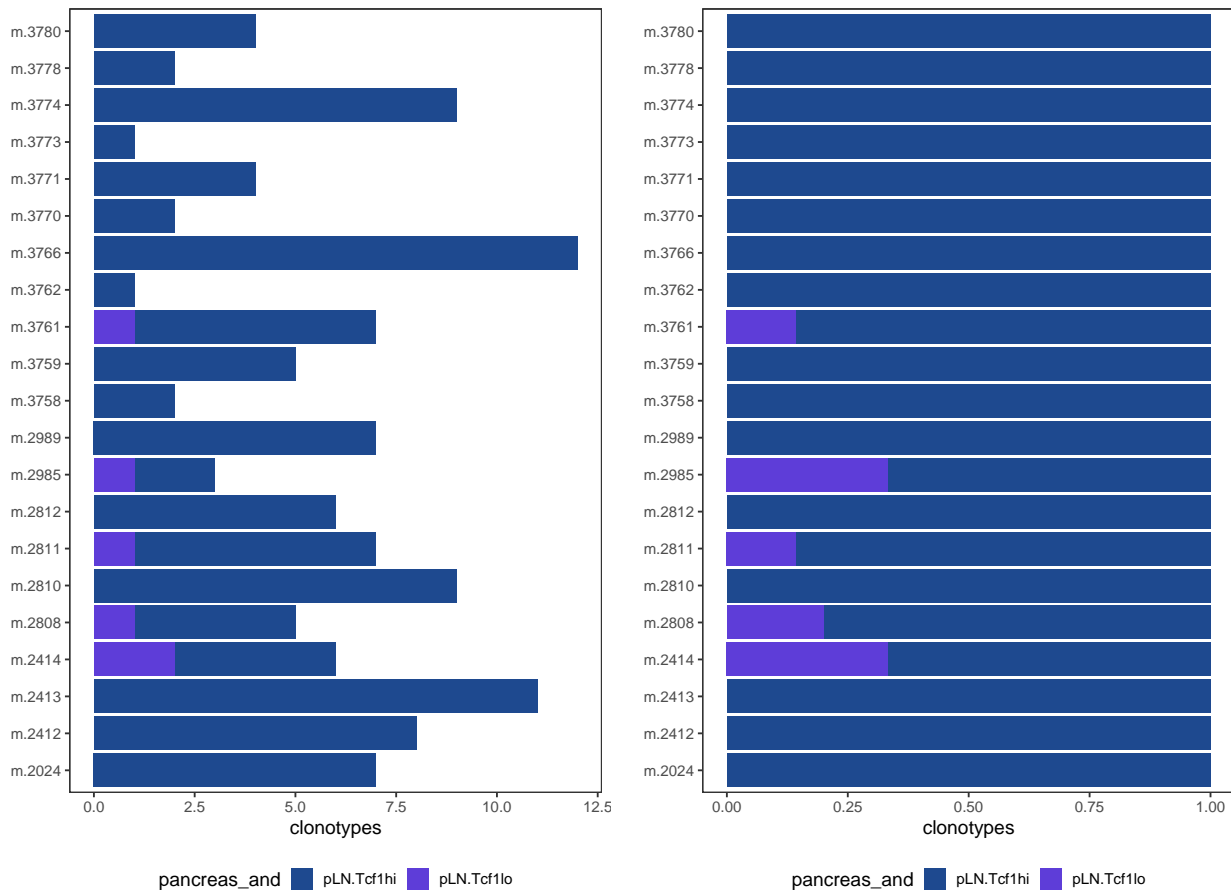
And, correspondingly, going back to the clonotypes: Pick the **clonotypes** found in the pancreas and color them based on whether they're also seen in the Tcf1-hi population.

Do we find clonotypes from the pancreas in our pLN.Tcf1–hi population?



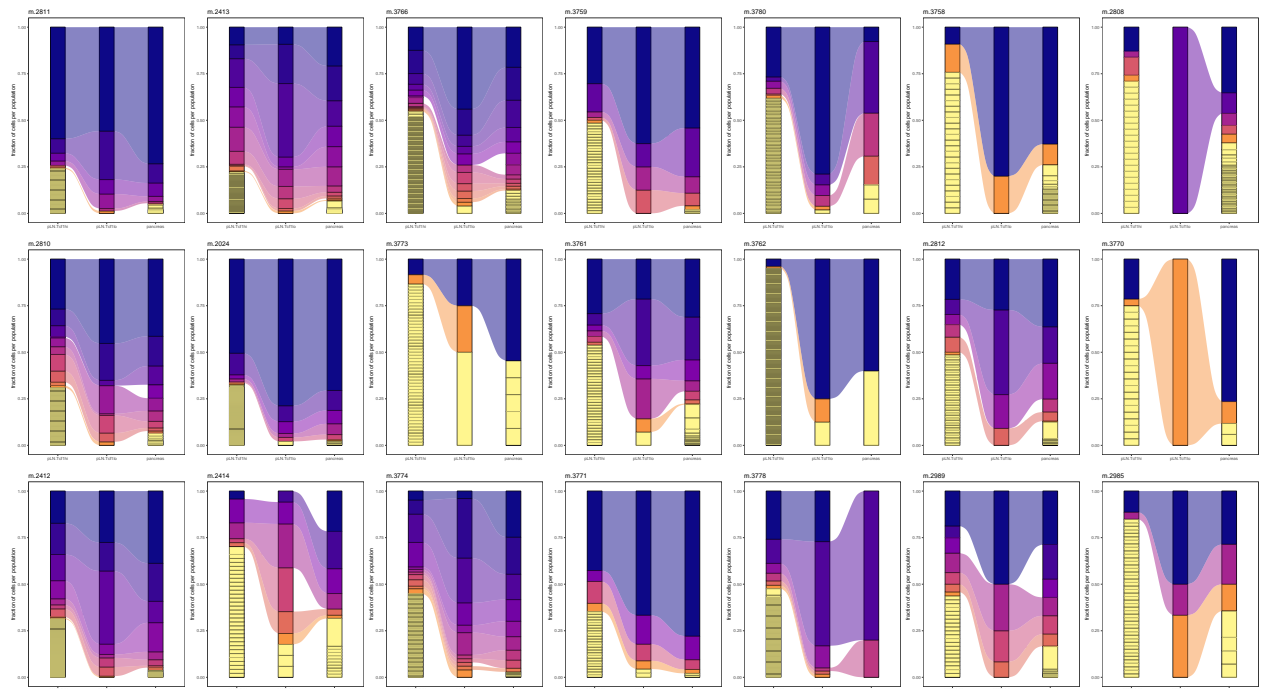
And focusing just on the clonotypes that we find in both pancreas and pLN, how many of those come from pLN-hi?

Do we find clonotypes from the pancreas that we also see in the pLN primarily in either Tcf population?
 focus on clonotypes found in panc + pLN

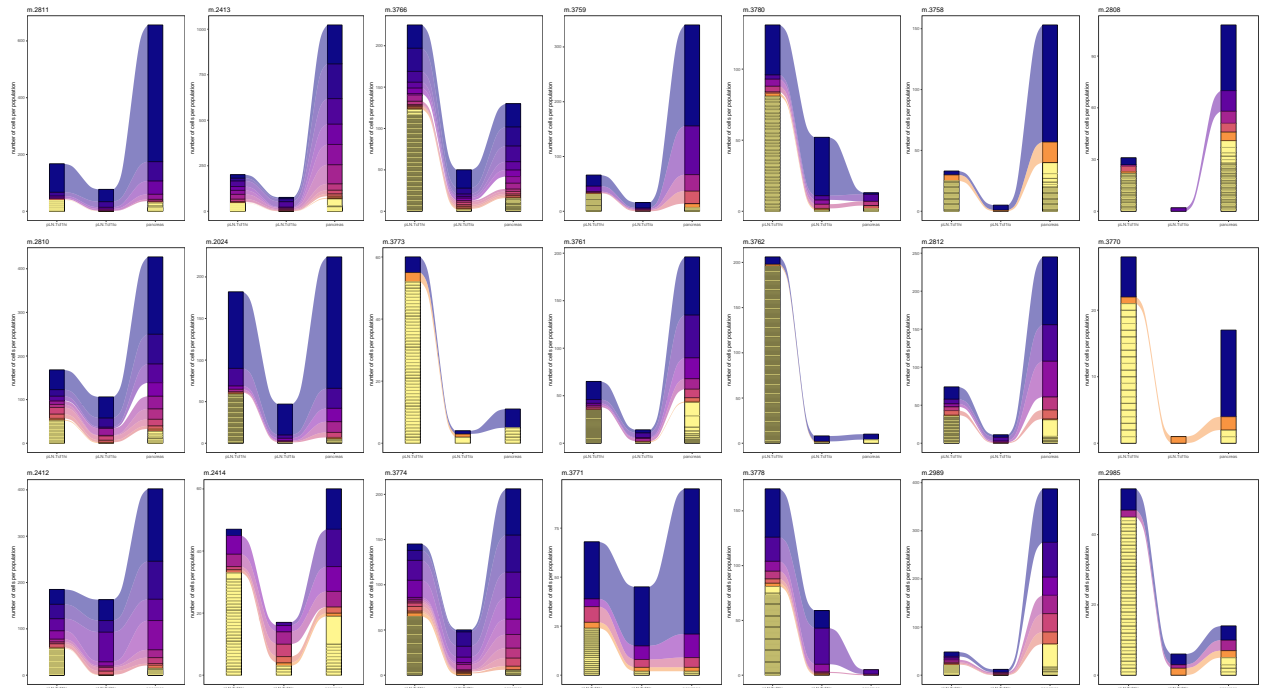


Alluvial plots

The plots below have all clonotypes that are NOT SHARED among different types of cells shown in YELLOW.
 All clonotypes that are NOT YELLOW are present in **>1 population of cells**.



Here are the absolute counts instead of the frequencies:



SessionInfo

```
## R version 4.0.2 (2020-06-22)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Catalina 10.15.7
##
```



```

## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] parallel stats4 stats graphics grDevices utils datasets
## [8] methods base
##
## other attached packages:
## [1] BiocFileCache_1.12.1 dbplyr_2.0.0
## [3] scales_1.1.1 ggalluvial_0.12.3
## [5] patchwork_1.1.1 ggplot2_3.3.3
## [7] data.table_1.13.6 magrittr_2.0.1
## [9] SingleCellExperiment_1.10.1 SummarizedExperiment_1.18.2
## [11] DelayedArray_0.14.1 matrixStats_0.57.0
## [13] Biobase_2.48.0 GenomicRanges_1.40.0
## [15] GenomeInfoDb_1.24.2 IRanges_2.22.2
## [17] S4Vectors_0.26.1 BiocGenerics_0.34.0
##
## loaded via a namespace (and not attached):
## [1] httr_1.4.2 tidyr_1.1.2 viridisLite_0.3.0
## [4] bit64_4.0.5 gtools_3.8.2 assertthat_0.2.1
## [7] blob_1.2.1 GenomeInfoDbData_1.2.3 vipor_0.4.5
## [10] yaml_2.2.1 pillar_1.4.7 RSQlite_2.2.2
## [13] lattice_0.20-41 glue_1.4.2 digest_0.6.27
## [16] XVector_0.28.0 rvest_0.3.6 colorspace_2.0-0
## [19] htmltools_0.5.1.1 Matrix_1.3-2 plyr_1.8.6
## [22] pkgconfig_2.0.3 zlibbioc_1.34.0 purrr_0.3.4
## [25] webshot_0.5.2 tibble_3.0.4 generics_0.1.0
## [28] farver_2.0.3 ellipsis_0.3.2 withr_2.3.0
## [31] crayon_1.3.4 memoise_1.1.0 evaluate_0.14
## [34] gplots_3.1.1 xml2_1.3.2 beeswarm_0.2.3
## [37] tools_4.0.2 lifecycle_0.2.0 stringr_1.4.0
## [40] drat_0.1.8 munsell_0.5.0 kableExtra_1.3.1
## [43] compiler_4.0.2 caTools_1.18.1 rlang_0.4.11
## [46] grid_4.0.2 ABCutilities_0.3.3 RCurl_1.98-1.2
## [49] ggridges_0.5.3 rstudioapi_0.13 rappdirs_0.3.3
## [52] bitops_1.0-6 labeling_0.4.2 rmarkdown_2.6
## [55] gtable_0.3.0 DBI_1.1.0 curl_4.3
## [58] R6_2.5.0 knitr_1.30 dplyr_1.0.2
## [61] bit_4.0.4 KernSmooth_2.23-18 stringi_1.5.3
## [64] ggbeeswarm_0.6.0 Rcpp_1.0.6 vctrs_0.3.8
## [67] tidyselect_1.1.1 xfun_0.20

```