

EDA organoid partition

Niklas Rindtorff

Loading packages

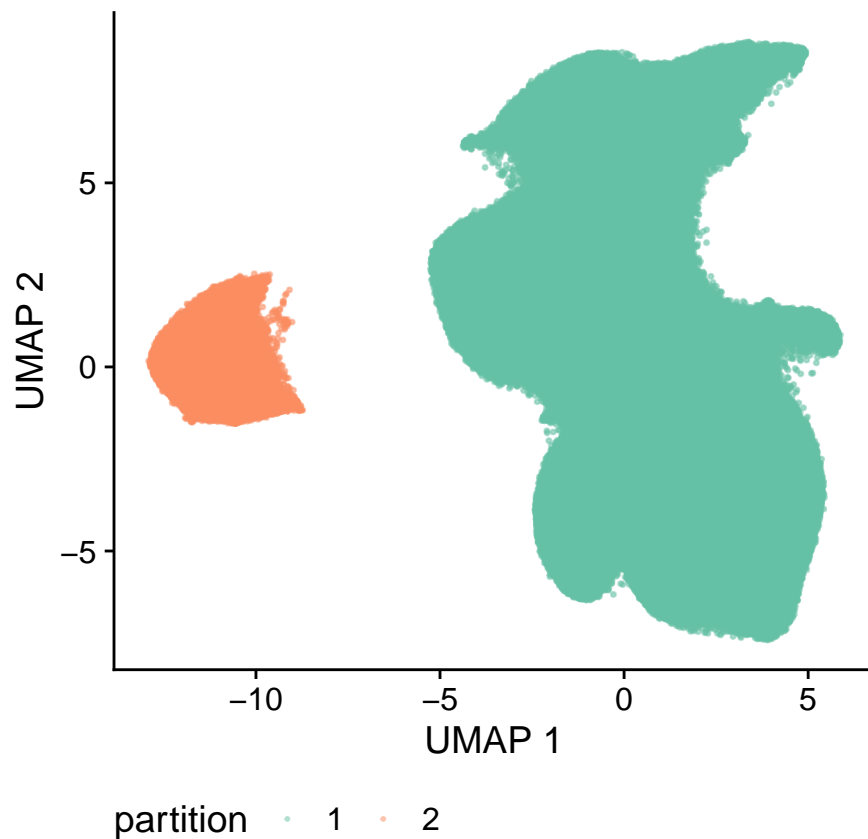
```
## [1] "parameter input:"
```

```
## [1] "data/processed/PhenotypeSpectrum/filtered_moderate/umap_absolute_all_drugs_sampled.Rds"
```

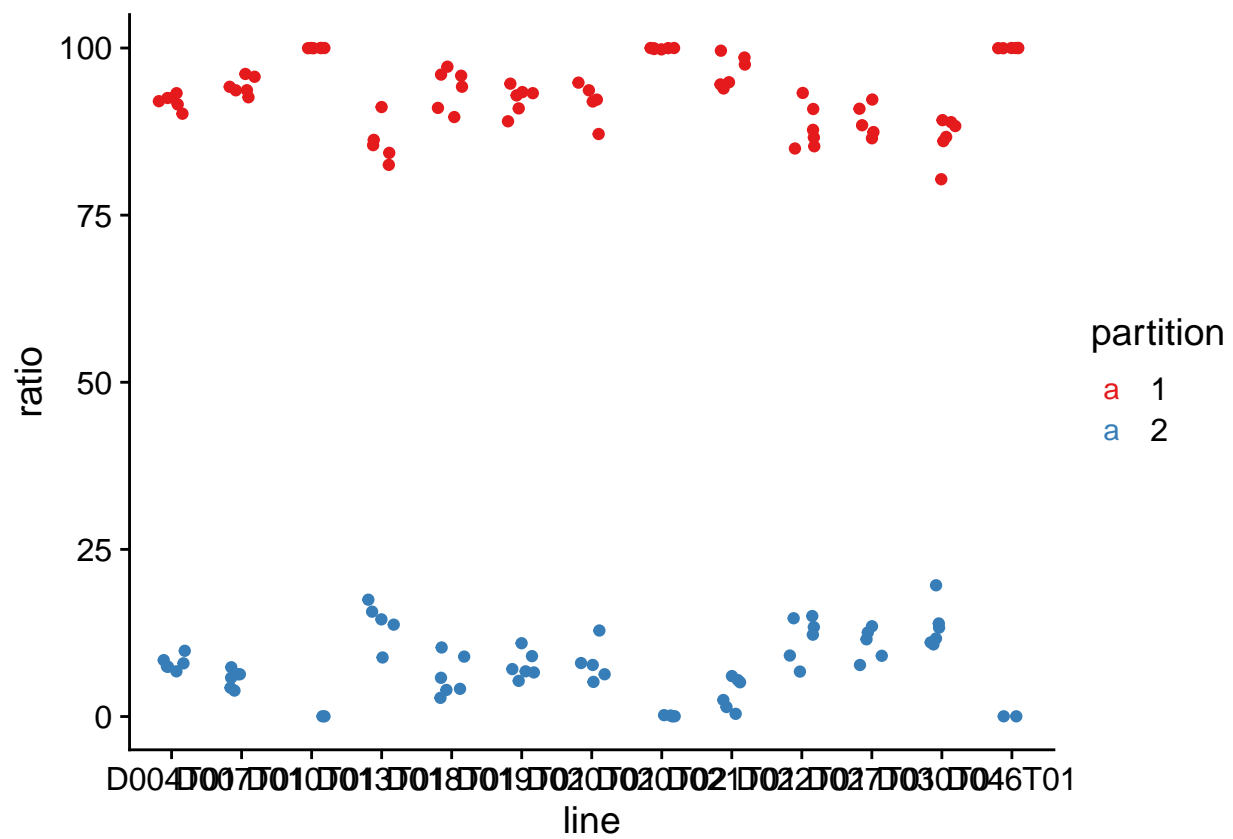
loading input data and annotation. Note that on the central cluster, with access to the complete data table, the definition of the input can easily be changed. For remote work, the subsampled dataset “umap_drugs_sampled.Rds” is the default choice.

Partition inspection

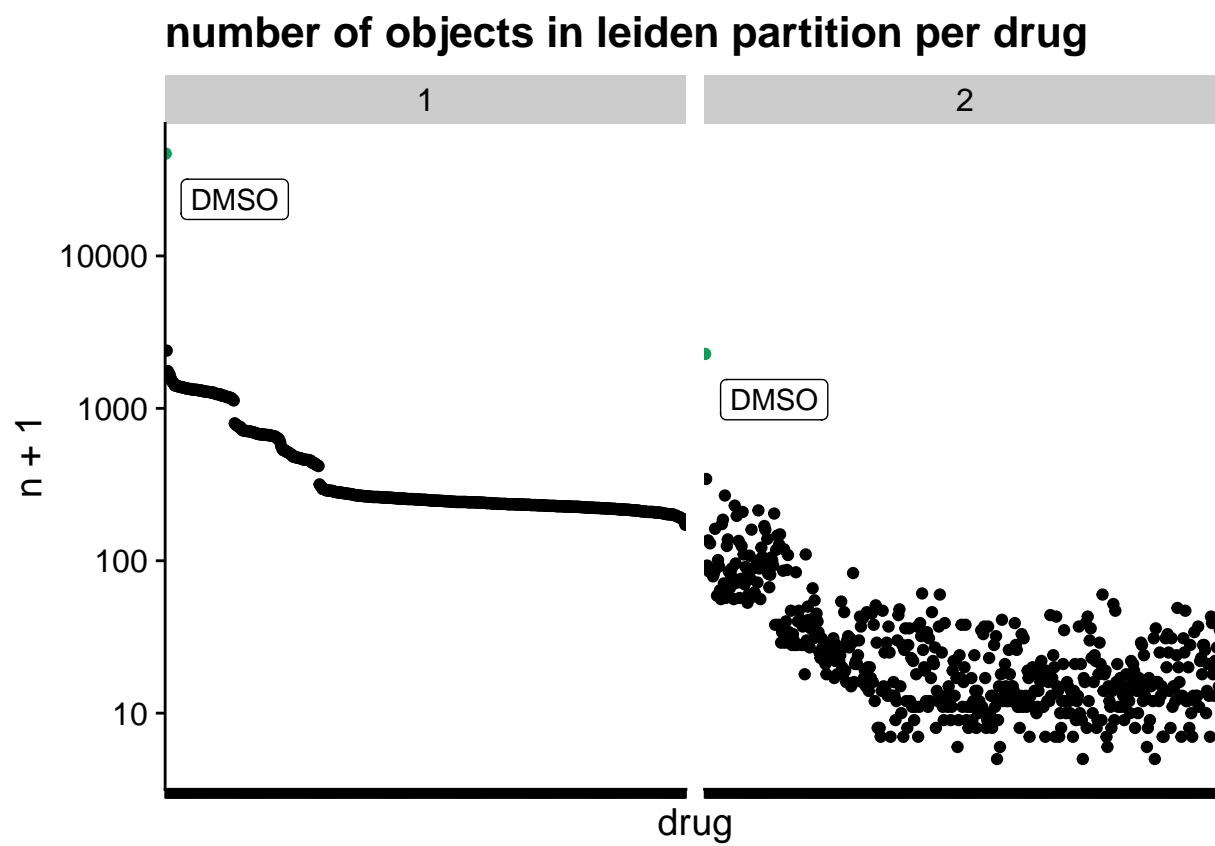
We are able to observe 4 partitions in our data. After manual inspection, it becomes clear that the two smallest partitions are mostly consisting of

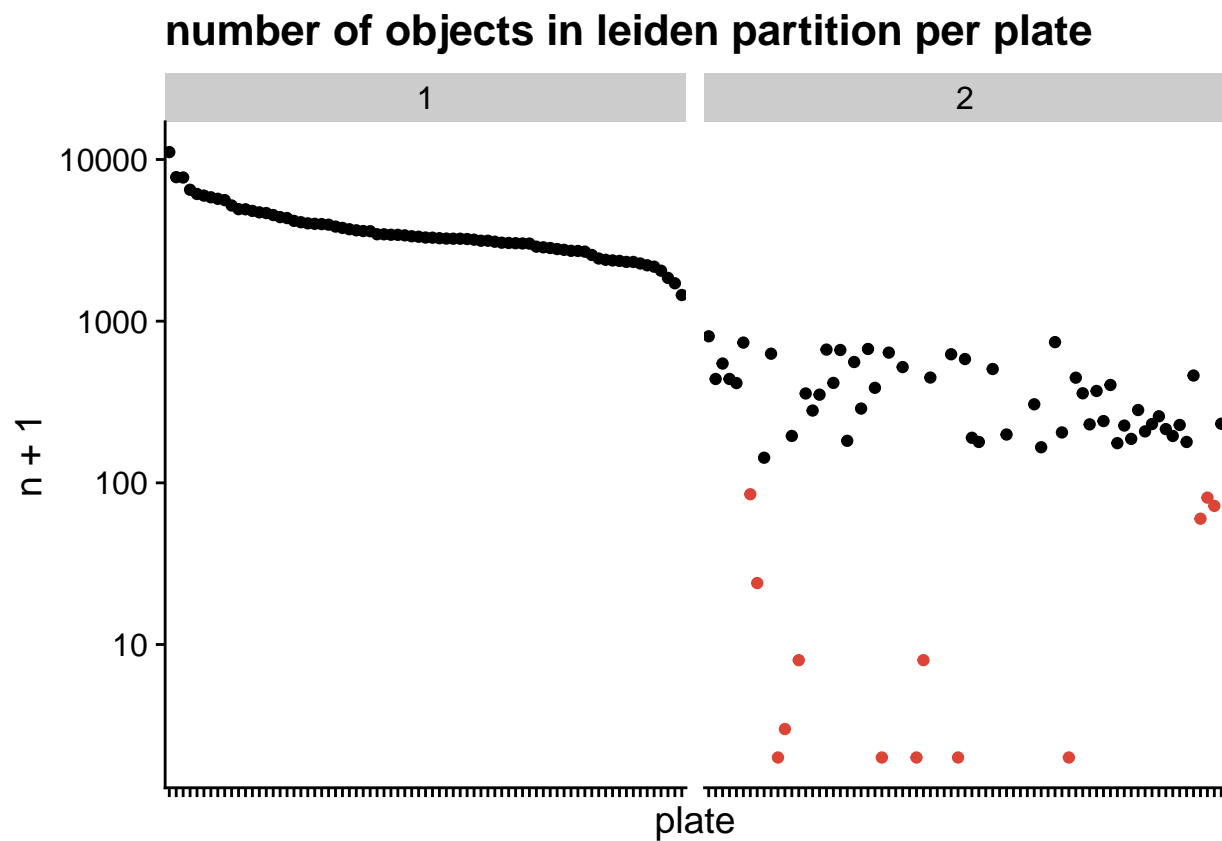


partition	n	ratio	min_ratio	max_ratio
1	278346	93.33	80.371	100.000
2	19891	6.67	0.020	19.629



drug overrepresentation





chi-square

I wonder whether certain batches or organoid lines are overrepresented in each section.

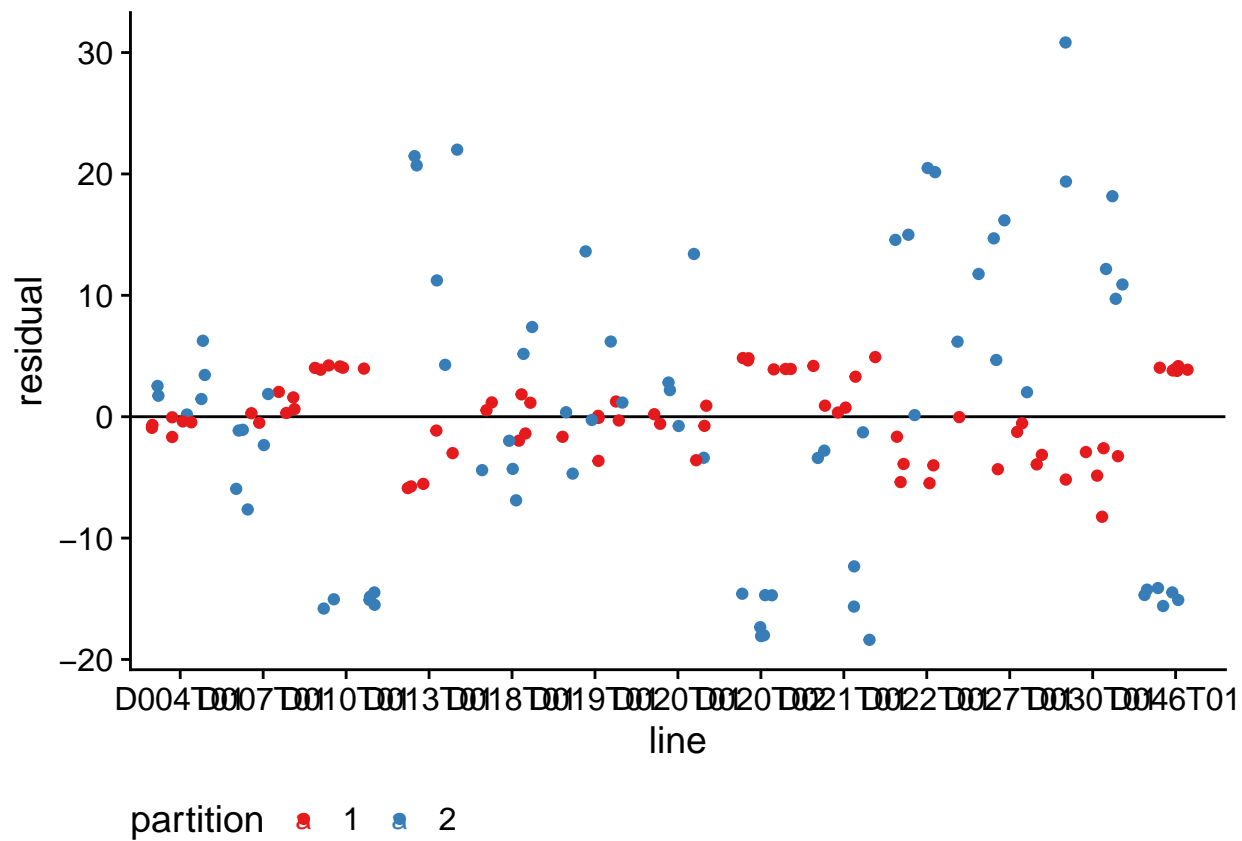
drug	partition	residual
DMSO	2	-17.574500
AT9283	1	-4.461688
Decernotinib (VX-509)	2	-4.257581
LGK974	2	-4.155685
Bortezomib	1	-4.073052
Irinotecan / SN-38	2	11.825451
Staurosporine_500nM	2	11.897281
Docetaxel	2	12.344762
Bortezomib	2	15.236464
AT9283	2	16.690270

line overrepresentation

plate	partition	residual
D030T01P906L03	2	30.83209
D013T01P003L08	2	22.00065
D013T01P906L03	2	21.46979
D013T01P002L03	2	20.70469
D022T01P907L08	2	20.48629
D010T01P019L08	2	-15.80439
D020T02P015L08	2	-17.34044
D020T02P014L03	2	-18.00610
D020T02P013L02	2	-18.07606
D021T01P003L08	2	-18.38152

I plot chisq residuals for each plate

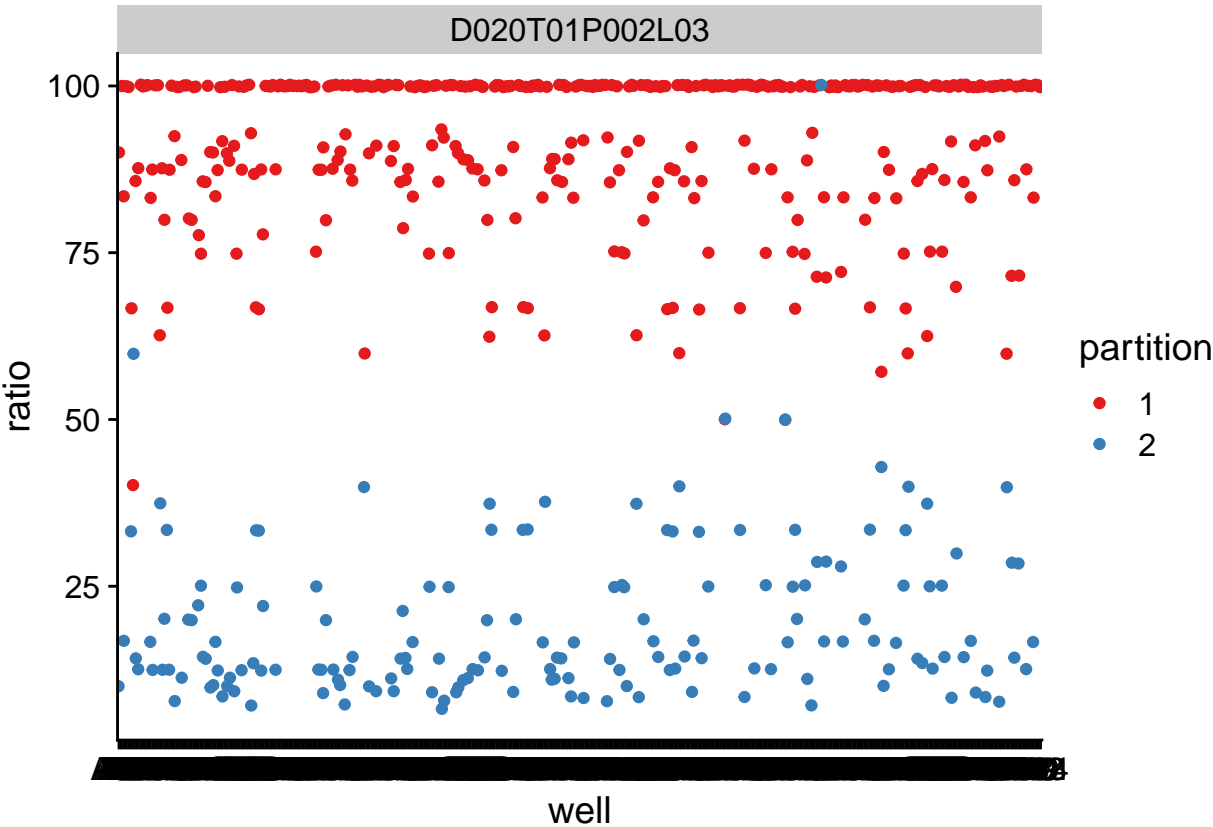
I recognize no difference between reimaged plates (leading digit is “9”, plates were reimaged due to errors during the first pass) and plates that were not reimaged.



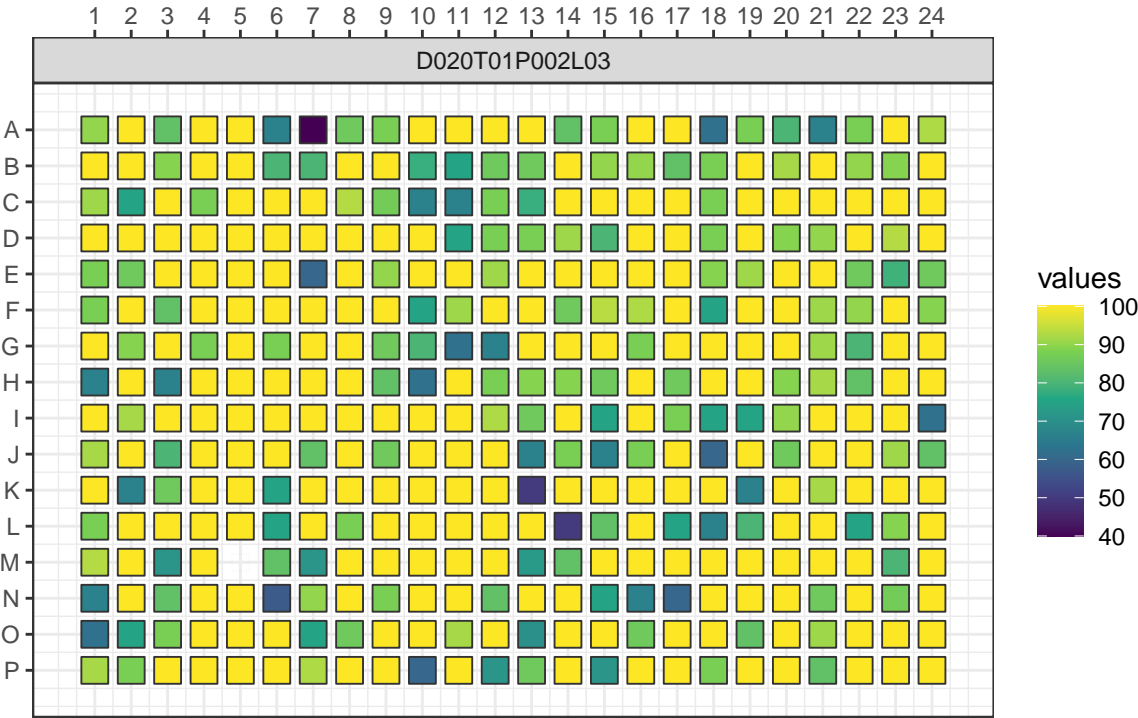
multinomial regression

We run a multinomial regression using the *nnet* package. I loaded the models in the beginning of the vignette.

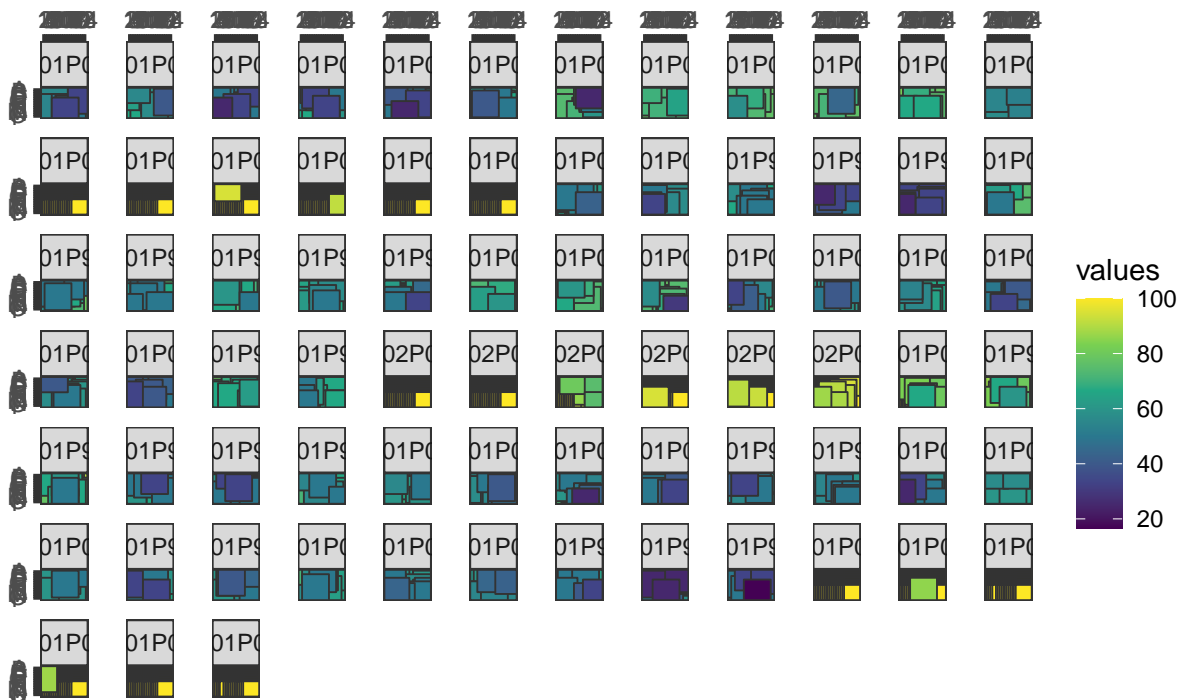
plate inspection



proportion of objects belonging to partition 1



D020T01P002L03 is a plate with one of the lowest chi-square residues



There appear to be three groups: * almost all organoids are within partition one, holds true for D010, D020 and D046 * 75% of objects are within partition one * almost no organoids are within partition one, this is true for the three plates previously identified and discussed below.

observation

- after processing, organoids organize in 4 distinct phenotype partitions
- the distribution of organoids across these partitions is non-random
 - the screening plate influences the distribution of organoids across partitions, 3 plates show strong deviation from the expected distribution, both in a chi-square test **and** in a multinomial regression. The feature **plate** is more predictive than the organoid **line** or **screen_id**
 - * D027T01P906L03
 - * D020T01P906L03
 - * D013T01P001L02
 - drug treatment influences the distribution of organoids across the partitions
 - * DMSO control treatment are depleted in sector 2, sector 2 has previously been shown to contain dead and small organoids
 - * SN38 and bortezomib are enriched in sector 2

conclusion

- given the patterns above, I believe it is most likely we are seeing systematic errors (not dependent on the biological axes of line and drug) in plates:

- D027T01P906L03
- D020T01P906L03
- D013T01P001L02

- in addition we are observing deviations that are drug dependent, that I consider signal

next steps

- inspect the three plates manually
- remove data in case there is more evidence for otherwise non-recoverable systematic noise

figure

```
## R version 4.0.0 (2020-04-24)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 20.04.2 LTS
##
## Matrix products: default
## BLAS/LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblas-p0.3.8.so
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=C
##  [7] LC_PAPER=en_US.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
##  [1] nnet_7.3-13      platetools_0.1.3 gridExtra_2.3    ggribes_0.5.2
##  [5] scico_1.1.0      princurve_2.1.4  cowplot_1.0.0    ggrastr_0.2.3
##  [9] here_0.1         forcats_0.5.0    stringr_1.4.0    dplyr_1.0.0
## [13] purrr_0.3.4      readr_1.3.1      tidyr_1.1.0      tibble_3.0.1
## [17] ggplot2_3.3.1    tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.4.6      lubridate_1.7.8    lattice_0.20-41    assertthat_0.2.1
##  [5] rprojroot_1.3-2   digest_0.6.25      R6_2.4.1           cellranger_1.1.0
##  [9] plyr_1.8.6        backports_1.1.7    reprex_0.3.0       evaluate_0.14
## [13] httr_1.4.1        pillar_1.4.4       rlang_0.4.6        readxl_1.3.1
## [17] rstudioapi_0.11   blob_1.2.1         rmarkdown_2.2      labeling_0.3
## [21] munsell_0.5.0     broom_0.5.6        compiler_4.0.0     vipor_0.4.5
## [25] modelr_0.1.8      xfun_0.14          pkgconfig_2.0.3    ggbeeswarm_0.6.0
## [29] htmltools_0.4.0   tidyselect_1.1.0   viridisLite_0.3.0  fansi_0.4.1
## [33] crayon_1.3.4      dbplyr_1.4.4       withr_2.2.0        grid_4.0.0
## [37] nlme_3.1-147      jsonlite_1.6.1     gtable_0.3.0       lifecycle_0.2.0
## [41] DBI_1.1.0         magrittr_1.5        scales_1.1.1       cli_2.0.2
## [45] stringi_1.4.6     farver_2.0.3       fs_1.4.1           xml2_1.3.2
```

## [49]	ellipsis_0.3.1	generics_0.0.2	vctrs_0.3.1	RColorBrewer_1.1-2
## [53]	Cairo_1.5-12	tools_4.0.0	glue_1.4.1	beeswarm_0.2.3
## [57]	hms_0.5.3	yaml_2.2.1	colorspace_1.4-1	rvest_0.3.5
## [61]	knitr_1.28	haven_2.3.1		