3 Tutorials

3.1 Tutorial-1: Plasma proteome (Bulk dataset)

This tutorial allows users to explore bulk plasma proteome measured from 6 healthy donors over 10 time-points. Plasma proteomic data available at github. 1. AIFI-Olink-NPX_log2_Protein.Rda (Normalized protein expression data) 2. AIFI-Metadata.Rda (clinical metadata). Longitudinal dataset includes 6 donors (3 male and 3 females). PBMC samples were collected from 6 donors over 10 weeks. To interrogate longitudinal data, please follow following steps.

3.1.1 Load Library

```
#Load Library and other vizualization packages
library("PALMO")
library("Hmisc")
library("ggpubr")
library("cowplot")
```

3.1.2 Load data and assign paramaters (Time ~ 10sec)

The annotation table `metadata` must consist of column `Sample` (Participant sample name), `PTID` (donor/Participant), `Time` (longitudinal time points) information, but not necessary same column names. Users can assign respective columns in subsequent steps. The datamatrix is an Expression data frame, where rows represents gene/proteins and column represents participant samples (same as annotation table `Sample` column).

```
#Load Plasma proteome data (longitudinal)
load("data/AIFI-Olink-NPX_log2_Protein.Rda")
#Load metadata
load("data/AIFI-Metadata.Rda")
```

3.1.3 Create PALMO object and merge annotation data (Time ~20sec)

The expression dataframe annotations merged with input annotation dataframe. Only overlapping samples kept. Missing annotations with Sample, Donor/participant, or Time columns are removed from downstream analysis.

```
#Create PALMO object
palmo_obj <- createPALMOobject(anndata=ann, data=data)

#Assign Sample, PTID and Time parameters
palmo_obj<- annotateMetadata(data_object=palmo_obj,
            sample_column= "Sample", donor_column= "PTID",
            time_column= "Time")

#Sample overlap and final matrix
palmo_obj <- mergePALMOdata(data_object=palmo_obj, datatype="bulk")

#Check for replicates
palmo_obj <- checkReplicates(data_object=palmo_obj, mergeReplicates = T)
```

3.1.4 Remove genes with >40%NAs (optional)

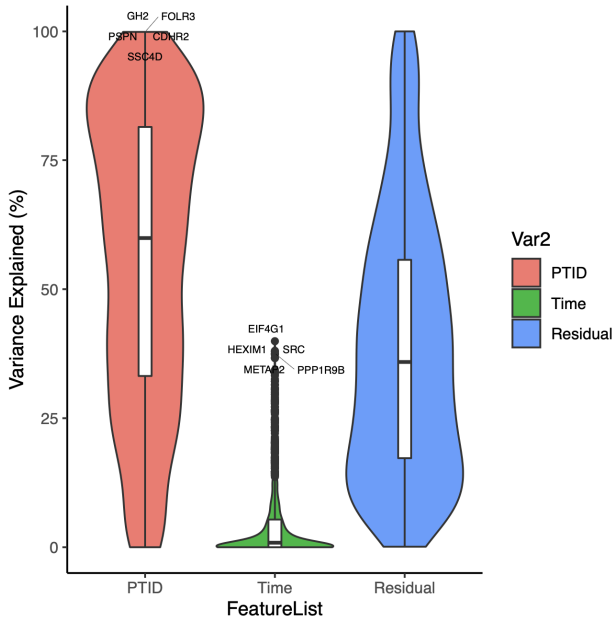
For downstream analysis select genes/proteins with less than 40% of missing values. Users can select cut-off for missing values as necessary.

```
palmo_obj <- naFilter(data_object=palmo_obj, na_cutoff=0.4)
```

3.1.5 Features contributing towards donor variations (Variance decomposition) (Time ~1min)

To perform variance decomposition apply `lmeVariance` function with input metadata, and datamatrix. The `featureSet` is a list of variables to which freaction variance explained by each gene is attributed. `meanThreshold` defines the minimum average expression threshold to be used for ongitudinal dataset. Here we used normalized protein expression 1 based on mean expression profile of each gene across longitudinal samples. Residuals suggest the variance can not be explained by available feature set. The variance explained by each gene towards the featureSet of interest given in percentage.

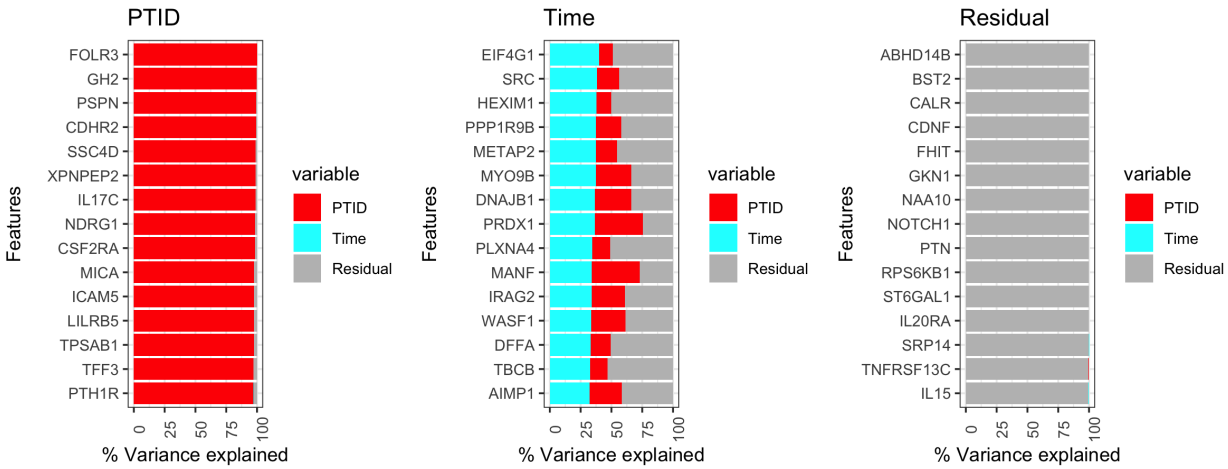
```
featureSet <- c("PTID","Time")
palmo_obj <- lmeVariance(data_object=palmo_obj, featureSet=featureSet,
                         meanThreshold=1, fileName="olink")
var_decomp <- palmo_obj@result$variance_decomposition
```




```
head(var_decomp[,c(featureSet, "Residual")])
#Features      donor      week  Residuals
#FOLR3   99.90070 0.0000000 0.09930098
#GH2     99.49856 0.0000000 0.50144042
#PSPN    99.26882 0.1021076 0.62906798
#CDHR2   99.07933 0.1157406 0.80493162
#SSC4D   98.82794 0.0000000 1.17206000
#XPNPEP2 98.67628 0.0000000 1.32372323
```

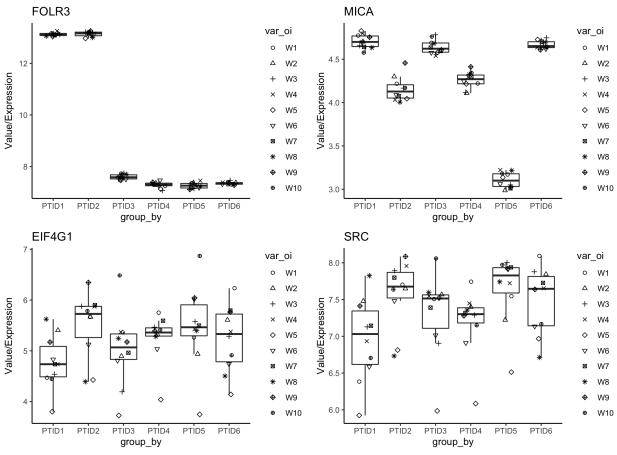
3.1.6 Donor-specific variance contrubuting features

```
plots <- variancefeaturePlot(vardata=var_decomp, featureSet=featureSet,
            Residual=T)
plot_grid(plotlist=plots, nrow = 1, align="hv")
```



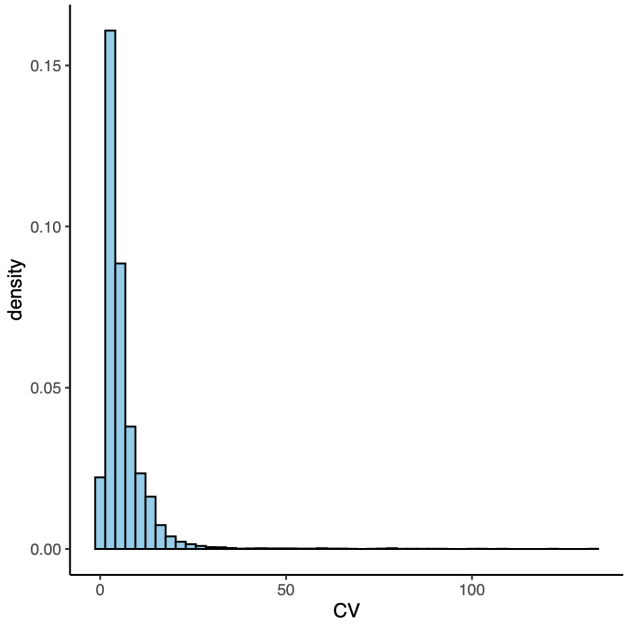
3.1.7 Plot top features

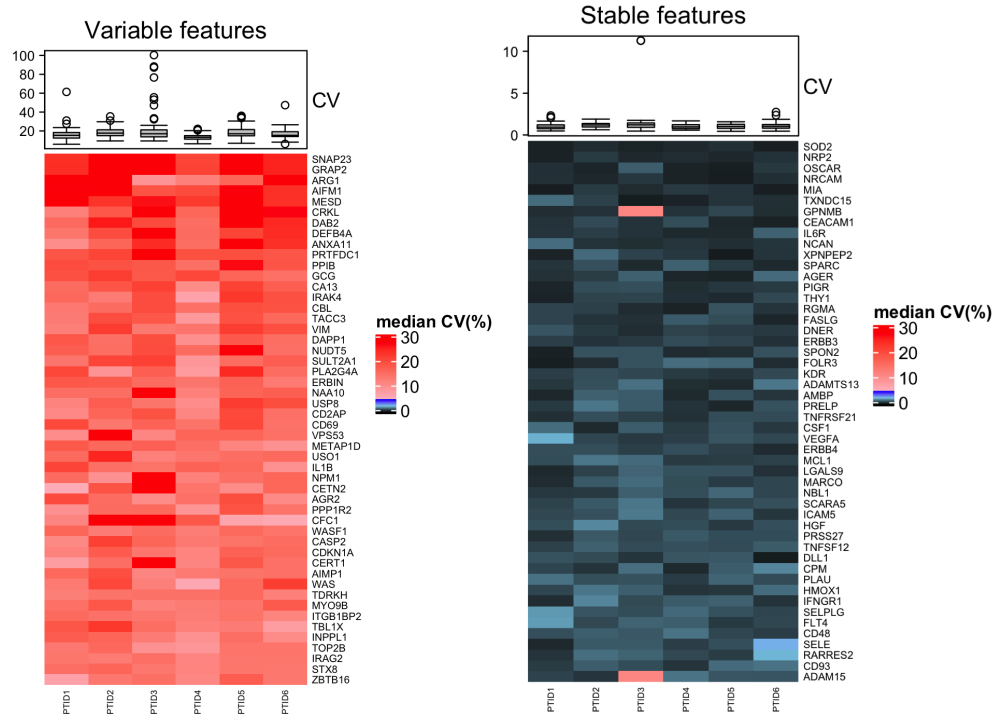
```
plots <- gene_featureplot(data_object=palmo_obj,
                          featureList=c("FOLR3", "MICA", "EIF4G1", "SRC"),
                          x_group_by="PTID", var_oi="Time")
plot_grid(plotlist=plots, ncol= 2, align="hv")
```



3.1.8 Intra-donor variations over time

```
#CV vs Mean
palmo_obj <- cvCalcBulk(data_object=palmo_obj, meanThreshold=1, cvThreshold=5)
```



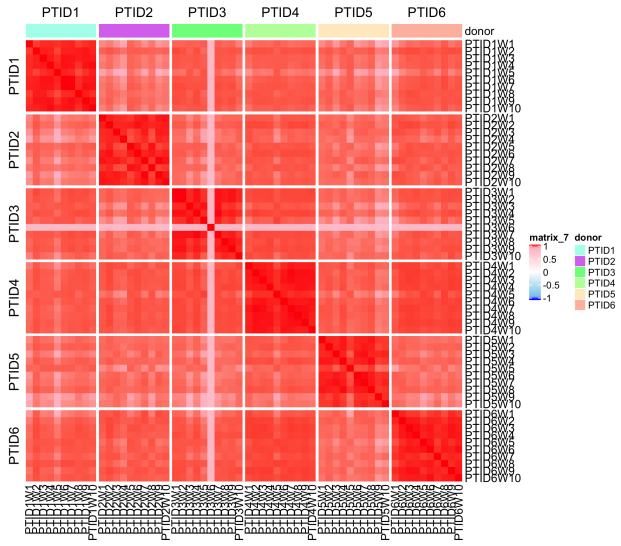



```
#Variable genes
head(palmo_obj@result[["variable_gene"]])
head(palmo_obj@result[["var_mat"]])
#Non-variable genes (stable)
head(palmo_obj@result[["non_variable_gene"]])
head(palmo_obj@result[["stable_mat"]])
```

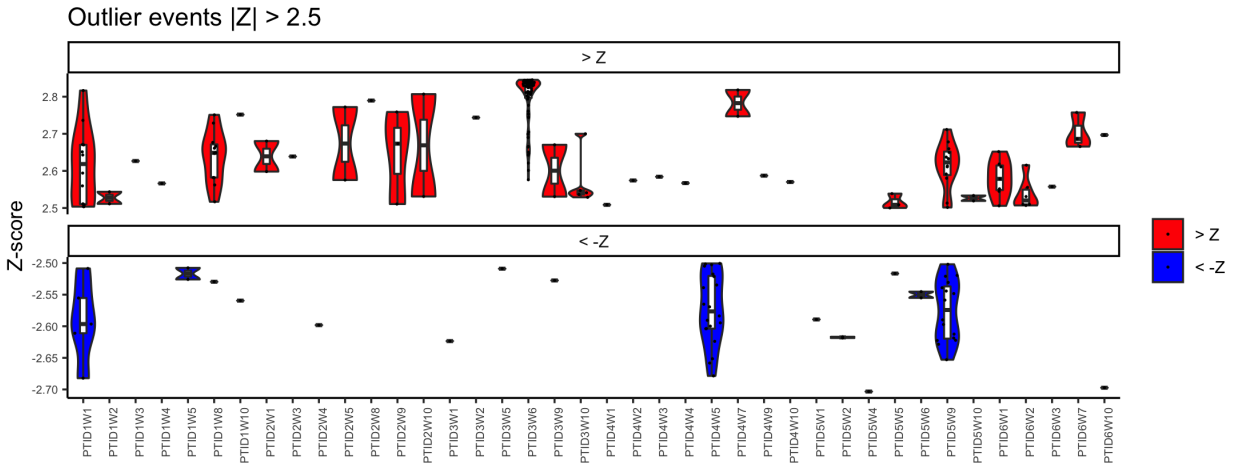
3.1.9 Outlier analysis (Time ~ 30sec)

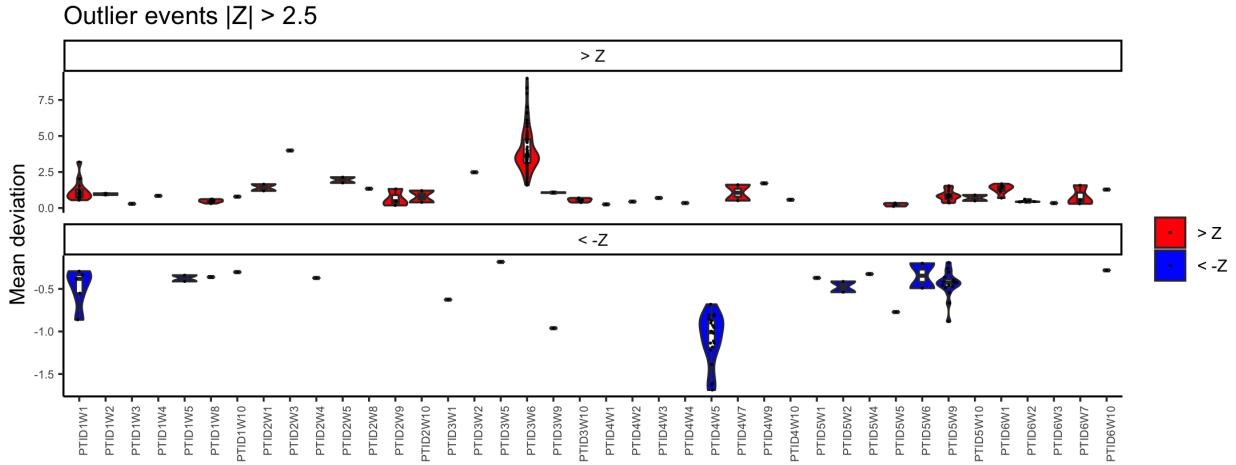
Perform the sample correlation to find out overall correlation between longitudinal samples.

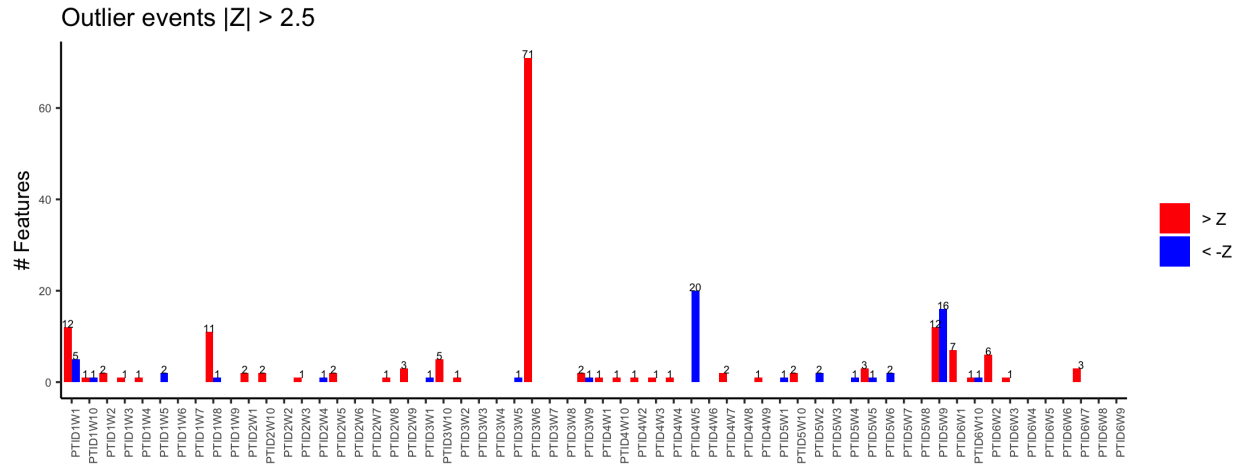
```
#Sample variability (Correlation)
palmo_obj <- sample_correlation(data_object=palmo_obj, donor_sep="W")
```



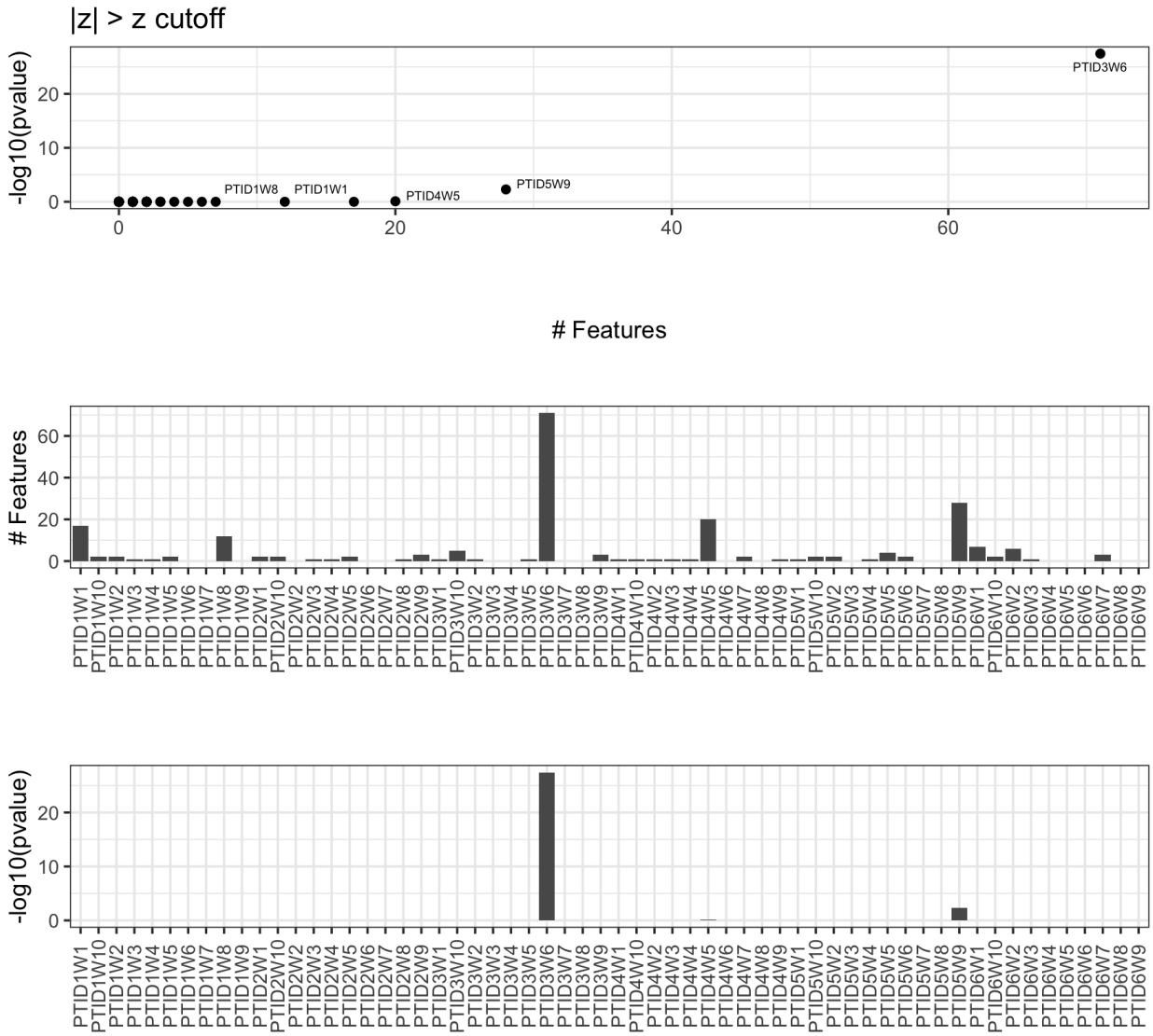

```
#Detect outliers (if any)
palmo_obj <- outlierDetect(data_object=palmo_obj, z_cutoff=2.5)
outlier_res <- palmo_obj@result[["outlier_res"]]
head(outlier_res)
#  exp  Sample  PTID Time    Sex      gene  meanDev        z
#PTID3W643 10.729880 PTID3W6 PTID3   W6 Female     IFI30 8.340474 2.845471
#PTID3W627 10.133645 PTID3W6 PTID3   W6 Female     DPEP2 6.595705 2.844629
#PTID3W631 10.188437 PTID3W6 PTID3   W6 Female      FCAR 7.004890 2.844607
#PTID3W626  7.656418 PTID3W6 PTID3   W6 Female     DPEP1 4.574449 2.844574
#PTID3W685  9.129149 PTID3W6 PTID3   W6 Female TNFRSF13C 6.605767 2.844410 2.844410
#PTID3W652  8.031215 PTID3W6 PTID3   W6 Female   KIR2DL3 5.156982 2.844018

#Number of outlier features
num_outlier <- data.frame(table(outlier_res$Sample))
num_outlier <- num_outlier[order(num_outlier$Freq, decreasing = T),]
head(num_outlier)
#Var1 Freq
#PTID3W6   71
#PTID5W9   28
#PTID4W5   20
#PTID1W1   17
#PTID1W8   12
#PTID6W1    7
```



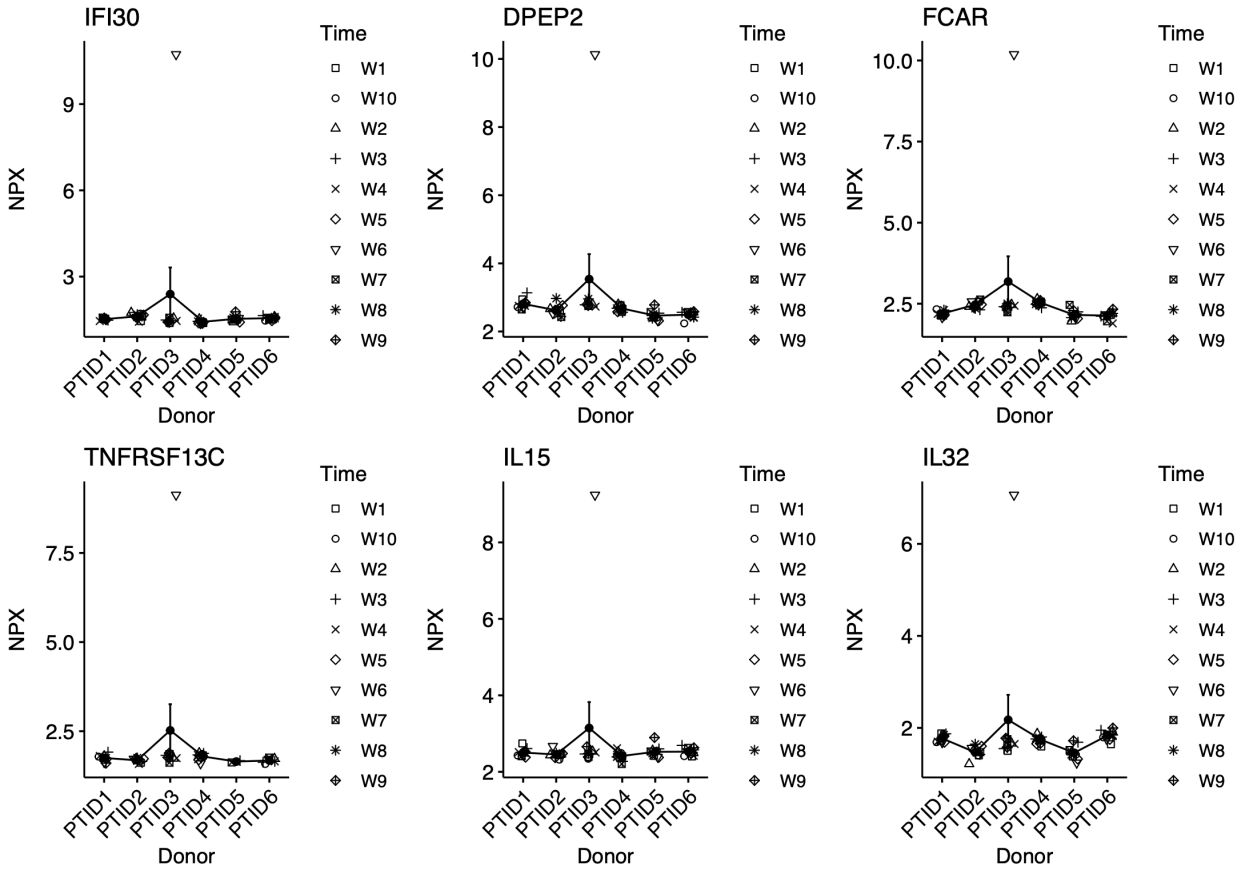





```
#Calculate p value
outlierDetectP(outlier_events=outlier_res, z_cutoff=2.5, nGenes=1042)
```




```
#Gene plot (probable outliers)
plots <- gene_featureplot(data_object=palmo_obj,
                featureList=c("IFI30", "DPEP2","FCAR",
                              "TNFRSF13C", "IL15", "IL32"),
                x_group_by="PTID", var_oi="Time")
plot_grid(plotlist = plots, ncol=3)
```


IFI30
DPEP2
FCAR
TNFRSF13C
IL15
IL32
NPX
Donor
PTID1
PTID2
PTID3
PTID4
PTID5
PTID6
Time
W1
W10
W2
W3
W4
W5
W6
W7
W8
W9

6 Session info

```
sessionInfo()
#> R version 4.0.3 (2020-10-10)
#> 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] grid      stats     graphics  grDevices utils     datasets  methods
#> [8] base
#>
#> other attached packages:
#> [1] PALMO_0.99.0
#>
#> loaded via a namespace (and not attached):
#>   [1] readxl_1.3.1                backports_1.2.0
#>   [3] circlize_0.4.11             plyr_1.8.6
#>   [5] igraph_1.2.8                lazyeval_0.2.2
#>   [7] splines_4.0.3               listenv_0.8.0
#>   [9] scattermore_0.7             GenomeInfoDb_1.24.2
#>  [11] ggplot2_3.3.5               digest_0.6.28
#>  [13] htmltools_0.5.2             fansi_0.5.0
#>  [15] magrittr_2.0.1              tensor_1.5
#>  [17] cluster_2.1.0               ROCR_1.0-11
#>  [19] ComplexHeatmap_2.4.3        globals_0.14.0
#>  [21] readr_1.4.0                 modelr_0.1.8
#>  [23] matrixStats_0.61.0          colorspace_2.0-2
#>  [25] rvest_0.3.6                 blob_1.2.1
#>  [27] ggrepel_0.9.1               haven_2.3.1
#>  [29] xfun_0.25                   dplyr_1.0.7
#>  [31] RCurl_1.98-1.2              crayon_1.4.2
#>  [33] jsonlite_1.7.2              lme4_1.1-25
#>  [35] spatstat_1.64-1             spatstat.data_2.1-0
#>  [37] survival_3.2-7              zoo_1.8-9
#>  [39] glue_1.5.0                  polyclip_1.10-0
#>  [41] gtable_0.3.0                zlibbioc_1.34.0
#>  [43] XVector_0.28.0              leiden_0.3.9
#>  [45] DelayedArray_0.14.1         GetoptLong_1.0.4
#>  [47] SingleCellExperiment_1.10.1 future.apply_1.8.1
#>  [49] shape_1.4.5                 BiocGenerics_0.34.0
#>  [51] abind_1.4-5                 scales_1.1.1
#>  [53] pheatmap_1.0.12             DBI_1.1.0
#>  [55] miniUI_0.1.1.1              Rcpp_1.0.7
#>  [57] viridisLite_0.4.0           xtable_1.8-4
#>  [59] clue_0.3-57                 reticulate_1.22
#>  [61] stats4_4.0.3                htmlwidgets_1.5.4
```

```
#>  [63] httr_1.4.2                 RColorBrewer_1.1-2
#>  [65] ellipsis_0.3.2             Seurat_4.0.0
#>  [67] factoextra_1.0.7.999       ica_1.0-2
#>  [69] farver_2.1.0               pkgconfig_2.0.3
#>  [71] uwot_0.1.10                dbplyr_1.4.4
#>  [73] deldir_1.0-6               utf8_1.2.2
#>  [75] tidyselect_1.1.1           rlang_0.4.12
#>  [77] reshape2_1.4.4             later_1.3.0
#>  [79] munsell_0.5.0              cellranger_1.1.0
#>  [81] tools_4.0.3                generics_0.1.1
#>  [83] broom_0.7.2                ggridges_0.5.3
#>  [85] evaluate_0.14              stringr_1.4.0
#>  [87] fastmap_1.1.0              yaml_2.2.1
#>  [89] goftest_1.2-3              knitr_1.30
#>  [91] fs_1.5.0                   fitdistrplus_1.1-6
#>  [93] purrr_0.3.4                RANN_2.6.1
#>  [95] pbapply_1.5-0              future_1.23.0
#>  [97] nlme_3.1-149               mime_0.12
#>  [99] xml2_1.3.2                 compiler_4.0.3
#> [101] plotly_4.10.0              png_0.1-7
#> [103] spatstat.utils_2.2-0       reprex_0.3.0
#> [105] tweenr_1.0.1               tibble_3.1.6
#> [107] statmod_1.4.35             stringi_1.7.5
#> [109] forcats_0.5.0              lattice_0.20-41
#> [111] Matrix_1.3-4               nloptr_1.2.2.2
#> [113] vctrs_0.3.8                pillar_1.6.4
#> [115] lifecycle_1.0.1            lmtest_0.9-39
#> [117] GlobalOptions_0.1.2        RcppAnnoy_0.0.19
#> [119] bitops_1.0-7               data.table_1.14.2
#> [121] cowplot_1.1.1              irlba_2.3.3
#> [123] GenomicRanges_1.40.0       httpuv_1.6.3
#> [125] patchwork_1.1.1            R6_2.5.1
#> [127] promises_1.2.0.1           KernSmooth_2.23-17
#> [129] gridExtra_2.3              IRanges_2.22.2
#> [131] parallelly_1.28.1          codetools_0.2-16
#> [133] boot_1.3-25                MASS_7.3-53
#> [135] assertthat_0.2.1           SummarizedExperiment_1.18.2
#> [137] MAST_1.14.0                rjson_0.2.20
#> [139] SeuratObject_4.0.2         sctransform_0.3.2
#> [141] GenomeInfoDbData_1.2.3     S4Vectors_0.26.1
#> [143] mgcv_1.8-33                parallel_4.0.3
#> [145] hms_0.5.3                  rpart_4.1-15
#> [147] tidyverse_1.3.0            tidyr_1.1.4
#> [149] minqa_1.2.4                rmarkdown_2.5
#> [151] Rtsne_0.15                 ggforce_0.3.2
#> [153] Biobase_2.48.0             shiny_1.7.1
#> [155] lubridate_1.7.9
```