

upset plot

2024-12-12

```
# install.packages("msigdbR")
library(msigdbR)

## For full functionality, please install the 'msigdbdf' package with:
## install.packages('msigdbdf', repos = 'https://igordot.r-universe.dev')
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(gofigR)
enable(TRUE)

## Analysis name: R_tips_17_upset_plot_for_gene_sets.Rmd
h_gene_sets<- msigdbR(species = "human", category = "H")

## Warning: The `category` argument of `msigdbR()` is deprecated as of msigdbR 9.0.0.
## i Please use the `collection` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

## The 'msigdbdf' package must be installed to access the full dataset.
h_gene_sets

## # A tibble: 7,325 x 22
##   gene_symbol ncbi_gene ensembl_gene db_gene_symbol db_ncbi_gene
##   <chr>      <chr>      <chr>      <chr>      <chr>
## 1 ABCA1      19        ENSG00000165029 ABCA1      19
## 2 ABCB8      11194     ENSG00000197150 ABCB8      11194
## 3 ACAA2      10449     ENSG00000167315 ACAA2      10449
## 4 ACADL      33        ENSG00000115361 ACADL      33
## 5 ACADM      34        ENSG00000117054 ACADM      34
## 6 ACADS      35        ENSG00000122971 ACADS      35
## 7 ACLY       47        ENSG00000131473 ACLY       47
## 8 ACO2       50        ENSG00000100412 ACO2       50
## 9 ACOX1      51        ENSG00000161533 ACOX1      51
## 10 ADCY6     112       ENSG00000174233 ADCY6     112
## # i 7,315 more rows
```

```
## # i 17 more variables: db_ensembl_gene <chr>, source_gene <chr>, gs_id <chr>,
## #   gs_name <chr>, gs_collection <chr>, gs_subcollection <chr>,
## #   gs_collection_name <chr>, gs_description <chr>, gs_source_species <chr>,
## #   gs_pmid <chr>, gs_geoid <chr>, gs_url <chr>, db_version <chr>,
## #   db_target_species <chr>, entrez_gene <chr>, gs_cat <chr>, gs_subcat <chr>
```

```
head(h_gene_sets)
```

```
## # A tibble: 6 x 22
##   gene_symbol ncbi_gene ensembl_gene db_gene_symbol db_ncbi_gene db_ensembl_gene
##   <chr>      <chr>      <chr>      <chr>      <chr>      <chr>
## 1 ABCA1      19        ENSG0000016~ ABCA1      19        ENSG00000165029
## 2 ABCB8      11194     ENSG0000019~ ABCB8      11194     ENSG00000197150
## 3 ACAA2      10449     ENSG0000016~ ACAA2      10449     ENSG00000167315
## 4 ACADL      33        ENSG0000011~ ACADL      33        ENSG00000115361
## 5 ACADM      34        ENSG0000011~ ACADM      34        ENSG00000117054
## 6 ACADS      35        ENSG0000012~ ACADS      35        ENSG00000122971
## # i 16 more variables: source_gene <chr>, gs_id <chr>, gs_name <chr>,
## #   gs_collection <chr>, gs_subcollection <chr>, gs_collection_name <chr>,
## #   gs_description <chr>, gs_source_species <chr>, gs_pmid <chr>,
## #   gs_geoid <chr>, gs_url <chr>, db_version <chr>, db_target_species <chr>,
## #   entrez_gene <chr>, gs_cat <chr>, gs_subcat <chr>
```

```
table(h_gene_sets$gs_name)
```

```
##
##           HALLMARK_ADIPOGENESIS
##                200
##   HALLMARK_ALLOGRAFT_REJECTION
##                200
##           HALLMARK_ANDROGEN_RESPONSE
##                101
##           HALLMARK_ANGIOGENESIS
##                36
##   HALLMARK_APICAL_JUNCTION
##                200
##           HALLMARK_APICAL_SURFACE
##                44
##           HALLMARK_APOPTOSIS
##                161
##   HALLMARK_BILE_ACID_METABOLISM
##                112
##   HALLMARK_CHOLESTEROL_HOMEOSTASIS
##                74
##           HALLMARK_COAGULATION
##                139
##           HALLMARK_COMPLEMENT
##                201
##           HALLMARK_DNA_REPAIR
##                150
##           HALLMARK_E2F_TARGETS
##                200
##   HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
##                200
##           HALLMARK_ESTROGEN_RESPONSE_EARLY
```

##		200
##	HALLMARK_ESTROGEN_RESPONSE_LATE	
##		200
##	HALLMARK_FATTY_ACID_METABOLISM	
##		158
##	HALLMARK_G2M_CHECKPOINT	
##		200
##	HALLMARK_GLYCOLYSIS	
##		200
##	HALLMARK_HEDGEHOG_SIGNALING	
##		36
##	HALLMARK_HEME_METABOLISM	
##		200
##	HALLMARK_HYPOXIA	
##		200
##	HALLMARK_IL2_STAT5_SIGNALING	
##		199
##	HALLMARK_IL6_JAK_STAT3_SIGNALING	
##		87
##	HALLMARK_INFLAMMATORY_RESPONSE	
##		201
##	HALLMARK_INTERFERON_ALPHA_RESPONSE	
##		97
##	HALLMARK_INTERFERON_GAMMA_RESPONSE	
##		200
##	HALLMARK_KRAS_SIGNALING_DN	
##		200
##	HALLMARK_KRAS_SIGNALING_UP	
##		200
##	HALLMARK_MITOTIC_SPINDLE	
##		199
##	HALLMARK_MTORC1_SIGNALING	
##		200
##	HALLMARK_MYC_TARGETS_V1	
##		200
##	HALLMARK_MYC_TARGETS_V2	
##		58
##	HALLMARK_MYOGENESIS	
##		200
##	HALLMARK_NOTCH_SIGNALING	
##		32
##	HALLMARK_OXIDATIVE_PHOSPHORYLATION	
##		200
##	HALLMARK_P53_PATHWAY	
##		200
##	HALLMARK_PANCREAS_BETA_CELLS	
##		40
##	HALLMARK_PEROXISOME	
##		104
##	HALLMARK_PI3K_AKT_MTOR_SIGNALING	
##		105
##	HALLMARK_PROTEIN_SECRETION	
##		96
##	HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY	

```
##                                49
##          HALLMARK_SPERMATOGENESIS
##                                135
##          HALLMARK_TGF_BETA_SIGNALING
##                                54
##          HALLMARK_TNFA_SIGNALING_VIA_NFKB
##                                200
##          HALLMARK_UNFOLDED_PROTEIN_RESPONSE
##                                113
##          HALLMARK_UV_RESPONSE_DN
##                                144
##          HALLMARK_UV_RESPONSE_UP
##                                158
##          HALLMARK_WNT_BETA_CATENIN_SIGNALING
##                                42
##          HALLMARK_XENOBIOTIC_METABOLISM
##                                200
```

```
gs_nest<- h_gene_sets %>%
  group_by(gs_name) %>%
  tidyr::nest()
```

```
gs_nest$data[[1]]
```

```
## # A tibble: 200 x 21
##   gene_symbol ncbi_gene ensembl_gene db_gene_symbol db_ncbi_gene
##   <chr>      <chr>    <chr>      <chr>      <chr>
## 1 ABCA1      19      ENSG00000165029 ABCA1      19
## 2 ABCB8     11194    ENSG00000197150 ABCB8     11194
## 3 ACAA2     10449    ENSG00000167315 ACAA2     10449
## 4 ACADL      33      ENSG00000115361 ACADL      33
## 5 ACADM      34      ENSG00000117054 ACADM      34
## 6 ACADS      35      ENSG00000122971 ACADS      35
## 7 ACLY       47      ENSG00000131473 ACLY       47
## 8 ACO2       50      ENSG00000100412 ACO2       50
## 9 ACOX1      51      ENSG00000161533 ACOX1      51
## 10 ADCY6     112      ENSG00000174233 ADCY6     112
## # i 190 more rows
## # i 16 more variables: db_ensembl_gene <chr>, source_gene <chr>, gs_id <chr>,
## #   gs_collection <chr>, gs_subcollection <chr>, gs_collection_name <chr>,
## #   gs_description <chr>, gs_source_species <chr>, gs_pmid <chr>,
## #   gs_geoid <chr>, gs_url <chr>, db_version <chr>, db_target_species <chr>,
## #   entrez_gene <chr>, gs_cat <chr>, gs_subcat <chr>
```

```
gene_list<- purrr::map(gs_nest$data, ~pull(.x, gene_symbol))
```

```
names(gene_list)<- gs_nest$gs_name %>% stringr::str_replace("HALLMARK_", "")
```

```
names(gene_list)
```

```
## [1] "ADIPOGENESIS"          "ALLOGRAFT_REJECTION"
## [3] "ANDROGEN_RESPONSE"     "ANGIOGENESIS"
## [5] "APICAL_JUNCTION"       "APICAL_SURFACE"
## [7] "APOPTOSIS"             "BILE_ACID_METABOLISM"
## [9] "CHOLESTEROL_HOMEOSTASIS" "COAGULATION"
## [11] "COMPLEMENT"           "DNA_REPAIR"
```

```

## [13] "E2F_TARGETS" "EPITHELIAL_MESENCHYMAL_TRANSITION"
## [15] "ESTROGEN_RESPONSE_EARLY" "ESTROGEN_RESPONSE_LATE"
## [17] "FATTY_ACID_METABOLISM" "G2M_CHECKPOINT"
## [19] "GLYCOLYSIS" "HEDGEHOG_SIGNALING"
## [21] "HEME_METABOLISM" "HYPOXIA"
## [23] "IL2_STAT5_SIGNALING" "IL6_JAK_STAT3_SIGNALING"
## [25] "INFLAMMATORY_RESPONSE" "INTERFERON_ALPHA_RESPONSE"
## [27] "INTERFERON_GAMMA_RESPONSE" "KRAS_SIGNALING_DN"
## [29] "KRAS_SIGNALING_UP" "MITOTIC_SPINDLE"
## [31] "MTORC1_SIGNALING" "MYC_TARGETS_V1"
## [33] "MYC_TARGETS_V2" "MYOGENESIS"
## [35] "NOTCH_SIGNALING" "OXIDATIVE_PHOSPHORYLATION"
## [37] "P53_PATHWAY" "PANCREAS_BETA_CELLS"
## [39] "PEROXISOME" "PI3K_AKT_MTOR_SIGNALING"
## [41] "PROTEIN_SECRETION" "REACTIVE_OXYGEN_SPECIES_PATHWAY"
## [43] "SPERMATOGENESIS" "TGF_BETA_SIGNALING"
## [45] "TNFA_SIGNALING_VIA_NFKB" "UNFOLDED_PROTEIN_RESPONSE"
## [47] "UV_RESPONSE_DN" "UV_RESPONSE_UP"
## [49] "WNT_BETA_CATENIN_SIGNALING" "XENOBIOTIC_METABOLISM"

gene_sub<- gene_list[c(13,14,15,16,18)]

names(gene_sub)

## [1] "E2F_TARGETS" "EPITHELIAL_MESENCHYMAL_TRANSITION"
## [3] "ESTROGEN_RESPONSE_EARLY" "ESTROGEN_RESPONSE_LATE"
## [5] "G2M_CHECKPOINT"

library(ComplexHeatmap)

## Loading required package: grid

## =====
## ComplexHeatmap version 2.22.0
## Bioconductor page: http://bioconductor.org/packages/ComplexHeatmap/
## Github page: https://github.com/jokergoo/ComplexHeatmap
## Documentation: http://jokergoo.github.io/ComplexHeatmap-reference
##
## If you use it in published research, please cite either one:
## - Gu, Z. Complex Heatmap Visualization. iMeta 2022.
## - Gu, Z. Complex heatmaps reveal patterns and correlations in multidimensional
## genomic data. Bioinformatics 2016.
##
##
## The new InteractiveComplexHeatmap package can directly export static
## complex heatmaps into an interactive Shiny app with zero effort. Have a try!
##
## This message can be suppressed by:
## suppressPackageStartupMessages(library(ComplexHeatmap))
## =====

m<- make_comb_mat(gene_sub)
m

## A combination matrix with 5 sets and 17 combinations.
## ranges of combination set size: c(1, 187).
## mode for the combination size: distinct.

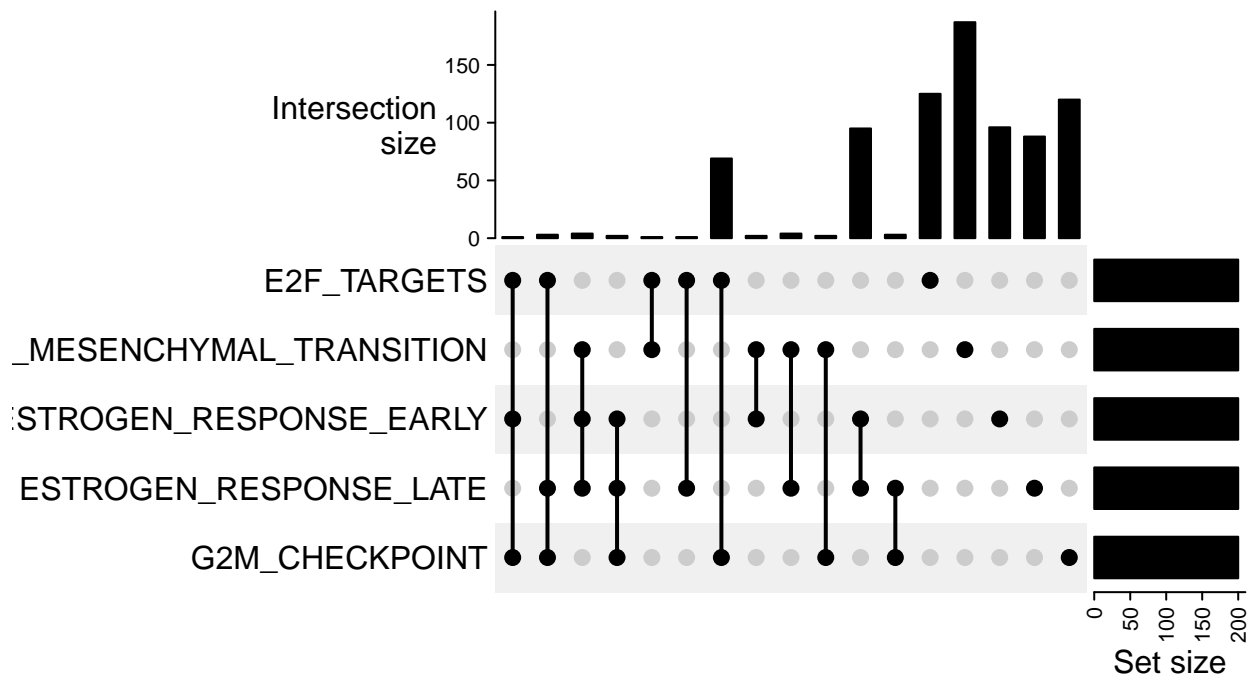
```

```

## sets are on rows.
##
## Top 8 combination sets are:
## E2F_TARGETS EPIHELIAL_MESENCHYMAL_TRANSITION ESTROGEN_RESPONSE_EARLY ESTROGEN_RESPONSE_LATE G2M_CHECKPOINT
##
##           x
##
##           x
##
##           x
##
##           x
##
##           x
##
##           x
##
## Sets are:
##
##           set size
##           E2F_TARGETS 200
## EPIHELIAL_MESENCHYMAL_TRANSITION 200
## ESTROGEN_RESPONSE_EARLY 200
## ESTROGEN_RESPONSE_LATE 200
## G2M_CHECKPOINT 200

```

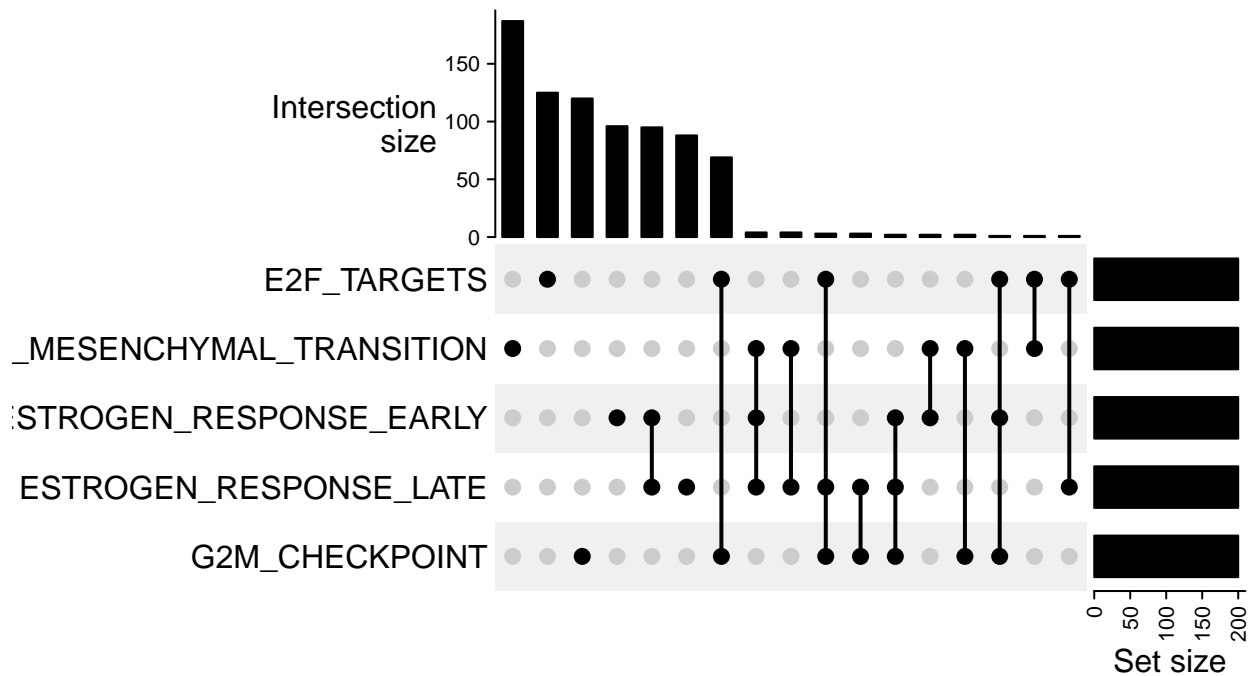
```
publish(UpSet(m), "Default UpSet")
```



<https://app.gofigr.io/r/9ee8aa98-89cd-456e-9d38-d6c61149bc8a>



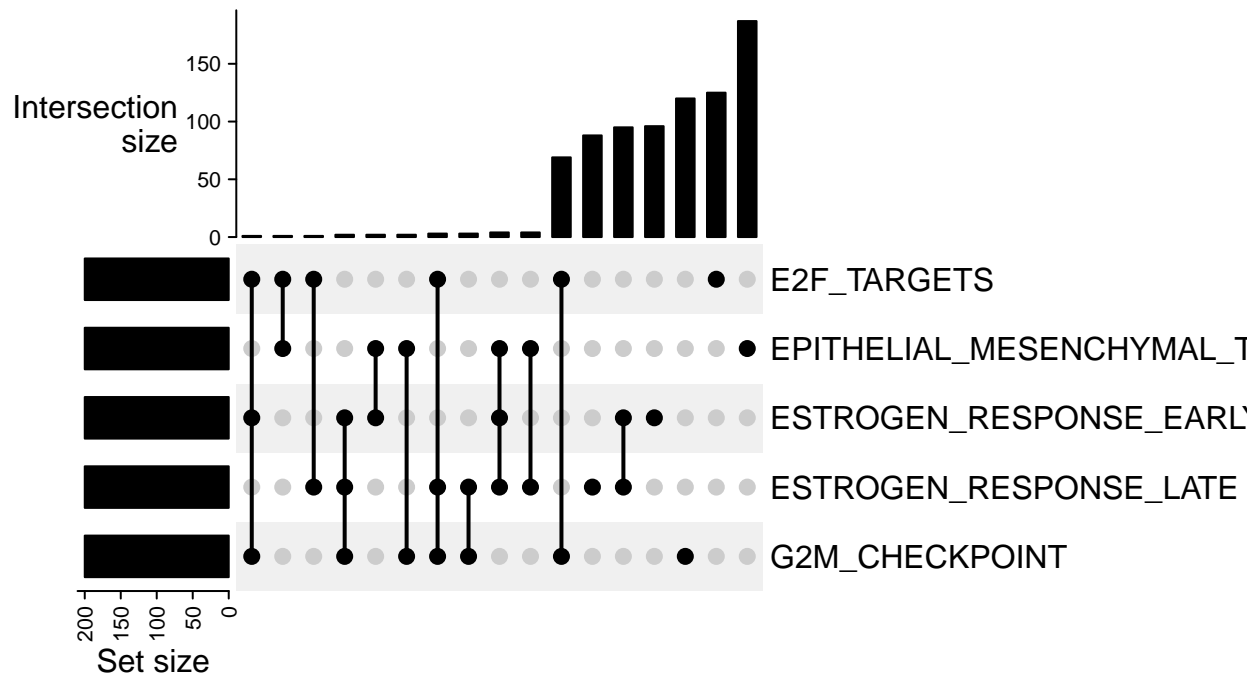
```
publish(UpSet(m, comb_order = order(-comb_size(m))), "UpSet with custom order")
```



<https://app.gofigr.io/r/e23ac32c-2583-49e8-bc79-830f3b7f923c>



```
publish(UpSet(m, left_annotation = upset_left_annotation(m),
  comb_order = order(comb_size(m))),
  "UpSet with custom order and left annotations")
```

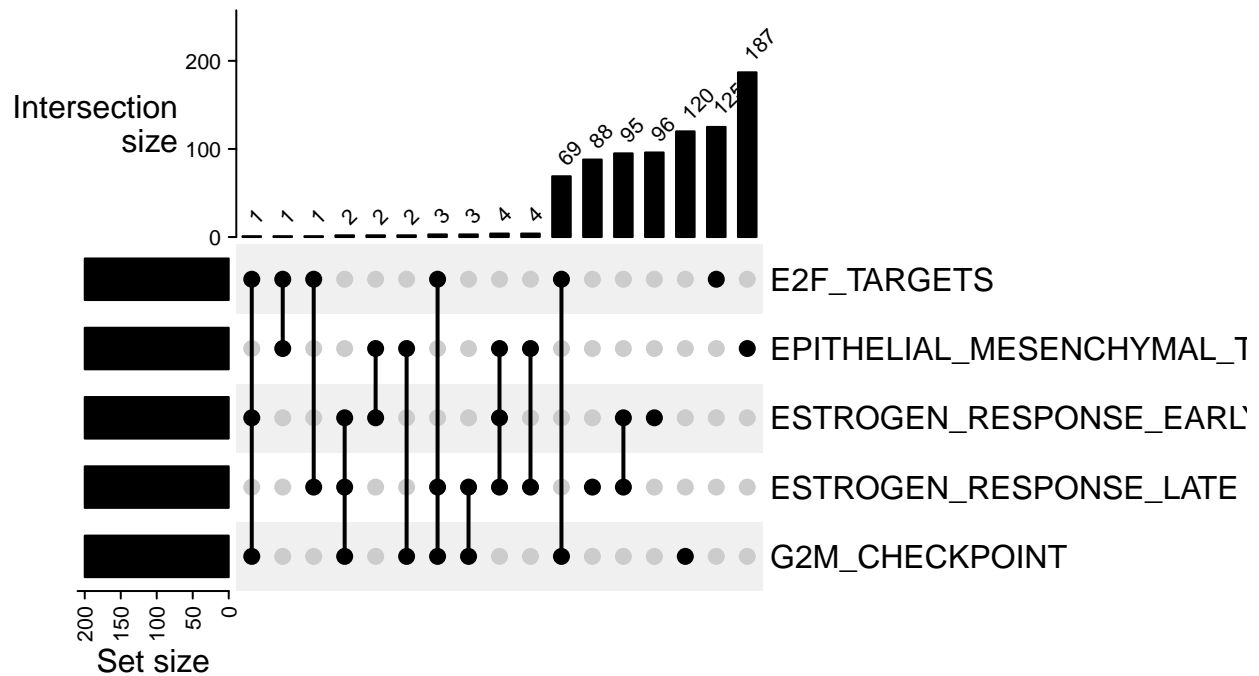


<https://app.gofigr.io/r/b69a288f-f417-4d11-8c47-8010d12f4905>



add numbers on the bar

```
publish(UpSet(m, left_annotation = upset_left_annotation(m),
  top_annotation = upset_top_annotation(m, add_numbers = TRUE),
  comb_order = order(comb_size(m))),
  "UpSet with custom order and right annotations")
```

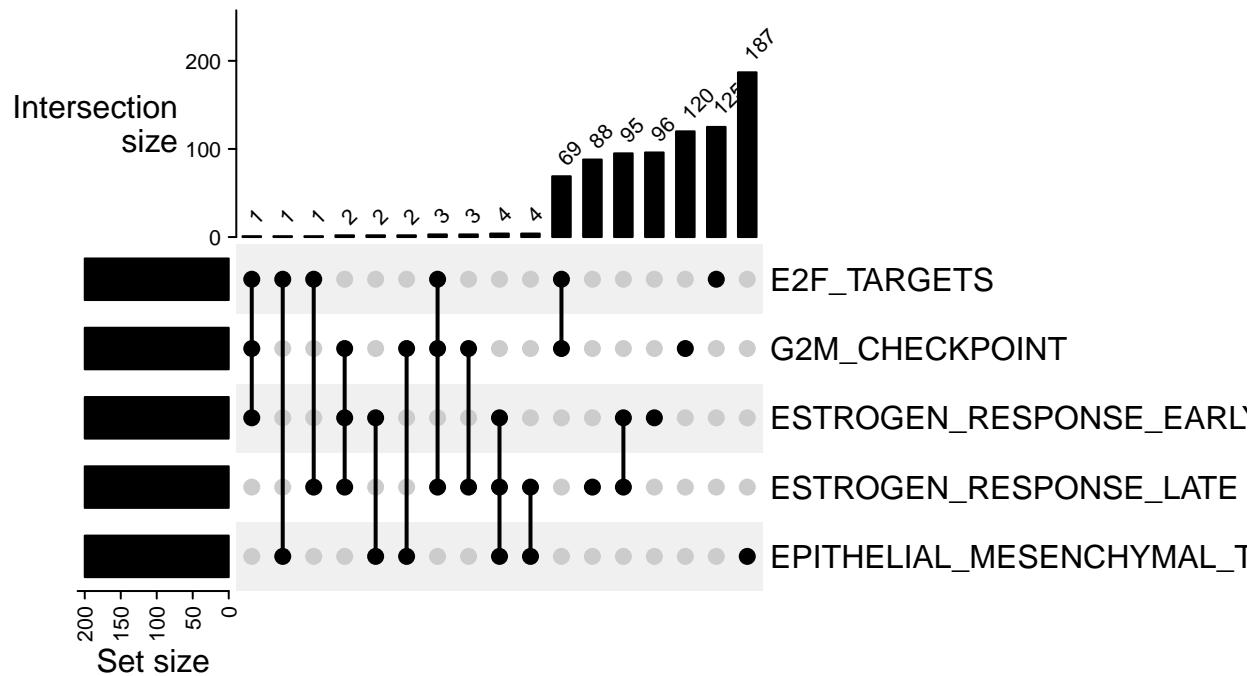



<https://app.gofigr.io/r/2eea2ed6-f456-4d0b-a847-aa72b8ff6e69>



change set order

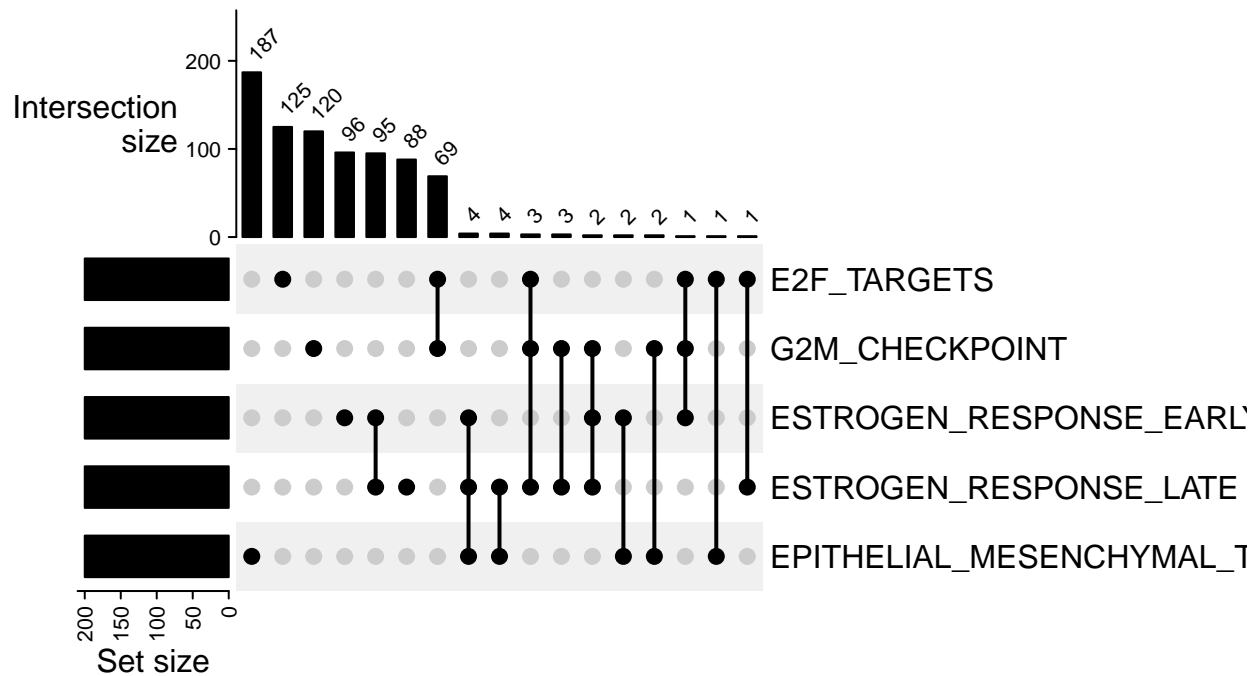
```
publish(UpSet(m,
  left_annotation = upset_left_annotation(m),
  top_annotation = upset_top_annotation(m, add_numbers = TRUE),
  set_order = c("E2F_TARGETS", "G2M_CHECKPOINT", "ESTROGEN_RESPONSE_EARLY",
    "ESTROGEN_RESPONSE_LATE", "EPITHELIAL_MESENCHYMAL_TRANSITION"),
  comb_order = order(comb_size(m))),
  "UpSet with custom set order")
```



<https://app.gofigr.io/r/5b7ccef4-947d-452b-a415-1be382a2a8b9>



```
publish(UpSet(m,
  left_annotation = upset_left_annotation(m),
  top_annotation = upset_top_annotation(m, add_numbers = TRUE),
  set_order = c("E2F_TARGETS", "G2M_CHECKPOINT", "ESTROGEN_RESPONSE_EARLY",
    "ESTROGEN_RESPONSE_LATE", "EPITHELIAL_MESENCHYMAL_TRANSITION"),
  comb_order = order(-comb_size(m)),
  "UpSet with reverse comb order")
```

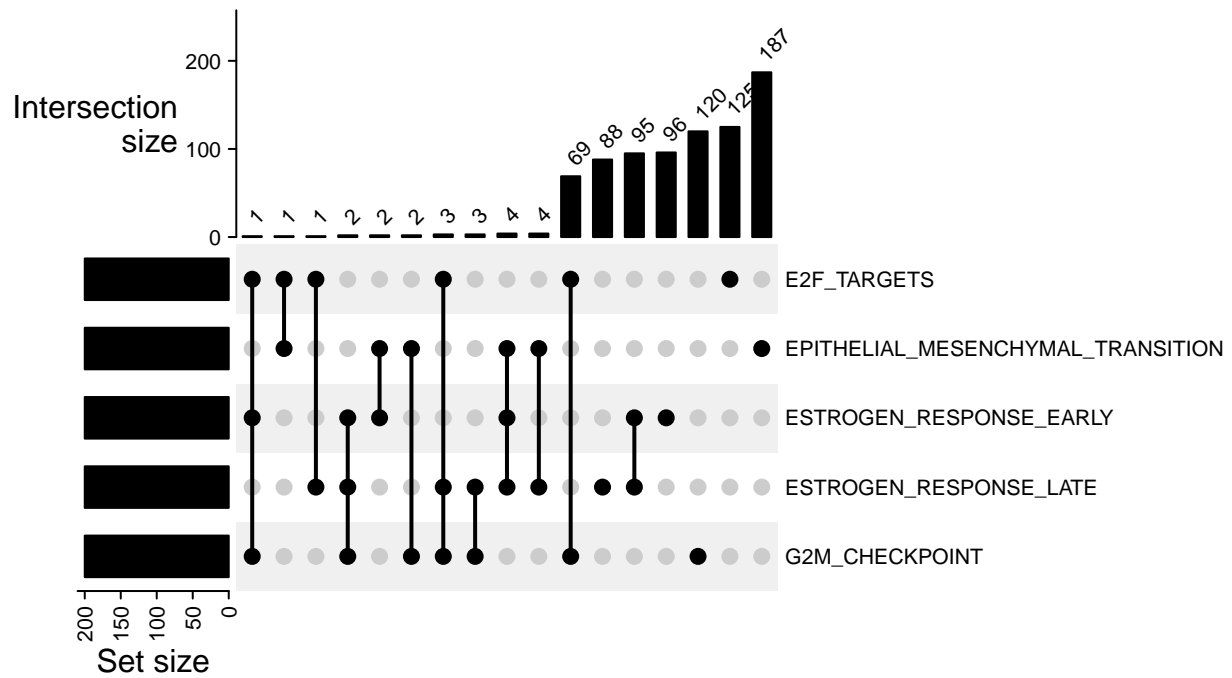


<https://app.gofigr.io/r/00ff620d-b5ab-4173-8896-937efc6f8b5c>



Change the rowname size

```
publish(UpSet(m, left_annotation = upset_left_annotation(m),
  top_annotation = upset_top_annotation(m, add_numbers = TRUE),
  comb_order = order(comb_size(m)),
  row_names_gp = gpar(fontsize = 8)),
  "UpSet with row names")
```



<https://app.gofigr.io/r/2a994f05-c736-4c04-8ea5-d2ac4a33147e>

