

Eplet Comparison Workflow

Patricia Hernandez

Nick Borcharding

2025-04-01

Loading Library

This will load the libraries necessary to run the analysis. If the libraries/packages are not installed, this code will install them before loading them.

```
required_libraries <- c(
  "dplyr", "data.table", "tidyr", "janitor", "stringr", "venn",
  "lubridate", "readxl", "openxlsx", "tidyverse", "hrbrthemes",
  "proustr", "VennDiagram", "ComplexUpset"
)

# Function to check, install if needed, and load libraries
load_libraries <- function(libraries) {
  for (lib in libraries) {
    if (!requireNamespace(lib, quietly = TRUE)) {
      install.packages(lib, dependencies = TRUE)
    }
    library(lib, character.only = TRUE)
  }
}

# Execute the function
load_libraries(required_libraries)
```

Loading Functions

These are custom functions written for the analysis to assist in loading and plotting the data. More information can be found in the github repository under the R folder.

```
source("./R/load_data.R")
source("./R/utils.R")
source("./R/calculate_sum.R")
source("./R/generate_venn_diagram.R")
source("./R/generate_upset_plot.R")
source("./R/summarise_eplet.R")
```

Data Formats

The most important step before running the analysis is to ensure that the data is consistently formatted. Much of the functions are specific to the column names and structure of the individual cells within the excel file or the document.

SAB Assay

```
SAB <- read_xlsx("data/SAB_Class_I_Lot_14_20221212.xlsx", sheet = 1)
head(SAB)
```

```
# A tibble: 6 x 11
  Column1 `Bead ID` `Antigen ID` `Molecular Typing` `Serological Typing`
    <dbl>    <dbl> <chr>          <chr>          <chr>
1      1      1 NC          NA             N/A
2      2      2 PC          NA             N/A
3      3      3 rA0101      A*01:01        A1
4      4      4 rA0201      A*02:01        A2
5      5      5 rA0203      A*02:03        A2
6      6      6 rA0206      A*02:06        A2
# i 6 more variables: `*W6/32` <dbl>, Results <lgl>, Column8 <chr>,
#   Column9 <chr>, Column10 <chr>, Column11 <chr>
```

PRA Assay

```
PRA <- read_xlsx("data/PRA_Class_I_Lot_020_20221212.xlsx", sheet = 1)
head(PRA)
```

```
# A tibble: 6 x 11
  Order `Bead ID` `Antigen ID` `A Left` `A Right` `B Left` `B Right` `C Left`
  <dbl> <chr>      <chr>      <chr>      <chr>      <chr>      <chr>      <chr>
1     1 1 1      NC          N/A        N/A        N/A        N/A        N/A
2     2 2 2      PC          N/A        N/A        N/A        N/A        N/A
3     3 3 3      C4966      A*02:01    A*02:07    B*46:01    X          C*01:02
4     4 4 4      E19109     A*01:01    A*23:01    B*49:01    B*55:01    C*03:03
5     5 5 5      G0142      A*11:02    A*24:02    B*27:06    B*40:01    C*03:04
6     6 6 6      E5482      A*11:01    A*24:02/50/~ B*54:01    B*59:01    C*01:02
# i 3 more variables: `C Right` <chr>, `Serological Typing` <chr>,
# Results <lgl>
```

Mix Assay

```
Mix <- read_xlsx("data/LSM12NC23_024_01.xlsx", sheet = 1)
head(Mix)
```

```
# A tibble: 6 x 8
  Bead `Pos Ctr` Abbreviated Specificit~1 ...4 `Recognition Site` ...6 Cutoff
  <dbl> <lgl>      <chr>      <lgl> <lgl>      <lgl> <lgl>
1     1 NA      -, -, -, -, -, -, - NA NA      NA NA
2     2 NA      -, -, -, -, -, -, - NA NA      NA NA
3     6 NA      A1,A80,B18,B50,Bw6,-,Cw~ NA NA      NA NA
4     6 NA      A1,A29,B8,B45,Bw6,-,Cw6~ NA NA      NA NA
5     6 NA      A1,A69,B8,B55,Bw6,-,Cw1~ NA NA      NA NA
6     9 NA      A11,A24,B35,B62,Bw6,-,C~ NA NA      NA NA
# i abbreviated name: 1: `Abbreviated Specificity`
# i 1 more variable: Specificity <chr>
```

ExPlex Assay

```
ExPlex <- read_xlsx("data/ExPlex_classI_II.xlsx", sheet = 1)
head(ExPlex)
```

```
# A tibble: 6 x 1
  Molecular.Typing
  <chr>
1 A*01:02
```

```
2 A*02:10
3 A*02:05
4 A*02:07
5 A*02:18
6 A*26:02
```

Figure 1: Allele Counting

The following code will load all the assay information and automatically format them with the `load_XX_data()` functions. From there, we can count the alleles.

Class I Alleles

```
sab <- process_sab_data(file = "./data/SAB_Class_I_Lot_14_20221212.xlsx", c(1,2))
pra <- process_pra_data(file = "./data/PRA_Class_I_Lot_020_20221212.xlsx", 1)
mix <- process_mix_data(file = "./data/LSM12NC23_024_01.xlsx", 1)
explex <- process_explex_data(file = "./data/ExPlex_classI_II.xlsx", 1)

#Remove non-type-I from Mix
mix <- mix[grepl("A_|B_|C_", mix$values),]

b <- list(
  SAB = unique(sab$Molecular.Typing),
  PRA = unique(pra$values),
  Mixed = unique(mix$values),
  ExPlex = unique(explex$Molecular.Typing)
)

# Create a binary membership matrix
all_elements <- unique(unlist(b))
membership_matrix <- data.frame(
  element = all_elements,
  SAB = all_elements %in% b$SAB,
  PRA = all_elements %in% b$PRA,
  Mixed = all_elements %in% b$Mix,
  ExPlex = all_elements %in% b$ExPlex
)

# Convert logical values to binary
membership_matrix <- membership_matrix %>%
```

```

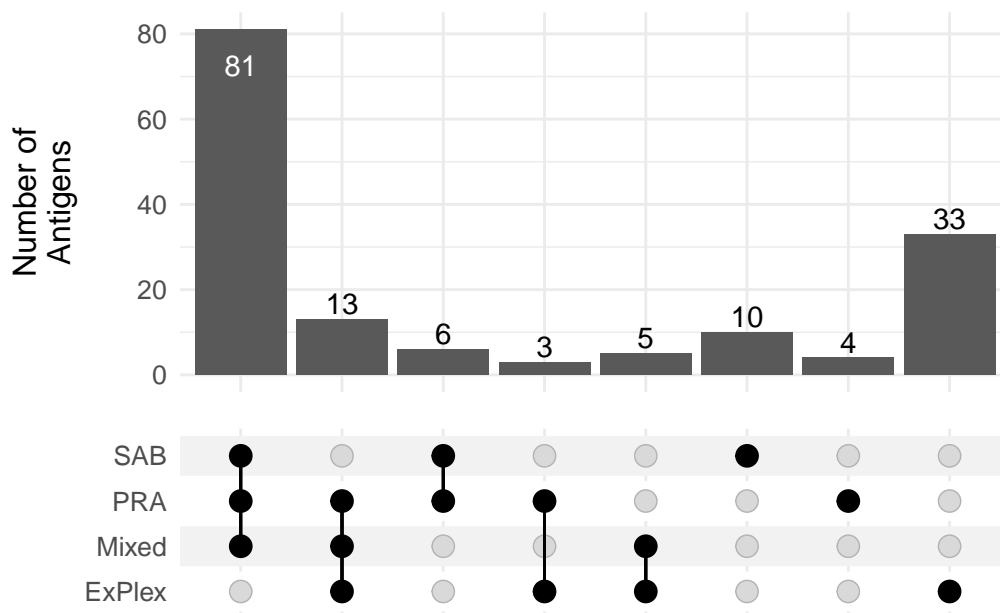
select(element, ExPlex, Mixed, PRA, SAB)

# Generate the UpSet plot
upset_data <- membership_matrix %>% select(-element)

upset_data <- upset_data[, c("ExPlex", "Mixed", "PRA", "SAB")]

# Then specify the same order in the `intersect` argument:
upset(
  upset_data,
  intersect = colnames(upset_data), # Now in alphabetical order
  sort_intersections_by='degree',
  base_annotations = list('Number of\nAntigens' = intersection_size()),
  themes = upset_default_themes(text = element_text(size = 12)),
  set_sizes = FALSE,
  sort_sets = FALSE
) +
  theme(axis.title.x = element_blank())

```



```

ggsave('outputs/viz/Figure1_ClassI.png', width = 8, height = 6)

# Summarizing results for supplemental table

```

```

summary_table <- membership_matrix %>%
  pivot_longer(cols = -element, names_to = "Assay", values_to = "Present") %>%
  filter(Present) %>%
  group_by(element) %>%
  summarize(Assay_Combination = paste(sort(Assay), collapse = "_")) %>%
  group_by(Assay_Combination) %>%
  summarize(
    Allele_Count = n(),
    Alleles = paste(sort(element), collapse = ", ")
  ) %>%
  arrange(desc(Allele_Count))
write.csv(summary_table, "outputs/files/SupplementalTable1_ClassI_Allele_Overlap_Table.csv",

```

Class II Alleles

```

sab <- process_sab_data(file = "./data/SAB_ClassII_Lot_15_20221212.xlsx", c(1:2))[-c(1:2),]
pra <- process_pra_data(file = "./data/PRA_Class_II_Lot_019_20221212.xlsx", 1)
mix <- process_mix_data(file = "./data/LSM12NC23_024_01.xlsx", 1)
explex <- process_explex_data(file = "./data/ExPlex_classI_II.xlsx", 2)

#Remove type-I from Mix
mix <- mix[!grepl("A_|B_|C_", mix$values),]

b <- list(
  SAB = unique(sab$Molecular.Typing),
  PRA = unique(pra$values),
  Mixed = unique(mix$values),
  ExPlex = unique(explex$Molecular.Typing)
)

# Create a binary membership matrix
all_elements <- unique(unlist(b))
membership_matrix <- data.frame(
  element = all_elements,
  SAB = all_elements %in% b$SAB,
  PRA = all_elements %in% b$PRA,
  Mixed = all_elements %in% b$Mix,
  ExPlex = all_elements %in% b$ExPlex
)

```

```

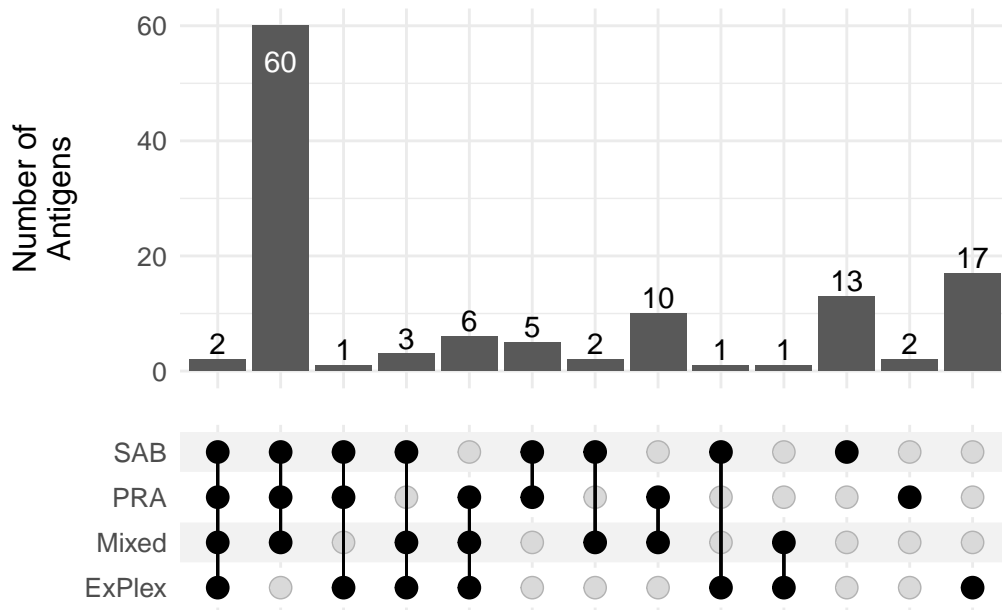
# Convert logical values to binary
membership_matrix <- membership_matrix %>%
  select(element, ExPlex, Mixed, PRA, SAB)

# Generate the UpSet plot
upset_data <- membership_matrix %>% select(-element)

upset_data <- upset_data[, c("ExPlex", "Mixed", "PRA", "SAB")]

# Then specify the same order in the `intersect` argument:
upset(
  upset_data,
  intersect = colnames(upset_data), # Now in alphabetical order
  sort_intersections_by='degree',
  base_annotations = list('Number of\nAntigens' = intersection_size()),
  themes = upset_default_themes(text = element_text(size = 12)),
  set_sizes = FALSE,
  sort_sets = FALSE
) +
  theme(axis.title.x = element_blank())

```



```

ggsave('outputs/viz/Figure1_ClassII.png', width = 8, height = 6)

# Summarizing results for supplemental table
summary_table <- membership_matrix %>%
  pivot_longer(cols = -element, names_to = "Assay", values_to = "Present") %>%
  filter(Present) %>%
  group_by(element) %>%
  summarize(Assay_Combination = paste(sort(Assay), collapse = "_")) %>%
  group_by(Assay_Combination) %>%
  summarize(
    Allele_Count = n(),
    Alleles = paste(sort(element), collapse = ", ")
  ) %>%
  arrange(desc(Allele_Count))
write.csv(summary_table, "outputs/files/SupplementalTable2_ClassII_Allele_Overlap_Table.csv")

```

Figure 2/3: Eplet Counting

Here we are going to calculate the respective eplet load within the individual assays. For each assay, we use the `calculate_sum()` function to add up the unique eplets represented by the alleles within the assay.

Class I Eplets

```

# Load and preprocess data
df <- process_hla_registry(file = "./data/HMM-CL01-0620-CAT-00.xlsx", 1)
sab <- process_sab_data(file = "./data/SAB_Class_I_Lot_14_20221212.xlsx", c(1,2))
pra <- process_pra_data(file = "./data/PRA_Class_I_Lot_020_20221212.xlsx", 1)
mix <- process_mix_data(file = "./data/LSM12NC23_024_01.xlsx", 1)
explex <- process_explex_data(file = "./data/ExPlex_classI_II.xlsx", 1)

# Add allele columns and calculate sums
df <- calculate_sum(df, sab$Molecular.Typing, "SAB_unique")
df <- calculate_sum(df, pra$values, "PRA_unique")
df <- calculate_sum(df, mix$values, "Mix_unique")
df <- calculate_sum(df, explex$Molecular.Typing, "EXPLEX_unique")

# Generate Venn diagrams
generate_venn_diagram(df, "outputs/viz/Figure2_classI.png",
  pattern = "_unique")

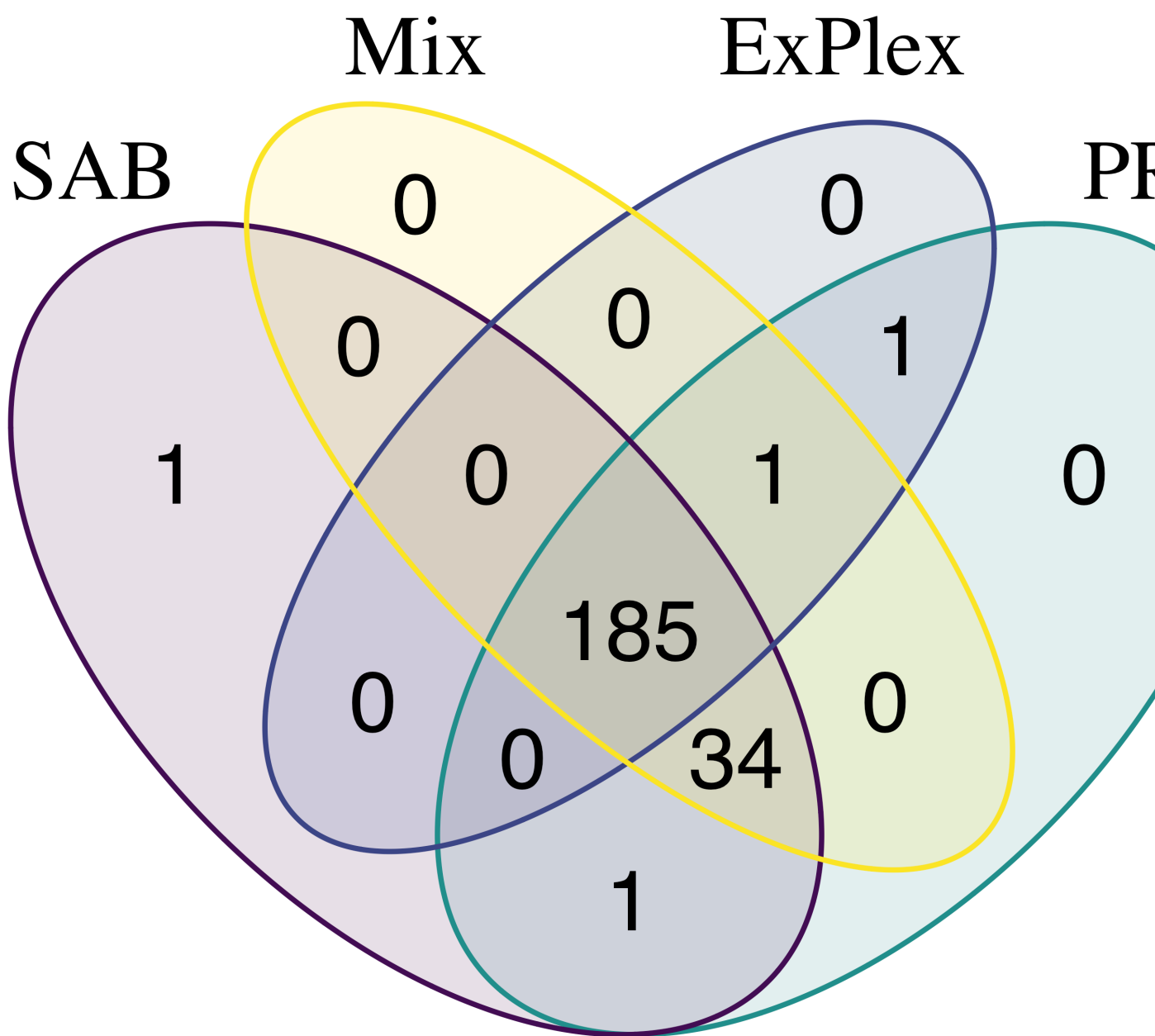
```

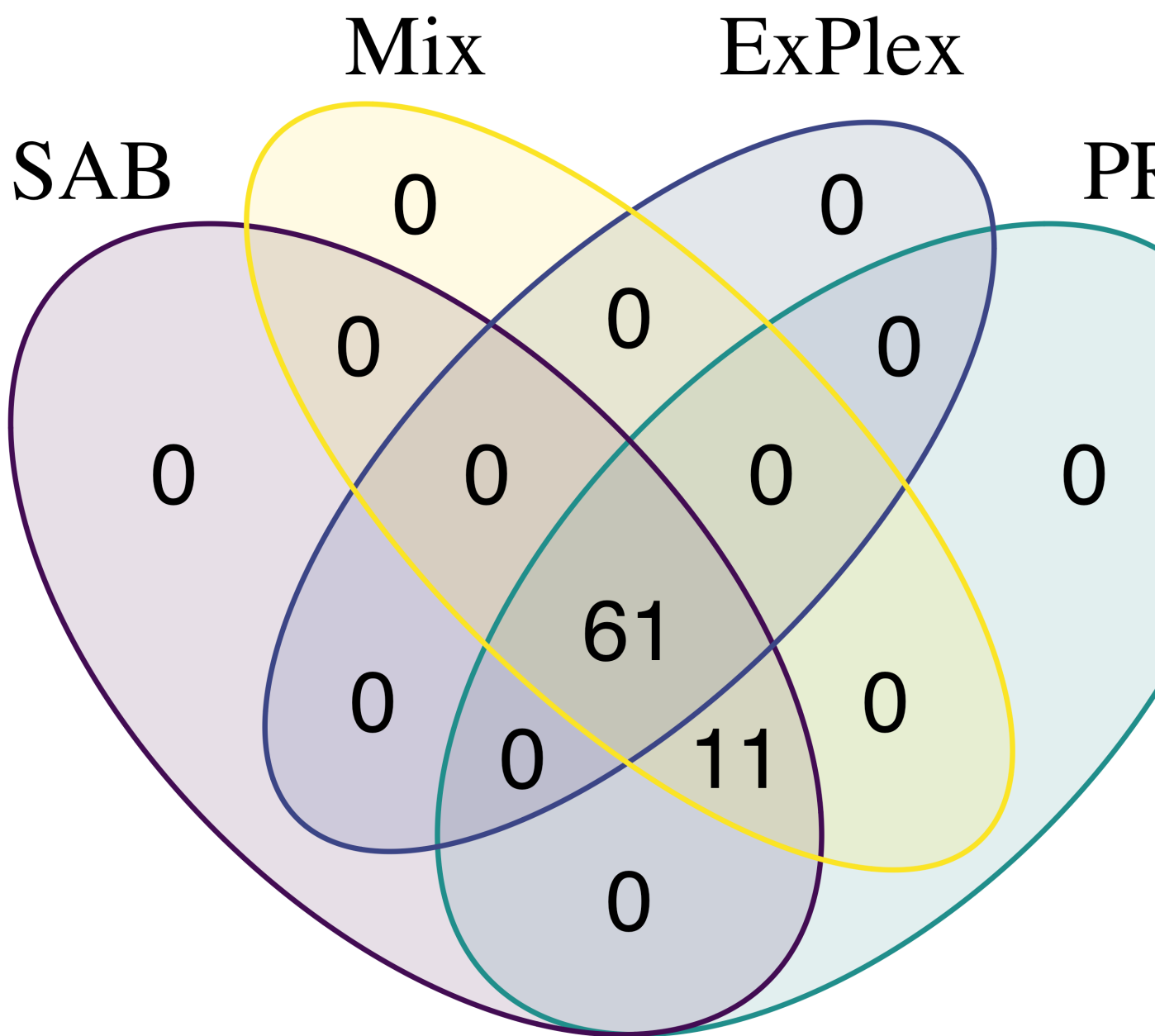

[1] 1

```
# Filter antibody-verified data and generate another Venn diagram
dfa <- df %>% filter(`Antibody Reactivity` == "Confirmed")
generate_venn_diagram(dfa, "outputs/viz/Figure3_classI.png",
                      pattern = "_unique")
```

[1] 1

```
# Generating supplemental tables
summary_table <- summarise_eplet(df)
write.csv(summary_table, "outputs/files/SupplementalTable3_ClassI_Eplet_Overlap_Table.csv", 1)
summary_table <- summarise_eplet(dfa)
write.csv(summary_table, "outputs/files/SupplementalTable5_ClassI_Eplet_Overlap_Table_AbConf")
```





Class II Eplets

```
df <- process_hla_registry(file = "./data/HMM-CL02-0620-CAT-00.xlsx", 1)
sab <- process_sab_data(file = "./data/SAB_ClassII_Lot_15_20221212.xlsx", c(1:2))[-c(1:2),]
pra <- process_pra_data(file = "./data/PRA_Class_II_Lot_019_20221212.xlsx", 1)
mix <- process_mix_data(file = "./data/LSM12NC23_024_01.xlsx", 1)
explex <- process_explex_data(file = "./data/ExPlex_classI_II.xlsx", 2)

# Add allele columns and calculate sums
df <- calculate_sum(df, sab$Molecular.Typing, "SAB_unique")
df <- calculate_sum(df, pra$values, "PRA_unique")
df <- calculate_sum(df, mix$values, "Mix_unique")
df <- calculate_sum(df, explex$Molecular.Typing, "EXPLEX_unique")

# Generate Venn diagrams
generate_venn_diagram(df, "outputs/viz/Figure2_classII.png", pattern = "_unique")
```

[1] 1

```
# Filter antibody-verified data and generate another Venn diagram
dfa <- df %>% filter(`Antibody Reactivity` == "Confirmed")
generate_venn_diagram(dfa, "outputs/viz/Figure3_classII.png", pattern = "_unique")
```

[1] 1

```
# Generating supplemental tables
summary_table <- summarise_eplet(df)
write.csv(summary_table, "outputs/files/SupplementalTable4_ClassII_Eplet_Overlap_Table.csv",
summary_table <- summarise_eplet(dfa)
write.csv(summary_table, "outputs/files/SupplementalTable6_ClassII_Eplet_Overlap_Table_AbCon
```

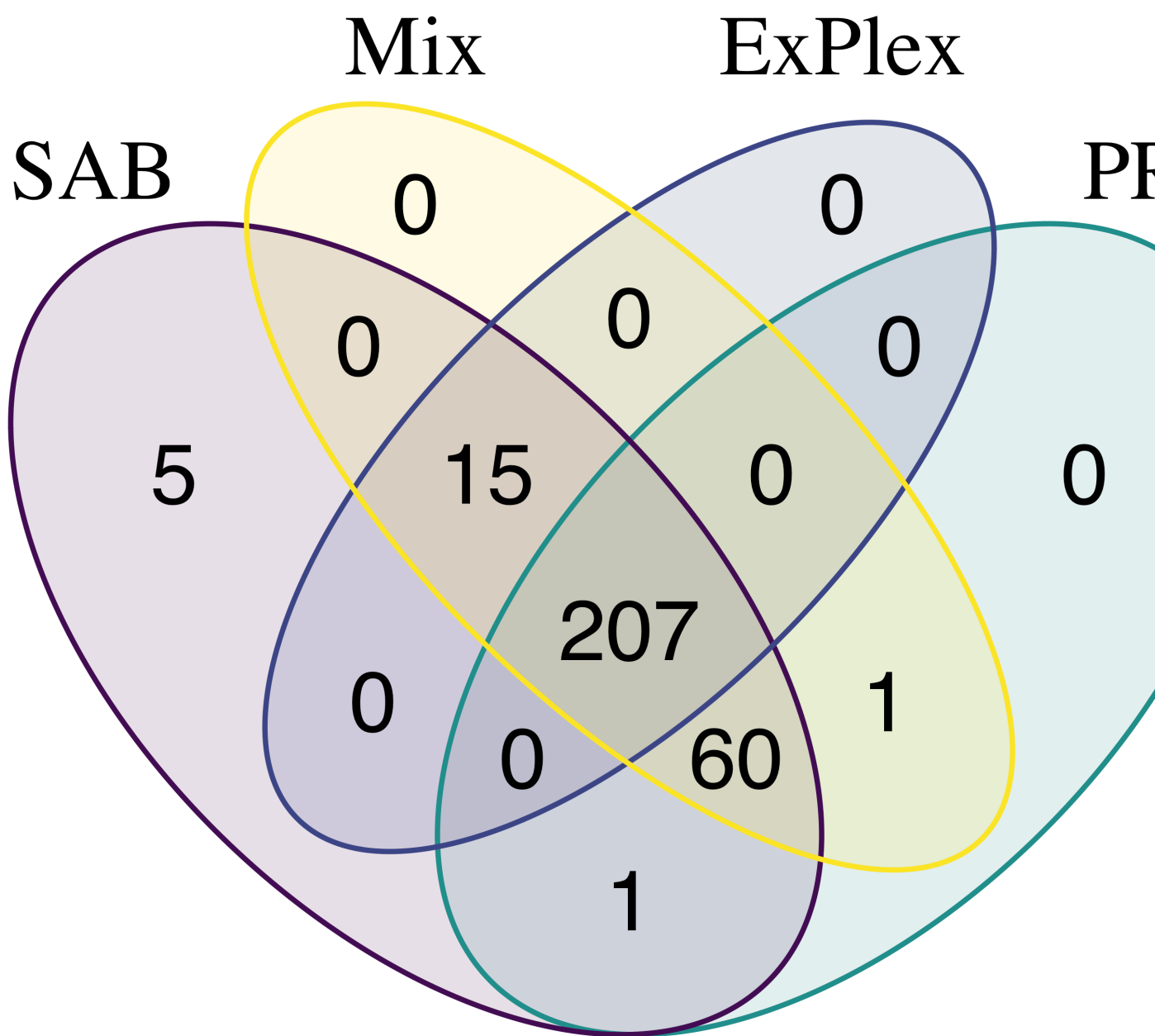


Figure 4: Unique alleles mapped

Class I

```
# Load and preprocess data
df <- process_hla_registry(file = "./data/HMM-CL01-0620-CAT-00.xlsx", 1)
sab <- process_sab_data(file = "./data/SAB_Class_I_Lot_14_20221212.xlsx",
                        c(1,2))
pra <- process_pra_data(file = "./data/PRA_Class_I_Lot_020_20221212.xlsx", 1)
mix <- process_mix_data(file = "./data/LSM12NC23_024_01.xlsx", 1)
explex <- process_explex_data(file = "./data/ExPlex_classI_II.xlsx", 1)

#Remove non-type-I from Mix
mix <- mix[grepl("A_|B_|C_", mix$values),]

#Making Data Frames
sab.frame <- calculate_sum(df, sab$Molecular.Typing, "SAB_sum",
                          return.sum = FALSE)
sab.frame <- sab.frame%>%
  select(-c('Alleles','Antibody Reactivity',
            'Class','Exposed','Sero Group',
            'AA Position')) %>%
  as.data.frame() %>%
  replace_with_colnames_and_col4() %>%
  select(-1) %>%
  unlist() %>%
  unique() %>%
  .[. != 0]

explex.frame <- calculate_sum(df, explex$Molecular.Typing, "EXPLEX_sum",
                             return.sum = F)
explex.frame <- explex.frame%>%
  select(-c('Alleles','Antibody Reactivity',
            'Class','Exposed','Sero Group',
            'AA Position')) %>%
  as.data.frame() %>%
  replace_with_colnames_and_col4() %>%
  select(-1) %>%
  unlist() %>%
  unique() %>%
  .[. != 0]
```

```

pra.frame <- calculate_sum(df, pra$values, "PRA_sum",
                          return.sum = F)
pra.frame <- pra.frame%>%
  select(-c('Alleles','Antibody Reactivity',
            'Class','Exposed','Sero Group',
            'AA Position')) %>%
  as.data.frame() %>%
  replace_with_colnames_and_col4() %>%
  select(-1) %>%
  unlist() %>%
  unique() %>%
  .[. != 0]

mix.frame <- calculate_sum(df, mix$values, "MIX_sum",
                          return.sum = F)
mix.frame <- mix.frame %>%
  select(-c('Alleles','Antibody Reactivity',
            'Class','Exposed','Sero Group',
            'AA Position')) %>%
  as.data.frame() %>%
  replace_with_colnames_and_col4() %>%
  select(-1) %>%
  unlist() %>%
  unique() %>%
  .[. != 0]

b=list(
  SAB = sab.frame,
  PRA = pra.frame,
  Mixed = mix.frame,
  ExPlex = explex.frame)

# Create a binary membership matrix
all_elements <- unique(unlist(b))
membership_matrix <- data.frame(
  element = all_elements,
  SAB = all_elements %in% b$SAB,
  PRA = all_elements %in% b$PRA,
  Mixed = all_elements %in% b$Mix,
  ExPlex = all_elements %in% b$ExPlex
)

```



```

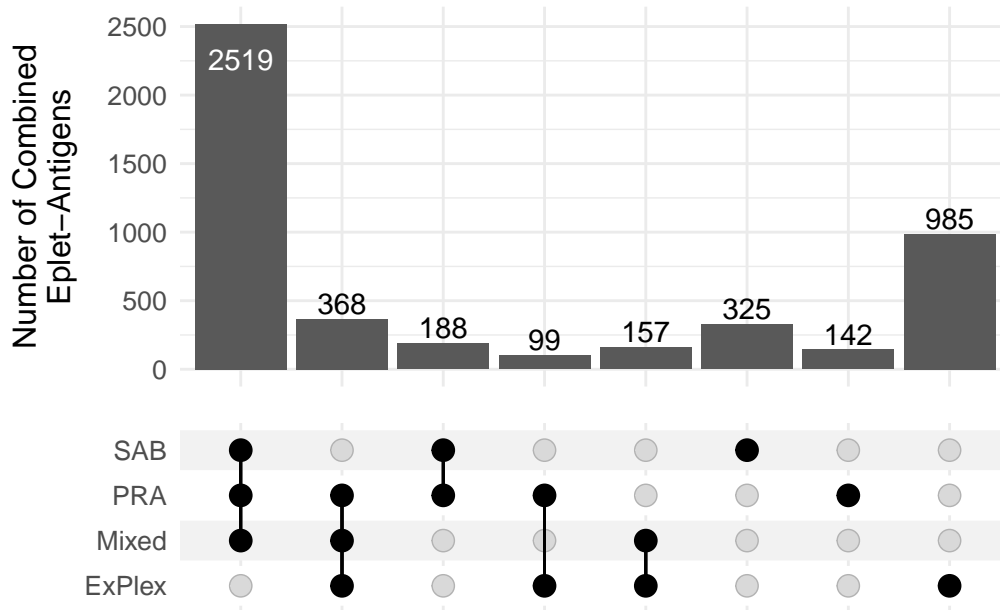
# Convert logical values to binary
membership_matrix <- membership_matrix %>%
  select(element, ExPlex, Mixed, PRA, SAB)

# Generate the UpSet plot
upset_data <- membership_matrix %>% select(-element)

upset_data <- upset_data[, c("ExPlex", "Mixed", "PRA", "SAB")]

# Then specify the same order in the `intersect` argument:
upset(
  upset_data,
  intersect = colnames(upset_data), # Now in alphabetical order
  base_annotations =
    list('Number of Combined\nEplet-Antigens' = intersection_size()),
  sort_intersections_by='degree',
  themes = upset_default_themes(text = element_text(size = 12)),
  set_sizes = FALSE,
  sort_sets = FALSE
) +
  theme(axis.title.x = element_blank())

```



```
ggsave('outputs/viz/Figure4_classI.png', width = 8, height = 6)
```

Class II

```
df <- process_hla_registry(file = "./data/HMM-CL02-0620-CAT-00.xlsx", 1)
sab <- process_sab_data(file = "./data/SAB_ClassII_Lot_15_20221212.xlsx",
                        c(1:2))[-c(1:2),]
pra <- process_pra_data(file = "./data/PRA_Class_II_Lot_019_20221212.xlsx", 1)
mix <- process_mix_data(file = "./data/LSM12NC23_024_01.xlsx", 1)
explex <- process_explex_data(file = "./data/ExPlex_classI_II.xlsx", 2)

#Remove non-type-I from Mix
mix <- mix[-grepl("A_|B_|C_", mix$values),]

#Making Data Frames
sab.frame <- calculate_sum(df, sab$Molecular.Typing, "SAB_sum",
                          return.sum = FALSE)
sab.frame <- sab.frame%>%
  select(-c('Alleles','Antibody Reactivity',
            'Class','Exposed','Sero Group',
            'AA Position')) %>%
  as.data.frame() %>%
  replace_with_colnames_and_col4() %>%
  select(-1) %>%
  unlist() %>%
  unique() %>%
  .[. != 0]

explex.frame <- calculate_sum(df, explex$Molecular.Typing, "EXPLEX_sum",
                             return.sum = F)
explex.frame <- explex.frame%>%
  select(-c('Alleles','Antibody Reactivity',
            'Class','Exposed','Sero Group',
            'AA Position')) %>%
  as.data.frame() %>%
  replace_with_colnames_and_col4() %>%
  select(-1) %>%
  unlist() %>%
  unique() %>%
```

```

      .[. != 0]

pra.frame <- calculate_sum(df, pra$values, "PRA_sum",
                          return.sum = F)
pra.frame <- pra.frame %>%
  select(-c('Alleles', 'Antibody Reactivity',
            'Class', 'Exposed', 'Sero Group',
            'AA Position')) %>%
  as.data.frame() %>%
  replace_with_colnames_and_col4() %>%
  select(-1) %>%
  unlist() %>%
  unique() %>%
  .[. != 0]

mix.frame <- calculate_sum(df, mix$values, "MIX_sum",
                          return.sum = F)
mix.frame <- mix.frame %>%
  select(-c('Alleles', 'Antibody Reactivity',
            'Class', 'Exposed', 'Sero Group',
            'AA Position')) %>%
  as.data.frame() %>%
  replace_with_colnames_and_col4() %>%
  select(-1) %>%
  unlist() %>%
  unique() %>%
  .[. != 0]

b=list(
  SAB = sab.frame,
  PRA = pra.frame,
  Mixed = mix.frame,
  ExPlex = explex.frame)

# Create a binary membership matrix
all_elements <- unique(unlist(b))
membership_matrix <- data.frame(
  element = all_elements,
  SAB = all_elements %in% b$SAB,
  PRA = all_elements %in% b$PRA,
  Mixed = all_elements %in% b$Mix,
  ExPlex = all_elements %in% b$ExPlex

```

```

)

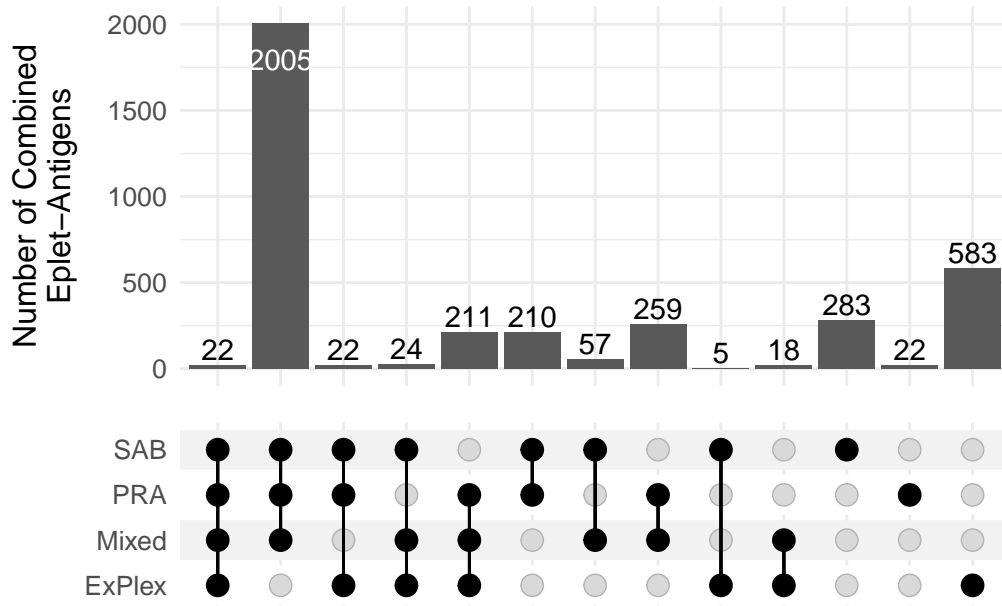
# Convert logical values to binary
membership_matrix <- membership_matrix %>%
  select(element, ExPlex, Mixed, PRA, SAB)

# Generate the UpSet plot
upset_data <- membership_matrix %>% select(-element)

upset_data <- upset_data[, c("ExPlex", "Mixed", "PRA", "SAB")]

# Then specify the same order in the `intersect` argument:
upset(
  upset_data,
  intersect = colnames(upset_data), # Now in alphabetical order
  base_annotations =
    list('Number of Combined\nEplet-Antigens' = intersection_size()),
  sort_intersections_by='degree',
  themes = upset_default_themes(text = element_text(size = 12)),
  set_sizes = FALSE,
  sort_sets = FALSE
) +
  theme(axis.title.x = element_blank())

```



```
ggsave('outputs/viz/Figure4_classII.png', width = 8, height = 6)
```

Conclusion

Here is the comprehensive summary and code of the analysis performed. The following package versions may be helpful in recreating the analysis or making your own.

```
sessionInfo()
```

```
R version 4.4.1 (2024-06-14)
Platform: aarch64-apple-darwin20
Running under: macOS Sonoma 14.6
```

```
Matrix products: default
```

```
BLAS: /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.4-arm64/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
```

```
time zone: America/Chicago
```

tzcode source: internal

attached base packages:

```
[1] grid      stats      graphics  grDevices  utils      datasets  methods
[8] base
```

other attached packages:

```
[1] ComplexUpset_1.3.3  VennDiagram_1.7.3  futile.logger_1.4.3
[4] proustr_0.4.0       hrbrthemes_0.8.7   forcats_1.0.0
[7] purrr_1.0.4         readr_2.1.5        tibble_3.2.1
[10] ggplot2_3.5.1       tidyverse_2.0.0    openxlsx_4.2.8
[13] readxl_1.4.5        lubridate_1.9.4    venn_1.12
[16] stringr_1.5.1       janitor_2.2.1      tidyr_1.3.1
[19] data.table_1.17.0   dplyr_1.1.4
```

loaded via a namespace (and not attached):

```
[1] gtable_0.3.6          xfun_0.51           tzdb_0.4.0
[4] vctrs_0.6.5           tools_4.4.1         generics_0.1.3
[7] pkgconfig_2.0.3       tokenizers_0.3.0    lifecycle_1.0.4
[10] farver_2.1.2          compiler_4.4.1      textshaping_1.0.0
[13] munsell_0.5.1         snakecase_0.11.1    fontquiver_0.2.1
[16] fontLiberation_0.1.0  htmltools_0.5.8.1   SnowballC_0.7.1
[19] yaml_2.3.10           Rttf2pt1_1.3.12     pillar_1.10.1
[22] extrafontdb_1.0       admisc_0.37         fontBitstreamVera_0.1.1
[25] tidyselect_1.2.1      zip_2.3.2           digest_0.6.37
[28] stringi_1.8.4         labeling_0.4.3      extrafont_0.19
[31] fastmap_1.2.0         colorspace_2.1-1    cli_3.6.4
[34] magrittr_2.0.3        patchwork_1.3.0     utf8_1.2.4
[37] withr_3.0.2           gdtools_0.4.1       scales_1.3.0
[40] timechange_0.3.0      lambda.r_1.2.4      rmarkdown_2.29
[43] cellranger_1.1.0     png_0.1-8           ragg_1.3.3
[46] hms_1.1.3            evaluate_1.0.3      knitr_1.49
[49] rlang_1.1.5          futile.options_1.0.1 Rcpp_1.0.14
[52] glue_1.8.0           formatR_1.14        attempt_0.3.1
[55] rstudioapi_0.17.1    jsonlite_1.9.1     R6_2.6.1
[58] systemfonts_1.2.1
```