Brain in a Blender cell type gene overlap

Kimberly Olney, Ph.D.

04/05/2023

Identify overlap of genes in the Brain in a Blender primary cell type list.

Setup

```
knitr::opts_knit$set(root.dir = ".")
```

User defined variables

```
source(here::here("scripts/R", "file_paths_and_colours.R"))
condition <- c("")
tool <- "star"
```

Function to create tables

```
fromList <- function (input) {
  # Same as original fromList()...
  elements <- unique(unlist(input))
  data <- unlist(lapply(input, function(x) {
      x <- as.vector(match(elements, x))
      }))
  data[is.na(data)] <- as.integer(0)
  data[data != 0] <- as.integer(1)
  data <- data.frame(matrix(data, ncol = length(input), byrow = F))
  data <- data[which(rowSums(data) != 0), ]
  names(data) <- names(input)
  # ... Except now it conserves your original value names!
  row.names(data) <- elements
  return(data)
}

makePaddedDataFrame <- function(l, ...) {
  maxlen <- max(sapply(l, length))
  data.frame(lapply(l, na.pad, len = maxlen), ...)
}
```

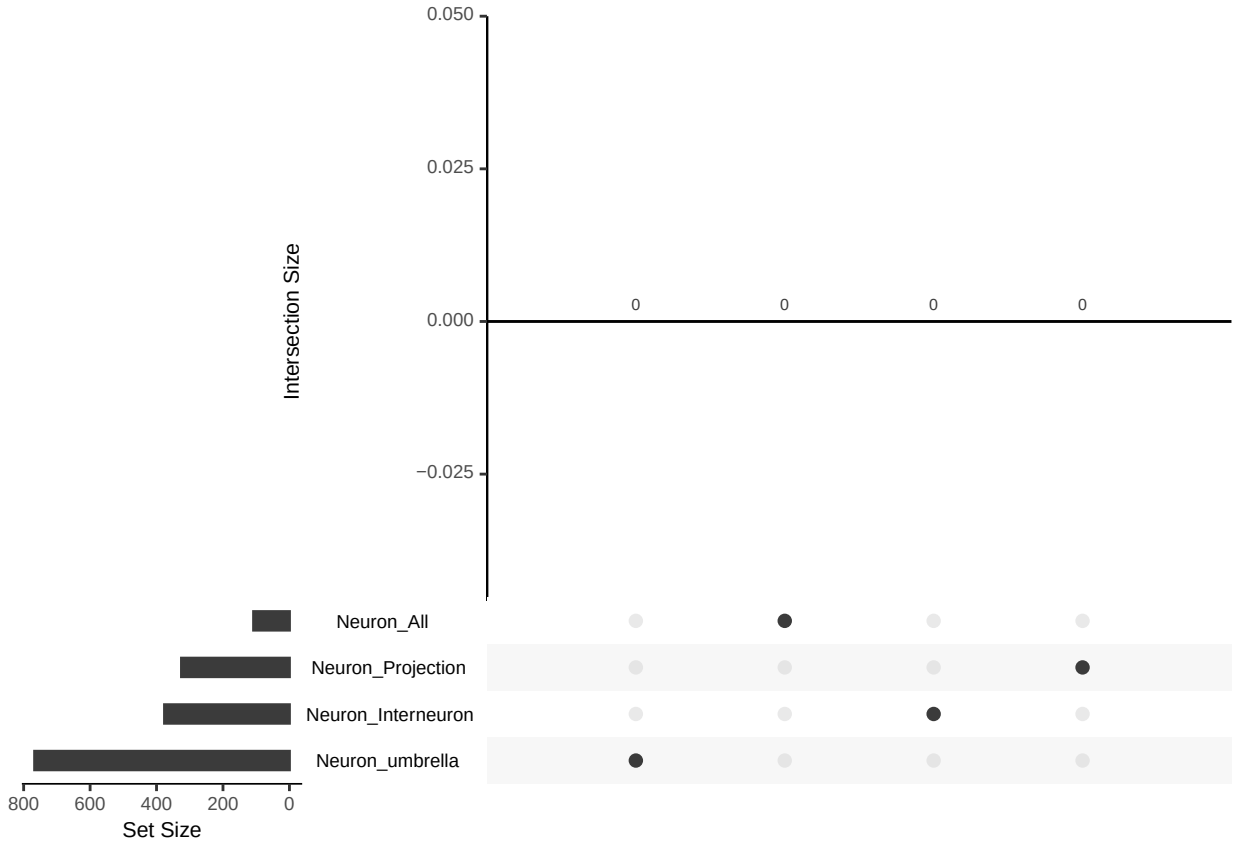
Read in cell marker table

```
#BinB <- read.delim("brain_blender_modified_cell_marker_table.tsv")
load("/research/labs/neurology/fryer/m239830/tools/BrainInABlender/data/CellTypeSpecificGenes_Master3.r
BinB <- as.data.frame(CellTypeSpecificGenes_Master3)
CellType_Primary <- levels(factor(BinB$CellType_Primary))
Umbrella.Cell.Type <- levels(factor(BinB$Umbrella.Cell.Type))

# split by primary cell types
CellType_splits_primary <- split(BinB, BinB$CellType_Primary) # Split data frame in list
for (i in 1:length(CellType_splits_primary)) {      # Run for-loop
  CellType_splits_primary[[i]] <- na.omit(CellType_splits_primary[[i]]) # remove rows with NA
  assign(CellType_Primary[i], CellType_splits_primary[[i]]) # assign cellType variable
}

# split by umbrella cell types
CellType_splits_umbrella <- split(BinB, BinB$Umbrella.Cell.Type) # Split data frame in list
for (i in 1:length(CellType_splits_umbrella)) {      # Run for-loop
  CellType_splits_umbrella[[i]] <- na.omit(CellType_splits_umbrella[[i]]) # remove rows with NA
  # assign cellType variable
  assign(paste0(Umbrella.Cell.Type[i], ".Umbrella"), CellType_splits_umbrella[[i]])
}
# rename choroid plexus
#Choroid_Plexus.Umbrella <- `Choroid Plexus.Umbrella`
```

Neurons

```
list_input <- list("Neuron_umbrella" = Neuron.Umbrella$GeneSymbol_Human,
                   "Neuron_All" = Neuron_All$GeneSymbol_Human,
                   "Neuron_Interneuron" = Neuron_Interneuron$GeneSymbol_Human,
                   "Neuron_Projection" = Neuron_Projection$GeneSymbol_Human
)
data <- fromList(list_input)
upset(data, intersect = c("Neuron_umbrella",
                          "Neuron_All",
                          "Neuron_Interneuron",
                          "Neuron_Projection"))
```

```
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
```




```
path <- paste0("../../results/star/cell_type_enrichment/BrainInABlender/upset_genes_Neurons")
saveToPDF(paste0(path, ".pdf"), width = 6.5, height = 5)
```

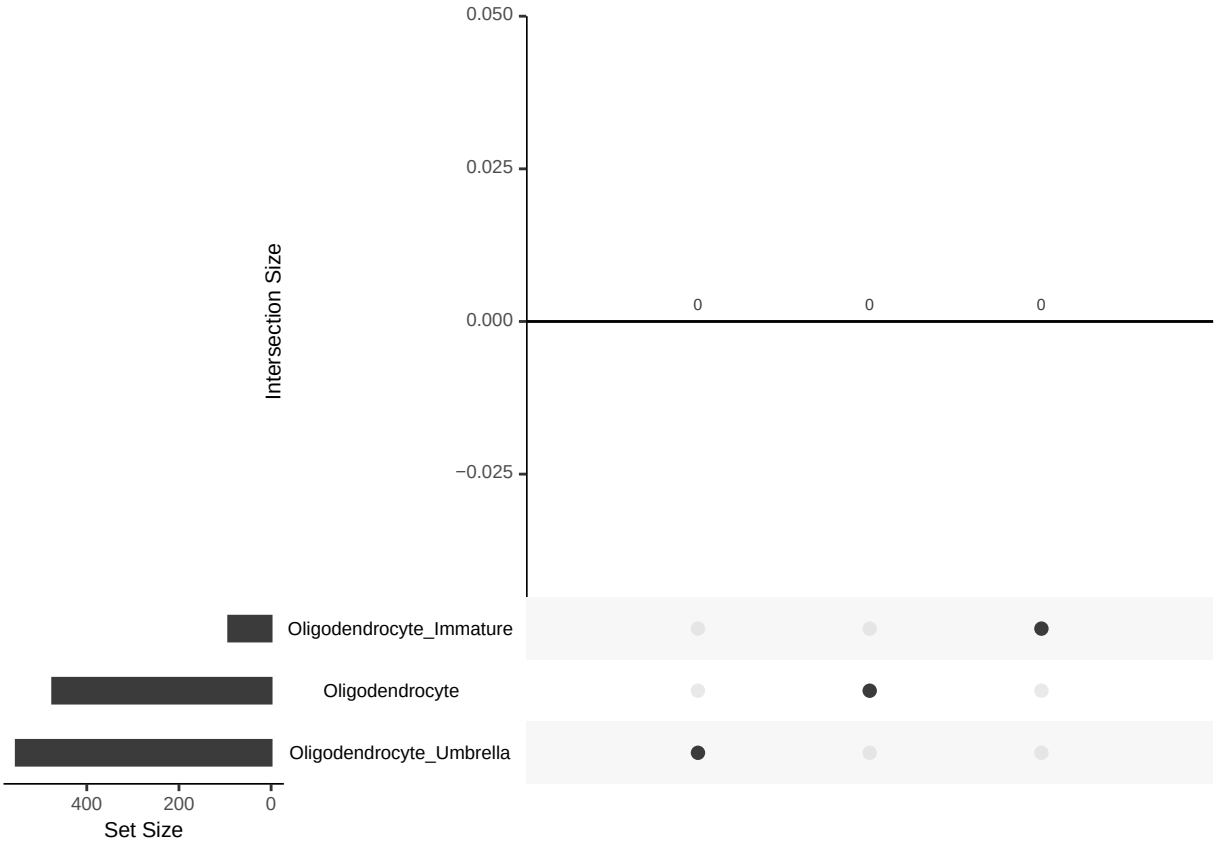
```
## pdf
##   2
```

```
remove(data, list_input)
```

Oligodendrocyte

```
list_input <- list("Oligodendrocyte_Umbrella" = Oligodendrocyte.Umbrella$GeneSymbol_Human,
                   "Oligodendrocyte" = Oligodendrocyte$GeneSymbol_Human,
                   "Oligodendrocyte_Immature" = Oligodendrocyte_Immature$GeneSymbol_Human
)
data <- fromList(list_input)
upset(data, intersect = c("Oligodendrocyte_Umbrella",
                          "Oligodendrocyte",
                          "Oligodendrocyte_Immature"))
```

```
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
```




```
path <- paste0("../../results/star/cell_type_enrichment/BrainInABlender/upset_genes_Oligodendrocyte")
saveToPDF(paste0(path, ".pdf"), width = 6.5, height = 5)
```

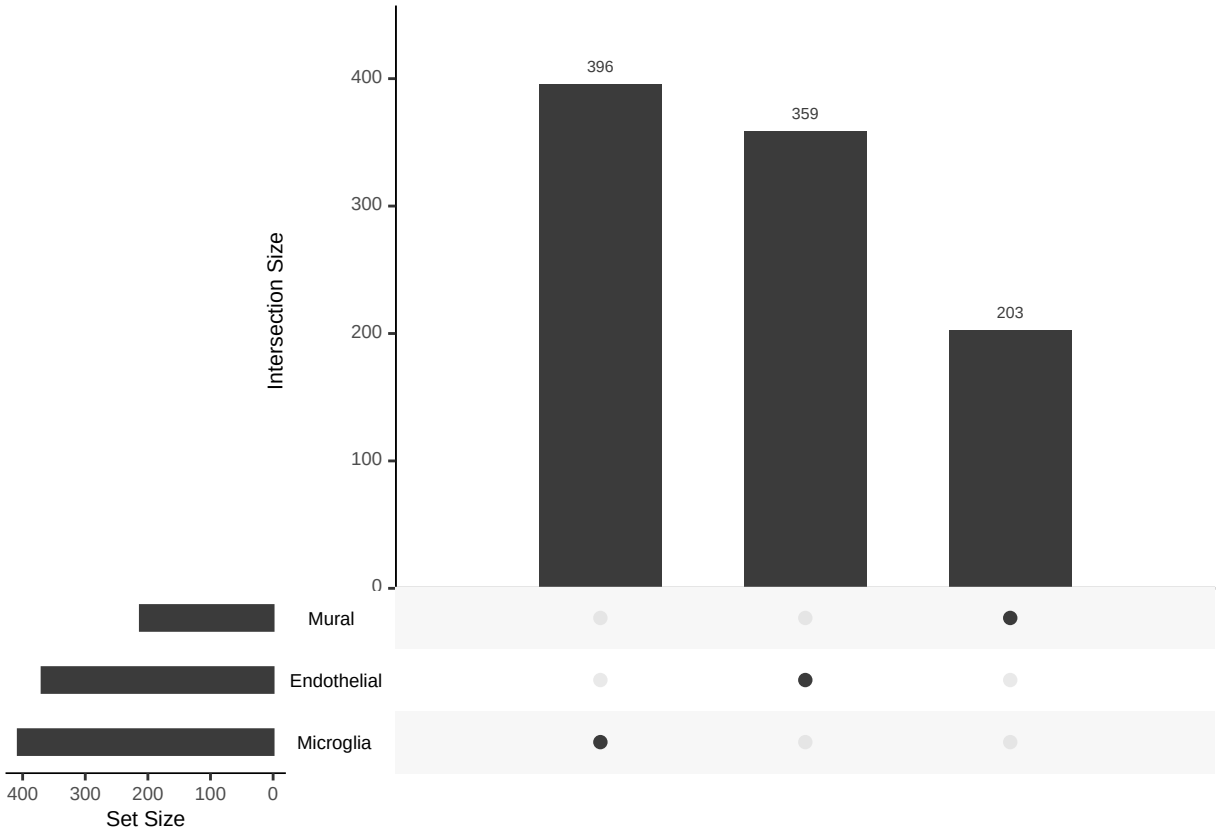
```
## pdf
##   2
```

```
remove(data, list_input)
```

Endothelial, Microglia, and Mural

```
list_input <- list("Endothelial" = Endothelial$GeneSymbol_Human,
                   "Microglia" = Microglia$GeneSymbol_Human,
                   "Mural" = Mural$GeneSymbol_Human
)
data <- fromList(list_input)
upset(
  data,
  intersect = c(
    "Endothelial",
    "Microglia",
    "Mural"
  )
)
```

```
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
```



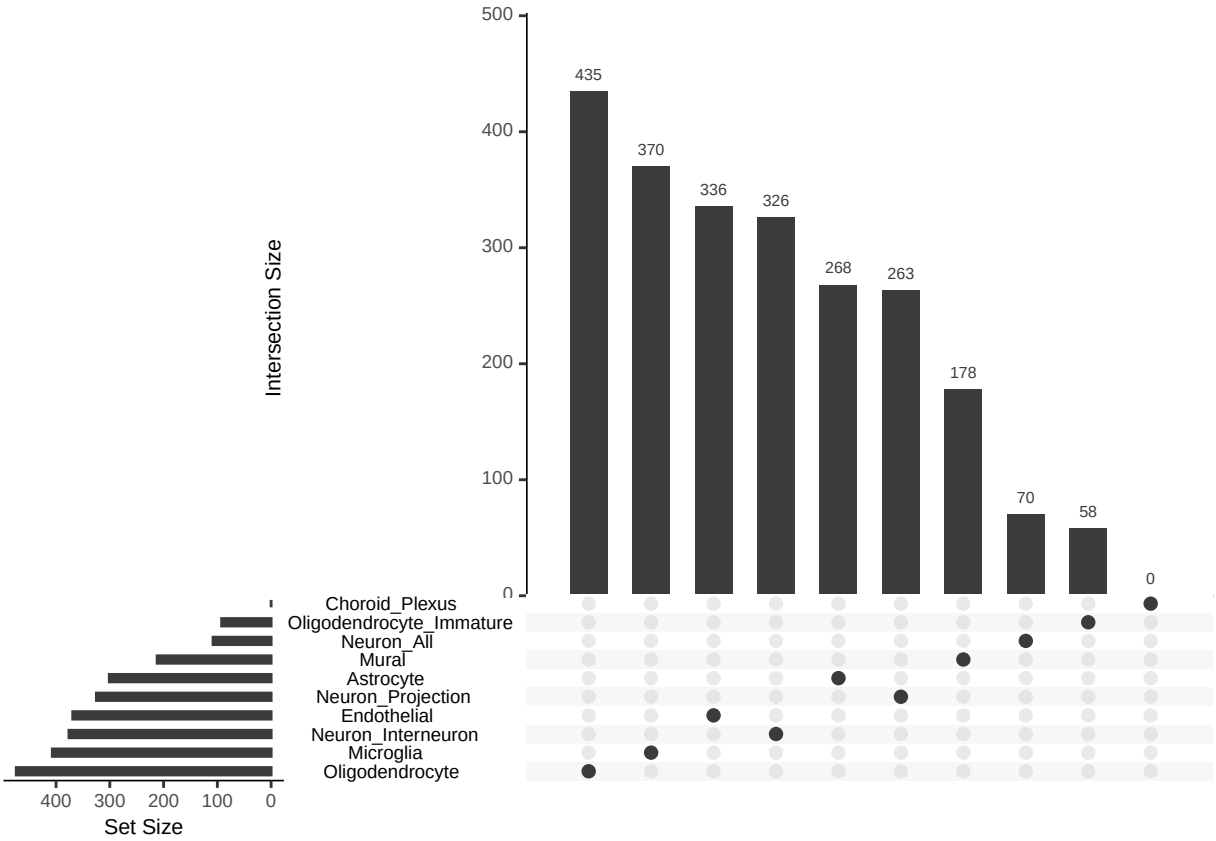

```
remove(data, list_input)
```

All primary cell types

```
list_input <- list("Astrocyte" = Astrocyte$GeneSymbol_Human,
                   "Oligodendrocyte" = Oligodendrocyte$GeneSymbol_Human,
                   "Oligodendrocyte_Immature" = Oligodendrocyte_Immature$GeneSymbol_Human,
                   "Choroid_Plexus" = Choroid_Plexus$GeneSymbol_Human,
                   "Endothelial" = Endothelial$GeneSymbol_Human,
                   "Microglia" = Microglia$GeneSymbol_Human,
                   "Mural" = Mural$GeneSymbol_Human,
                   "Neuron_All" = Neuron_All$GeneSymbol_Human,
                   "Neuron_Interneuron" = Neuron_Interneuron$GeneSymbol_Human,
                   "Neuron_Projection" = Neuron_Projection$GeneSymbol_Human
)
data <- fromList(list_input)
upset(
  data,
  intersect = c(
    "Astrocyte",
    "Oligodendrocyte",
    "Oligodendrocyte_Immature",
    "Choroid_Plexus",
    "Endothelial",
    "Microglia",
    "Mural",
```

```
    "Neuron_All",
    "Neuron_Interneuron",
    "Neuron_Projection"
  )
)
```

```
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
```




```
path <- paste0("../../results/star/cell_type_enrichment/BrainInABlender/upset_genes_all_primary_cell_ty
saveToPDF(paste0(path, ".pdf"), width = 12, height = 6)
```

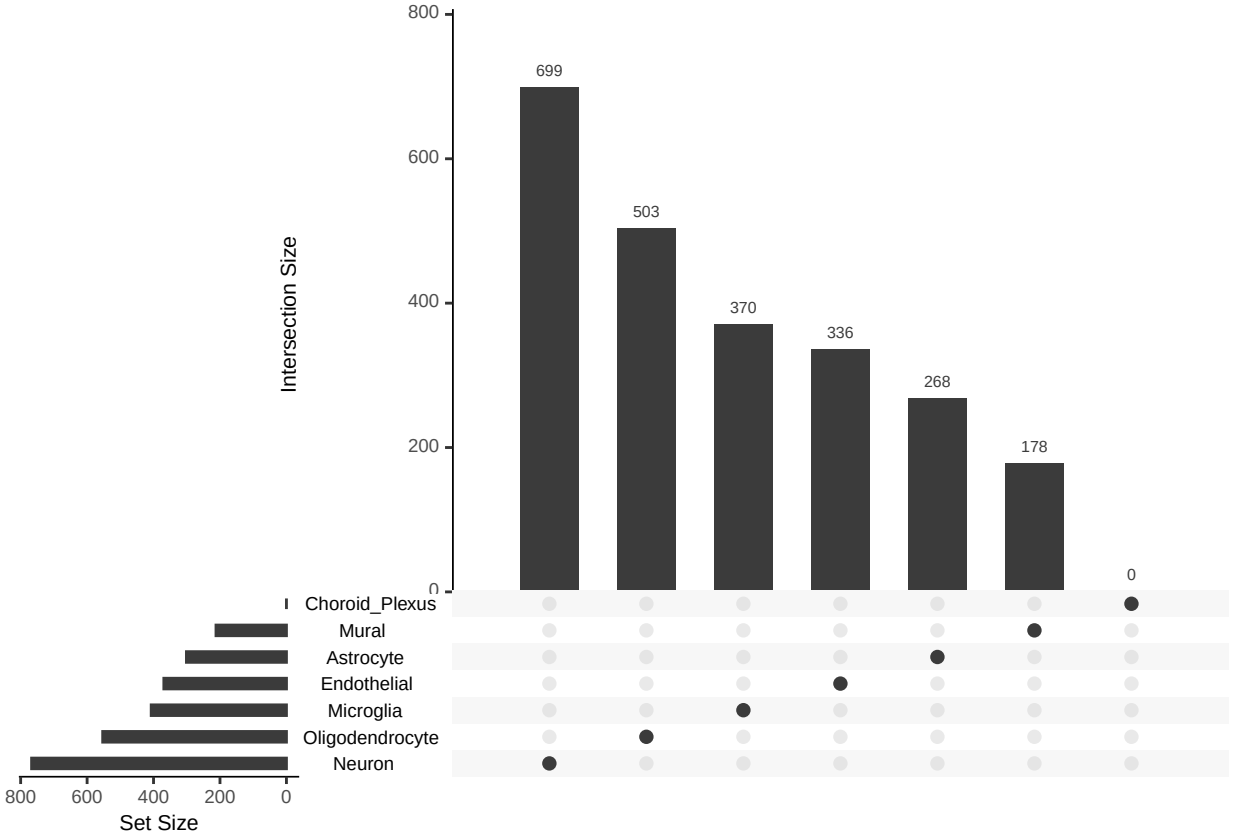
```
## pdf
##   2
```

```
# Binary table with colnames:
write.table(
  data,
  paste0("../../results/",
  tool,
  "/cell_type_enrichment/BrainInABlender/genes_shared_bt_primary_cell_types.txt"),
  sep = "\t",
  quote = FALSE
)
remove(data, list_input)
```

All umbrella cell types

```
list_input <- list("Astrocyte" = Astrocyte.Umbrella$GeneSymbol_Human,
                   "Oligodendrocyte" = Oligodendrocyte.Umbrella$GeneSymbol_Human,
                   "Choroid_Plexus" = Choroid_Plexus.Umbrella$GeneSymbol_Human,
                   "Endothelial" = Endothelial.Umbrella$GeneSymbol_Human,
                   "Microglia" = Microglia.Umbrella$GeneSymbol_Human,
                   "Mural" = Mural.Umbrella$GeneSymbol_Human,
                   "Neuron" = Neuron.Umbrella$GeneSymbol_Human
)
data <- fromList(list_input)
upset(
  data,
  intersect = c(
    "Astrocyte",
    "Oligodendrocyte",
    "Choroid_Plexus",
    "Endothelial",
    "Microglia",
    "Mural",
    "Neuron"
  )
)
```

```
## `geom_line()`: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
```



```
path <- paste0("../../results/star/cell_type_enrichment/BrainInABlender/upset_genes_all_umbrella_cell_ty
saveToPDF(paste0(path, ".pdf"), width = 12, height = 6)
```

```
## pdf
##   2
```

```
# Binary table with colnames:
write.table(
  data,
  paste0("../../results/",
  tool,
  "/cell_type_enrichment/BrainInABlender/genes_shared_bt_umbrella_cell_types.txt"),
  sep = "\t",
  quote = FALSE
)
remove(list_input)
```