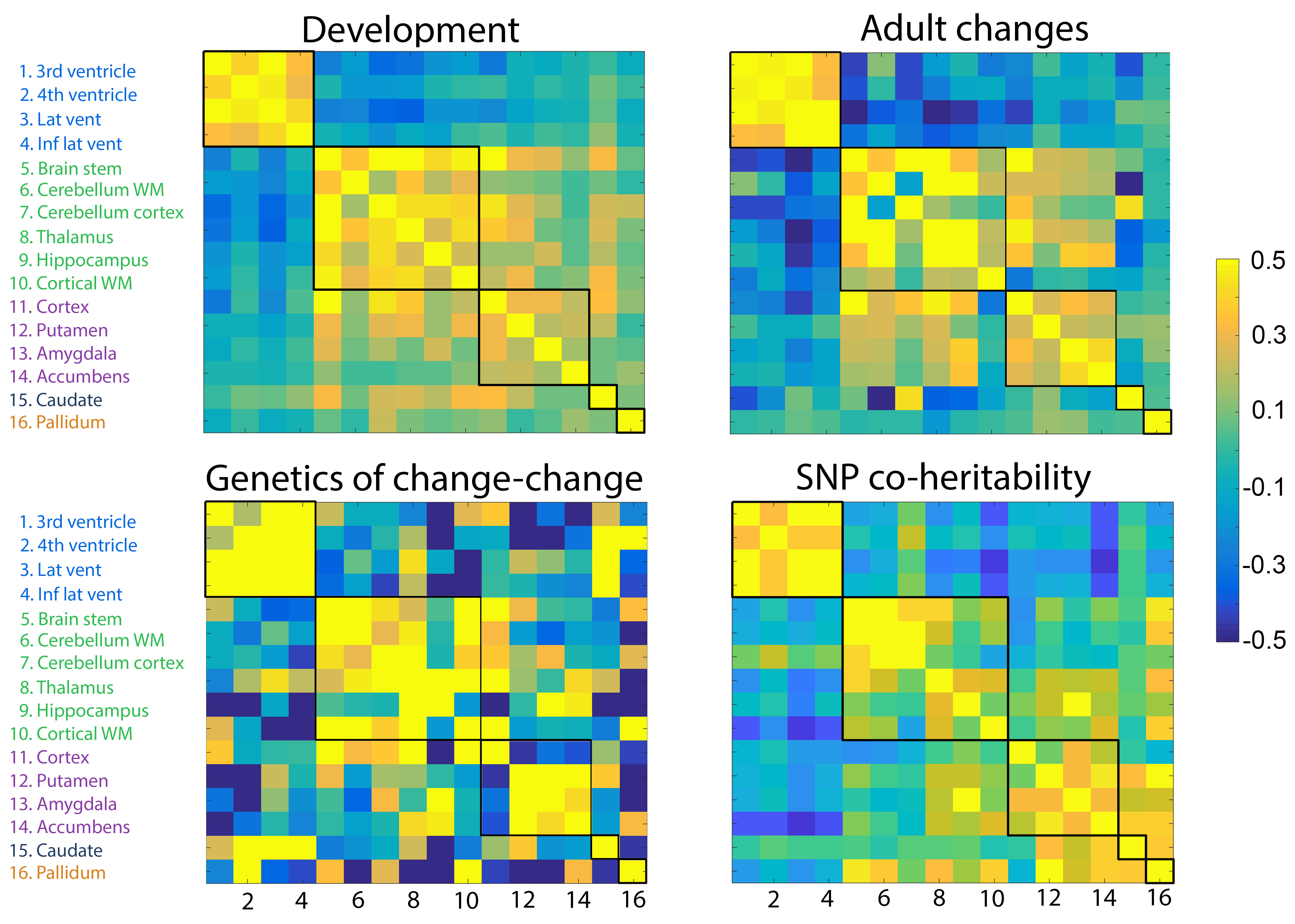
**Intra- vs. extra-cluster correlation tests**

Cluster Tests

library(readxl)  
library(tidyverse)

We define the clusters in a list.

clusters <- list(  
 Cluster1 = c("Bilatv3rdVentricle", "Bilatv4thVentricle", "BilatLatVentricle",   
 "BilatInfLatVentricle"),  
 Cluster2 = c("BilatBrainStem", "BilatCerebellumWM", "BilatCerebellumCortex",  
 "BilatThalamus", "BilatHippocampus", "BilatCerebralWM"),  
 Cluster3 = c("BilatCerebralCortex", "BilatPutamen", "BilatAmygdala",  
 "BilatAccumbens"),  
 Cluster4 = "BilatCaudate",  
 Cluster5 = "BilatPallidum"  
)



Function for testing correlations.

test\_correlations <- function(dataset){  
 imap(clusters, function(cl, nm){  
 if(length(cl) > 1){  
 correlations <- dataset %>%   
 filter(Structure %in% cl) %>% ## Subset to the right rows  
 pivot\_longer(cols = -Structure, names\_to = "Structure2",   
 values\_to = "correlation") %>% # Reshape data  
 filter(Structure != Structure2) %>% # Remove diagonal  
 mutate( # Define whether it is within or outside cluster  
 within\_cluster = Structure2 %in% cl  
 ) %>%   
 select(within\_cluster, correlation) %>%   
 group\_by(within\_cluster) %>%   
 summarise(correlations = list(correlation))  
   
 within\_correlations <- correlations %>%   
 filter(within\_cluster) %>%   
 pull(correlations) %>%   
 unlist()  
   
 between\_correlations <- correlations %>%   
 filter(!within\_cluster) %>%   
 pull(correlations) %>%   
 unlist()  
   
 cat("Testing in ", nm, "\n")  
   
 test <- t.test(  
 x = within\_correlations,  
 y = between\_correlations,  
 alternative = "greater"  
 )  
   
 print(test)  
 cat("\n\n")  
   
 return(test$p.value)  
 } else {  
 cat(nm, "has only a single member. No test performed.\n\n")  
 return(NA\_real\_)  
 }  
   
 })  
}

## UK Biobank

Read in the data.

ukb <- read\_excel("UKB\_GeneticCorr\_Change\_symmetric.xlsx") %>%   
 rename(Structure = "...1") %>%   
 rename\_at(vars(-Structure), ~ paste0("Bilat", .)) %>%   
 mutate(Structure = paste0("Bilat", Structure))

## New names:  
## \* `` -> ...1

Perform two-group t-tests in each cluster, testing if the correlation is larger in within than between. Information about the test is printed below.

ukb\_tests <- test\_correlations(ukb)

## Testing in Cluster1   
##   
## Welch Two Sample t-test  
##   
## data: within\_correlations and between\_correlations  
## t = 12.148, df = 15.745, p-value = 1.04e-9  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.5237763 Inf  
## sample estimates:  
## mean of x mean of y   
## 0.5092477 -1.1025374   
##   
##   
##   
## Testing in Cluster2   
##   
## Welch Two Sample t-test  
##   
## data: within\_correlations and between\_correlations  
## t = 5.805, df = 60.243, p-value = 1.283e-07  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.1745083 Inf  
## sample estimates:  
## mean of x mean of y   
## 0.29909213 0.05407273   
##   
##   
##   
## Testing in Cluster3   
##   
## Welch Two Sample t-test  
##   
## data: within\_correlations and between\_correlations  
## t = 5.5518, df = 41.305, p-value = 9.169e-07  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.157487 Inf  
## sample estimates:  
## mean of x mean of y   
## 0.28288517 0.05691312   
##   
##   
##   
## Cluster4 has only a single member. No test performed.  
##   
## Cluster5 has only a single member. No test performed.

## VETSA / Genetics of change-change

Read in the data.

# VETSA  
vetsa <- read\_delim("Subcort\_GeneticCorr\_Change\_ed.csv", delim = ";",  
 locale = locale(decimal\_mark = ",")) %>%   
 rename(Structure = X1)

## Warning: Missing column names filled in: 'X1' [1]

## Parsed with column specification:  
## cols(  
## X1 = col\_character(),  
## BilatThalamus = col\_double(),  
## BilatCaudate = col\_double(),  
## BilatPutamen = col\_double(),  
## BilatPallidum = col\_double(),  
## BilatHippocampus = col\_double(),  
## BilatAmygdala = col\_double(),  
## BilatAccumbens = col\_double(),  
## BilatInfLatVentricle = col\_double(),  
## BilatLatVentricle = col\_double(),  
## BilatCerebralWM = col\_double(),  
## BilatCerebralCortex = col\_double(),  
## BilatCerebellumCortex = col\_double(),  
## BilatCerebellumWM = col\_double(),  
## Bilatv3rdVentricle = col\_double(),  
## Bilatv4thVentricle = col\_double(),  
## BilatBrainStem = col\_double()  
## )

Perform two-group t-tests in each cluster, testing if the correlation is larger in within than between. Information about the test is printed below.

vetsa\_tests <- test\_correlations(vetsa)

## Testing in Cluster1   
##   
## Welch Two Sample t-test  
##   
## data: within\_correlations and between\_correlations  
## t = 8.2789, df = 36.711, p-value = 3.21e-10  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.5412866 Inf  
## sample estimates:  
## mean of x mean of y   
## 0.5668262 -0.1130322   
##   
##   
##   
## Testing in Cluster2   
##   
## Welch Two Sample t-test  
##   
## data: within\_correlations and between\_correlations  
## t = 5.5531, df = 67.32, p-value = 2.578e-07  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.3130808 Inf  
## sample estimates:  
## mean of x mean of y   
## 0.39946907 -0.04800671   
##   
##   
##   
## Testing in Cluster3   
##   
## Welch Two Sample t-test  
##   
## data: within\_correlations and between\_correlations  
## t = 0.98457, df = 13.426, p-value = 0.1711  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## -0.1400012 Inf  
## sample estimates:  
## mean of x mean of y   
## 0.15888474 -0.01736666   
##   
##   
##   
## Cluster4 has only a single member. No test performed.  
##   
## Cluster5 has only a single member. No test performed.

## Lifebrain

Translation of names for Lifebrain.

translate\_lifebrain <- function(old\_name){  
 lifebrain\_translation <- list(  
 Accumbensarea\_APC = "BilatAccumbens",  
 Amygdala\_APC = "BilatAmygdala",  
 BrainStem\_APC = "BilatBrainStem",  
 Caudate\_APC = "BilatCaudate",  
 CerebellumCortex\_APC = "BilatCerebellumCortex",  
 CerebellumWhiteMatter\_APC = "BilatCerebellumWM",  
 CortexVol\_APC = "BilatCerebralCortex",  
 CerebralWhiteMatterVol\_APC = "BilatCerebralWM",  
 Hippocampus\_APC = "BilatHippocampus",  
 InfLatVent\_APC = "BilatInfLatVentricle",  
 LateralVentricle\_APC = "BilatLatVentricle",  
 Pallidum\_APC = "BilatPallidum",  
 Putamen\_APC = "BilatPutamen",  
 Thalamus\_APC = "BilatThalamus",  
 X3rdVentricle\_APC = "Bilatv3rdVentricle",  
 X4thVentricle\_APC = "Bilatv4thVentricle"  
 )  
   
 if(!old\_name %in% names(lifebrain\_translation)){  
 stop(old\_name, "not found")  
 }   
   
 lifebrain\_translation[[old\_name]]  
}

Read in the data and translate.

lifebrain <- read\_excel("Lifebrain\_change\_change\_n836.xlsx") %>%   
 rename(Structure = "...1") %>%   
 rename\_at(vars(-Structure), ~ map\_chr(., translate\_lifebrain)) %>%   
 mutate(Structure = map\_chr(Structure, translate\_lifebrain))

## New names:  
## \* `` -> ...1

Perform two-group t-tests in each cluster, testing if the correlation is larger in within than between. Information about the test is printed below.

lifebrain\_tests <- test\_correlations(lifebrain)

## Testing in Cluster1   
##   
## Welch Two Sample t-test  
##   
## data: within\_correlations and between\_correlations  
## t = 11.283, df = 15.321, p-value = 3.985e-09  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.5347269 Inf  
## sample estimates:  
## mean of x mean of y   
## 0.4726826 -0.1602484   
##   
##   
##   
## Testing in Cluster2   
##   
## Welch Two Sample t-test  
##   
## data: within\_correlations and between\_correlations  
## t = 3.7976, df = 50.148, p-value = 0.0001978  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.1045314 Inf  
## sample estimates:  
## mean of x mean of y   
## 0.1363119 -0.0507810   
##   
##   
##   
## Testing in Cluster3   
##   
## Welch Two Sample t-test  
##   
## data: within\_correlations and between\_correlations  
## t = 2.9104, df = 23.332, p-value = 0.003904  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.06001752 Inf  
## sample estimates:  
## mean of x mean of y   
## 0.139147746 -0.006711137   
##   
##   
##   
## Cluster4 has only a single member. No test performed.  
##   
## Cluster5 has only a single member. No test performed.

## LCBC

This correlation matrix is sorted after the order of regions appearing in the figure above. The order is as follows.

(lcbc\_names <- unname(unlist(clusters)))

## [1] "Bilatv3rdVentricle" "Bilatv4thVentricle" "BilatLatVentricle"   
## [4] "BilatInfLatVentricle" "BilatBrainStem" "BilatCerebellumWM"   
## [7] "BilatCerebellumCortex" "BilatThalamus" "BilatHippocampus"   
## [10] "BilatCerebralWM" "BilatCerebralCortex" "BilatPutamen"   
## [13] "BilatAmygdala" "BilatAccumbens" "BilatCaudate"   
## [16] "BilatPallidum"

Read in the data.

lcbc <- read\_delim("corrmat\_adults\_sorted.txt", delim = ",",  
 col\_names = lcbc\_names) %>%   
 mutate(Structure = lcbc\_names)

## Parsed with column specification:  
## cols(  
## Bilatv3rdVentricle = col\_double(),  
## Bilatv4thVentricle = col\_double(),  
## BilatLatVentricle = col\_double(),  
## BilatInfLatVentricle = col\_double(),  
## BilatBrainStem = col\_double(),  
## BilatCerebellumWM = col\_double(),  
## BilatCerebellumCortex = col\_double(),  
## BilatThalamus = col\_double(),  
## BilatHippocampus = col\_double(),  
## BilatCerebralWM = col\_double(),  
## BilatCerebralCortex = col\_double(),  
## BilatPutamen = col\_double(),  
## BilatAmygdala = col\_double(),  
## BilatAccumbens = col\_double(),  
## BilatCaudate = col\_double(),  
## BilatPallidum = col\_double()  
## )

Perform two-group t-tests in each cluster, testing if the correlation is larger in within than between. Information about the test is printed below.

lcbc\_tests <- test\_correlations(lcbc)

## Testing in Cluster1   
##   
## Welch Two Sample t-test  
##   
## data: within\_correlations and between\_correlations  
## t = 17.141, df = 33.991, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.5812216 Inf  
## sample estimates:  
## mean of x mean of y   
## 0.4456900 -0.1991452   
##   
##   
##   
## Testing in Cluster2   
##   
## Welch Two Sample t-test  
##   
## data: within\_correlations and between\_correlations  
## t = 6.2174, df = 71.134, p-value = 1.534e-08  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.2575573 Inf  
## sample estimates:  
## mean of x mean of y   
## 0.29512893 -0.05674885   
##   
##   
##   
## Testing in Cluster3   
##   
## Welch Two Sample t-test  
##   
## data: within\_correlations and between\_correlations  
## t = 6.1225, df = 50.806, p-value = 6.677e-08  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## 0.166707 Inf  
## sample estimates:  
## mean of x mean of y   
## 0.28430333 0.05479108   
##   
##   
##   
## Cluster4 has only a single member. No test performed.  
##   
## Cluster5 has only a single member. No test performed.

## Summary

The table below summarizes the p-values for each sample.

bind\_rows(  
 UKB = ukb\_tests,  
 VETSA = vetsa\_tests,  
 Lifebrain = lifebrain\_tests,  
 LCBC = lcbc\_tests,  
 .id = "Dataset"  
) %>%   
 knitr::kable()

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dataset | Cluster1 | Cluster2 | Cluster3 | Cluster4 | Cluster5 |
| UKB | 0 | 0.0000001 | 0.0000009 | NA | NA |
| VETSA | 0 | 0.0000003 | 0.1711224 | NA | NA |
| Lifebrain | 0 | 0.0001978 | 0.0039035 | NA | NA |
| LCBC | 0 | 0.0000000 | 0.0000001 | NA | NA |