In [2]:

```r
library('Seurat')
library('data.table')
library('Matrix')
library('tidyverse')
library('ggplot2')
library('ggpubr')
library('biomaRt')
library('cluster')
library('pheatmap')
library(org.Hs.eg.db)
library(stringr)
`%notin%` <- Negate(`%in%`)
```

```
Warning message:
“Strategy 'multiprocess' is deprecated in future (>= 1.20.0) [2020-10-30]. Instead, explicitly specify either 'multisession' (recommended) or 'multicore'. In the current R session, 'mu
ltiprocess' equals 'multicore'.”
```

SPP1+ Mac matrix initialization

In [209…

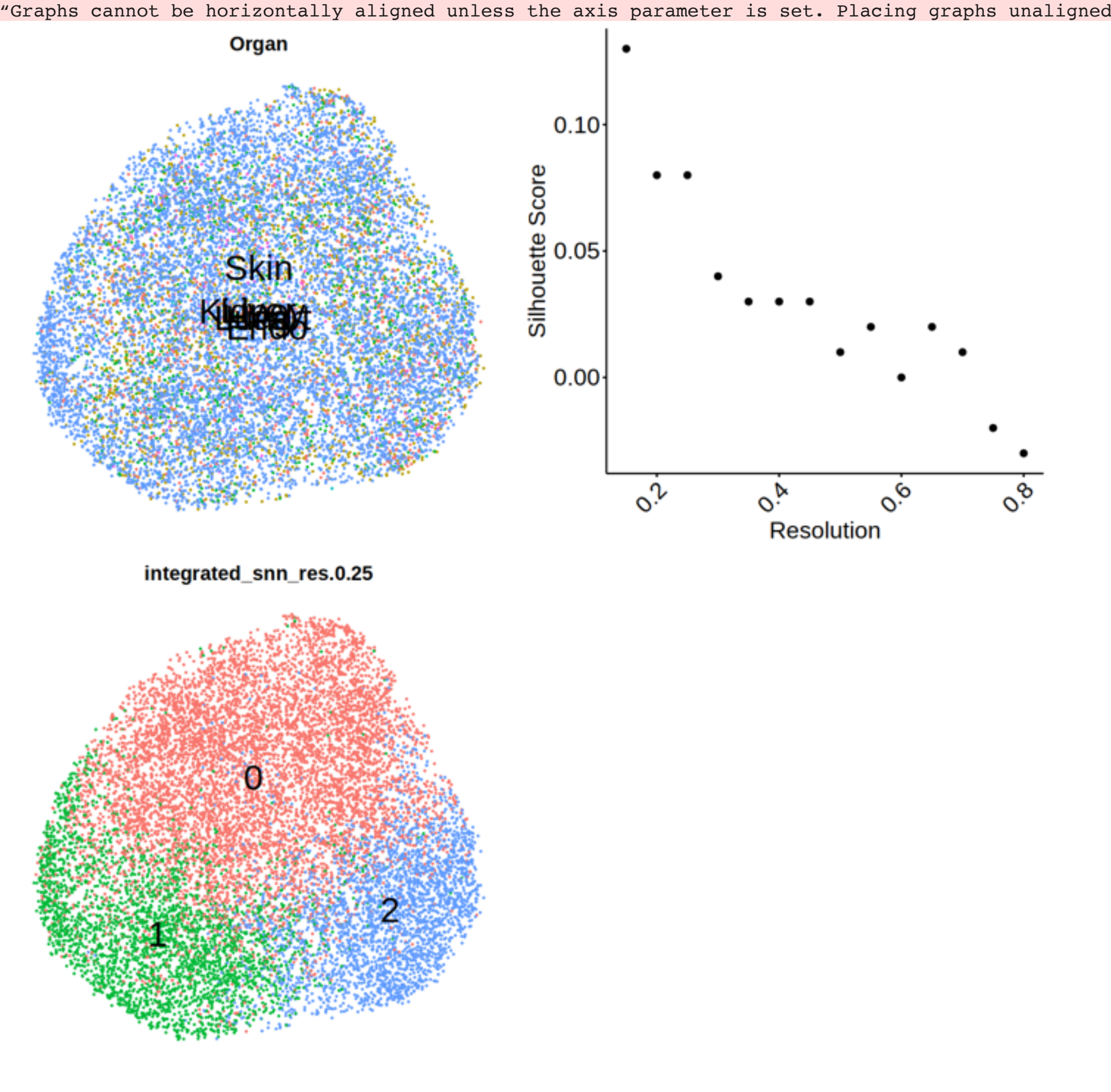
```r
Skin_all <- readRDS('~/Skin_Code_Final.RDS')
Kidney_all <- readRDS( 'Kidney_Code_Final.RDS')
All_endo <- readRDS('~/Endo_Final_Code.RDS')
Lung_all <- readRDS('~/Lung_all_Feb_20.RDS')
All_Heart <- readRDS('~/Heart_Code.RDS')
Liver <- readRDS( 'Liver_code_final.RDS')

transcriptome_SPP1 <- list(All_Heart, Liver, Lung_all, All_endo, Kidney_all, Skin_all)
organ <- c('Heart', 'Liver', 'Lung', 'Endo', 'Kidney', 'Skin')
extract <- c('Diagnosis', 'orig.ident', 'nCount_RNA', 'nFeature_RNA', 'ribo.ratio', 'heam.ratio', 'mito.ratio', 'cluster_d', 'Organ')

for(i in 1:length(transcriptome_SPP1)) {
    transcriptome_SPP1[[i]]$Organ <- organ[[i]]
    transcriptome_SPP1[[i]]@meta.data <- transcriptome_SPP1[[i]]@meta.data[, extract]
    transcriptome_SPP1[[i]] <- RenameCells(object = transcriptome_SPP1[[i]], add.cell.id = paste0(organ[[i]], '_'))
}

common_g <- rownames(transcriptome_SPP1[[1]]$RNA@counts)
for(i in 2:length(transcriptome_SPP1)) {
    common_g <- intersect( common_g , rownames(transcriptome_SPP1[[i]]$RNA@counts) )
}

# SPP1 <- list()
# for(i in 1:length(transcriptome_SPP1)) {
#     SPP1[[i]] <- transcriptome_SPP1[[i]][, transcriptome_SPP1[[i]]$cluster_d == 'SPP1']
#     DefaultAssay(SPP1[[i]]) <- 'RNA'
#     SPP1[[i]] <- SPP1[[i]][ common_g, ]

# }
```

Integrative SPP1+ macophage atlas

In [109…

```r
# set.seed(10)
# SPP1[[7]] <- subset( SPP1[[3]], downsample = dim(SPP1[[3]])[2]* 2/ 3)
# SPP1[[3]] <- SPP1[[3]][, colnames(SPP1[[3]]) %notin% colnames(SPP1[[7]]) ]

# for(i in 1:length(SPP1)) {
#     SPP1[[i]] <- SCTransform( SPP1[[i]], vars.to.regress = c('orig.ident', 'mito.ratio'))
# }

# features <- SelectIntegrationFeatures(object.list = SPP1, nfeatures = 3000)
# SPP1 <- PrepSCTIntegration(object.list = SPP1, anchor.features = features)

# anchors <- FindIntegrationAnchors(object.list = SPP1, normalization.method = "SCT", anchor.features = features, reference = c(1, 2, 3, 4, 5, 6), dims = 1:30)
# SPP1 <- IntegrateData(anchorset = anchors, normalization.method = "SCT")
# SPP1 <- RunPCA(SPP1)

# DefaultAssay( SPP1) <- "integrated"
# SPP1 <- FindNeighbors(SPP1, dims = 1:12)
# SPP1 <- FindClusters(SPP1, resolution = seq(from = 0.15, to = .8, by = .05))

# DefaultAssay(SPP1) <- 'integrated'
# resolution <- colnames( SPP1@meta.data )[ startsWith(x =colnames( SPP1@meta.data ), prefix = 'integrated_snn_res')]

# silhouetteee <- list()
# i <- 1
# dist.matrix <- dist(x = Embeddings(object = SPP1[['pca']])[,1:12])
# for(sil in sort(resolution)) {
#     s <- silhouette(x = as.numeric(x = as.factor(x = SPP1@meta.data[, sil])), dist = dist.matrix)
#     silhouetteee[[i]] <- summary(s)
#     i <- i + 1
# }

# sil_score <- c()
# for( sil in silhouetteee) {
#     sil_score <- c(sil_score, mean(sil$clus.avg.widths))
# }

# to_plot <- data.frame('Score' = sil_score[1: length(sil_score)], 'Res' = seq(from = 0.15, to = .8, by = .05))
# to_plot$Score <- round(to_plot$Score,digits = 2)

p1 <- ggpubr::ggscatter(data = to_plot, x = 'Res', y = 'Score', xlab = 'Resolution', ylab = 'Silhouette Score') +
 rotate_x_text(angle = 45) +
   theme(text = element_text(size = 19), plot.title = element_text(hjust = 0.5))

# SPP1 <- RunUMAP(SPP1, dims = 1:8, verbose = F)

p3 <- DimPlot(SPP1, group.by = 'Organ', label = T, label.size = 10) + NoLegend() + NoAxes()
p2 <- DimPlot(SPP1, group.by = 'integrated_snn_res.0.25', label = T, label.size = 10) + NoLegend() + NoAxes()

options(repr.plot.width=12, repr.plot.height=12)
CombinePlots(list(p3, p1, p2), ncol = 2)
```

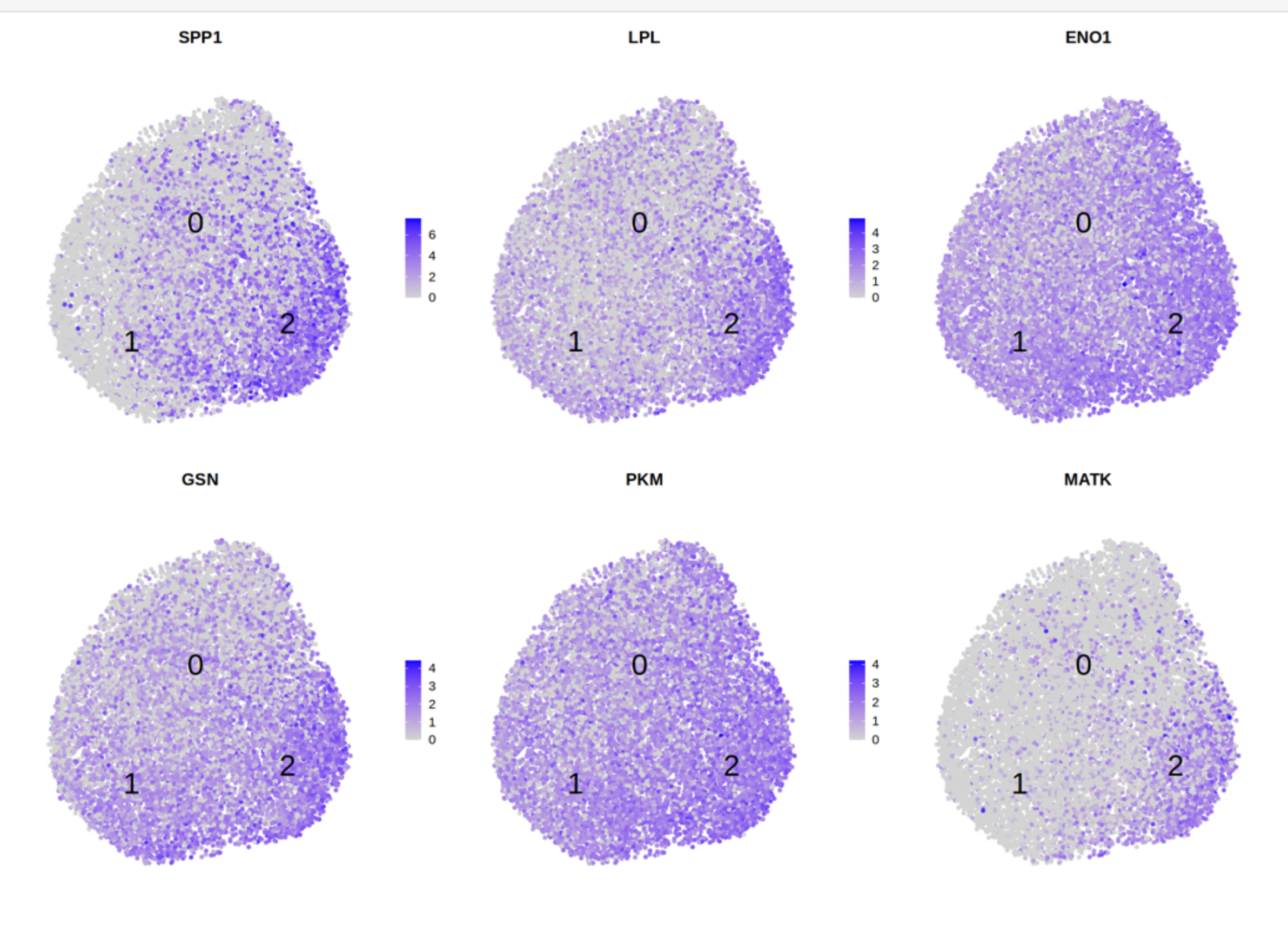
```
Warning message:
“CombinePlots is being deprecated. Plots should now be combined using the patchwork system.”
Warning message:
“Graphs cannot be vertically aligned unless the axis parameter is set. Placing graphs unaligned.”
Warning message:
“Graphs cannot be horizontally aligned unless the axis parameter is set. Placing graphs unaligned.”
```



Signature of matrisome-associated macrophage (MAM)

In [116…

```r
# DefaultAssay(SPP1) <- 'RNA'
# Idents(SPP1) <- SPP1$integrated_snn_res.0.25

# i <- 1
# MAM.marker <- list()
# for( org in organ) {
#     SPP1_tmp <- SPP1[, SPP1$Organ == org]
#     DefaultAssay( SPP1_tmp) <- 'RNA'
#     SPP1_tmp <- NormalizeData(SPP1_tmp)
#     MAM.marker[[i]] <- FindMarkers( SPP1_tmp, ident.1 = 2, logfc.threshold = -Inf, min.pct = -Inf, min.diff.pct = -Inf)
#     i <- i + 1
# }

# MAM_FC_avg <-  MAM.marker[[1]]$avg_log2FC
# MAM_diff <-  MAM.marker[[1]]$pct.1 -  MAM.marker[[1]]$pct.2
# gene <- rownames( MAM.marker[[1]])

# for(i in 2:5) {
#     MAM.marker[[i]] <- MAM.marker[[i]][ gene, ]
#     MAM_FC_avg  <- cbind(MAM_FC_avg, MAM.marker[[i]][, 'avg_log2FC'])
#     MAM_diff  <- cbind(MAM_diff, MAM.marker[[i]]$pct.1 -  MAM.marker[[i]]$pct.2)
# }
# rownames(MAM_FC_avg) <- gene

# MAM_signature <- data.frame( 'log2_FC_avg' = rowMeans( MAM_FC_avg), 'Diff_prop' = rowMeans( MAM_diff))
# MAM_signature$Diff_prop_z <- ( MAM_signature$Diff_prop  - mean(MAM_signature$Diff_prop )) /
#         sd(MAM_signature$Diff_prop )

MAM_signature_FINAL <- MAM_signature[ MAM_signature$Diff_prop_z > 1.645 & MAM_signature$log2_FC_avg > .25, ]
#MAM_signature_FINAL <- MAM_signature_FINAL[ order(MAM_signature_FINAL$, decreasing = T),]

options(repr.plot.width=18, repr.plot.height=12)
DefaultAssay(SPP1) <- 'RNA'
FeaturePlot( SPP1, features = rownames(MAM_signature_FINAL[1:6,]), ncol = 3,  label = T, label.size = 10, pt.size = 1, combine = TRUE) & NoAxes()
```



Top 10 MAM markers by fold change (log2)

In [117…

```r
MAM_signature_FINAL <- MAM_signature_FINAL[ order(MAM_signature_FINAL$log2_FC_avg, decreasing = T),]
MAM_signature_FINAL[1:10, ]
```

A data.frame: 10 × 3

	log2_FC_avg	Diff_prop	Diff_prop_z
	<dbl>	<dbl>	<dbl>
SPP1	2.2890512	0.3728	7.700796
MMP9	1.4910151	0.2450	4.815045
LPL	1.4528990	0.3460	7.095646
TIMP3	1.2384094	0.1586	2.864115
GSN	1.1404324	0.2544	5.027299
MATK	1.0056239	0.3348	6.842748
CCL2	0.9529313	0.1584	2.859599
IL1RN	0.9207779	0.2474	4.869237
RGCC	0.9087962	0.2132	4.096994
TM4SF19	0.8441229	0.2358	4.607307

Gene-set pathway enrichment analysis of MAM

In [125…

```r
# GO <- rbind( clusterProfiler::read.gmt('/mnt/storage/projects/huang_macrophage/gene_set/Human/Pathways/c2.cp.v7.5.1.symbols.gmt.txt'),
#   clusterProfiler::read.gmt('/mnt/storage/projects/huang_macrophage/gene_set/Human/Pathways/h.all.v7.5.1.symbols.gmt.txt'))
# GO$prefix <- str_split_fixed(GO$term, pattern = '_', n = 2)[,1]
# GO <- GO[GO$prefix %notin% c( 'WP'), ]
# GO$prefix <- NULL
# Term <- c()
# term <- str_split_fixed(GO$term, pattern = '_', n = 20)
# for( i  in  1:dim(term )[1]) {
#     Term <- c(Term, str_trim(paste0(term[i, 2:20], collapse = ' ')))
# }
# GO$term <- Term


# MAM_signature <- MAM_signature[ order( MAM_signature$log2_FC_avg, decreasing = T), ]
# FC_avg <- MAM_signature$log2_FC_avg
# names(FC_avg) <- rownames(MAM_signature)
# FC_avg <- sort(FC_avg[!startsWith(names(FC_avg), prefix =  'PSM')], decreasing = T)


# set.seed(15)
# mac_1 <- clusterProfiler::GSEA(geneList= FC_avg, TERM2GENE = GO, pAdjustMethod = 'BH',
#                minGSSize = 5, maxGSSize = 500, pvalueCutoff = 0.1, verbose = TRUE)
# mac_1@result  <- mac_1@result[ order(mac_1@result$NES, decreasing = T), ]
mac_1@result[, c('NES', 'p.adjust')][1:15, ]
```

```
preparing geneSet collections...

GSEA analysis...

Warning message in fgseaMultilevel(...):
“For some pathways, in reality P-values are less than 1e-10. You can set the `eps` argument to zero for better estimation.”
leading edge analysis...

done...
```

A data.frame: 15 × 2

	NES	p.adjust
	<dbl>	<dbl>
VIBRIO CHOLERAE INFECTION	2.150931	2.515163e-06
DEGRADATION OF THE EXTRACELLULAR MATRIX	2.113289	5.568602e-06
L1CAM INTERACTIONS	2.109446	2.176599e-06
EXTRACELLULAR MATRIX ORGANIZATION	2.055029	3.557786e-07
RECYCLING PATHWAY OF L1	2.049970	1.307001e-04
GOLGI TO ER RETROGRADE TRANSPORT	2.049897	1.655479e-06
COPI DEPENDENT GOLGI TO ER RETROGRADE TRAFFIC	2.048792	1.925594e-05
EPH EPHRIN SIGNALING	2.038894	7.173166e-06
COPI MEDIATED ANTEROGRADE TRANSPORT	2.011594	2.006020e-05
TRANSLOCATION OF SLC2A4 GLUT4 TO THE PLASMA MEMBRANE	2.001046	8.938920e-05
LYSOSOME	1.991261	3.052530e-06
COLLAGEN DEGRADATION	1.961644	1.139816e-04
CELL CELL COMMUNICATION	1.953614	2.450618e-04
GLUCONEOGENESIS	1.938557	2.014740e-03
APICAL JUNCTION	1.933335	1.891224e-05

SPP1+ MAM+ enriched matrisome processes (NABA lab, UIC)

In [143…

```r
# NABA <- rbind( clusterProfiler::read.gmt('/mnt/storage/projects/huang_macrophage/gene_set/Human/Pathways/c2.cp.v7.5.1.symbols.gmt.txt'),
#   clusterProfiler::read.gmt('/mnt/storage/projects/huang_macrophage/gene_set/Human/Pathways/h.all.v7.5.1.symbols.gmt.txt'))
# NABA <- NABA[ str_split_fixed( NABA$term , pattern = '_', n = 2)[,1] == 'NABA',]
# NABA$term <- as.character(NABA$term)
# NABA <- NABA[ NABA$term %notin% c('NABA_MATRISOME', 'NABA_MATRISOME_ASSOCIATED', 'NABA_CORE_MATRISOME'), ]


mam_naba <- rownames( MAM_signature )[ MAM_signature$log2_FC_avg > .25 & MAM_signature$Diff_prop_z > 1.645]
mam_naba <- clusterProfiler::enricher(gene = mam_naba, TERM2GENE = NABA, pAdjustMethod = 'BH',
               pvalueCutoff = 1)

mam_n_naba <- rownames( MAM_signature )[ MAM_signature$log2_FC_avg < -.25 & MAM_signature$Diff_prop_z < -1.645]
mam_n_naba <- clusterProfiler::enricher(gene = mam_n_naba, TERM2GENE = NABA, pAdjustMethod = 'BH',
               pvalueCutoff = 1)

mam_naba@result[, c('qvalue','pvalue', 'geneID')]
```

A data.frame: 4 × 3

	qvalue	pvalue	geneID
	<dbl>	<dbl>	<chr>
NABA_ECM_REGULATORS	0.05358308	0.01696798	MMP9/TIMP3/FAM20C/ADAM9/TGM2/MMP19/CD109/A2M/PLOD3
NABA_ECM_AFFILIATED	0.12413950	0.07862168	SDC2/ANXA11/GPC4/ANXA4/LGALS9/CLEC5A
NABA_ECM_GLYCOPROTEINS	0.78603772	0.90439625	SPP1/CRELD2
NABA_SECRETED_FACTORS	0.78603772	0.99564778	CCL2/IL1RN

SPP1+ MAM- enriched matrisome processes (NABA lab, UIC)

In [142…

```r
mam_n_naba@result[, c('qvalue','pvalue', 'geneID')]
```

A data.frame: 4 × 3

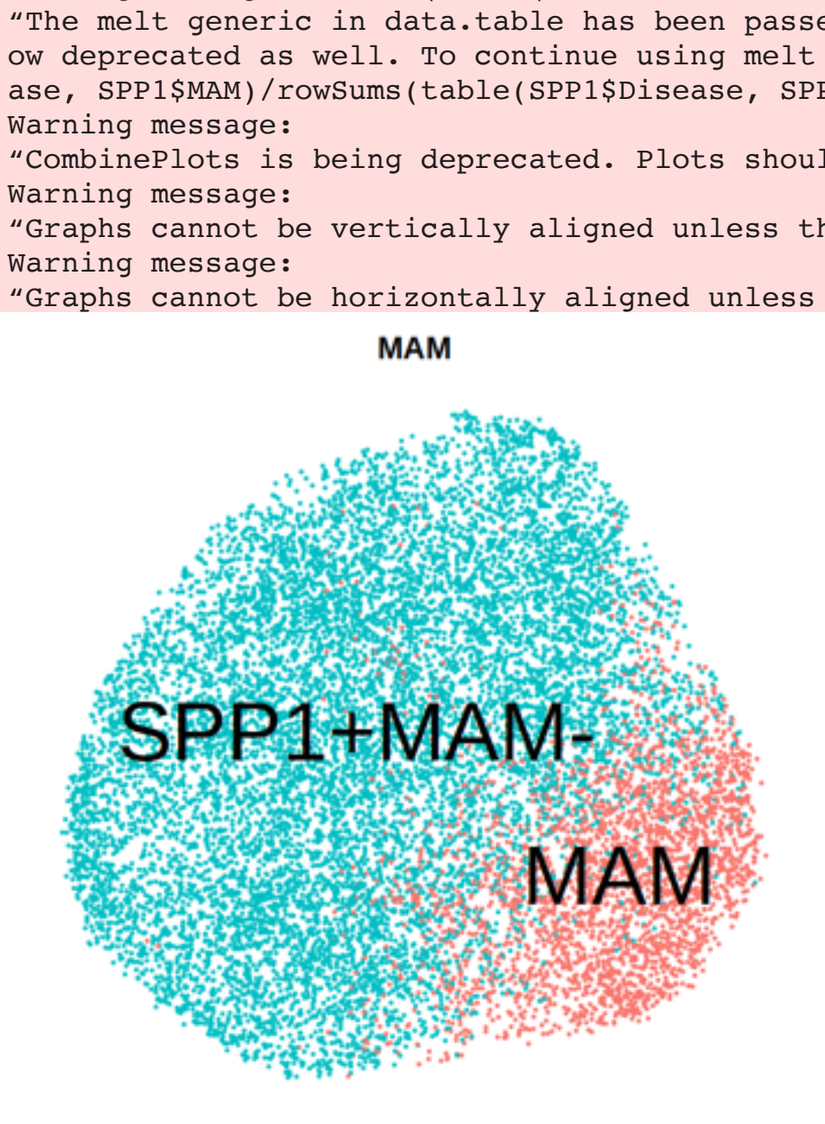
	qvalue	pvalue	geneID
	<dbl>	<dbl>	<chr>
NABA_ECM_AFFILIATED	0.04452238	0.01409876	CLEC12A/CLEC4E/CLEC7A/C1QC/CLEC10A/FCN1/C1QB/C1QA
NABA_SECRETED_FACTORS	0.73619133	0.58638121	EREG/CCL20/AREG/CXCL2/CXCL3/IL1B/CXCL8
NABA_ECM_REGULATORS	0.73619133	0.75538119	SERPINB9/SERPING1/SERPINA1/CSTA
NABA_ECM_GLYCOPROTEINS	0.73619133	0.93250902	FGL2/THBS1

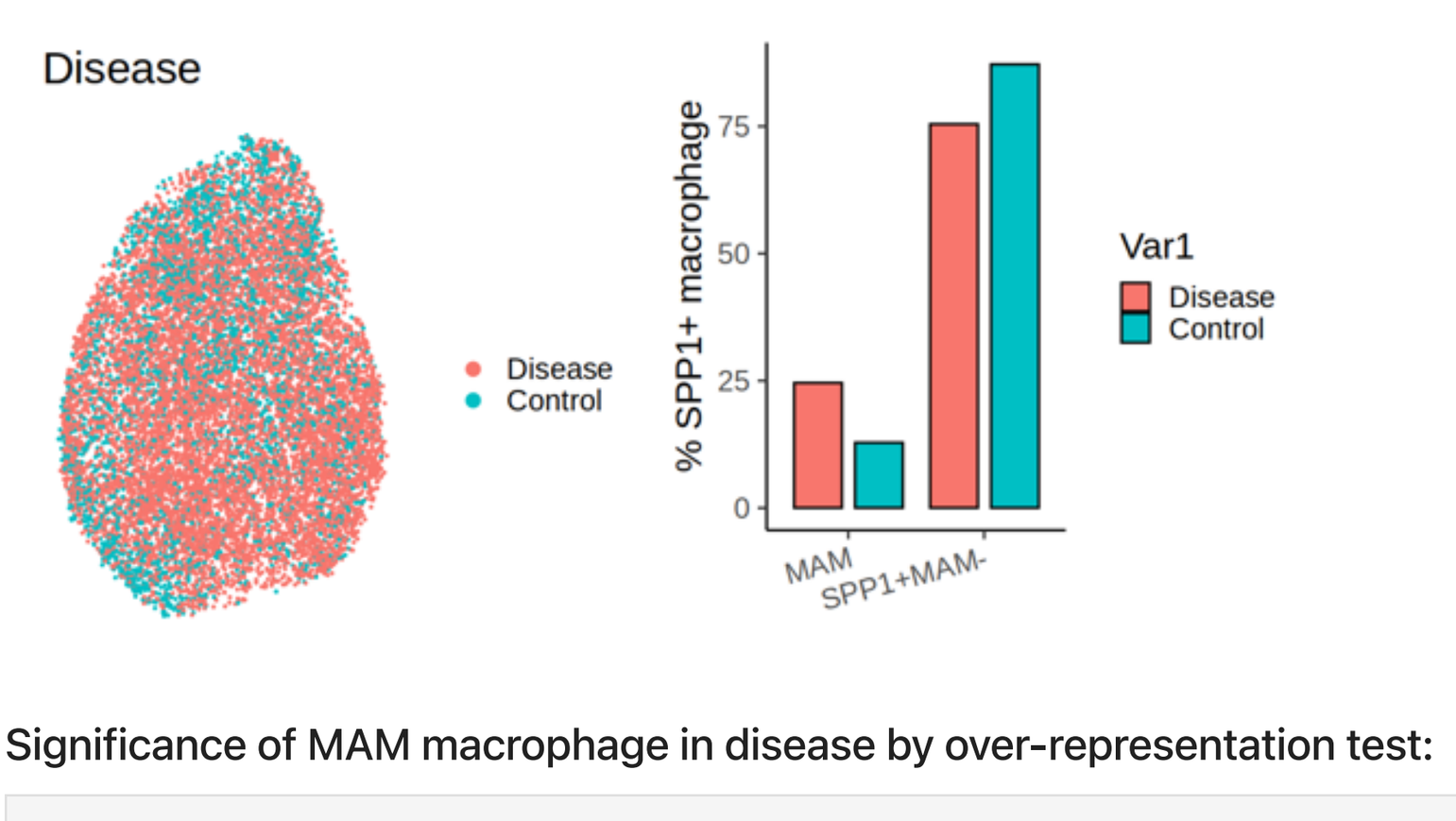
MAM in disease v. control

In [186…

```r
SPP1$Disease <- SPP1$Diagnosis
SPP1$Disease[ SPP1$Disease!= 'Control'] <- 'Disease'
SPP1$Disease <- factor( SPP1$Disease, levels = c('Disease', 'Control'))

SPP1$MAM <- 'MAM'
SPP1$MAM[SPP1$integrated_snn_res.0.25 != 2 ] <- 'SPP1+MAM-'


p1 <- DimPlot(SPP1, group.by = 'Disease') + theme_classic( base_line_size = .5, base_size = 20) +  NoAxes()

to_plot <- melt(table(SPP1$Disease, SPP1$MAM) / rowSums(table(SPP1$Disease, SPP1$MAM)) * 100)
p2 <- ggbarplot(data = to_plot, x= 'Var2', y = 'value', fill = 'Var1', position = position_dodge(0.9)) +
    xlab('') + ylab('% SPP1+ macrophage') + theme_classic( base_line_size = .5, base_size = 20) +
    rotate_x_text(15)

options(repr.plot.width=6, repr.plot.height=6)
DimPlot(SPP1, group.by = 'MAM', label = T, label.size = 15) + NoLegend() + NoAxes()

options(repr.plot.width=10, repr.plot.height=5)
CombinePlots(list( p1, p2), ncol = 2)
```

```
Warning message in melt(table(SPP1$Disease, SPP1$MAM)/rowSums(table(SPP1$Disease, :
“The melt generic in data.table has been passed a table and will attempt to redirect to the relevant reshape2 method; please note that reshape2 is deprecated, and this redirection is n
ow deprecated as well. To continue using melt methods from reshape2 while both libraries are attached, e.g. melt.list, you can prepend the namespace like reshape2::melt(table(SPP1$Dise
ase, SPP1$MAM)/rowSums(table(SPP1$Disease, SPP1$MAM)) *     100). In the next version, this warning will become an error.”
Warning message:
“CombinePlots is being deprecated. Plots should now be combined using the patchwork system.”
Warning message:
“Graphs cannot be vertically aligned unless the axis parameter is set. Placing graphs unaligned.”
Warning message:
“Graphs cannot be horizontally aligned unless the axis parameter is set. Placing graphs unaligned.”
```





Significance of MAM macrophage in disease by over-representation test:

In [191…

```r
over_t <- phyper( sum( SPP1$Disease == 'Disease' & SPP1$MAM == 'MAM') -1, sum(SPP1$Disease == 'Disease'),
       length(SPP1$MAM ) -sum(SPP1$Disease == 'Disease'), sum(SPP1$MAM == 'MAM'),lower.tail= FALSE)

sprintf('p = %s', over_t)
```

'p = 1.06491553964182e-49'

by Organ

In [ ]:

```r
for( i in 1: length(transcriptome_SPP1) ) {

    MAM <- (table( transcriptome_SPP1[[i]]$orig.ident, transcriptome_SPP1[[i]]$cluster_d) / rowSums(table(transcriptome_SPP1[[i]]$orig.ident, transcriptome_SPP1[[i]]$cluster_d)))[table
    Diagnosis <-  plyr::mapvalues( names(table( transcriptome_SPP1[[i]]$orig.ident ))[ table(transcriptome_SPP1[[i]]$orig.ident) > N[[i]] ],
        from = transcriptome_SPP1[[i]]$orig.ident[ !duplicated( transcriptome_SPP1[[i]]$orig.ident )], to = transcriptome_SPP1[[i]]$Diagnosis[ !duplicated(transcriptome_SPP1[[i]]$orig.
    tmp <- data.frame('MAM' = MAM, 'Diagnosis' = Diagnosis)
    tmp <- tmp[ tmp$Diagnosis %in% D[[i]], ]
    tmp$Diagnosis <- factor( tmp$Diagnosis , levels = D[[i]])


}
```

In [327…

```r
# for(i in 1:length(transcriptome_SPP1)) {
#     transcriptome_SPP1[[i]]$cluster_d <- as.character( transcriptome_SPP1[[i]]$cluster_d)
#     transcriptome_SPP1[[i]]$cluster_d [ colnames(transcriptome_SPP1[[i]]) %in% colnames(SPP1)[ SPP1$MAM == 'MAM']] <- 'MAM'

# }

# to_plot <- list()
# N <- c(0, 0, 0, 10, 25, 10)
# D <- list( c('Control', 'DCM', 'ICM'), c('Control', 'Cirrhosis'), c('Control', 'IPF', 'SSc'),
#          c('Control', 'Endometriosis'), c('Control', 'AKI', 'CKD'), c('Control', 'SSC', 'Keloid'))


D <- list( c('Control', 'DCM', 'ICM'), c('Control', 'Cirrhosis'), c('Control', 'IPF', 'SSc'),
         c('Control', 'Endometriosis'), c('Control', 'AKI', 'CKD'), c('Control', 'SSC', 'Keloid'))

for( i in 1: length(transcriptome_SPP1) ) {

    MAM <- (table( transcriptome_SPP1[[i]]$orig.ident, transcriptome_SPP1[[i]]$cluster_d) / rowSums(table(transcriptome_SPP1[[i]]$orig.ident, transcriptome_SPP1[[i]]$cluster_d)))[table
    Diagnosis <-  plyr::mapvalues( names(table( transcriptome_SPP1[[i]]$orig.ident ))[ table(transcriptome_SPP1[[i]]$orig.ident) > N[[i]] ],
        from = transcriptome_SPP1[[i]]$orig.ident[ !duplicated( transcriptome_SPP1[[i]]$orig.ident )], to = transcriptome_SPP1[[i]]$Diagnosis[ !duplicated(transcriptome_SPP1[[i]]$orig.
    tmp <- data.frame('MAM' = MAM, 'Diagnosis' = Diagnosis)
    tmp <- tmp[ tmp$Diagnosis %in% D[[i]], ]
    tmp$Diagnosis[ tmp$Diagnosis == 'Control'] <- paste0(tmp$Diagnosis[ tmp$Diagnosis == 'Control'], '_', organ[[i]])

    if(i == 1) {
        all_pat <- tmp
    } else {
        all_pat <- rbind(all_pat, tmp)
    }
}

D <- c('Control_Heart', 'DCM', 'ICM','Control_Liver', 'Cirrhosis', 'Control_Lung', 'IPF', 'SSc',
 'Control_Endo', 'Endometriosis', 'Control_Kidney', 'AKI', 'CKD','Control_Skin', 'SSC', 'Keloid')

Com <- list( c('DCM', 'Control_Heart'), c('ICM', 'Control_Heart'),
           c('Cirrhosis', 'Control_Liver'), c('IPF', 'Control_Lung'), c('SSc', 'Control_Lung'),
           c('Endometriosis', 'Control_Endo'), c('AKI', 'Control_Kidney'), c('CKD', 'Control_Kidney'),
           c('SSC', 'Control_Skin'), c('Keloid', 'Control_Skin'))

all_pat$Diagnosis <- factor( all_pat$Diagnosis, levels = D)
all_pat$MAM <- 100 * all_pat$MAM

options(repr.plot.width=20, repr.plot.height=8)
ggboxplot( data = all_pat, x = 'Diagnosis', y = 'MAM',  add = "jitter" ) +
    theme_classic(base_line_size = .5, base_size = 25) + ylab('% SPP1+MAM+') + rotate_x_text(30) + xlab('')  +
    stat_compare_means( comparisons = Com, method = 'wilcox.test', size = 6, bracket.size = 0.1, tip.length = 0, step.increase = 0.15)
```

```
The following `from` values were not present in `x`: 1

The following `from` values were not present in `x`: 30-10034, 34-10184, 34-10187, 27-10039, 28-10051, 31-10001, 29-10008, 31-10000, PRE018-1, PRE019-4, PRE027, PRE038, PRE055-1, PRE06
2-1, PRE19-025, PRE19-05, PRE98sc, Sample1153-EO1, Sample1153-EO2, Sample1157-EO1, Sample1157-EO2, Sample1157-EO3, Sample1158-EO1, Sample1158-EO2, Sample1158-EO3, Sample1162-EO1, Sampl
e1162-EO2, CDm7, CDm9, CDm11, CDp1, CDp2

The following `from` values were not present in `x`: pt03032, GVHD_pt005, pt01004, pt01007, Ctrl002, Ctrl004, Ctrl005, Ctrl007, pt02003, pt03001, pt01011, pt01010, pt02004, Ctrl008, pt
02006, Ctrl009, Ctrl010, pt03005, pt01019, pt03016, pt03017, pt01030, Ctrl023, pt01014, pt01046

Warning message in wilcox.test.default(c(0, 0.854700854700855, 0.72463768115942, :
“cannot compute exact p-value with ties”
Warning message in wilcox.test.default(c(0.970873786407767, 1.01010101010101, 1.86170212765957, :
“cannot compute exact p-value with ties”
Warning message in wilcox.test.default(c(1.81818181818182, 2.5, 3.79310344827586, :
“cannot compute exact p-value with ties”
Warning message in wilcox.test.default(c(3.7037037037037, 1.42857142857143, 0.632911392405063, :
“cannot compute exact p-value with ties”
Warning message in wilcox.test.default(c(2.38095238095238, 2.56410256410256, 0), :
“cannot compute exact p-value with ties”
```