Mixscape Vignette

Compiled: July 01, 2020

Install mixscape branch from Seurat.

```
remotes::install_github("satijalab/seurat", ref = "mixscape")
```

Load required packages.

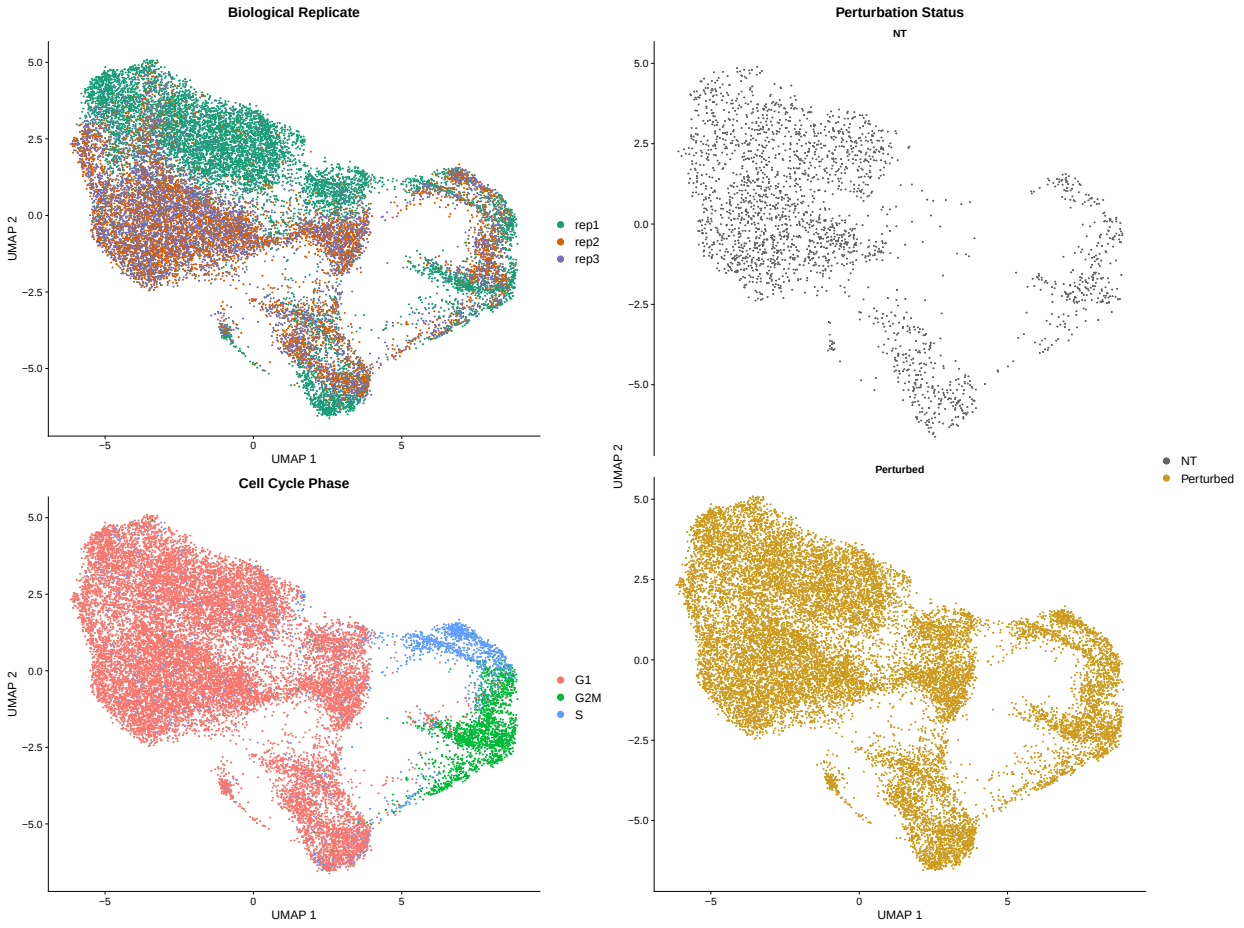
```
# Load packages.
library(Seurat)
library(SeuratData)
library(ggplot2)
library(patchwork)
library(scales)
library(dplyr)

# Download dataset using SeuratData.
InstallData(ds = "thp1.eccite")

# Setup custom theme for plotting.
custom_theme <- theme(
  plot.title = element_text(size=16, hjust = 0.5),
  legend.key.size = unit(0.7, "cm"),
  legend.text = element_text(size = 14))
```

Load Seurat object containing ECCITE-seq dataset.

```
# Load object.
eccite <- LoadData(ds = "thp1.eccite")

# Normalize protein.
eccite <- NormalizeData(
  object = eccite,
  assay = "ADT",
  normalization.method = "CLR",
  margin = 2)
```

RNA-based clustering is driven by confounding sources of variation.

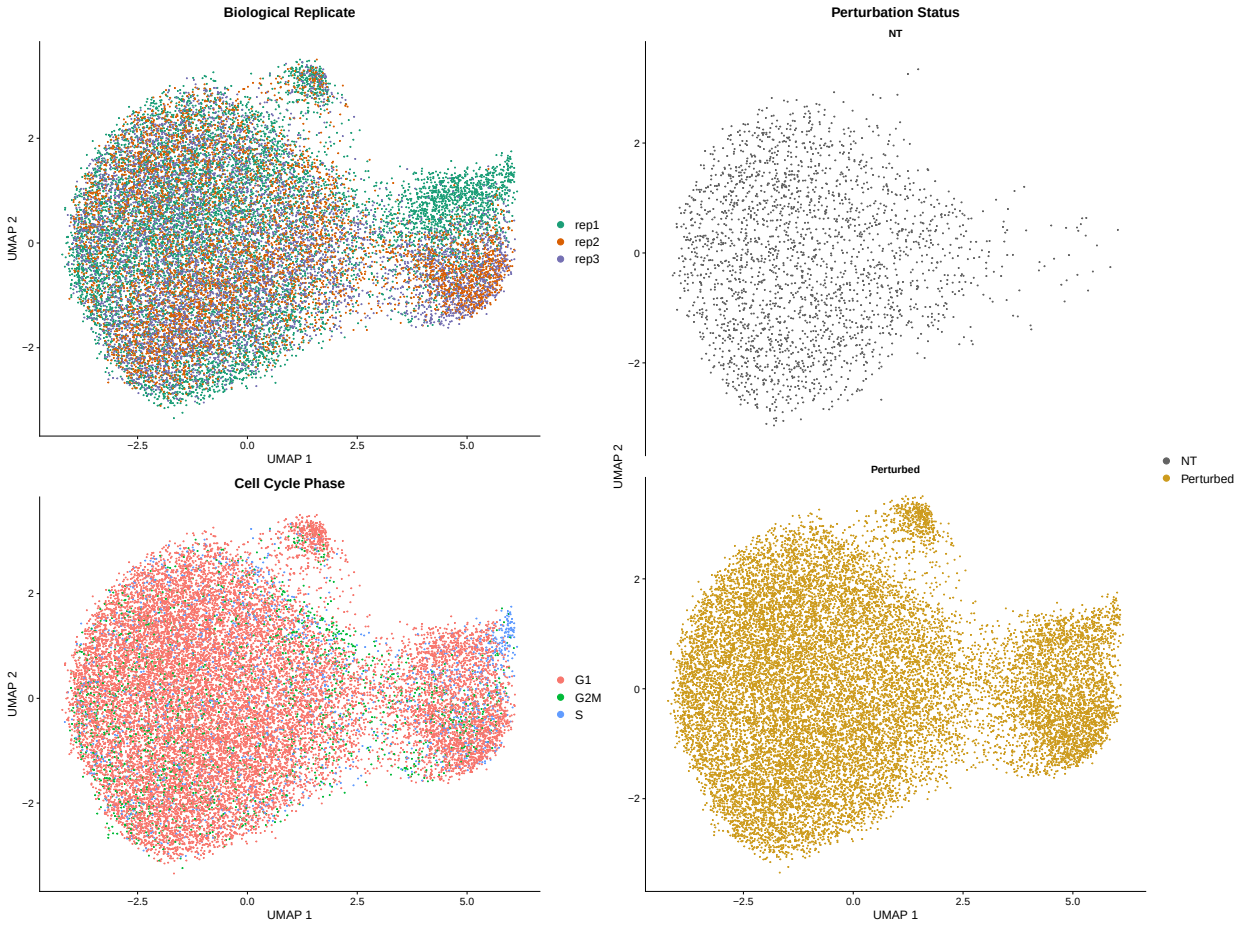
```
# Prepare RNA assay for dimensionality reduction:
# Normalize data, find variable features and scale data.
DefaultAssay(object = eccite) <- 'RNA'
eccite <- NormalizeData(object = eccite) %>% FindVariableFeatures() %>% ScaleData()

# Run Principle Component Analysis (PCA) to reduce the dimensionality of the data.
eccite <- RunPCA(object = eccite)

# Run Uniform Manifold Approximation and Projection (UMAP) to visualize clustering in 2-D.
eccite <- RunUMAP(object = eccite, dims = 1:40)

# Generate plots to check if clustering is driven by biological replicate ID,
# cell cycle phase or target gene class.
p1 <- DimPlot(
  object = eccite,
  group.by = 'replicate',
  label = F,
  pt.size = 0.2,
  reduction = "umap") +
  scale_color_brewer(palette = "Dark2") +
  ggtitle("Biological Replicate") +
  xlab("UMAP 1") +
  ylab("UMAP 2") +
  custom_theme

p2 <- DimPlot(
  object = eccite,
  group.by = 'Phase',
  label = F, pt.size = 0.2,
  reduction = "umap") +
  ggtitle("Cell Cycle Phase") +
  ylab("UMAP 2") +
  xlab("UMAP 1") +
  custom_theme

p3 <- DimPlot(
  object = eccite,
  group.by = 'crispr',
  pt.size = 0.2,
  reduction = "umap",
  split.by = "crispr",
  ncol = 1,
  cols = c("grey39","goldenrod3")) +
  ggtitle("Perturbation Status") +
  ylab("UMAP 2") +
  xlab("UMAP 1") +
  custom_theme

# Visualize plots.
((p1 / p2 + plot_layout(guides = 'auto')) | p3 )
```



Calculating local perturbation signatures mitigates confounding effects.

```
# Calculate local perturbation signature.
eccite<- CalcPerturbSig(
  object = eccite,
  assay = "RNA",
  slot = "data",
  gd.class ="gene",
  nt.cell.class = "NT",
  reduction = "pca",
  ndims = 40,
  num.neighbors = 20,
  split.by = "replicate",
  new.assay.name = "PRTB")

# Prepare PRTB assay for dimensionality reduction:
# Normalize data, find variable features and center data.
DefaultAssay(object = eccite) <- 'PRTB'

# Use variable features from RNA assay.
VariableFeatures(object = eccite) <- VariableFeatures(object = eccite[["RNA"]])
eccite <- ScaleData(object = eccite, do.scale = F, do.center = T)
```

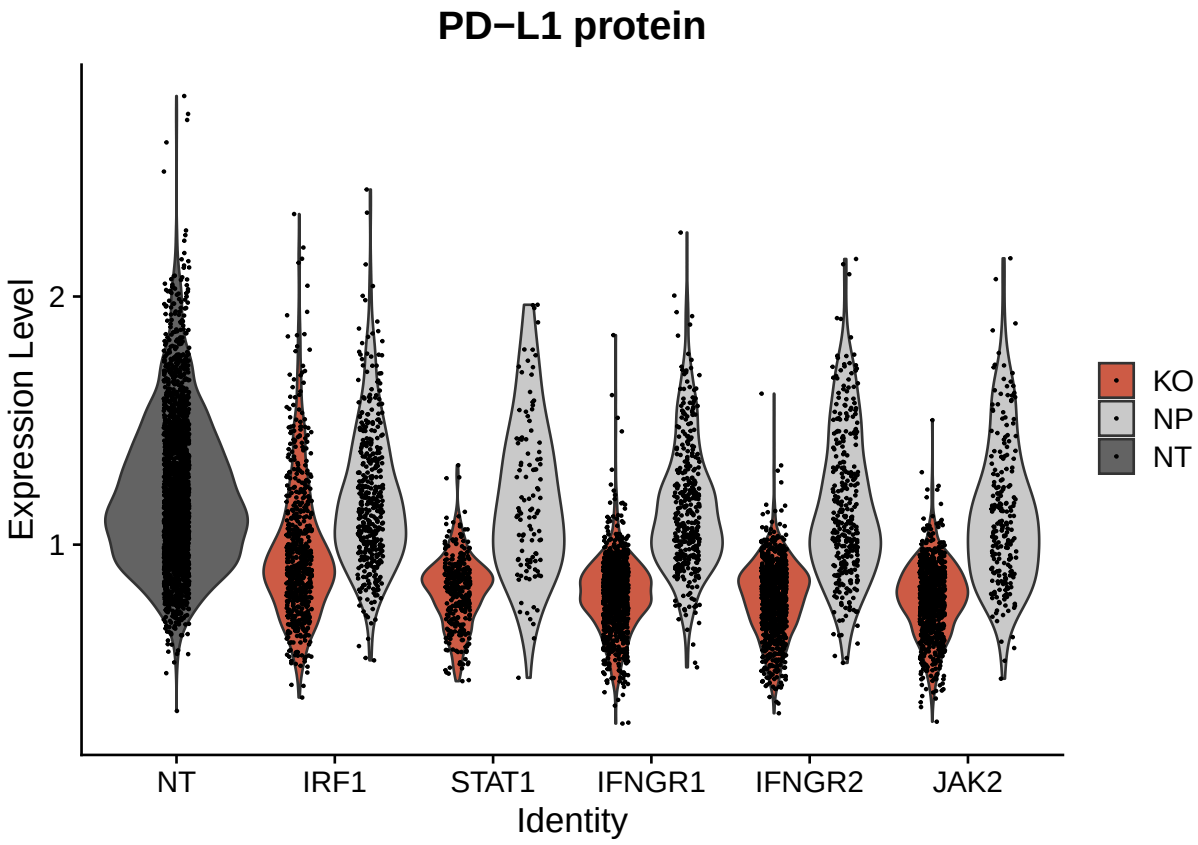
```r
# Run PCA to reduce the dimensionality of the data.
eccite <- RunPCA(object = eccite, reduction.key = 'prtbpca', reduction.name = 'prtbpca')

# Run UMAP to visualize clustering in 2-D.
eccite <- RunUMAP(
  object = eccite,
  dims = 1:40,
  reduction = 'prtbpca',
  reduction.key = 'prtbumap',
  reduction.name = 'prtbumap')

# Generate plots to check if clustering is driven by biological replicate ID,
# cell cycle phase or target gene class.
q1 <- DimPlot(
  object = eccite,
  group.by = 'replicate',
  reduction = 'prtbumap',
  pt.size = 0.2) +
  scale_color_brewer(palette = "Dark2") +
  ggtitle("Biological Replicate") +
  ylab("UMAP 2") +
  xlab("UMAP 1") +
  custom_theme

q2 <- DimPlot(
  object = eccite,
  group.by = 'Phase',
  reduction = 'prtbumap',
  pt.size = 0.2) +
  ggtitle("Cell Cycle Phase") +
  ylab("UMAP 2") +
  xlab("UMAP 1") +
  custom_theme

q3 <- DimPlot(
  object = eccite,
  group.by = 'crispr',
  reduction = 'prtbumap',
  split.by = "crispr",
  ncol = 1,
  pt.size = 0.2,
  cols = c("grey39","goldenrod3")) +
  ggtitle("Perturbation Status") +
  ylab("UMAP 2") +
  xlab("UMAP 1") +
  custom_theme

# Visualize plots.
(q1 / q2 + plot_layout(guides = 'auto') | q3)
```



Mixscape identifies cells with no detectable perturbation.

```
# Run mixscape to classify cells based on their perturbation status.
eccite <- RunMixscape(
  object = eccite,
  assay = "PRTB",
  slot = "scale.data",
  labels = "gene",
  nt.class.name = "NT",
  min.de.genes = 5,
  iter.num = 10,
  de.assay = "RNA",
  verbose = F)

# Show that only IFNG pathway KO cells have a reduction in PD-L1 protein expression.
Idents(object = eccite) <- "gene"

VlnPlot(
  object = eccite,
  features = "adt_PDL1",
  idents = c("NT","JAK2","STAT1","IFNGR1","IFNGR2", "IRF1"),
  group.by = "gene",
```

```
  pt.size = 0.2,
  sort = T,
  split.by = "mixscape_class.global",
  cols = c("coral3","grey79","grey39")) +
  ggtitle("PD-L1 protein") +
  theme(axis.text.x = element_text(angle = 0, hjust = 0.5))
```



Visualizing perturbation responses with Linear Discriminant Analysis (LDA).

```
# Remove non-perturbed cells and run LDA to reduce the dimensionality of the data.
Idents(eccite) <- "mixscape_class.global"
sub <- subset(eccite, idents = c("KO", "NT"))

sub <- MixscapeLDA(
  object = sub,
  assay = "RNA",
  pc.assay = "PRTB",
  labels = "gene",
  nt.label = "NT",
  npcs = 10,
  logfc.threshold = 0.25,
  verbose = F)
```

```r
# Use LDA results to run UMAP and visualize cells on 2-D.
sub <- RunUMAP(
  object = sub,
  dims = 1:11,
  reduction = 'lda',
  reduction.key = 'ldaumap',
  reduction.name = 'ldaumap')

# Visualize UMAP clustering results.
Idents(sub) <- "mixscape_class"
sub$mixscape_class <- as.factor(sub$mixscape_class)
p <- DimPlot(object = sub, reduction = "ldaumap", label = T, repel = T, label.size = 5)

col = setNames(object = hue_pal()(12),nm = levels(sub$mixscape_class))
names(col) <- c(names(col)[1:7], "NT", names(col)[9:12])
col[8] <- "grey39"

p +
  scale_color_manual(values=col, drop=FALSE) +
  ylab("UMAP 2") +
  xlab("UMAP 1") +
  custom_theme
```