# 03_modulescore

Pooja Paresh Savla

## Overall purpose

This pipeline evaluates **pathway activation signatures** (STAT, MAPK, and PI3K/mTOR) from single-cell RNA-seq data after oIL-9R stimulation.
It summarizes **module scores** — averaged expression of key genes representing each signaling pathway — and visualizes them as heatmaps to reveal:

- which pathways are **preferentially activated** across cytokine conditions or cell populations,
- how pathway activation **varies between samples**, and
- whether certain **cell types** are especially responsive to IL-9R signaling.

```
suppressPackageStartupMessages(library(dplyr))
suppressPackageStartupMessages(library(purrr))
suppressPackageStartupMessages(library(tibble))
suppressPackageStartupMessages(library(SeuratObject))
suppressPackageStartupMessages(library(Seurat))
suppressPackageStartupMessages(library(patchwork))
suppressPackageStartupMessages(library(ggplot2))
suppressPackageStartupMessages(library(pheatmap))
suppressPackageStartupMessages(library(tidyr))

# Function to save plot as PNG
save_plot_as_png <- function(plot, filename) {
  png(filename, res = 250, width = 4000, height = 2000)
  print(plot)
  dev.off()
}
```

```r
load("../data/demo_data_annotated_biomart.RData")
```

```r
# Define gene sets
stat_genes <- c("Stat1", "Stat3", "Stat4", "Stat5a", "Stat5b", "Stat6")
mapk_genes <- c("Mapk1", "Mapk3", "Map2k1", "Map2k2", "Fos", "Jun")
pi3k_mtor_genes <- c("Pik3ca", "Akt1", "Akt2", "Mtor", "Rps6", "Rheb", "Tsc1", "Tsc2")

# Add module scores for each pathway
demo_data <- AddModuleScore(demo_data, features = list(stat_genes), name = "STAT_Module")
```

```
Warning: The `slot` argument of `GetAssayData()` is deprecated as of SeuratObject 5.0.0.
i Please use the `layer` argument instead.
i The deprecated feature was likely used in the Seurat package.
  Please report the issue at <https://github.com/satijalab/seurat/issues>.
```

```r
demo_data <- AddModuleScore(demo_data, features = list(mapk_genes), name = "MAPK_Module")
demo_data <- AddModuleScore(demo_data, features = list(pi3k_mtor_genes), name = "PI3K_Module
```

```r
# Summarize module scores
module_scores <- demo_data@meta.data %>%
  dplyr::group_by(samples) %>%
  dplyr::summarise(
    STAT = mean(STAT_Module1, na.rm = TRUE),
    MAPK = mean(MAPK_Module1, na.rm = TRUE),
    PI3K = mean(PI3K_Module1, na.rm = TRUE)
  ) %>%
  column_to_rownames("samples") %>%
  as.matrix()
```

## File 1: `modulescore_pathway.png`

**"Pathway Module Scores Across Samples"**

### What it contains

This heatmap displays **average pathway scores (STAT, MAPK, PI3K/mTOR)** across **samples** (columns).
Each pathway is represented as one row.
The color gradient (blue → white → red) corresponds to low → intermediate → high mean pathway activity.

### How it's generated

Then `pheatmap(t(module_scores))` transposes the data so **pathways are rows** and **samples are columns.**

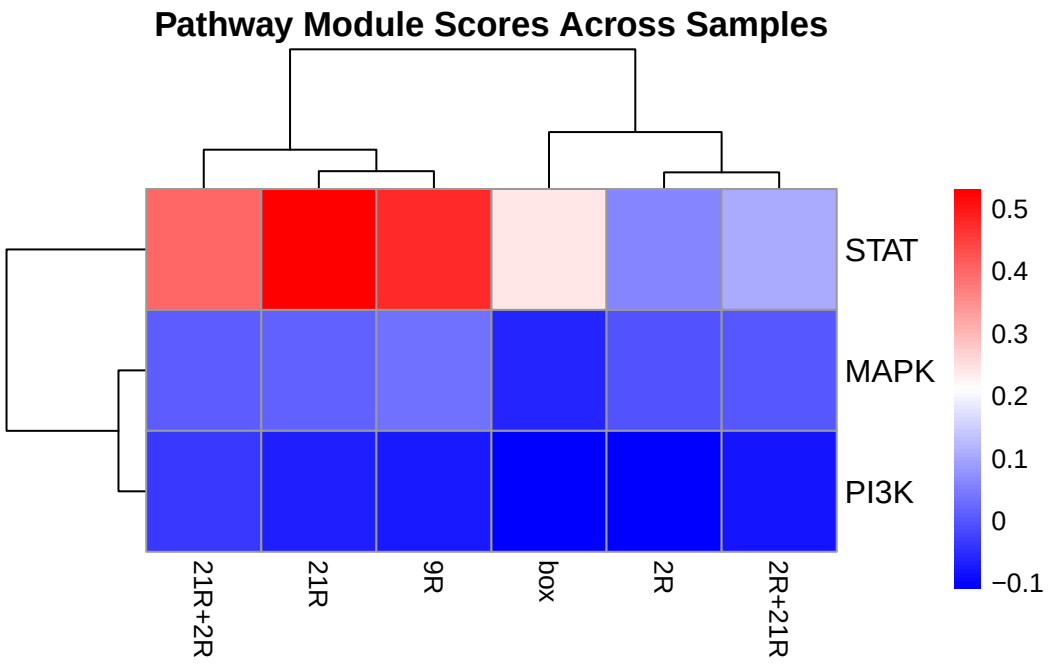
### What it's asking biologically

This heatmap helps you see:

- Which cytokine conditions (samples) **strongly activate** each signaling axis.
- Whether IL-9R stimulation predominantly engages the **STAT** pathway (as expected for many interleukin receptors), or if it co-activates **MAPK** or **PI3K/mTOR** signaling.
- How different experimental conditions (e.g., wild type vs knockout, treated vs control) differ in their **downstream transcriptional activation**.

**Example interpretation:**
If the "STAT" row is deep red for oIL-9R samples but white or blue for controls, this supports that **IL-9R signaling drives a JAK/STAT transcriptional program**.
If MAPK or PI3K are also activated, it indicates **cross-talk** or **multi-pathway engagement**, relevant for immune-cell differentiation or survival.

```
heatmap_1 <- pheatmap(
  t(module_scores),  # Transpose: pathways as rows, samples as columns
  cluster_rows = TRUE,
  cluster_cols = TRUE,
  color = colorRampPalette(c("blue", "white", "red"))(100),
  main = "Pathway Module Scores Across Samples",
  fontsize_row = 12,
  fontsize_col = 10
)
```




```
save_plot_as_png(heatmap_1,"modulescore_pathway.png")
```

```
pdf
  2
```

**File 2:** `heatmap_STAT_Module1.png`

**"STAT module by Sample and Cell Type"**

### What it contains

A heatmap of **STAT pathway module scores** (y-axis = cell types, x-axis = samples). Each cell type's mean STAT activity is plotted across all samples.

### Biological reasoning

The **STAT family** (Stat1, Stat3, Stat4, Stat5a/b, Stat6) are core transcription factors downstream of cytokine receptors, including IL-9R.
This heatmap shows:

- Which **cell populations** (e.g., T cells, NK cells, epithelial cells) have strong STAT activation in response to IL-9R signaling.
- Whether **only select subsets** (e.g., IL-9-responsive T cells) show upregulation, reflecting **cell-type-specific responsiveness**.

**Biological question addressed:**

> Does oIL-9R activate STAT signaling uniformly across the immune landscape, or are certain populations (e.g., Th9-like or memory T cells) more responsive?

## File 3: `heatmap_MAPK_Module1.png`

**"MAPK module by Sample and Cell Type"**

### What it contains

A heatmap of **MAPK pathway activity** (genes: Mapk1/3, Map2k1/2, Fos, Jun) across cell types and samples.
The color scale (white → yellow → red) indicates increasing activation intensity.

### Biological reasoning

The **MAPK/ERK** pathway mediates cytokine-driven proliferation and stress responses.
In the IL-9R context, this module tells you:

- Whether **oIL-9R** signaling also triggers **MAPK/ERK** activation in addition to STAT, implying **parallel signaling branches**.
- Which cell types rely on **ERK-dependent programs** (e.g., proliferation, differentiation).

**Possible biological insights:**

- High MAPK activity in proliferative clusters (e.g., dividing T cells) suggests **IL-9R-induced proliferation**.
- Weak MAPK activity but strong STAT activity indicates **transcriptional rather than proliferative signaling dominance.**

---

### File 4: `heatmap_PI3K_Module1.png`

**"PI3K/mTOR module by Sample and Cell Type"**

#### What it contains

A heatmap of **PI3K/mTOR pathway scores** (genes: Pik3ca, Akt1/2, Mtor, Rps6, Rheb, Tsc1/2) across samples and annotated cell types.

#### Biological reasoning

The **PI3K-Akt-mTOR** axis controls **metabolic reprogramming**, **cell survival**, and **protein synthesis**.
For IL-9R, this pathway may reflect:

- **Survival signaling** downstream of receptor engagement.
- **Metabolic adaptation** (e.g., glycolytic shift or biosynthetic activation) in IL-9-responsive immune subsets.

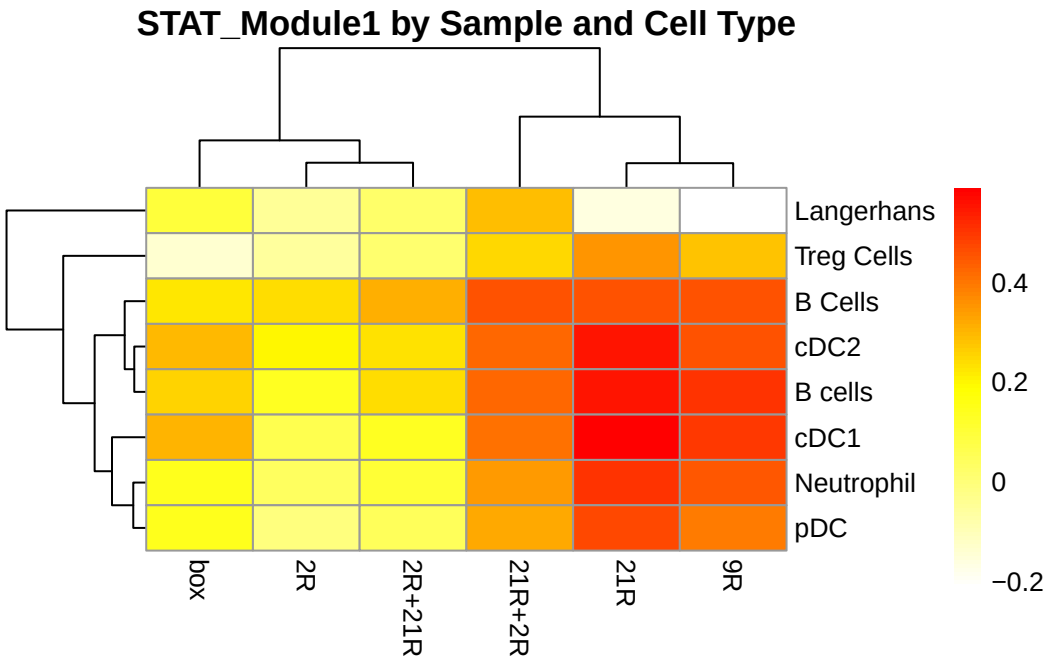
**Biological question addressed:**

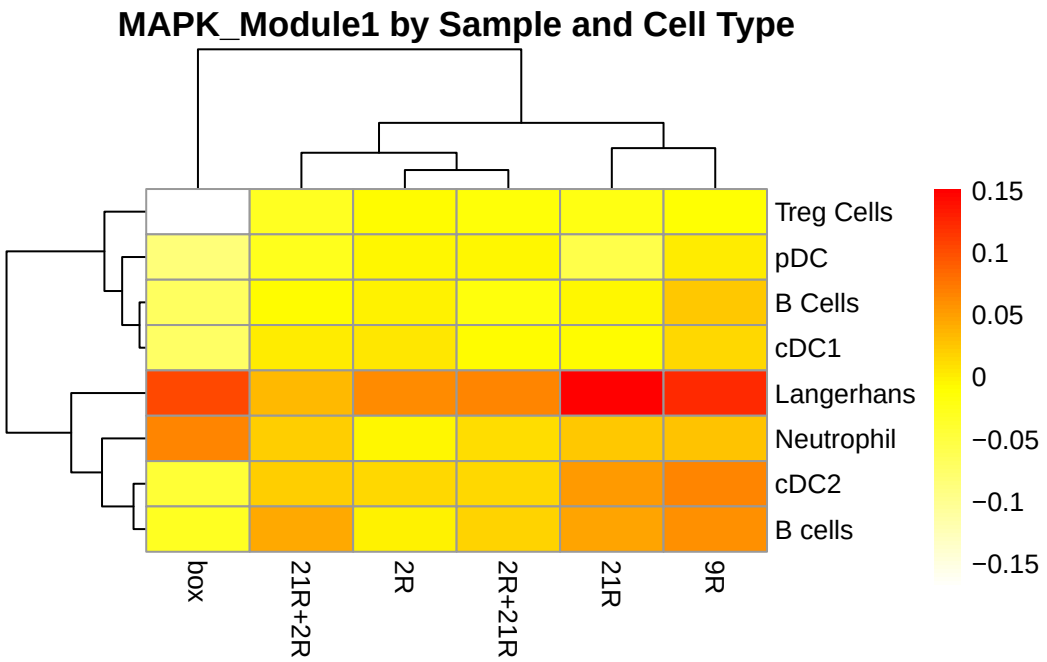
> Does IL-9R signaling promote metabolic activation via the PI3K/mTOR axis, and in which cell populations is this strongest?

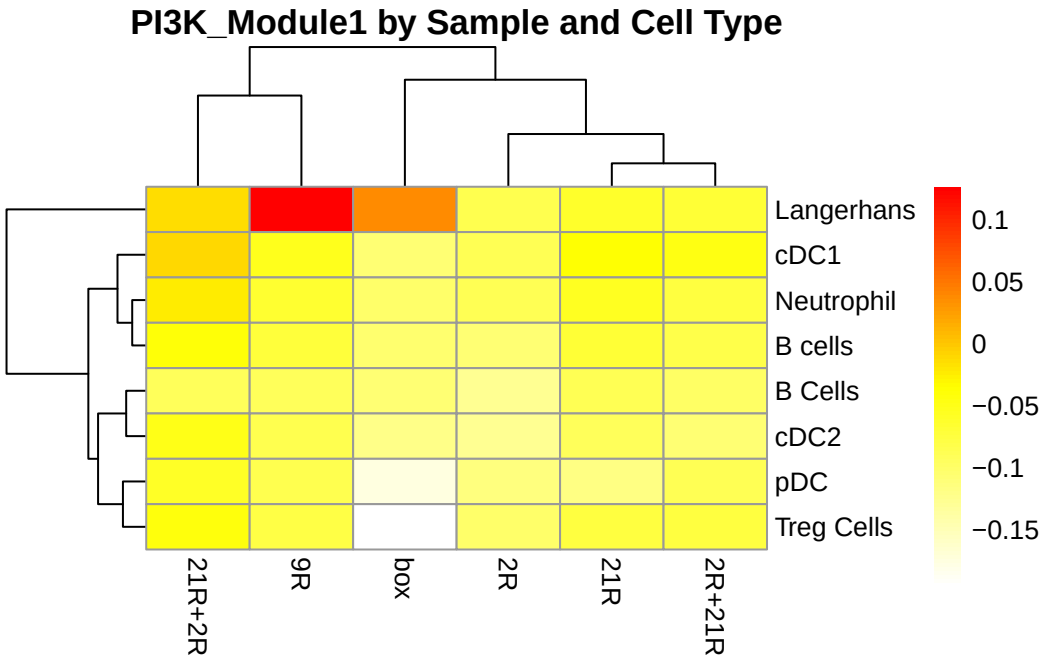
**Interpretation:**
High PI3K/mTOR scores in effector or activated cell types suggest **enhanced anabolic metabolism** and **cell growth responses** to IL-9R.

```
# Define the module scores to visualize
module_scores <- c("STAT_Module1", "MAPK_Module1", "PI3K_Module1")

# Function to create a heatmap for each module
generate_module_heatmap <- function(seurat_object, module_col) {
  # 1. Summarize score by sample and cell annotation
  summary_df <- seurat_object@meta.data %>%
    group_by(samples, cell_annotation) %>%
    summarise(score = mean(.data[[module_col]], na.rm = TRUE), .groups = 'drop')

  # 2. Pivot to wide matrix
  score_matrix <- summary_df %>%
    tidyr::pivot_wider(names_from = samples, values_from = score) %>%
    column_to_rownames("cell_annotation") %>%
```

```
    as.matrix()

  # 3. Plot heatmap
  heatmap <- pheatmap(
    score_matrix,
    cluster_rows = TRUE,
    cluster_cols = TRUE,
    color = colorRampPalette(c("white", "yellow", "red"))(100),
    main = paste(module_col, "by Sample and Cell Type"),
    fontsize_row = 10,
    fontsize_col = 10
  )
  save_plot_as_png(heatmap,paste0("heatmap_",module_col,".png"))
}

# Loop through each module score and plot
for (module in module_scores) {
  generate_module_heatmap(demo_data, module)
}
```






## Conceptual summary

| Module | Key genes | Biological role | What the heatmap reveals |
|---|---|---|---|
| **STAT** | Stat1–6 | Canonical transcriptional effectors of cytokine signaling | Transcriptional responsiveness to IL-9R |
| **MAPK** | Mapk1/3, Fos, Jun | Mitogenic and stress pathway, proliferation | Whether IL-9R drives ERK activation |
| **PI3K/mTOR** | Pik3ca, Akt1/2, Mtor, Rps6, Tsc1/2 | Growth/metabolism pathway | Metabolic activation and survival signals |

### Why this analysis matters

- **Mechanistic validation:** Confirms which intracellular cascades IL-9R primarily uses (STAT-biased vs multi-pathway).
- **Cell-type specificity:** Identifies which immune subsets are functionally responsive to IL-9.
- **Cross-talk mapping:** Reveals overlap between pathways, hinting at coordinated transcriptional and metabolic rewiring.
- **Therapeutic insight:** If certain cell types rely on specific pathways (e.g., STAT5 or mTOR), inhibitors could modulate IL-9-driven immune responses or inflammation.

### File 5: `scaled_heatmap_pSTAT.png`

**Title in plot:** "Scaled (0–100) — pSTAT Signature across Cytokine Receptors"

**What's inside**
Rows: pSTAT1, pSTAT3, pSTAT4, pSTAT5a, pSTAT5b, pSTAT6
Columns: your *samples* (cytokines/conditions)
Each row is min–max scaled to [0, 100] across columns, so you see where each STAT gene is *relatively* highest within its own row.

**What it's asking biologically**
Which conditions maximally engage each STAT axis?
Typical cues: IFNs → STAT1; IL-6 family/IL-21 → STAT3; IL-12 → STAT4; c cytokines (IL-2/7/9/15/21) → STAT5; IL-4/13 → STAT6. Peaks suggest ligand–receptor biasing along these axes.

**When to trust vs. be cautious**

- **Useful for:** pattern recognition per gene; ranking which *condition* best activates a given STAT transcript program.
- **Be cautious:** not for comparing *magnitudes between genes* (scaling removes gene-to-gene baseline differences).

**Why keep both views?**
Scaled view highlights *which condition* is the top hit per gene, independent of absolute expression.

## File 6: `unscaled_avgexp_heatmap_pSTAT.png`

**Title in plot:** "pSTAT Signature across Cytokine Receptors"

### What's inside

- Same rows/columns as above, but values are the **raw average expression** from `AverageExpression()` (no scaled).

### What it's asking biologically

- **How strongly expressed are STAT transcripts per condition, in absolute terms?**
  If `STAT5a/b` are both high in IL-9R samples relative to control, that's consistent with a **STAT5-driven** program at the transcriptional level.

**When to trust vs. be cautious**

- **Useful for:** magnitude comparisons *between* genes and conditions; showing real effect size.
- **Be cautious:** gene-length/GC/UMI depth and cell-type composition can affect raw averages; interpret alongside QC and cell-state proportions.

**Why keep both views?**
Unscaled preserves absolute differences; paired with the scaled view, you get both magnitude and pattern.**Unscaled** = preserves absolute differences (gene A really is higher than gene B).

- **Scaled** = highlights **which condition** is "best" for each gene, irrespective of gene-to-gene baseline.

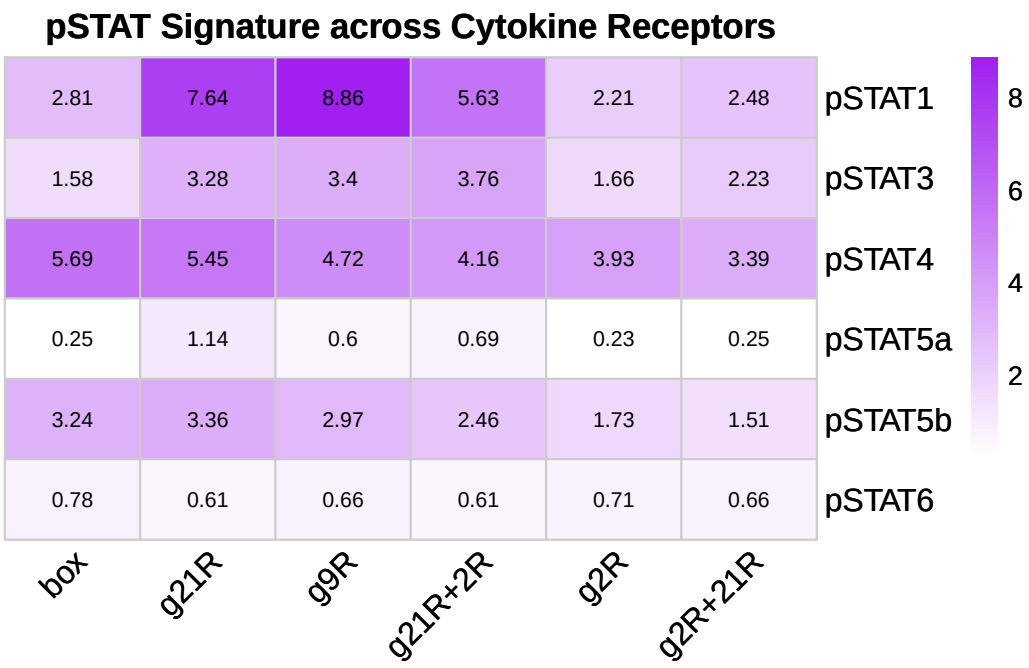
---

```r
# --- Settings ---
Idents(demo_data) <- "samples"  # set your grouping
pal <- colorRampPalette(c("white", "purple"))(100)

# --- Helpers ---
avg_mat_for <- function(obj, genes, rename_map, assay = "RNA") {
  genes_present <- intersect(genes, rownames(obj[[assay]]))
  if (length(genes_present) == 0) stop("None of the requested genes are in the assay.")
  mat <- AverageExpression(
    obj, assays = assay, features = genes_present, return.seurat = FALSE
  )[[assay]]
  # rename rows if a mapping is provided
  rn <- rownames(mat)
  rn2 <- ifelse(rn %in% names(rename_map), rename_map[rn], rn)
  rownames(mat) <- rn2
  as.matrix(mat)
}

scale_rows_0_100 <- function(mat) {
  s <- t(apply(mat, 1, function(x) {
    rng <- range(x, na.rm = TRUE)
    if (diff(rng) == 0) rep(0, length(x)) else 100 * (x - rng[1]) / diff(rng)
  }))
  # keep dimnames
  rownames(s) <- rownames(mat); colnames(s) <- colnames(mat)
  s
}

pheatmap_create <- function(mat, title, palette = pal) {
  p1 <- pheatmap(
    mat,
    color = palette,
    cluster_rows = FALSE, cluster_cols = FALSE,
    fontsize_row = 12, fontsize_col = 12,
    angle_col = 45,
    display_numbers = round(mat, 2),
    number_color = "black",
    border_color = "grey80",
    legend = TRUE,
    main = title,
    treeheight_row = 0, treeheight_col = 0
  )
```

```
}

# =============== pSTAT set ===============
pstat_genes <- c("Stat1","Stat3","Stat4","Stat5a","Stat5b","Stat6")
pstat_map <- c(
  Stat1 = "pSTAT1", Stat3 = "pSTAT3", Stat4 = "pSTAT4",
  Stat5a = "pSTAT5a", Stat5b = "pSTAT5b", Stat6 = "pSTAT6"
)

pstat_avg <- avg_mat_for(demo_data, pstat_genes, pstat_map, assay = "RNA")
```

```
As of Seurat v5, we recommend using AggregateExpression to perform pseudo-bulk analysis.
First group.by variable `ident` starts with a number, appending `g` to ensure valid variable
This message is displayed once per session.
```

```
pstat_scaled <- scale_rows_0_100(pstat_avg)


# Unscaled
p1 <- pheatmap_create(pstat_avg,
  "pSTAT Signature across Cytokine Receptors")
print(p1)
```



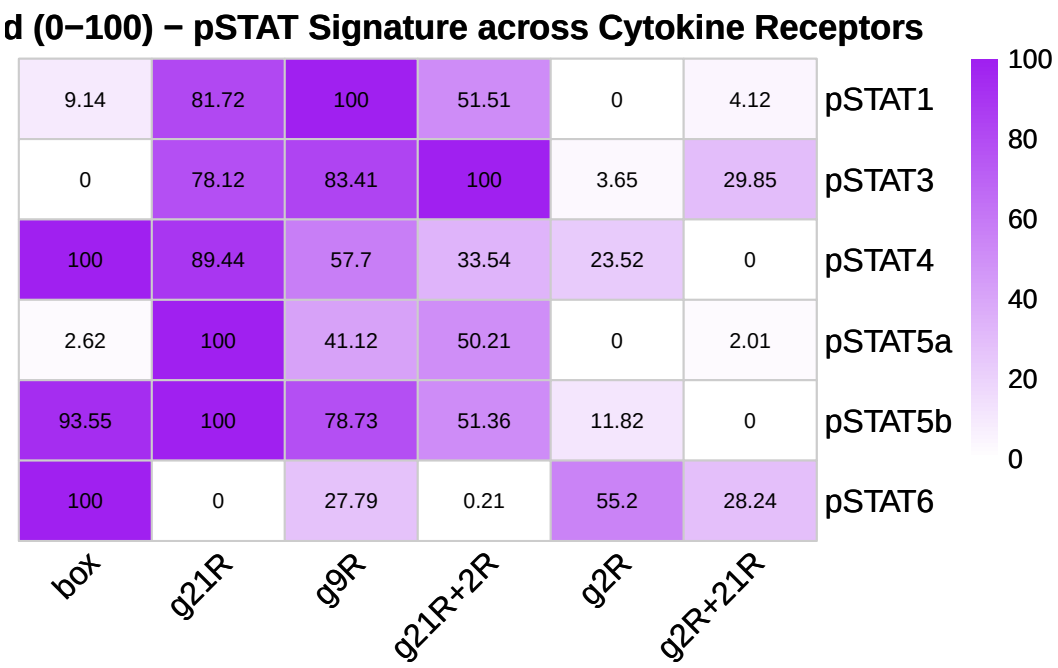
```
save_plot_as_png(p1, "unscaled_avgexp_heatmap_pSTAT.png")
```

```
pdf
  2
```

```
# Scaled 0-100
p2 <- pheatmap_create(pstat_scaled,
  "Scaled (0-100) - pSTAT Signature across Cytokine Receptors")

save_plot_as_png(p2, "scaled_heatmap_pSTAT.png")
```

```
pdf
  2
```

```
print(p2)
```




**File 7:** `scaled_heatmap_nonSTAT.png`

**Title in plot:** "Scaled (0–100) — Non-STAT Signature across Cytokine Receptors"

### What's inside

- Rows: pErk1 (Mapk3), pErk2 (Mapk1), pAkt1 (Akt1), pAkt2 (Akt2), pAkt3 (Akt3), pS6 (Rps6)
  Columns: your *samples* (cytokines/conditions)
  Row-wise min–max scaling to [0, 100] emphasizes *relative peaks* per pathway component.

### What it's asking biologically

- Which conditions preferentially engage the MAPK (ERK1/2), PI3K–Akt (Akt1/2/3), and mTOR (S6) axes? Cytokine receptors can co-activate these pathways via JAK-dependent adapters or parallel signaling; peaks can reveal biased engagement (e.g., strong pS6 suggests mTOR output; strong pErk suggests MAPK bias).

**When to trust vs. be cautious**

- **Useful for:** identifying *which condition* most up-regulates each pathway readout relative to that gene's own baseline.
- **Be cautious:** scaling obscures whether pS6 is globally higher than pErk, etc.—use the unscaled plot to judge magnitude.

**Why keep both views?**
Scaled view is for *pattern discovery* (who peaks where), complementary to the unscaled magnitude view.

## File 8: `unscaled_avgexp_heatmap_nonSTAT.png`

**Title in plot:** "Non-STAT Signature across Cytokine Receptors"

### What's inside

- Same rows/columns, but raw average expression (no scaling).

### What it's asking biologically

- What is the *absolute* expression level of ERK/Akt/S6 transcripts per condition? This helps judge whether a pathway is broadly higher across conditions (e.g., uniformly elevated Rps6 → ribosomal/mTOR program) versus condition-specific spikes.

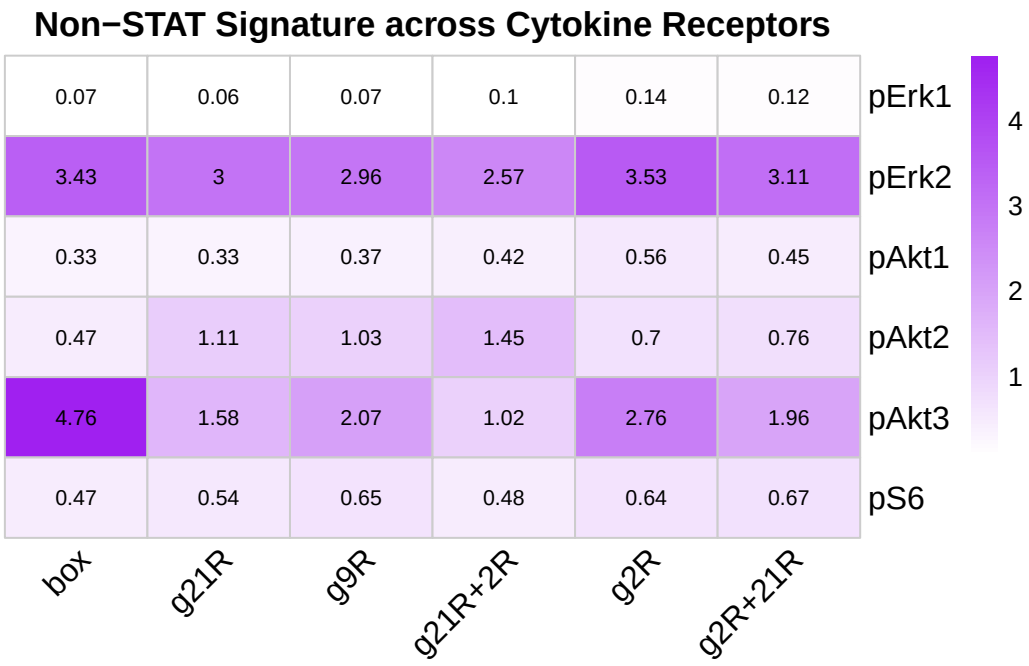
### When to trust vs. be cautious

- **Useful for:** real effect sizes across both genes and conditions; comparing MAPK vs PI3K–Akt vs mTOR axes directly.
- **Be cautious:** transcript abundance is an *indirect* proxy of pathway activity—post-translational phosphorylation drives signaling. Interpret with phosphorylation/protein data if available.

### Why keep both views?

Unscaled shows magnitude; together with the scaled plot you get both "how much" and "where it peaks."

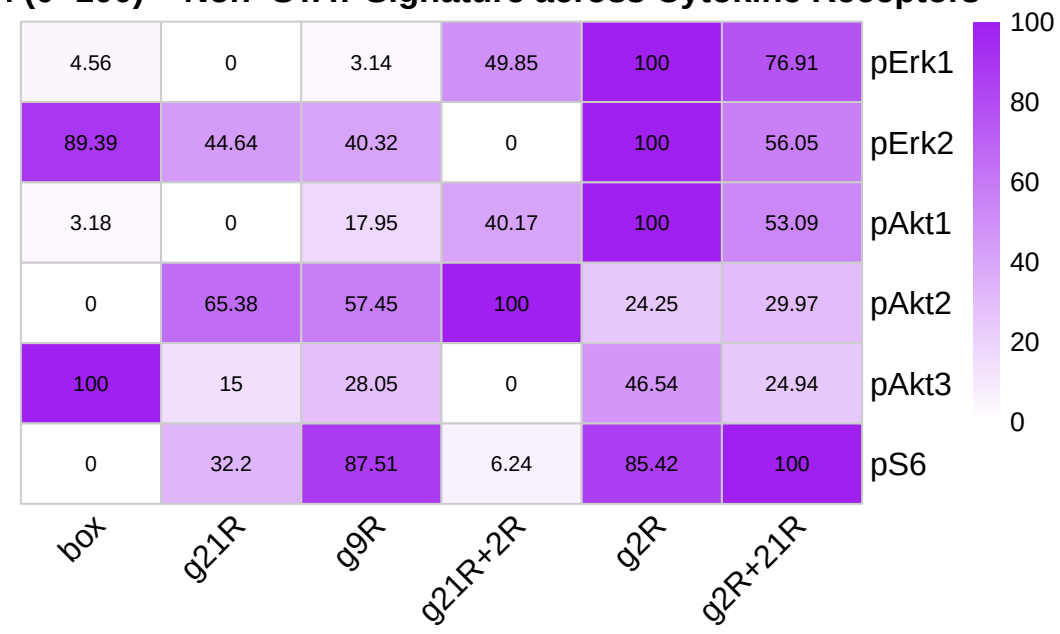
```
# =============== non-STAT set ===============
non_pstat_genes <- c("Mapk3","Mapk1","Akt1","Akt2","Akt3","Rps6")
non_pstat_map <- c(
  Mapk3 = "pErk1", Mapk1 = "pErk2",
  Akt1 = "pAkt1", Akt2 = "pAkt2", Akt3 = "pAkt3",
  Rps6 = "pS6"
)

nonp_avg <- avg_mat_for(demo_data, non_pstat_genes, non_pstat_map, assay = "RNA")
nonp_scaled <- scale_rows_0_100(nonp_avg)

# Unscaled
p3 <- pheatmap_create(nonp_avg,
 "Non-STAT Signature across Cytokine Receptors")
```




```
save_plot_as_png(p3, "unscaled_avgexp_heatmap_nonSTAT.png")
```

```
pdf
  2
```

```
# Scaled 0-100
p4 <- pheatmap_create(nonp_scaled,
  "Scaled (0-100) - Non-STAT Signature across Cytokine Receptors")
```




```
save_plot_as_png(p4, "scaled_heatmap_nonSTAT.png")
```

```
pdf
  2
```

### Biological context & caveats

- These are **STAT gene transcripts**, not phospho-STAT proteins. True pathway activation is phosphorylation and nuclear translocation; RNA captures **downstream transcriptional consequences** and/or **feedback expression** of STATs.
- If you want to directly reflect signaling activity, consider:
  - A curated **target-gene signature** for STAT pathways (e.g., known STAT5 targets) rather than STAT transcripts themselves.
  - If you have CITE-seq or phospho-flow/Ab-seq: include protein/phospho features (e.g., `pSTAT5`, `pERK`, `pAKT`, `pS6`) and compute module scores from those assays.