

# AD vs Psoriasis Transcriptomic Reproduction (with figure explainers)

Jae

2025-12-24

## Contents

<b>Introduction</b>	<b>1</b>
<b>Differential analysis - setup</b>	<b>3</b>
<b>Figure 1</b>	<b>4</b>
Fig 1a - PCA (PC1-PC3) . . . . .	4
Fig 1b — Transcriptomic heterogeneity . . . . .	6
Fig 1c — DEG counts (lesional vs healthy) . . . . .	8
Fig 1d — Venn overlap . . . . .	8
Fig 1e-f — log2FC concordance (lesional + non-lesional) . . . . .	10
<b>Figure 2 — Functional enrichment (paper Fig 2)</b>	<b>11</b>
<b>Figure 3 — Cytokines (paper Fig 3)</b>	<b>15</b>
<b>Figure 4 — H3K27ac cell-type enrichment</b>	<b>17</b>
<b>Conclusion</b>	<b>20</b>
<b>References</b>	<b>20</b>
<b>Session info</b>	<b>20</b>

## Introduction

This notebook reproduces transcriptome-level comparisons from Tsoi et al. (2019), a large RNA-seq study comparing atopic dermatitis (AD), psoriasis (PSO), and healthy control skin. The paper's central conclusion is that AD is predominantly IL-13–driven, while psoriasis is dominated by IL-17/IL-36–associated inflammatory responses. Also, another conclusion was that AD shows greater molecular heterogeneity than psoriasis

across samples. We test on reproducibility in this notebook, using the authors' public RNA-seq cohort (GSE121212).

First, we will load the data and remove some impurities.

```
counts <- read.delim("GSE121212_readcount.txt.gz", comment.char = "#") %>%
  as.data.frame() %>%
  # Drop those weird Excel date rows that are not genes
  filter(!X %in% c(
    "1-Dec", "1-Mar", "1-Sep", "10-Mar", "10-Sep", "11-Sep",
    "15-Sep", "2-Mar", "2-Sep", "3-Mar", "3-Sep", "4-Mar",
    "4-Sep", "5-Mar", "5-Sep", "6-Mar", "6-Sep", "7-Mar",
    "7-Sep", "8-Mar", "8-Sep", "9-Mar", "9-Sep"
  )) %>%
  column_to_rownames("X")
```

Next, we will generate metadata based off the column names. Samples are grouped by disease (AD vs PSO) and lesion status (lesional vs non-lesional). ‘Lesional’ indicates active inflamed plaques; ‘non-lesional’ indicates clinically unaffected skin from the same disease cohort, which may still show a mild ‘pre-inflammatory’ transcriptomic shift compared to healthy controls.

```
meta <- tibble(sample = colnames(counts)) %>%
  mutate(
    group = str_extract(sample, "^(AD|PSO|CTRL)"),
    subject = str_extract(sample, "(?=<^AD|PSO|CTRL)_)[0-9]+") %>%
    replace_na("NA"),
    status = sample %>%
      sub("^(AD|PSO|CTRL)_[0-9]+_"," ", .) %>%
      tolower() %>%
      gsub("\\.", "_", .)
  ) %>%
  filter(
    (group == "CTRL" & status == "healthy") |
    (group == "AD" & status %in% c("lesional", "non_lesional")) |
    (group == "PSO" & status %in% c("lesional", "non_lesional"))
  ) %%%
  mutate(
    condition = case_when(
      group == "CTRL" & status == "healthy" ~ "healthy",
      group == "AD" & status == "non_lesional" ~ "ad_non_lesion",
      group == "AD" & status == "lesional" ~ "ad_lesion",
      group == "PSO" & status == "non_lesional" ~ "pso_non_lesion",
      group == "PSO" & status == "lesional" ~ "pso_lesion"
    ),
    condition = factor(condition, levels = c("healthy", "ad_non_lesion",
                                              "ad_lesion", "pso_non_lesion",
                                              "pso_lesion")),
    subject_id = factor(paste(group, subject, sep = "_"))
  ) %>%
  select(sample, condition, group, status, subject_id)

counts <- counts[, meta$sample]
counts <- check_sample_alignment(counts, meta)

## [1] "Sample alignment: "
```

```




---



```

## Differential analysis - setup

For differential analysis, we follow the paper's instructions regarding filtering and design. We used a limma-vroom workflow. Please see the paper for more details. The `get_de` function is setup to extract contrast-based DEGs.

```

dge <- DGEList(counts = counts, group = meta$condition)
keep <- rowSums(dge$counts > 1) >= 2
dge <- dge[keep, , keep.lib.sizes = FALSE]
dge <- calcNormFactors(dge)

design <- model.matrix(~ 0 + condition, data = meta)
colnames(design) <- levels(meta$condition)

v <- voom(dge, design)
fit <- lmFit(v, design)

cont.matrix <- makeContrasts(
  ad_non_lesion_vs_healthy = ad_non_lesion - healthy,
  ad_lesion_vs_healthy = ad_lesion - healthy,
  pso_non_lesion_vs_healthy = pso_non_lesion - healthy,
  pso_lesion_vs_healthy = pso_lesion - healthy,
  levels = design
)

fit2 <- eBayes(contrasts.fit(fit, cont.matrix))

get_de <- function(coef_name) {
  topTable(fit2, coef = coef_name, number = Inf) %>%
    rownames_to_column("gene") %>%
    as_tibble() %>%
    mark_deg()
}

res_ad_non_lesion <- get_de("ad_non_lesion_vs_healthy")
res_ad_lesion <- get_de("ad_lesion_vs_healthy")
res_pso_non_lesion <- get_de("pso_non_lesion_vs_healthy")
res_pso_lesion <- get_de("pso_lesion_vs_healthy")

logexpr <- v$E

```

---

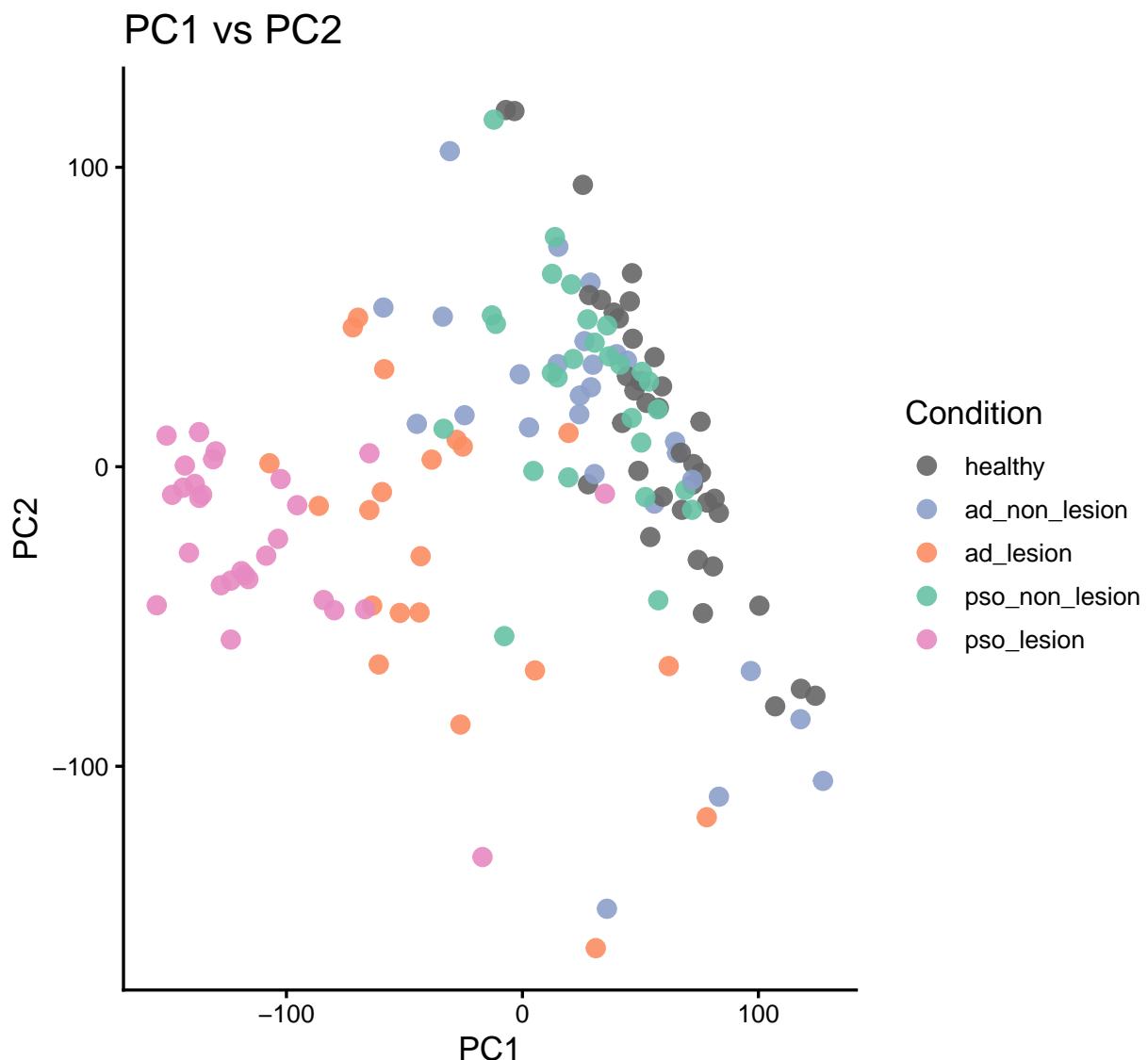
**Figure 1**

**Fig 1a - PCA (PC1-PC3)**

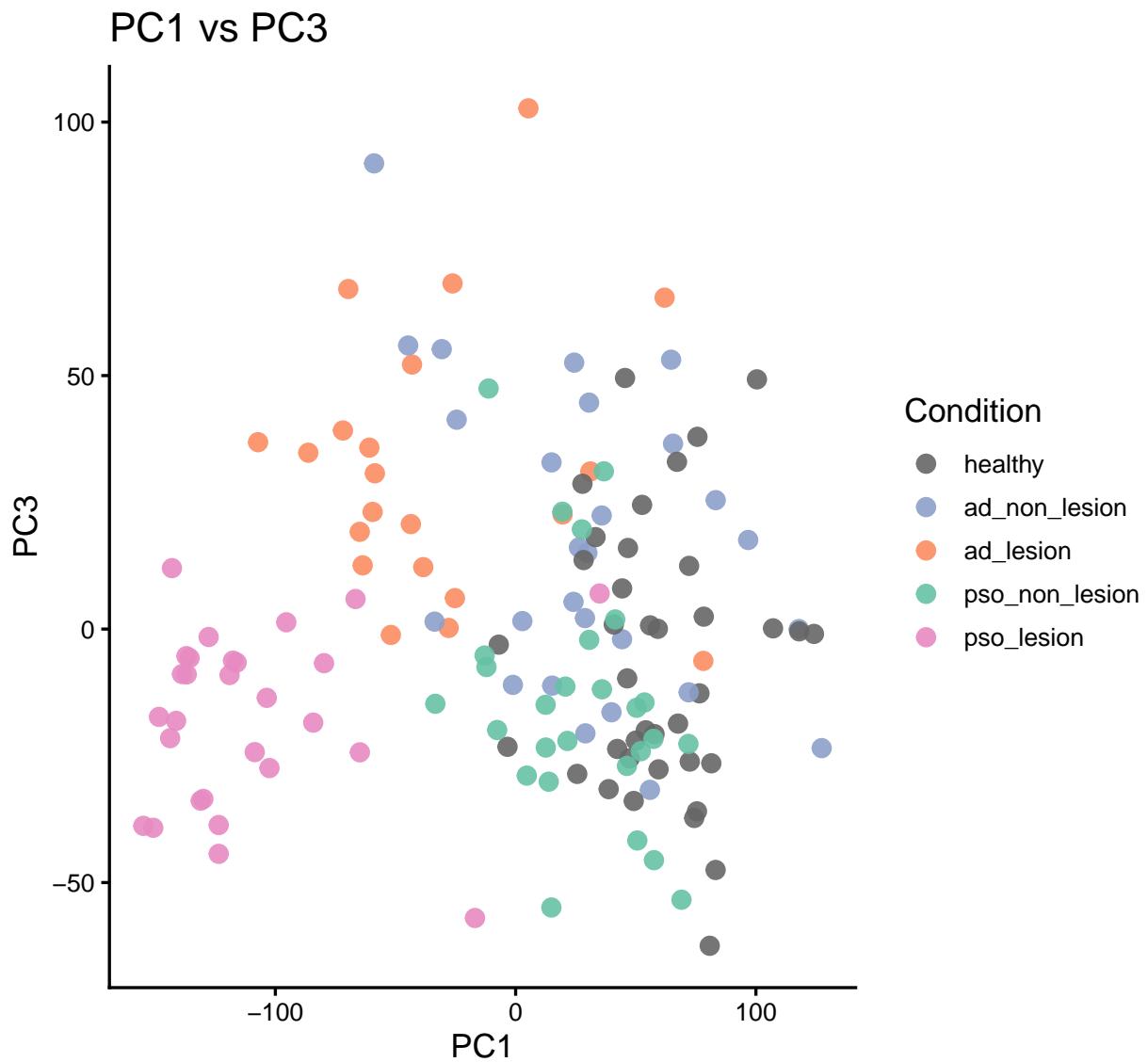
```
pca <- prcomp(t(logexpr))

pca_df <- as.data.frame(pca$x[, 1:3]) %>%
  rownames_to_column("sample") %>%
  left_join(meta, by = "sample")

plot_pca(pca_df, "PC1", "PC2", "PC1 vs PC2")
```



```
plot_pca(pca_df, "PC1", "PC3", "PC1 vs PC3")
```

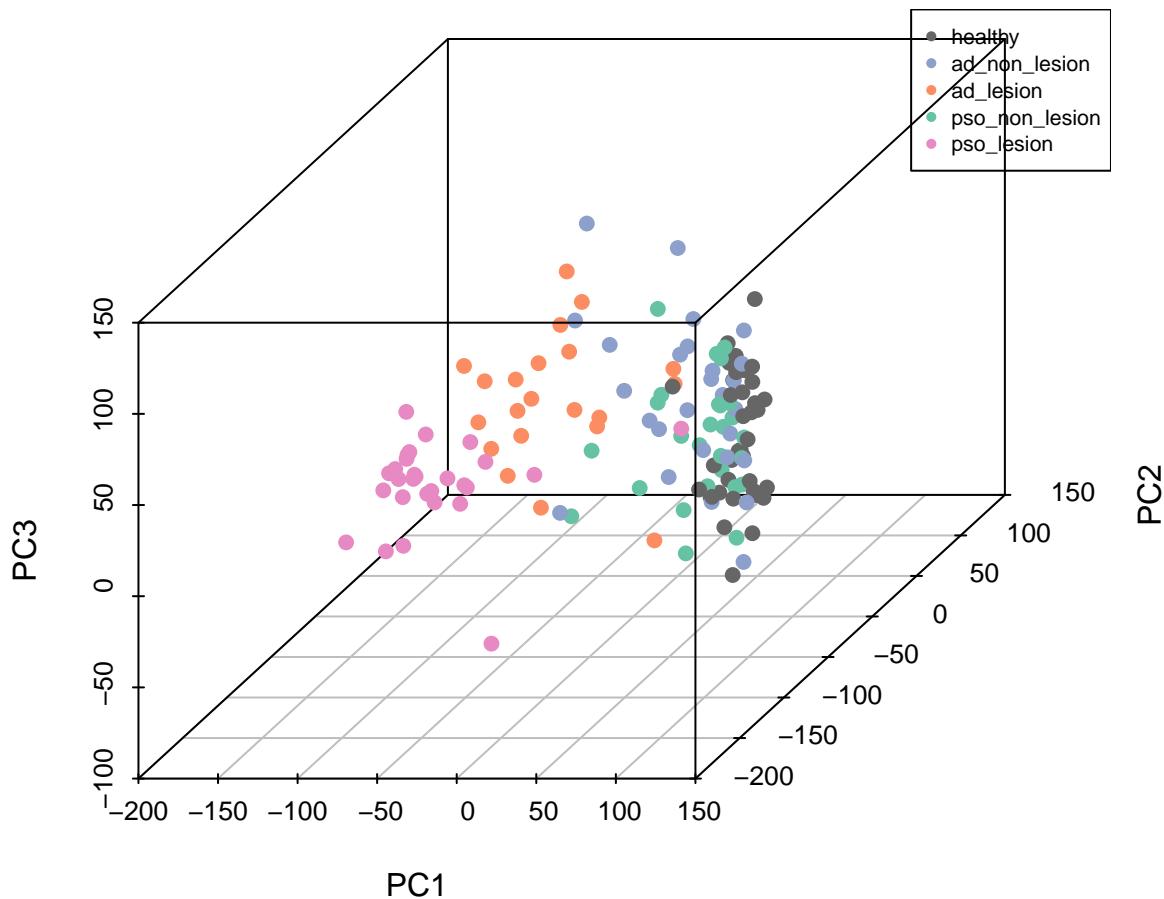


```

cols <- cond_cols[as.character(pca_df$condition)]
scatterplot3d(pca_df$PC1, pca_df$PC2, pca_df$PC3,
              color = cols, pch = 16, cex.symbols = 1.1,
              xlab = "PC1", ylab = "PC2", zlab = "PC3",
              main = "PCA: PC1-PC3")
legend("topright", legend = names(cond_cols), col = cond_cols,
      pch = 16, cex = 0.7)

```

## PCA: PC1–PC3



This PCA summarizes global expression variation across samples (each point is one biopsy), where distances reflect transcriptome similarity. Lesional samples separate from healthy/non-lesional skin, with psoriasis lesional typically most distinct and AD lesional more intermediate, matching the original paper's qualitative clustering pattern.

---

**Fig 1b — Transcriptomic heterogeneity**

```
pcs <- pca$x[, 1:3]

get_group_distances <- function(cond_label) {
  idx <- which(meta$condition == cond_label)
  d <- dist(pcs[idx, , drop = FALSE])
  tibble(distance = as.vector(d), group = cond_label)
```

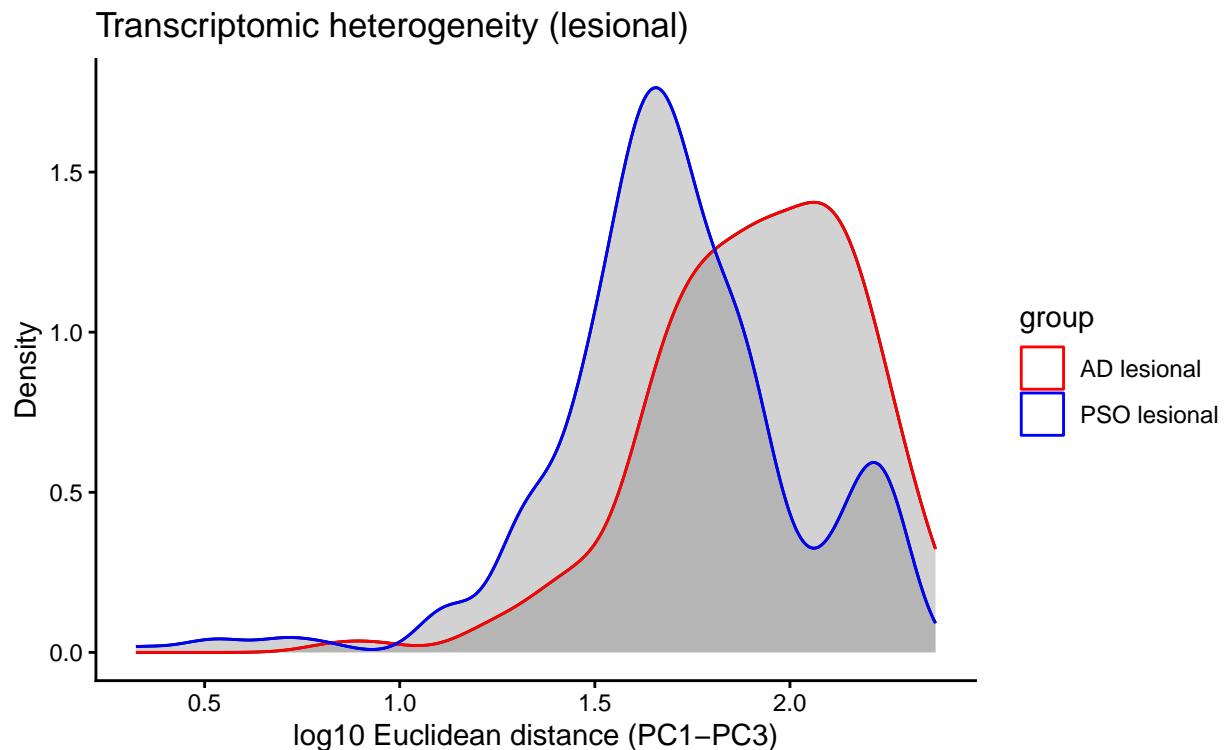
```

}

dist_df <- bind_rows(get_group_distances("ad_lesion"),
                      get_group_distances("pso_lesion")) %>%
  mutate(log_distance = log10(distance),
        group = recode(group, ad_lesion = "AD lesional",
                       pso_lesion = "PSO lesional"))

ggplot(dist_df, aes(x = log_distance)) +
  geom_density(aes(fill = group), alpha = 0.35) +
  geom_density(aes(color = group)) +
  scale_fill_manual(values = c("AD lesional" = cond_cols["ad_lesion"],
                               "PSO lesional" = cond_cols["pso_lesion"])) +
  scale_color_manual(values = c("AD lesional" = "red",
                               "PSO lesional" = "blue")) +
  theme_classic() +
  labs(x = "log10 Euclidean distance (PC1-PC3)", y = "Density",
       title = "Transcriptomic heterogeneity (lesional)")

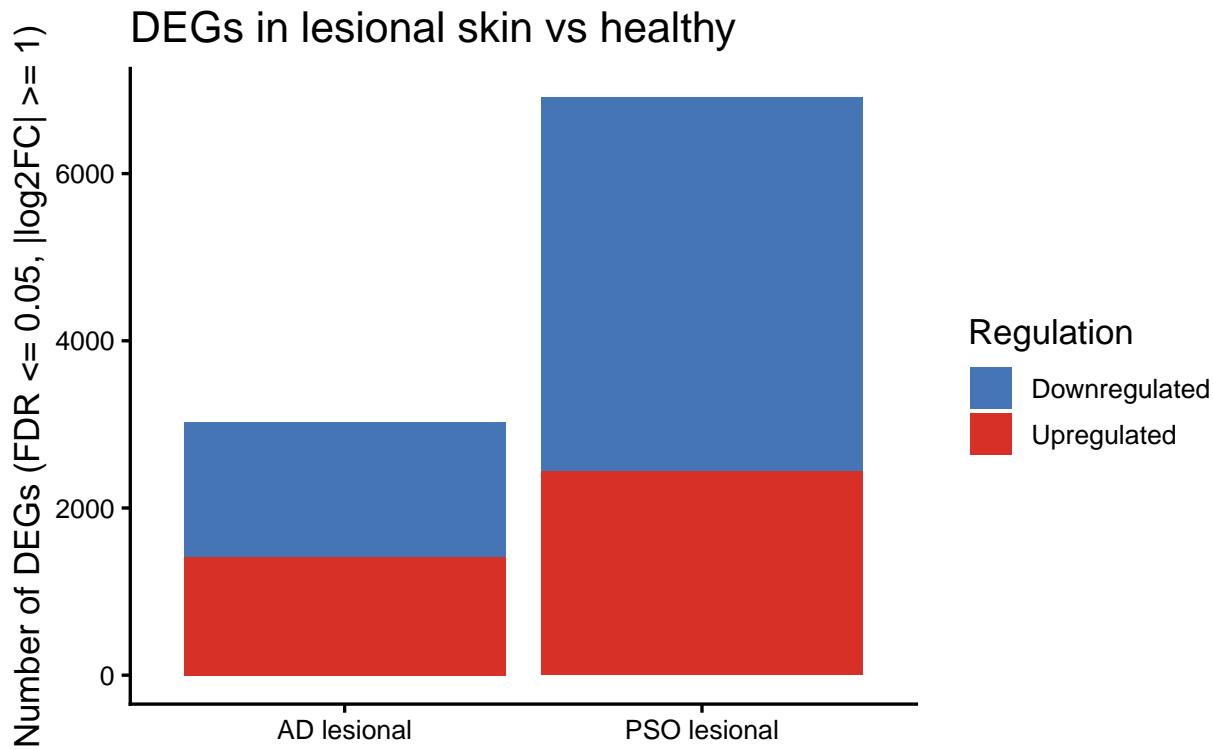
```



This plot shows the distribution of within-group pairwise Euclidean distances between lesional samples for PC1-PC3, where a broader or right-shifted curve indicates greater transcriptomic heterogeneity. The AD and PSO lesional distributions overlap but are not identical, suggesting slightly different within-disease variability, which matches the original paper's reported pattern of overlapping curves with a subtle shift between AD and psoriasis.

Fig 1c — DEG counts (lesional vs healthy)

```
plot_deg_counts(list(res_ad_lesion, res_pso_lesion),
                c("AD lesional", "PSO lesional"),
                "DEGs in lesional skin vs healthy")
```



This bar plot summarizes the number of significantly up- and down-regulated genes in lesional skin (using the DEG thresholds defined in `mark_deg()`). Psoriasis lesional samples show a larger overall DEG burden than AD lesional samples, matching the original paper's result that psoriasis exhibits stronger transcriptomic disruption in lesions.

Fig 1d — Venn overlap

```
ad_lesion_deg <- get_deg_genes(res_ad_lesion)
pso_lesion_deg <- get_deg_genes(res_pso_lesion)

cont_ad_vs_pso <- makeContrasts(ad_vs_pso_lesion = ad_lesion-pso_lesion,
                                    levels = design)
fit_ad_vs_pso <- eBayes(contrasts.fit(fit, cont_ad_vs_pso))
res_ad_vs_pso_lesion <- topTable(fit_ad_vs_pso, coef = "ad_vs_pso_lesion",
                                    number = Inf) %>%
  rownames_to_column("gene") %>%
  as_tibble() %>%
```

```

mark_deg()

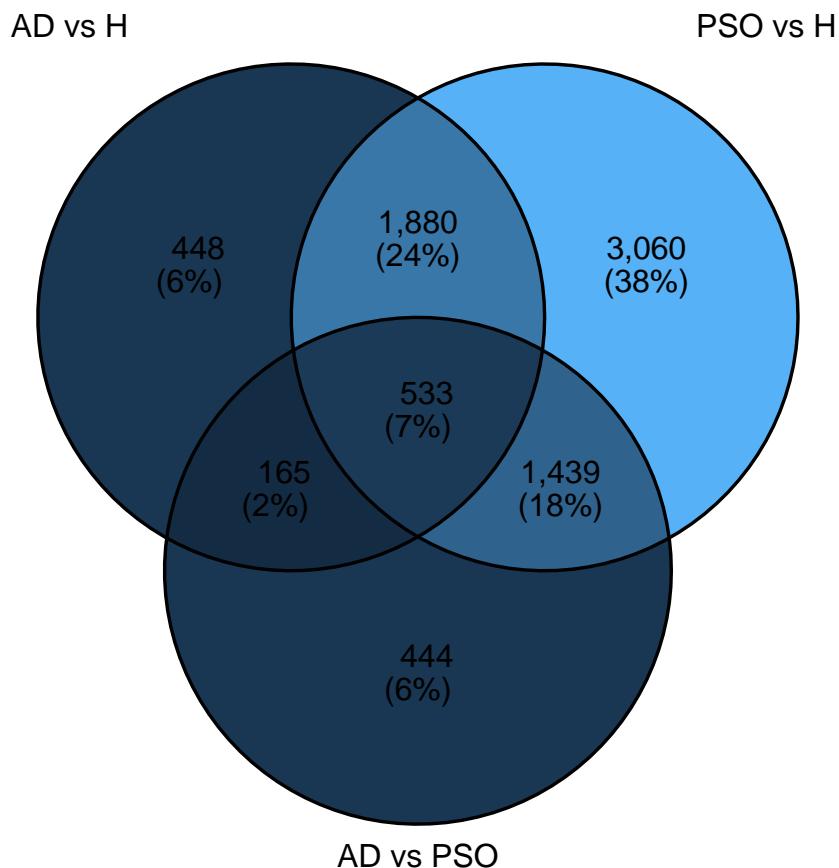
ad_vs_pso_deg <- get_deg_genes(res_ad_vs_pso_lesion)

lesion_sets <- list(
  "AD vs H" = ad_lesion_deg,
  "PSO vs H" = pso_lesion_deg,
  "AD vs PSO" = ad_vs_pso_deg
)

ggVennDiagram(lesion_sets, label_alpha = 0, edge_size = 0.6) +
  theme(legend.position = "none") +
  labs(title = "Overlap of DEG sets in lesional skin")

```

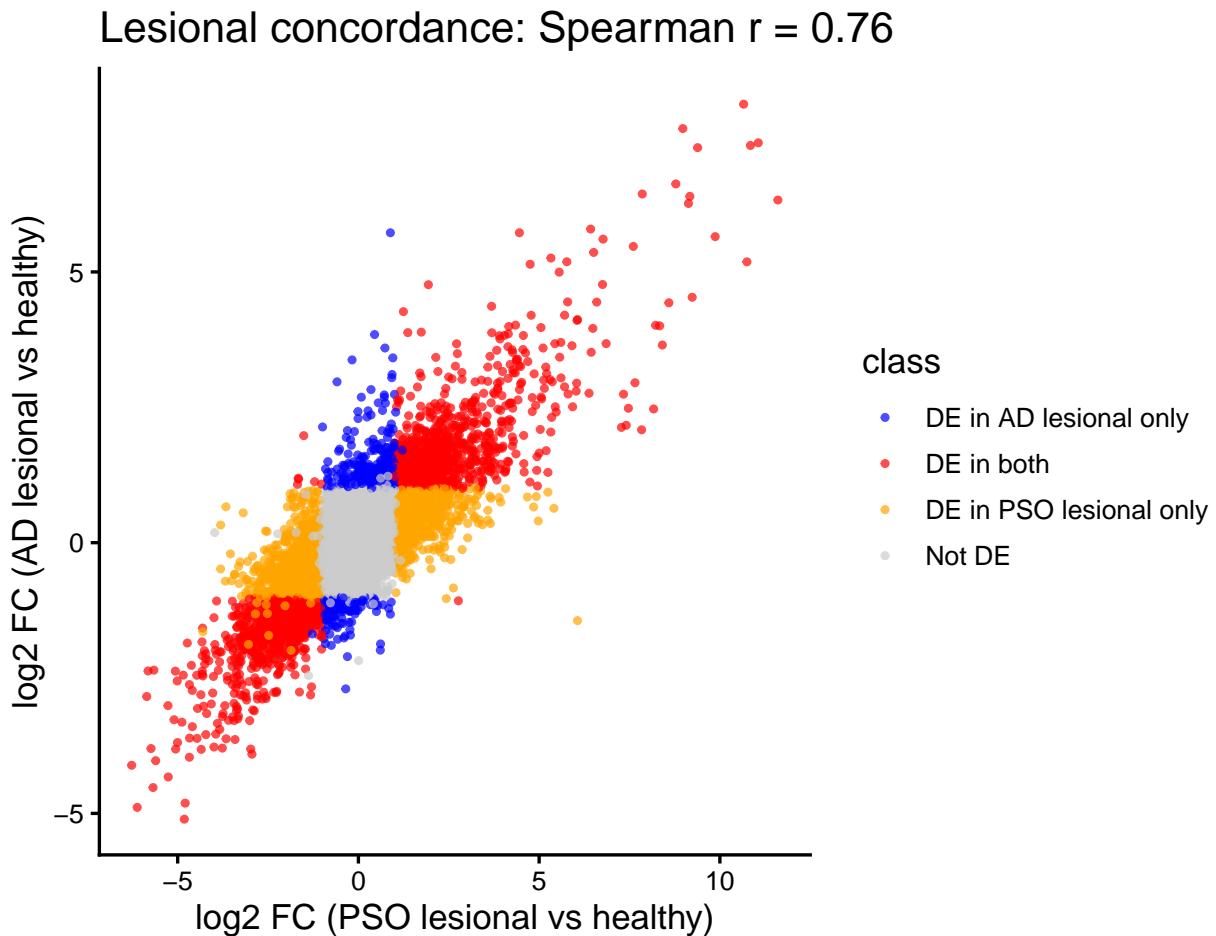
Overlap of DEG sets in lesional skin



This Venn diagram shows how the lesional DEG lists overlap for psoriasis vs healthy, AD vs healthy, and psoriasis vs AD. In the paper, the unique counts are 2911, 494, and 311; in our run they're 3060, 448, and 444. The takeaway is the same in both: psoriasis has more lesion-specific changes, and AD shares a big chunk of its lesion signature with psoriasis. The numbers aren't identical because a bunch of genes sit right on the “significant/not significant” edge, and small choices like filtering or gene ID handling can nudge them over the line.

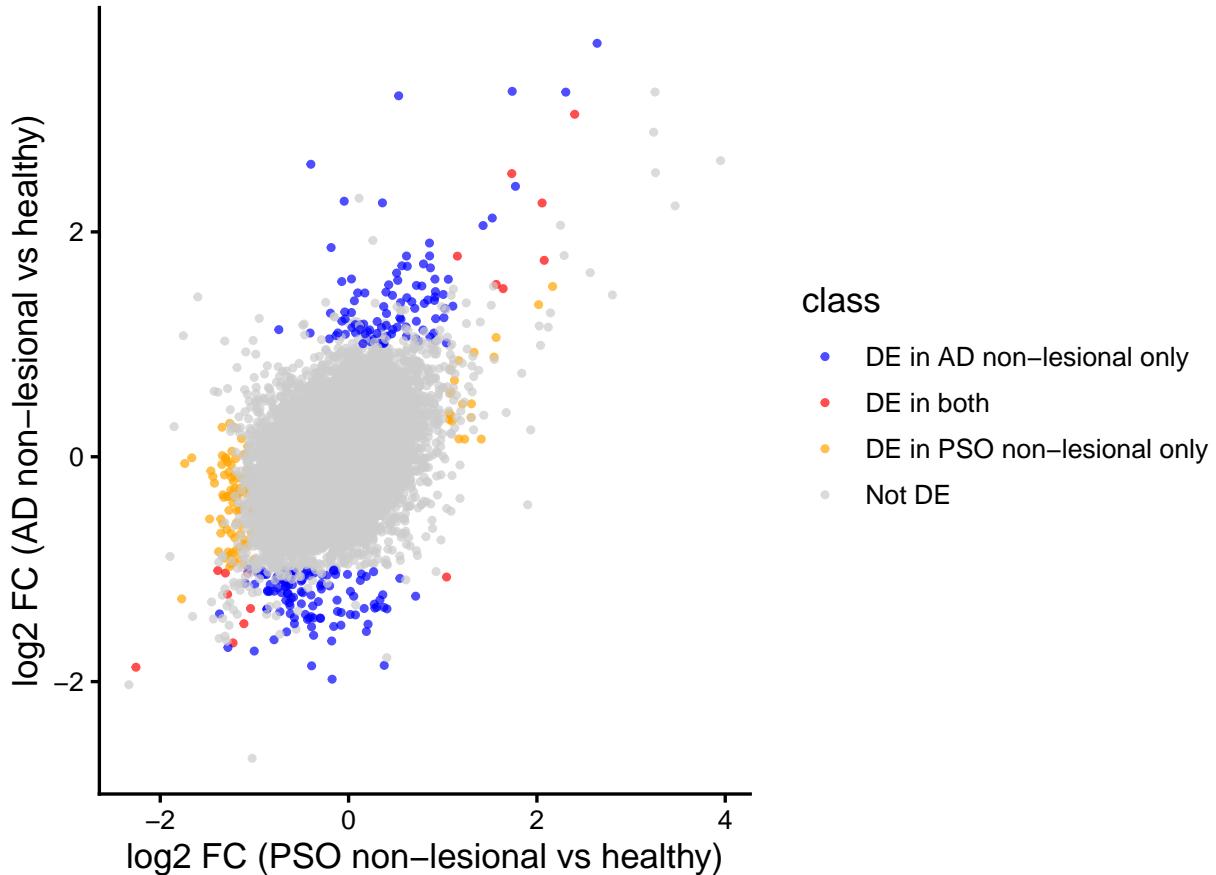
Fig 1e–f — log2FC concordance (lesional + non-lesional)

```
plot_fc_concordance(res_pso_lesion, res_ad_lesion,
  "PSO lesional vs healthy", "AD lesional vs healthy", "Lesional concordance")
```



```
plot_fc_concordance(res_pso_non_lesion, res_ad_non_lesion,
  "PSO non-lesional vs healthy", "AD non-lesional vs healthy",
  "Non-lesional concordance")
```

## Non-lesional concordance: Spearman $r = 0.42$



These concordance plots compare gene-wise log2 fold changes between psoriasis (x-axis) and AD (y-axis). In lesional skin there's a clear positive trend, meaning many genes shift in the same direction in both diseases, while the non-lesional plot is much tighter around zero but still shows a mild positive relationship. This matches the original paper's takeaway: strong concordance in lesions and weaker, but still detectable, concordance in non-lesional skin.

---

**Figure 2 — Functional enrichment (paper Fig 2)**

```

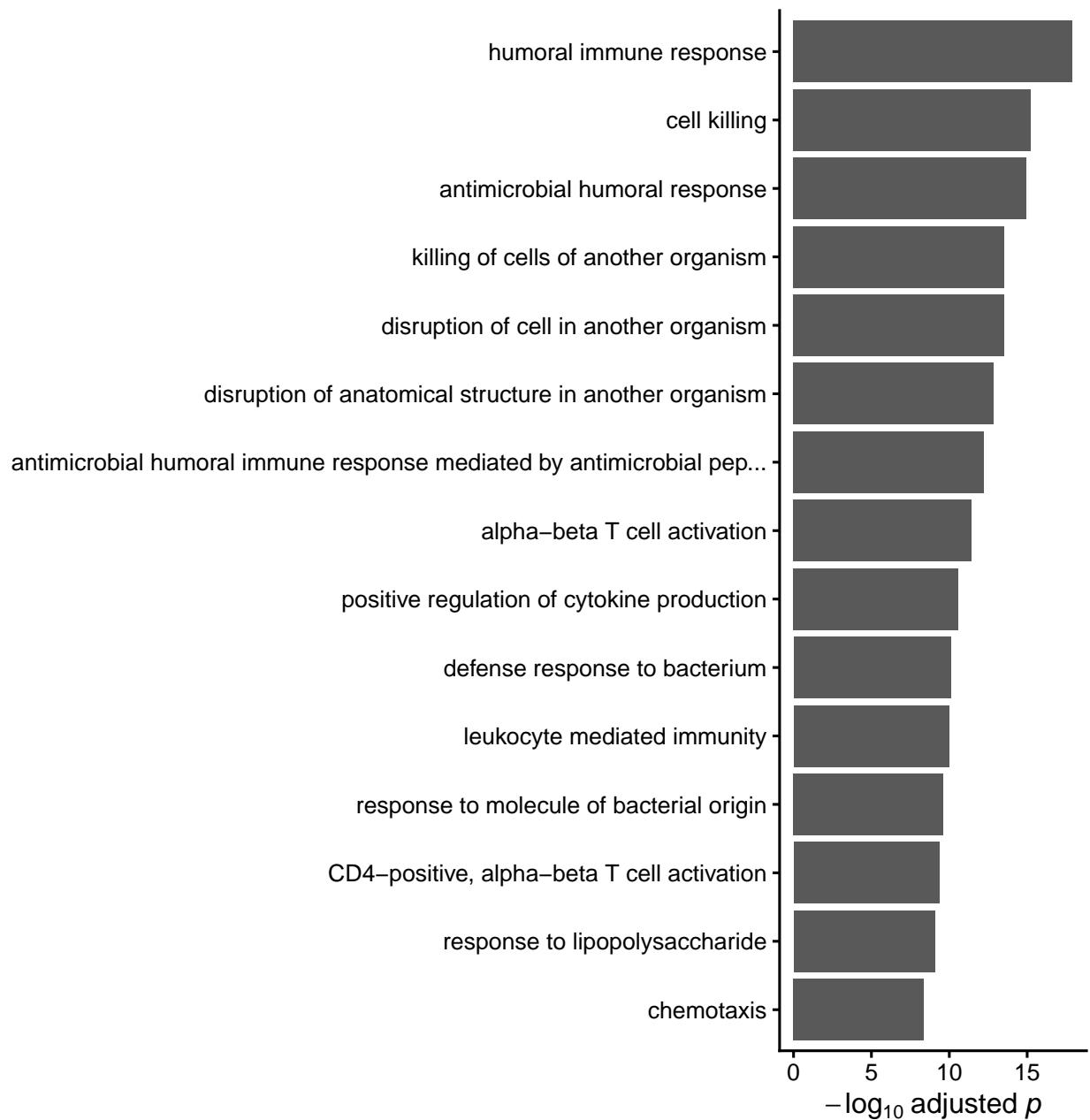
ad_lesion_deg <- get_deg_genes(res_ad_lesion)
psoriasis_lesion_deg <- get_deg_genes(res_psoriasis_lesion)

common_lesion <- intersect(ad_lesion_deg, psoriasis_lesion_deg)
ad_only_lesion <- setdiff(ad_lesion_deg, psoriasis_lesion_deg)
psoriasis_only_lesion <- setdiff(psoriasis_lesion_deg, ad_lesion_deg)
universe_genes <- res_ad_lesion$gene

go_common <- run_go_enrich(common_lesion, universe_genes)
plot_go_bar(go_common, "Common lesional DEGs: top GO BP terms")

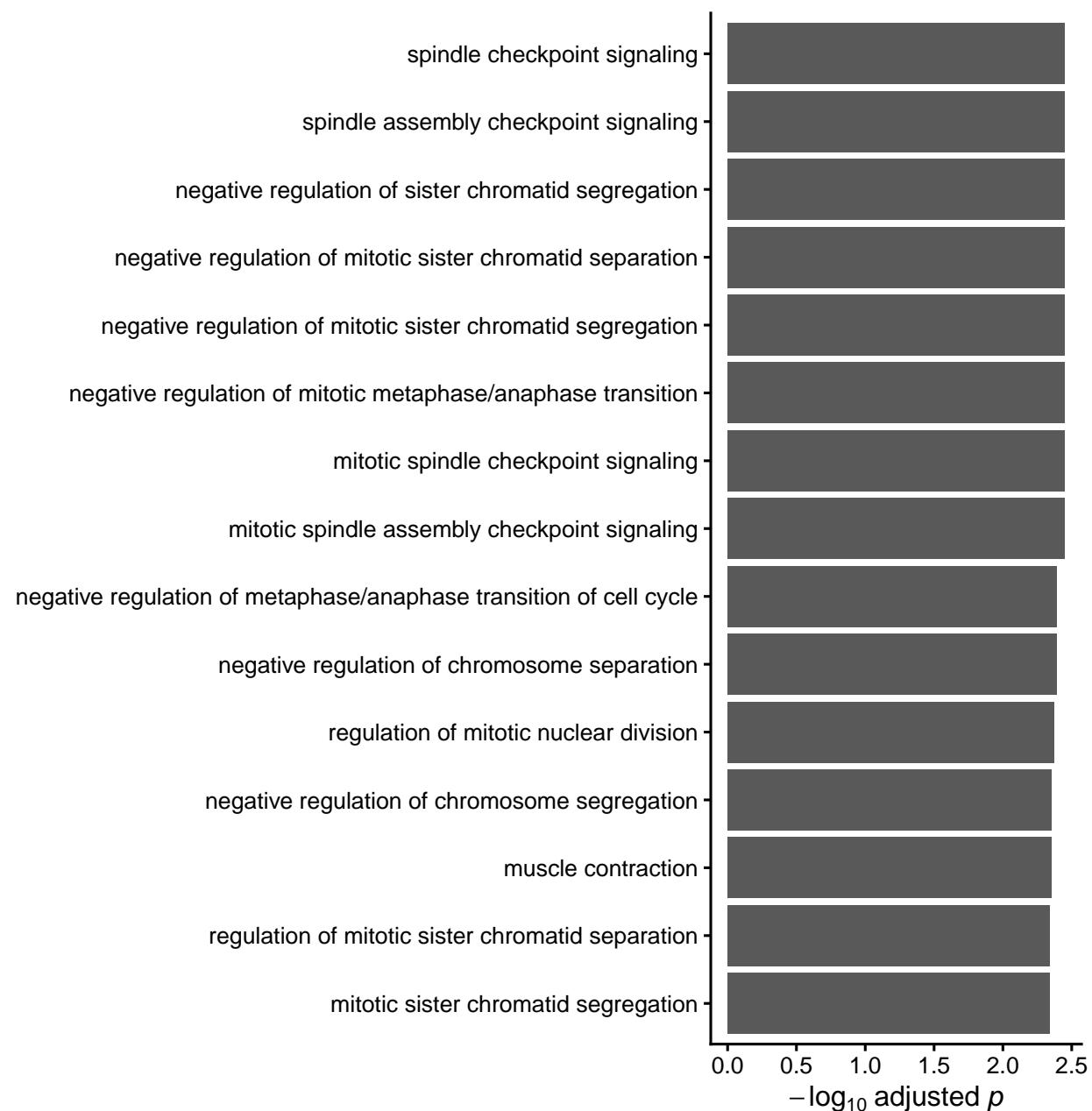
```

### Common lesional DEG

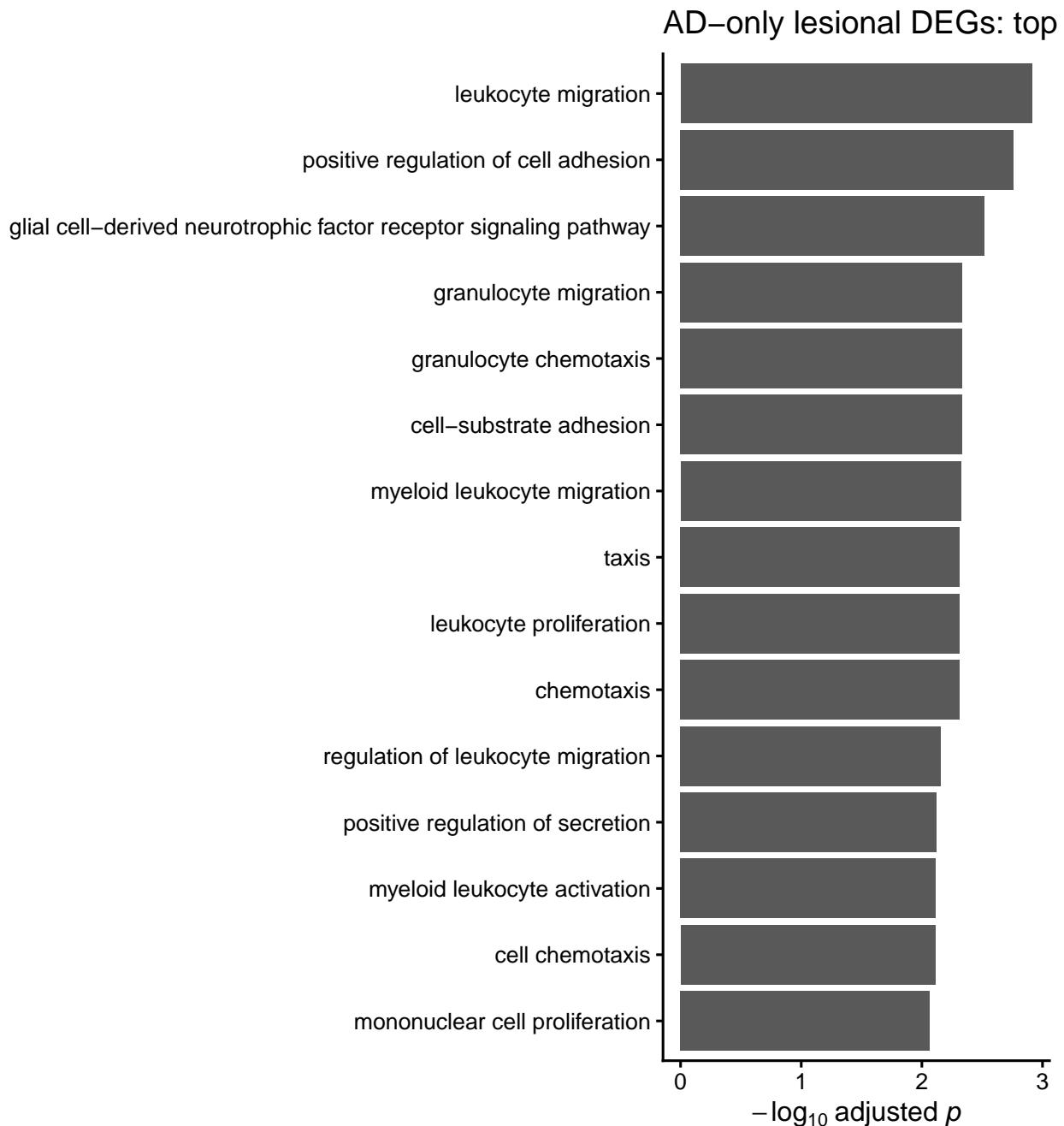


```
go_pso_only <- run_go_enrich(pso_only_lesion, universe_genes)
plot_go_bar(go_pso_only, "PSO-only lesional DEGs: top GO BP terms")
```

### PSO-only lesional DEGs:



```
go_ad_only <- run_go_enrich(ad_only_lesion, universe_genes)
plot_go_bar(go_ad_only, "AD-only lesional DEGs: top GO BP terms")
```



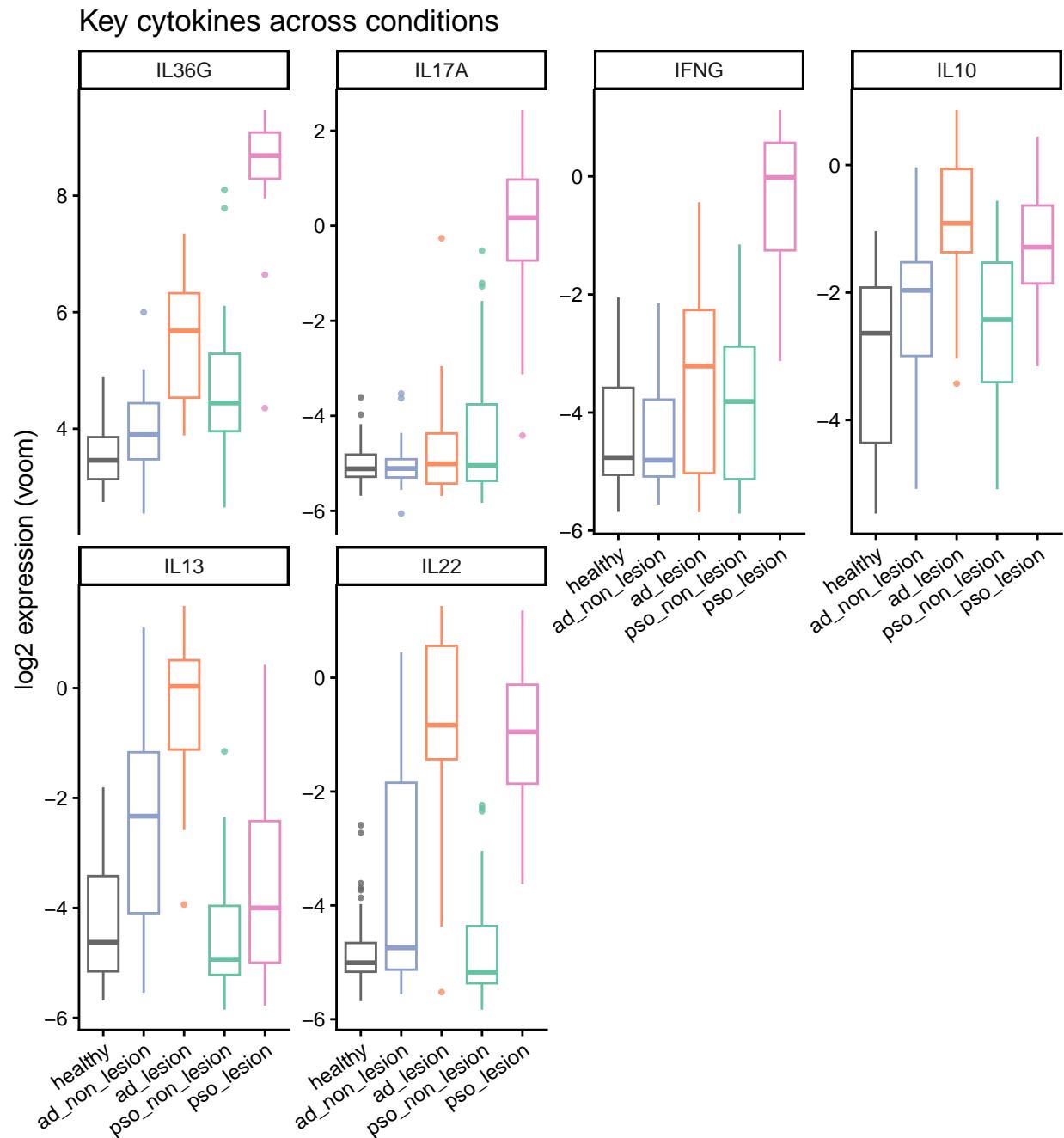
These enrichment plots summarize which biological processes are overrepresented in three DEG sets: genes shared by both diseases, genes unique to psoriasis lesions, and genes unique to AD lesions. The shared set is dominated by immune and inflammatory terms, psoriasis-only genes strongly point to cell cycle and mitosis (consistent with hyperproliferation), and AD-only genes lean more toward immune regulation/signaling and barrier-related biology. This matches the paper's main themes, but we can't reproduce the exact annotation labels because the study uses a broader function/pathway annotation framework than plain GO, and we're limited here to GO-based enrichment (without the same proprietary/curated mapping the authors used).

---

Figure 3 — Cytokines (paper Fig 3)

```
# Match the paper's Fig 3b boxplot set (6 genes)
cytokine_key <- c("IL36G", "IL17A", "IFNG", "IL10", "IL13", "IL22")

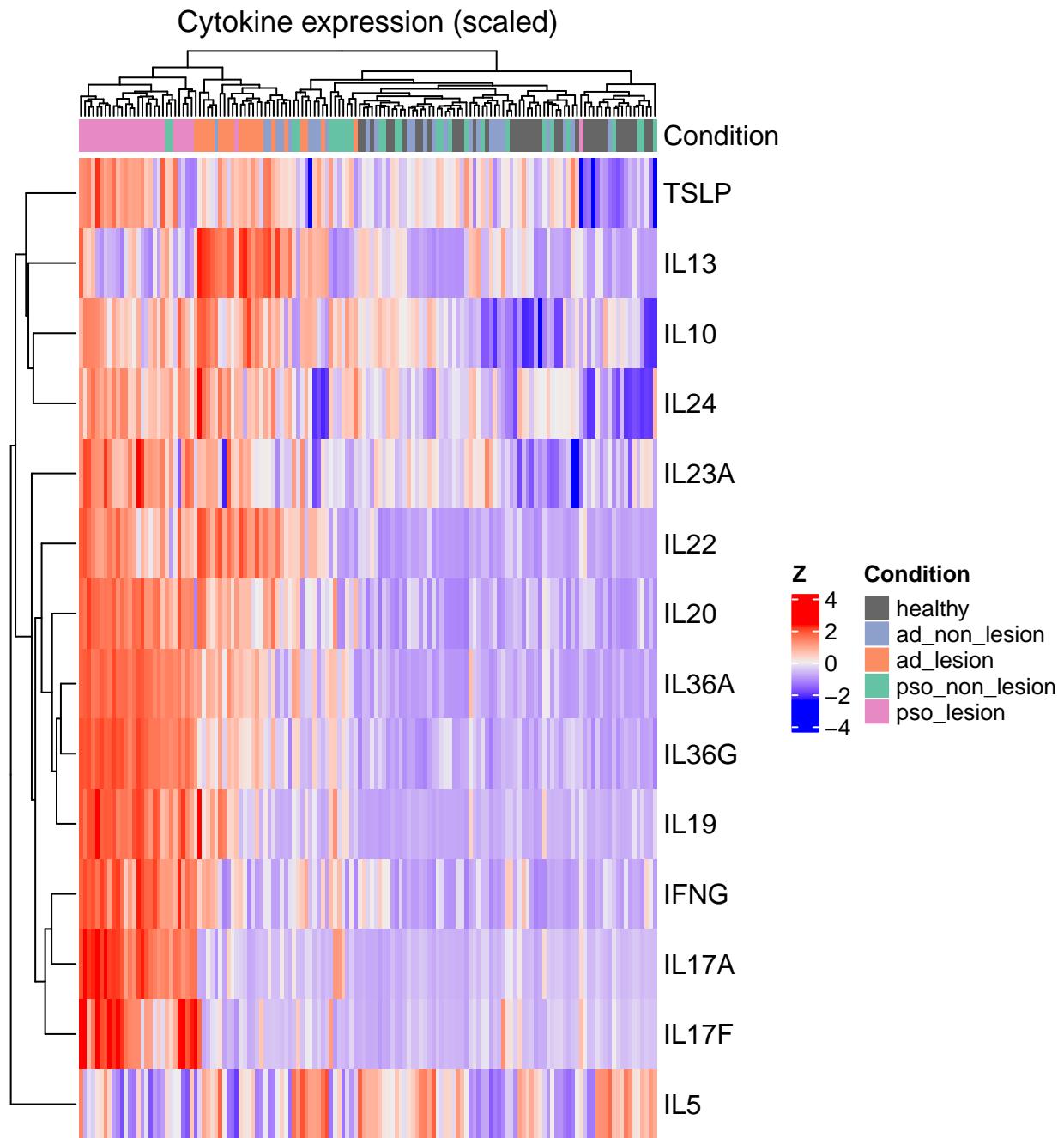
present_key <- intersect(cytokine_key, rownames(logexpr))
plot_gene_boxplot(logexpr, present_key, meta, "Key cytokines across conditions")
```



```
cytokines_heat <- unique(c(cytokine_key, "IL17F", "IL36A", "IL4", "IL5",
                           "IL19", "IL20", "IL24", "IL23A", "TSLP"))
present_heat <- intersect(cytokines_heat, rownames(logexpr))
mat <- logexpr[present_heat, meta$sample, drop = FALSE]
matz <- t(scale(t(mat)))

ha <- HeatmapAnnotation(Condition = meta$condition,
                        col = list(Condition = cond_cols))

Heatmap(matz, name = "Z", top_annotation = ha,
        show_column_names = FALSE, column_title = "Cytokine expression (scaled)")
```



Overall, the direction matches the paper: psoriasis is driven more by IL-17/IL-36 signals, and AD by IL-13. Both lesions tend to cluster together, while non-lesions cluster more with the healthy samples.

---

**Figure 4 — H3K27ac cell-type enrichment**

```

enhancer_gene_sets <- readRDS("celltype_h3k27ac_gene_sets.rds")

# Recompute AD vs PSO (lesional) DEGs for directional sets
cont_ad_vs_pso <- makeContrasts(ad_vs_pso_lesion = ad_lesion-pso_lesion, levels = design)
fit_ad_vs_pso_full <- eBayes(contrasts.fit(fit, cont_ad_vs_pso))
res_ad_vs_pso_lesion_full <- topTable(fit_ad_vs_pso_full, coef = "ad_vs_pso_lesion", number = Inf) %>%
  rownames_to_column("gene") %>%
  as_tibble() %>%
  mark_deg()

ad_higher <- res_ad_vs_pso_lesion_full %>% filter(DEG, logFC > 0) %>% pull(gene)
pso_higher <- res_ad_vs_pso_lesion_full %>% filter(DEG, logFC < 0) %>% pull(gene)

deg_sets <- list(
  "Common lesional" = common_lesion,
  "AD-only lesional" = ad_only_lesion,
  "PSO-only lesional" = pso_only_lesion,
  "AD > PSO" = ad_higher,
  "PSO > AD" = pso_higher
)

enrich_tbl <- purrr::map_dfr(names(enhancer_gene_sets), function(ct) {
  purrr::map_dfr(names(deg_sets), function(comp) {
    res <- fisher_enrichment(deg_sets[[comp]], enhancer_gene_sets[[ct]], universe_genes)
    tibble(cell_type = ct, comparison = comp, p_value = res$p_value)
  })
}) %>%
  mutate(p_adj = p.adjust(p_value, method = "BH"),
        neglog10_p = -log10(p_value))

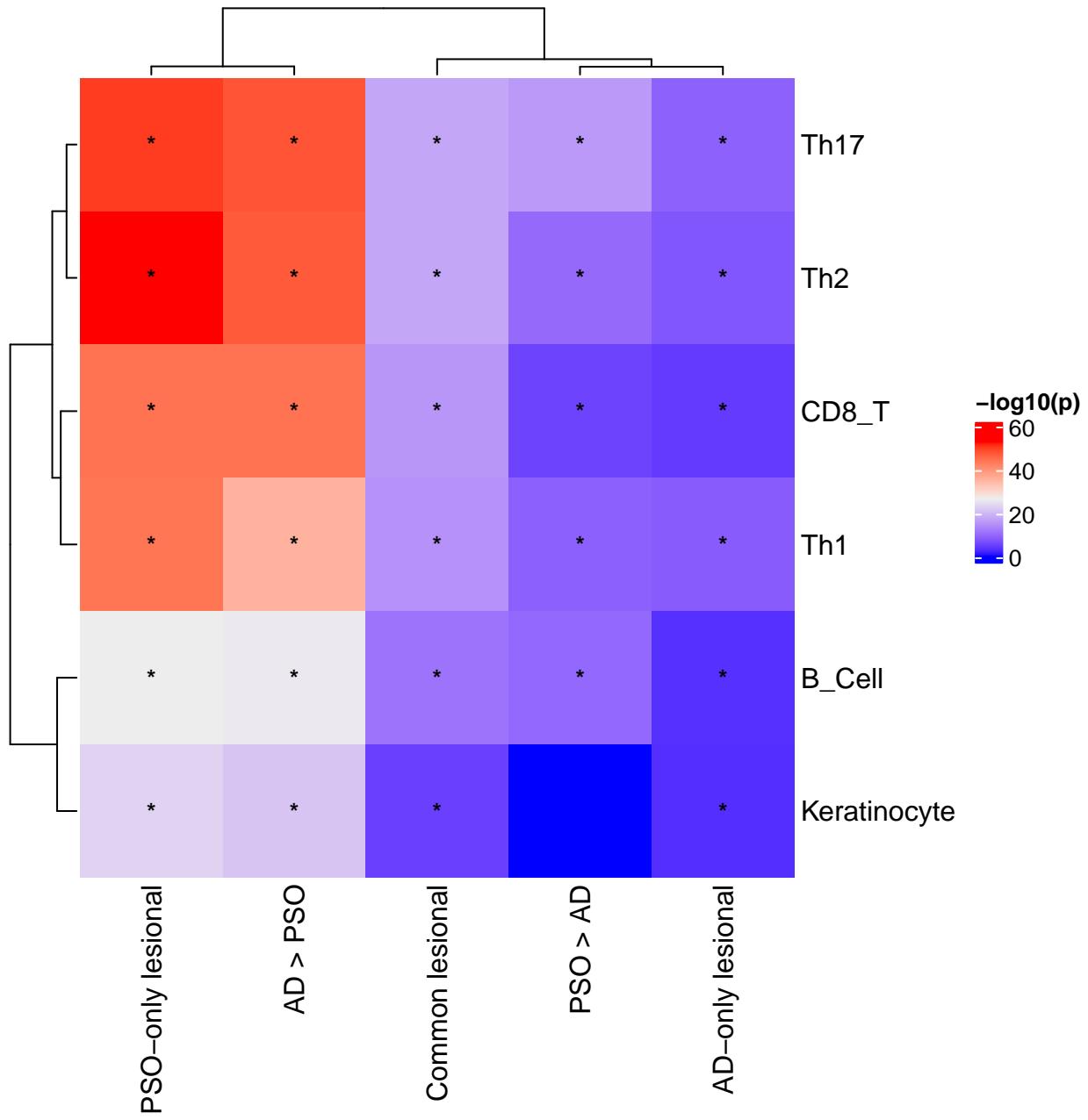
mat_p <- enrich_tbl %>%
  select(cell_type, comparison, neglog10_p) %>%
  pivot_wider(names_from = comparison, values_from = neglog10_p) %>%
  column_to_rownames("cell_type") %>%
  as.matrix()
mat_p[is.na(mat_p)] <- 0

mat_fdr <- enrich_tbl %>%
  select(cell_type, comparison, p_adj) %>%
  pivot_wider(names_from = comparison, values_from = p_adj) %>%
  column_to_rownames("cell_type") %>%
  as.matrix()
mat_fdr[is.na(mat_fdr)] <- 1

Heatmap(mat_p,
  name = "-log10(p)",
  column_title = "H3K27ac enrichment (paper Fig 4a structure)",
  cell_fun = function(j, i, x, y, width, height, fill) {
    if (mat_fdr[i, j] <= 0.05) grid.text("*", x, y, gp = gpar(fontsize = 10, fontface = "bold"))
  }
)

```

### H3K27ac enrichment (paper Fig 4a structure)



This panel asks a simple question: do the genes that change in AD/psoriasis sit near enhancer regions (H3K27ac) that are typical for certain cell types? If yes, that hints at which cell types are most “driving” each DEG set. Here, the shared lesional DEGs tend to light up across multiple immune-related signatures, while the AD-only vs PSO-only sets (and the AD>PSO vs PSO>AD sets) show different skews, which is the same high-level interpretation as the paper. The exact cell-type names and which ones pop the strongest can differ because our H3K27ac gene sets come from the specific `celltype_h3k27ac_gene_sets.rds` file we are using, not necessarily the exact same reference the authors used.

We load the cell-type enhancer gene sets, then recompute the AD vs PSO lesional differential expression so we can split DEGs into genes higher in AD ( $\text{logFC} > 0$ ) versus higher in PSO ( $\text{logFC} < 0$ ). Next we define five DEG lists (common, AD-only, PSO-only, AD>PSO, PSO>AD).

---

## Conclusion

Overall, we end up with the same takeaway as the paper: psoriasis lesions show stronger gene-expression changes than AD, and AD and psoriasis share a big chunk of the same inflammation signal. So yes, the authors' main conclusion holds up. The exact numbers don't match perfectly because tiny choices like gene filtering can make "borderline" genes drop in or out of the DEG lists. We also didn't redo every figure because a few panels depend on extra resources or author-specific annotation/classifier setups that aren't fully reproducible from the public files alone.

## References

Tsoi LC, Rodriguez E, Degenhardt F, et al. Atopic Dermatitis Is an IL-13–Dominant Disease with Greater Molecular Heterogeneity Compared to Psoriasis. *Journal of Investigative Dermatology* (2019). doi:10.1016/j.jid.2018.12.018.

## Session info

```
sessionInfo()

## R version 4.5.1 (2025-06-13)
## Platform: aarch64-apple-darwin20
## Running under: macOS Tahoe 26.2
##
## Matrix products: default
## BLAS:    /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/lib/libRblas.0.dylib
## LAPACK:  /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/lib/libRlapack.dylib;  LAPACK v...
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: Europe/Amsterdam
## tzcode source: internal
##
## attached base packages:
## [1] stats4      grid        stats       graphics   grDevices  utils      datasets
## [8] methods     base
##
## other attached packages:
## [1] scatterplot3d_0.3-44  pROC_1.19.0.1      ggVennDiagram_1.5.4
## [4] clusterProfiler_4.16.0 org.Hs.eg.db_3.21.0  AnnotationDbi_1.70.0
## [7] IRanges_2.42.0        S4Vectors_0.46.0   Biobase_2.68.0
## [10] BiocGenerics_0.54.1   generics_0.1.4    ggrepel_0.9.6
## [13] circlize_0.4.16      ComplexHeatmap_2.24.1 edgeR_4.6.3
## [16] limma_3.64.3         lubridate_1.9.4   forcats_1.0.1
## [19] stringr_1.6.0        dplyr_1.1.4       purrr_1.2.0
## [22] readr_2.1.5          tidyverse_1.3.1    tibble_3.3.0
```

```

## [25] ggplot2_4.0.0           tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##   [1] DBI_1.2.3                  gson_0.1.0
##   [4] magrittr_2.0.4             clue_0.3-66
##   [7] DOSE_4.2.0                 matrixStats_1.5.0
##  [10] RSQLite_2.4.4              reshape2_1.4.5
##  [13] vctrs_0.6.5                pkgconfig_2.0.3
##  [16] crayon_1.5.3              fastmap_1.2.0
##  [19] labeling_0.4.3             rmarkdown_2.30
##  [22] tzdb_0.5.0                UCSC.utils_1.4.0
##  [25] xfun_0.54                 cachem_1.1.0
##  [28] GenomeInfoDb_1.44.3       jsonlite_2.0.0
##  [31] BiocParallel_1.42.2       parallel_4.5.1
##  [34] R6_2.6.1                  stringi_1.8.7
##  [37] GOSemSim_2.34.0          Rcpp_1.1.0
##  [40] knitr_1.50                 ggtangle_0.0.8
##  [43] igraph_2.2.1              splines_4.5.1
##  [46] timechange_0.3.0          tidyselect_1.2.1
##  [49] rstudioapi_0.17.1        yaml_2.3.10
##  [52] codetools_0.2-20          plyr_1.8.9
##  [55] treeio_1.32.0             withr_3.0.2
##  [58] S7_0.2.0                  evaluate_1.0.5
##  [61] Biostrings_2.76.0          ggtree_3.16.3
##  [64] foreach_1.5.2              ggfun_0.2.0
##  [67] tidytree_0.4.6             scales_1.4.0
##  [70] lazyeval_0.2.2            tools_4.5.1
##  [73] fgsea_1.34.2              locfit_1.5-9.12
##  [76] fastmatch_1.1-6            cowplot_1.2.0
##  [79] colorspace_2.1-2           nlme_3.1-168
##  [82] GenomeInfoDbData_1.2.14    cli_3.6.5
##  [85] gtable_0.3.6               R.methodsS3_1.8.2
##  [88] digest_0.6.38              ggplotify_0.1.3
##  [91] farver_2.1.2               memoise_2.0.1
##  [94] R.oo_1.27.1                lifecycle_1.0.4
##  [97] GlobalOptions_0.1.2        GO.db_3.21.0
## [100] bit64_4.6.0-1

```