

Comparaciones estadísticas

Elena Eyre Sánchez, PhD

2025-01-01

Contents

1	Introducción y Objetivo	1
2	Paquetes y datos	1
2.1	Metadata original	1
2.2	Metadata harmonizada	2
3	Comparación estadística	2
3.1	GSE22155_GPL6947	2
3.2	GSE22155_GPL6102	4
3.3	GSE35640	6
3.4	GSE61992	8
3.5	GSE91061	10
3.6	TCGA-SKCM	12
3.7	Todas las cohortes	14
4	Heatmaps	17
5	scRNA-seq vs deconvoluted bulk RNA-seq	17
5.1	GSE120575	17
5.2	GSE91061	19
6	DEGs	22

1 Introducción y Objetivo

2 Paquetes y datos

2.1 Metadata original

Para asegurar que examinamos todas las opciones, examinamos las variables originales.

Los datos a comparar son los resultados obtenidos de los análisis previos de deconvolución:

- GSE22155 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>) con dos plataformas: GPL6102 y GPL6947.
- GSE35640 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35640>)
- GSE61992 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE61992>)
- GSE91061 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE91061>)
- TCGA-SKCM (<https://portal.gdc.cancer.gov/projects/TCGA-SKCM>)

De manera excepcional, tenemos GSE120575 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120575>), que consiste en un dataset scRNA-seq en el que han analizado muestras tumorales y controles.

```
setwd("~/Desktop/ELENA_UOC/TFM")
GSE22155_GPL6102 <- read_csv("Deconvolution_analysis/Quantiseq_analyses/quantiseqr_GSE22155_GPL6102.csv")
GSE22155_GPL6947 <- read_csv("Deconvolution_analysis/Quantiseq_analyses/quantiseqr_GSE22155_GPL6947.csv")
GSE35640 <- read_csv("Deconvolution_analysis/Quantiseq_analyses/quantiseq_GSE35640.csv")
GSE61992 <- read_csv("Deconvolution_analysis/Quantiseq_analyses/quantiseq_GSE61992.csv")
#GSE91061 <- read_csv("Deconvolution_analysis/Quantiseq_analyses/quantiseq_GSE91061.csv")
GSE91061 <- read_delim("Deconvolution_analysis/Quantiseq_analyses/quantiseqr_GSE91061.csv",
  delim = ",")
tcga_skcm <- read_csv("Deconvolution_analysis/Quantiseq_analyses/quantiseq_tcga_skcm.csv")
TCGA_SKCM_clinical_data <- readxl::read_excel("Datasets/TCGA_SKCM_clinical_data.xlsx")
```

2.2 Metadata harmonizada

Para poder estudiar si podemos comparar

- Cohorte GSE22155 y plataforma GPL6947:

```
## Warning: Unknown or uninitialised column: `sample_type`.
```

- Cohorte GSE22155 y plataforma GPL6102:

```
## Warning: Unknown or uninitialised column: `sample_type`.
```

- Cohorte GSE35640:
- Cohorte GSE61992:
- Cohorte GSE91061:
- Cohorte TCGA-SKCM:

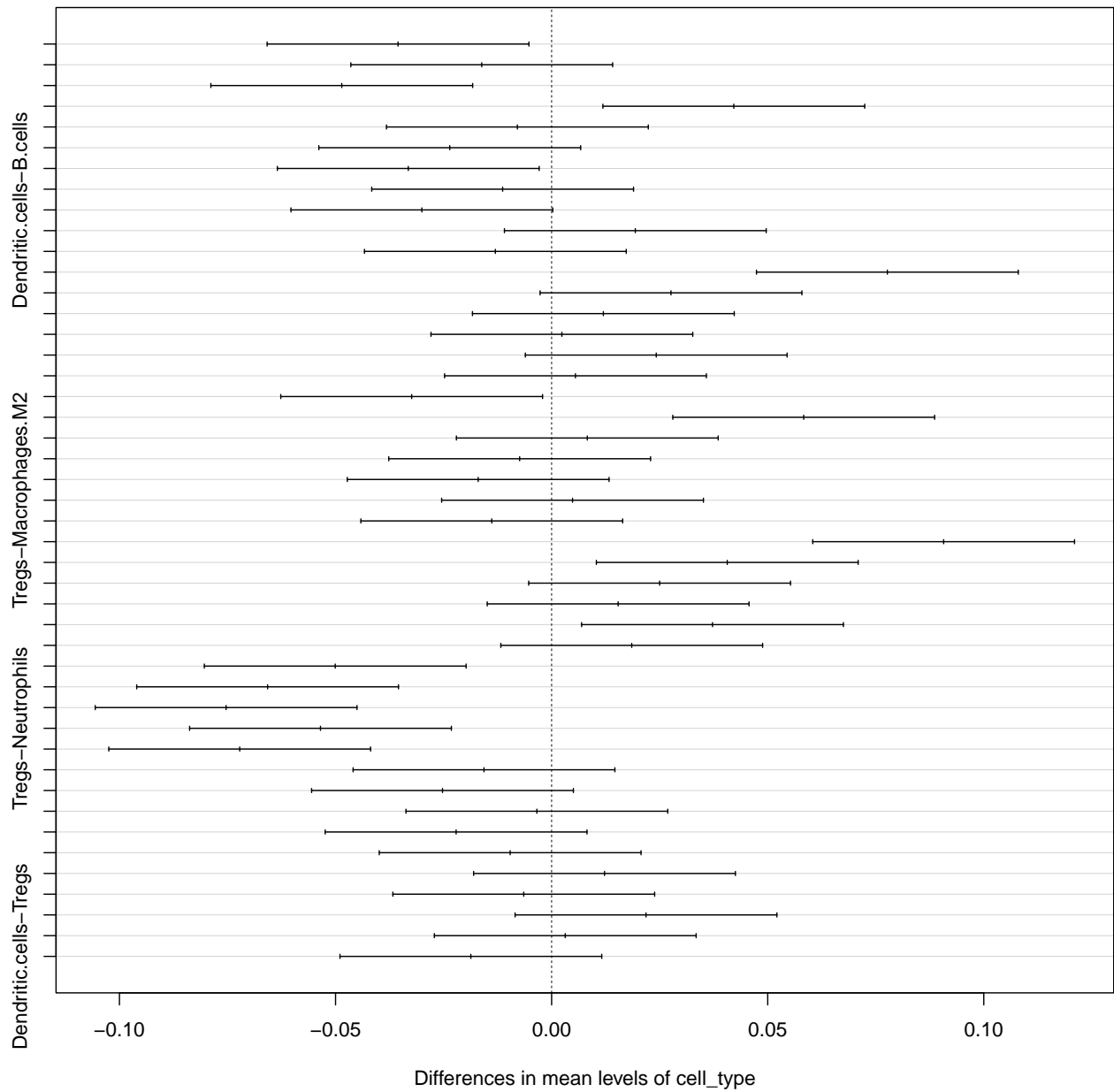
3 Comparación estadística

3.1 GSE22155_GPL6947

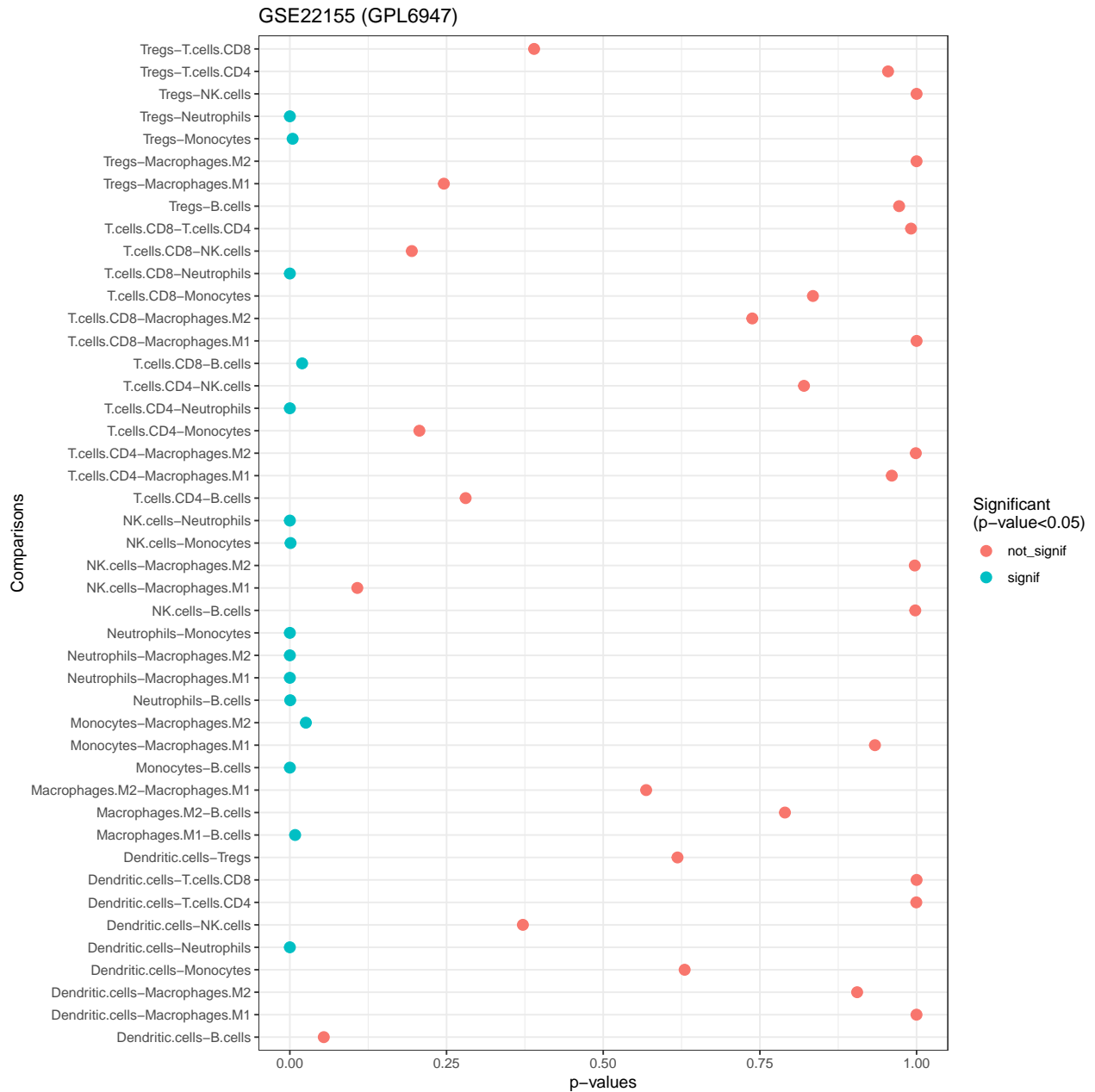
```
## No id variables; using all as measure variables
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## cell_type    9 0.1257  0.013967   14.16 <2e-16 ***
## Residuals  210 0.2072  0.000987
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

95% family-wise confidence level



```
## Warning: Use of `` tk$p adj`` is discouraged.
## i Use `p adj` instead.
```

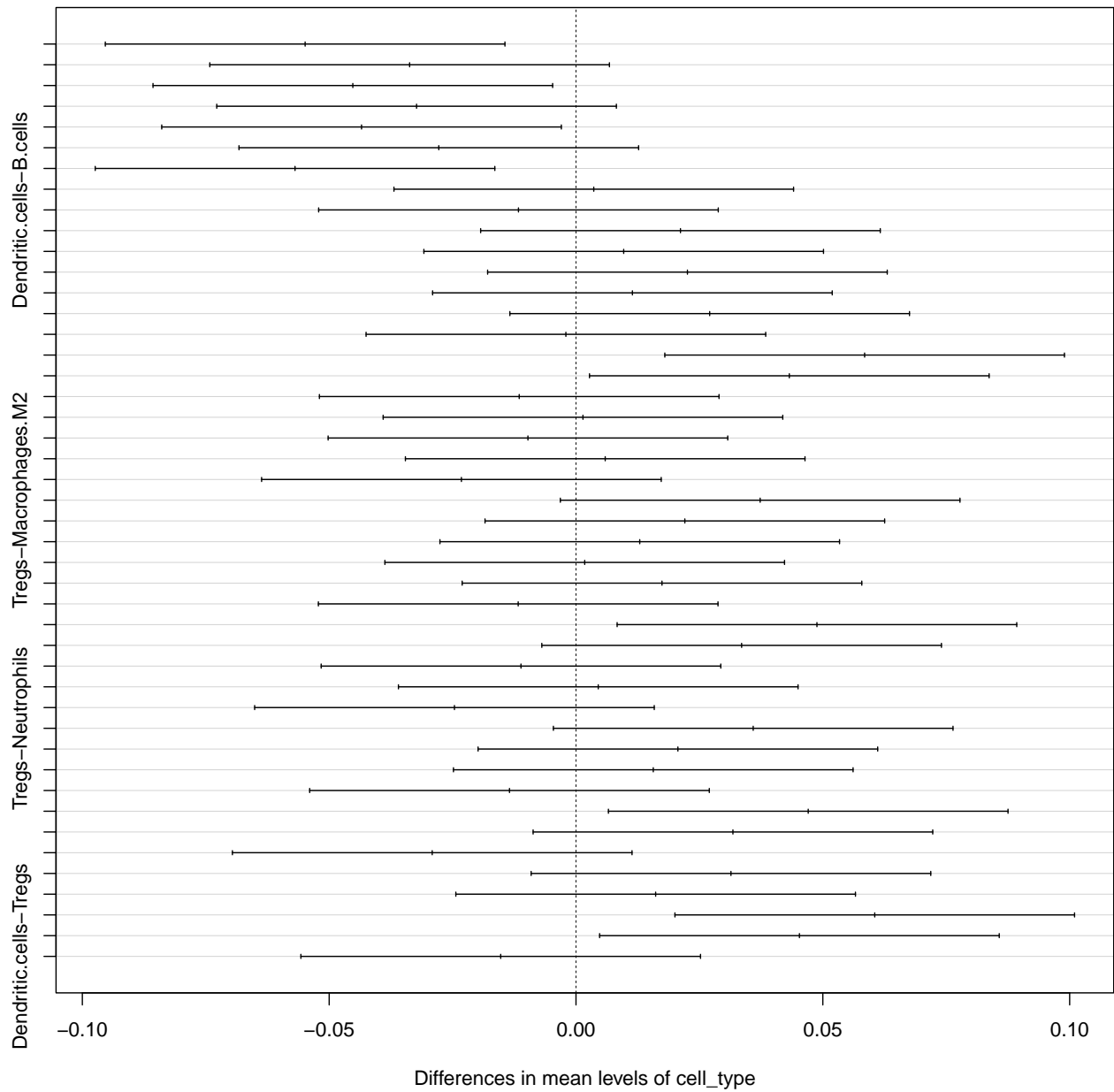


3.2 GSE22155_GPL6102

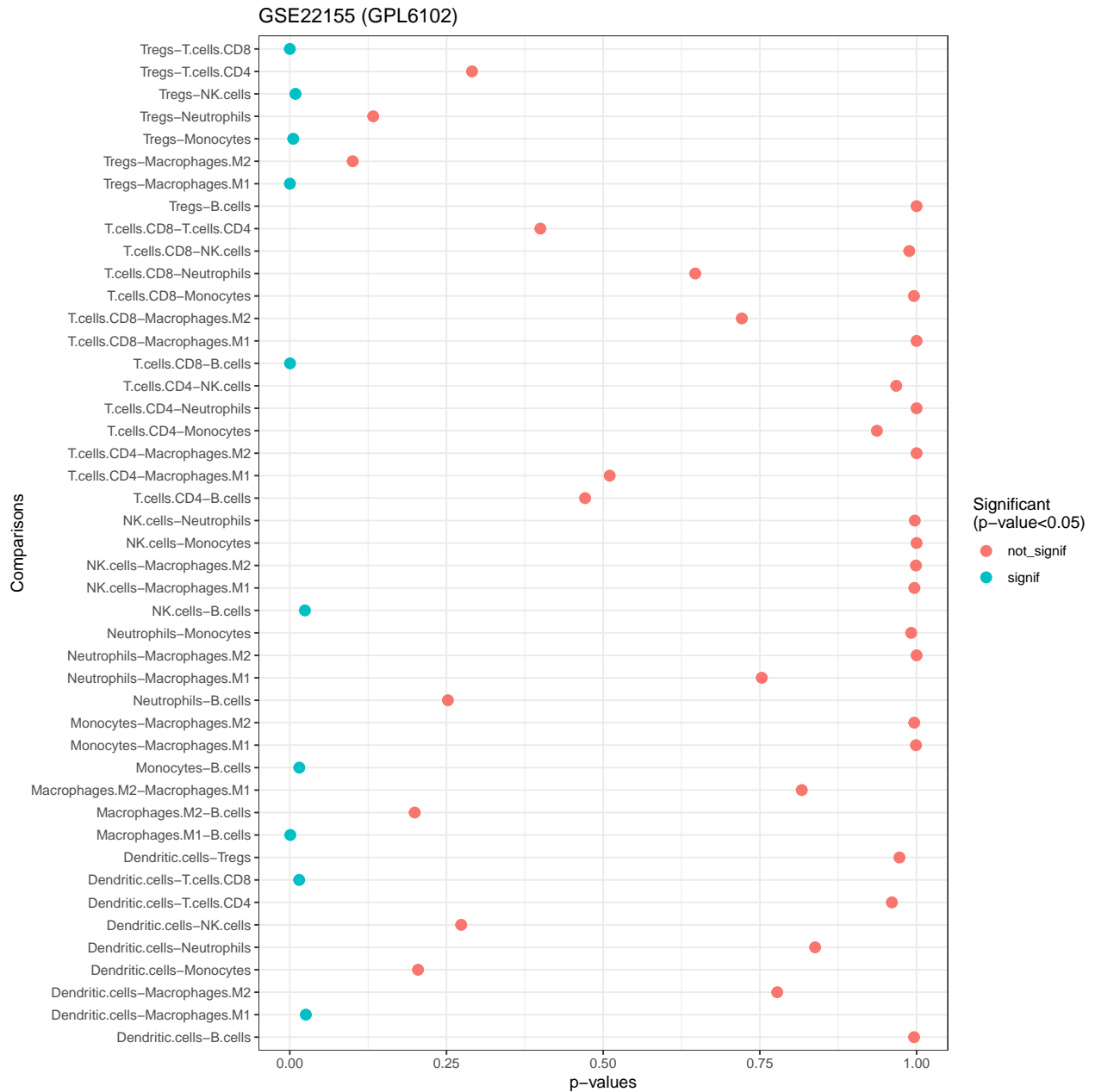
No id variables; using all as measure variables

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## cell_type    9  0.2362  0.026245    5.672 1.55e-07 ***
## Residuals 560  2.5913  0.004627
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

95% family-wise confidence level



```
## Warning: Use of `` tk$p adj`` is discouraged.
## i Use `p adj` instead.
```



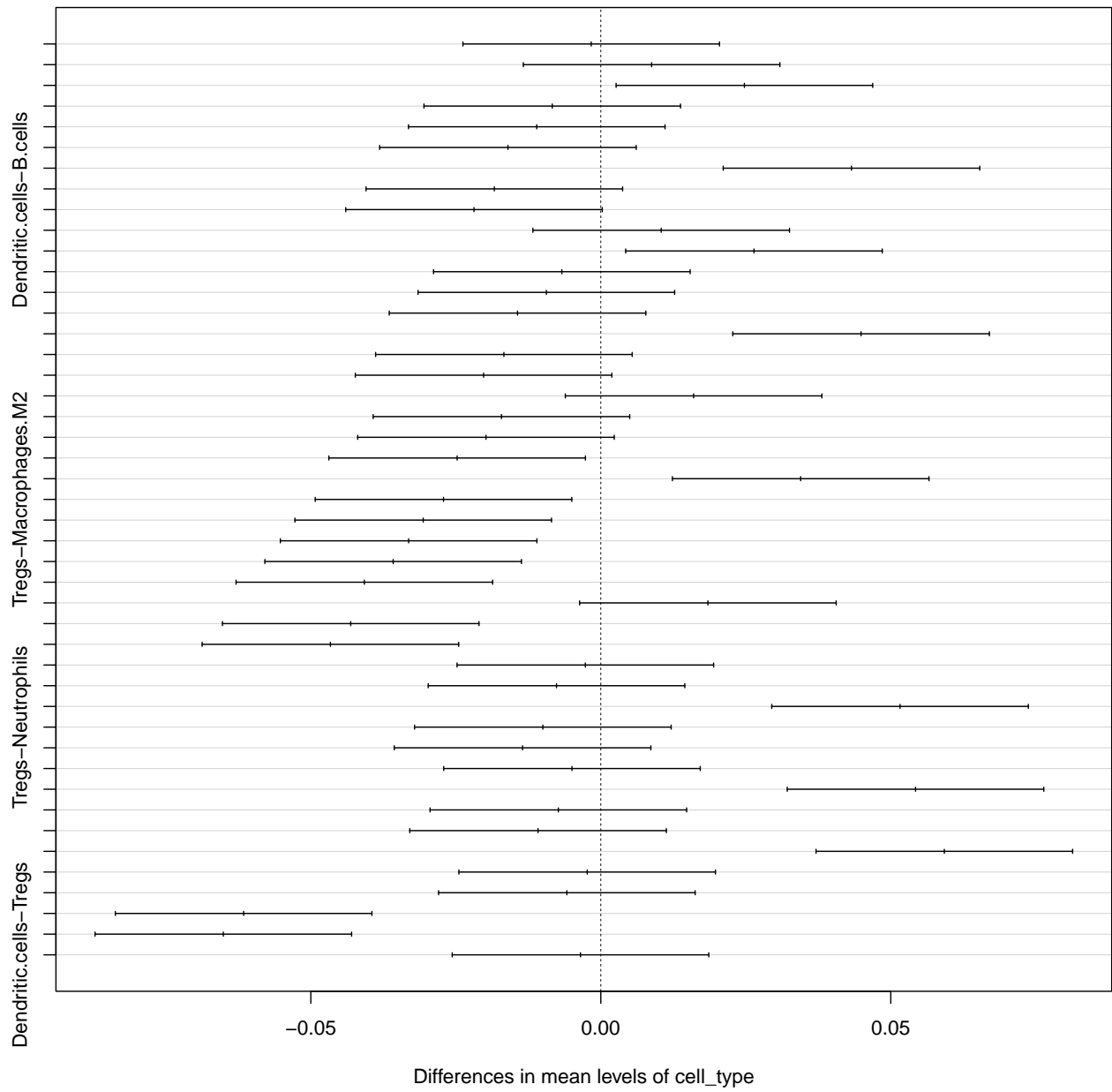
3.3 GSE35640

No id variables; using all as measure variables

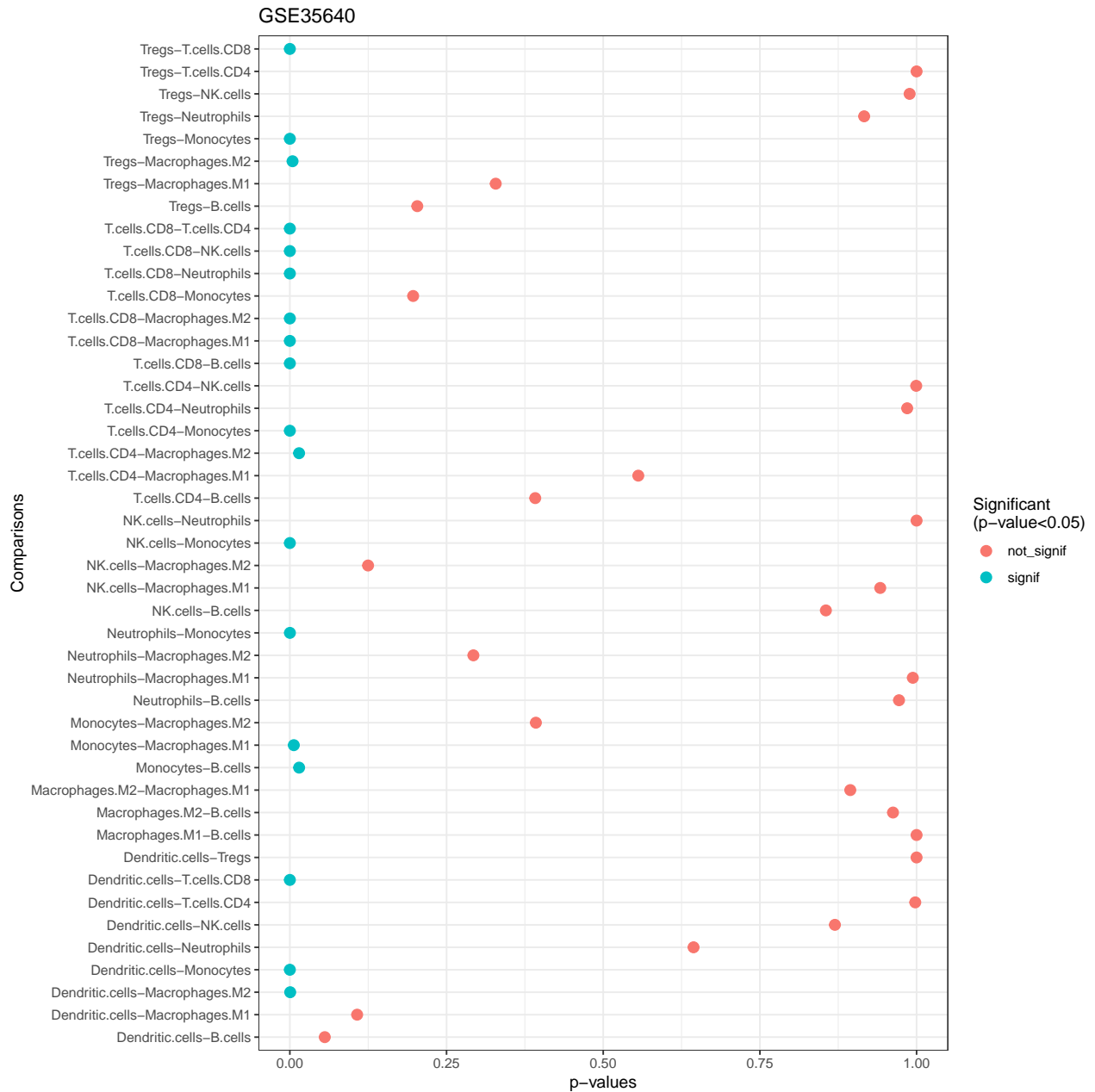
```
##           Df Sum Sq Mean Sq F value Pr(>F)
## cell_type   9  0.2486  0.027621   17.51 <2e-16 ***
## Residuals 640  1.0096  0.001577
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

95% family-wise confidence level



```
## Warning: Use of `` tk$p adj`` is discouraged.
## i Use `p adj` instead.
```

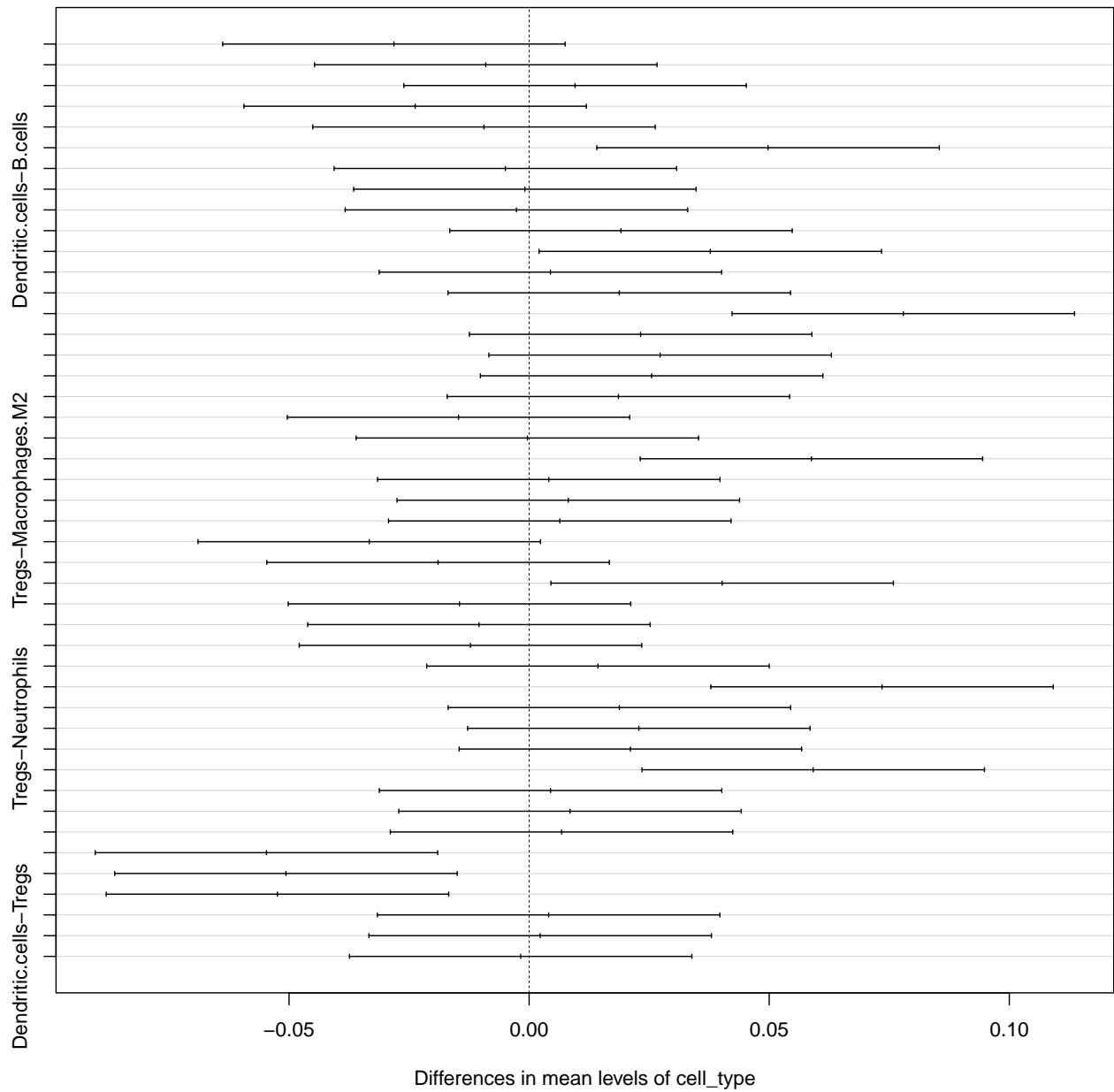


3.4 GSE61992

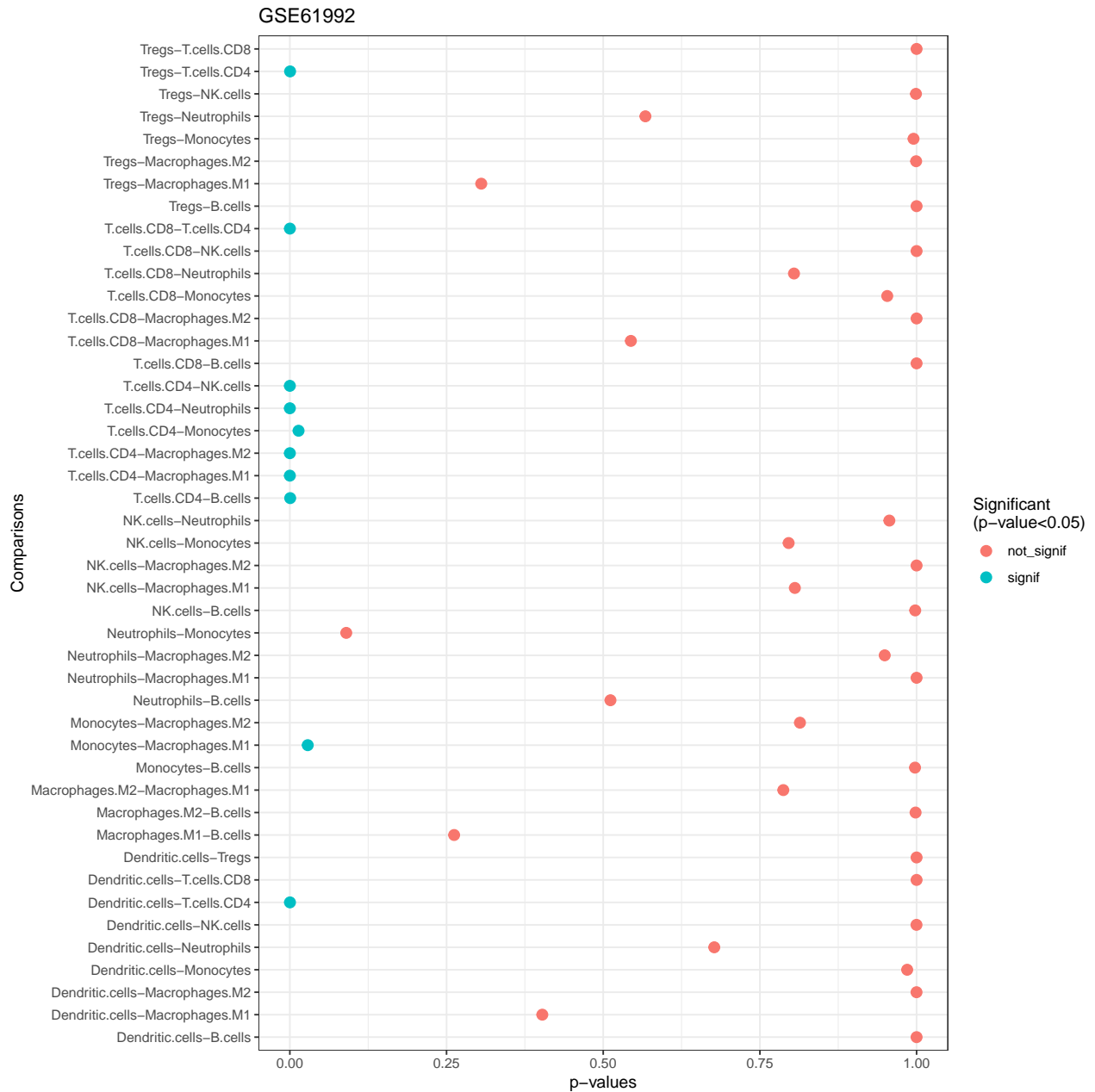
No id variables; using all as measure variables

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## cell_type    9 0.1063  0.011812   7.283 2.08e-09 ***
## Residuals  250 0.4055  0.001622
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


95% family-wise confidence level



```
## Warning: Use of `` tk$p adj`` is discouraged.
## i Use `p adj` instead.
```



3.5 GSE91061

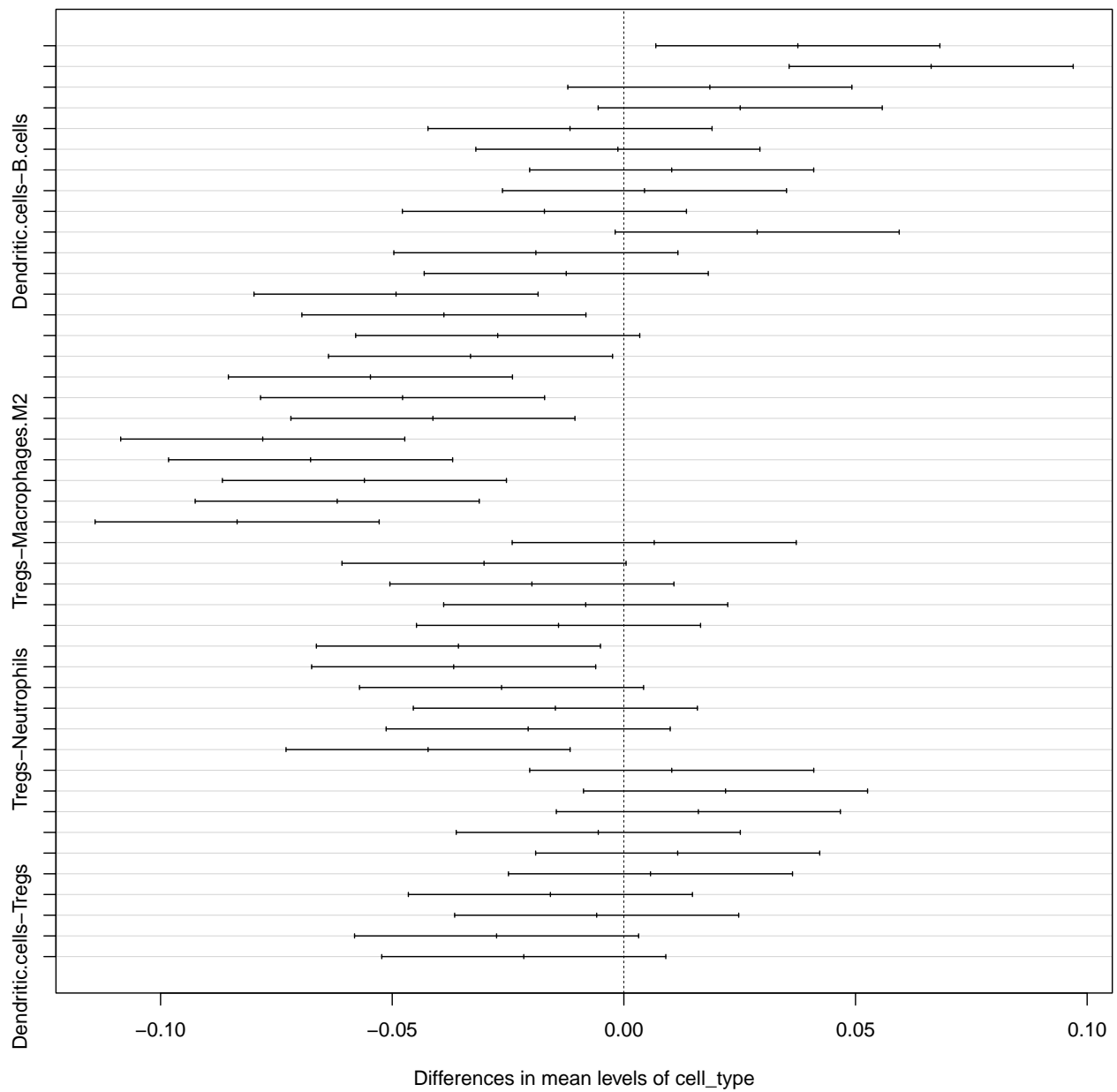
No id variables; using all as measure variables

No id variables; using all as measure variables

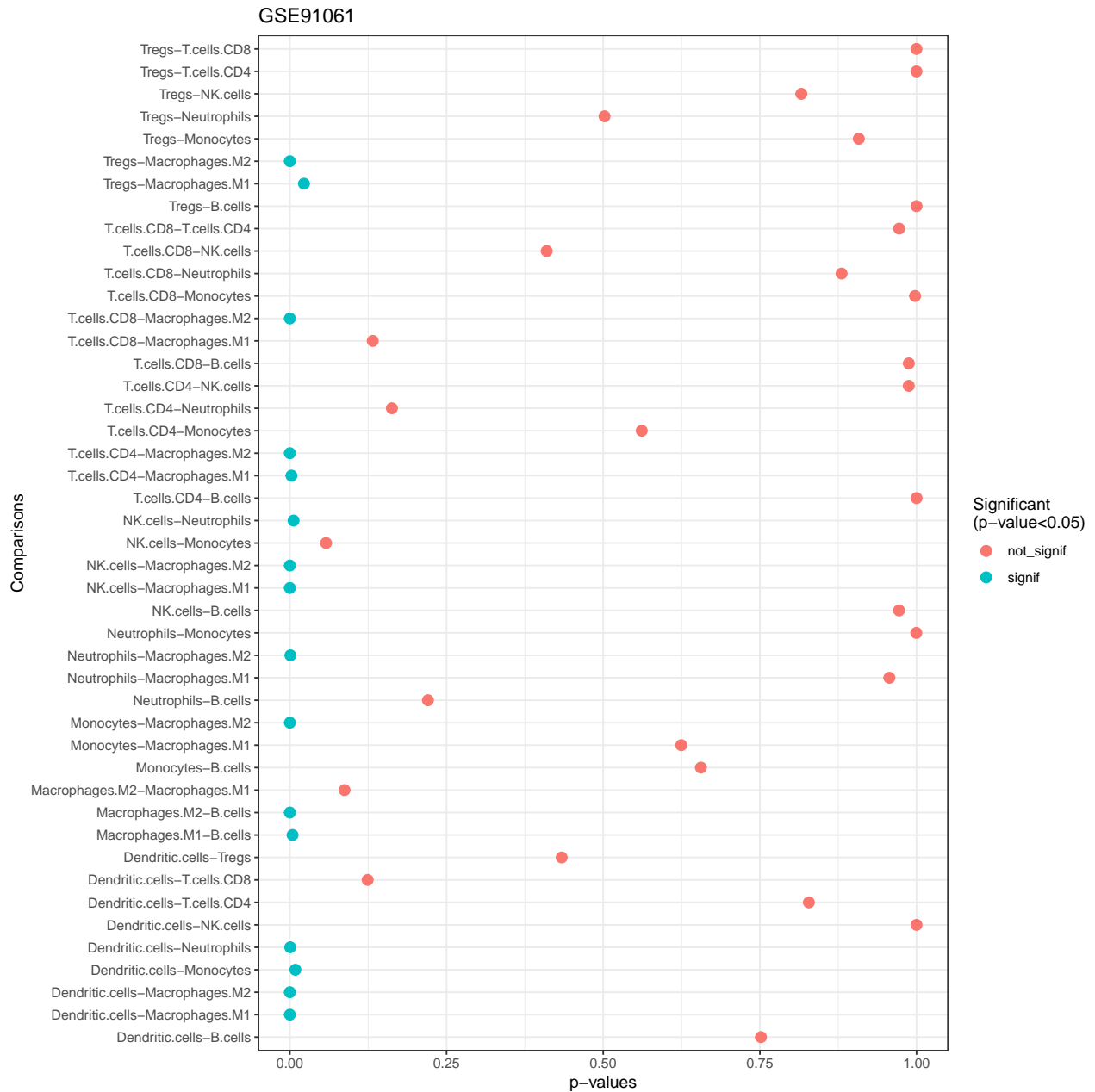
```
##           Df Sum Sq Mean Sq F value Pr(>F)
## cell_type    9  0.610  0.06773   13.29 <2e-16 ***
## Residuals 1080  5.504  0.00510
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

95% family-wise confidence level



```
## Warning: Use of `` tk$p adj`` is discouraged.
## i Use `p adj` instead.
```

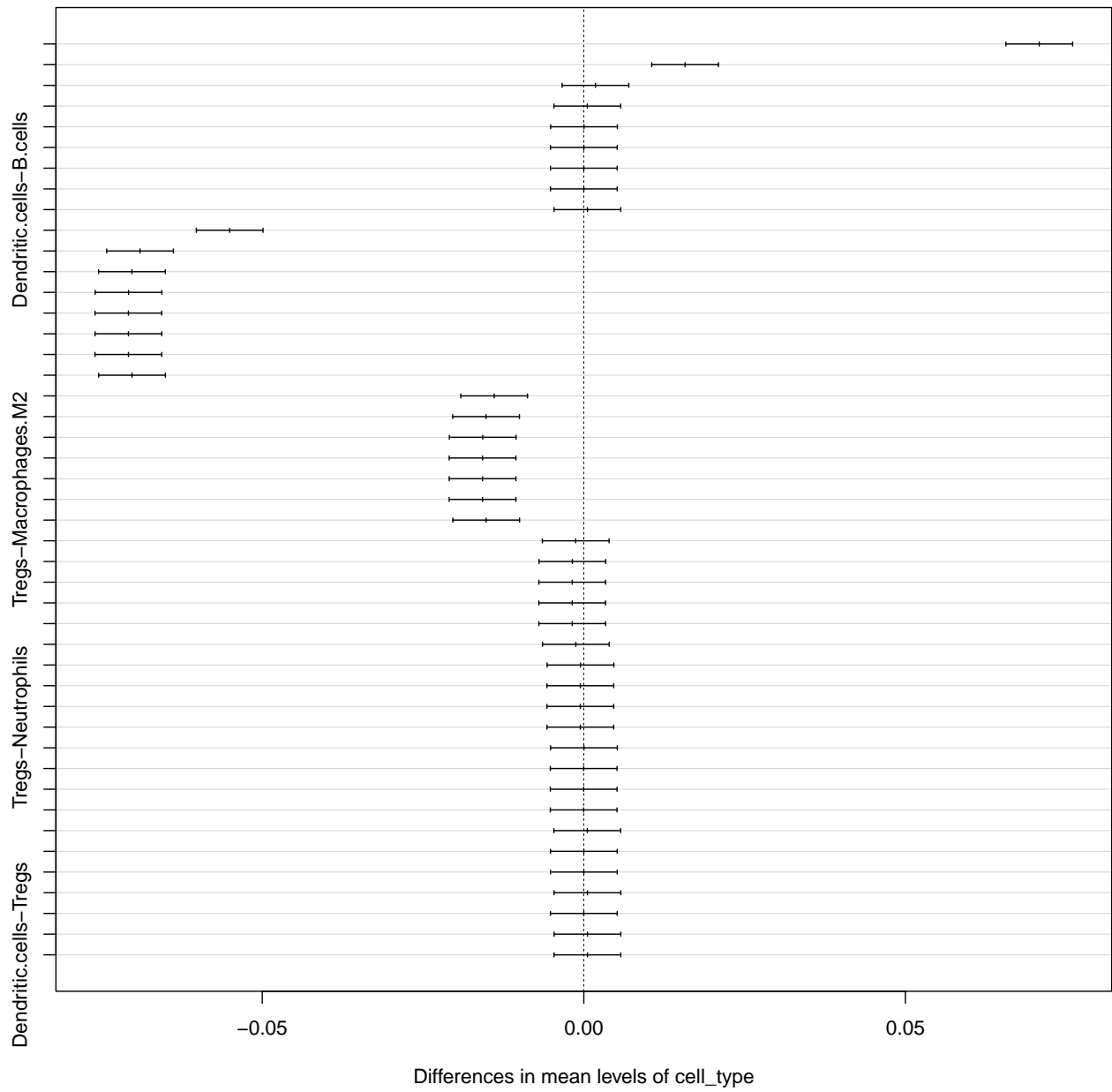


3.6 TCGA-SKCM

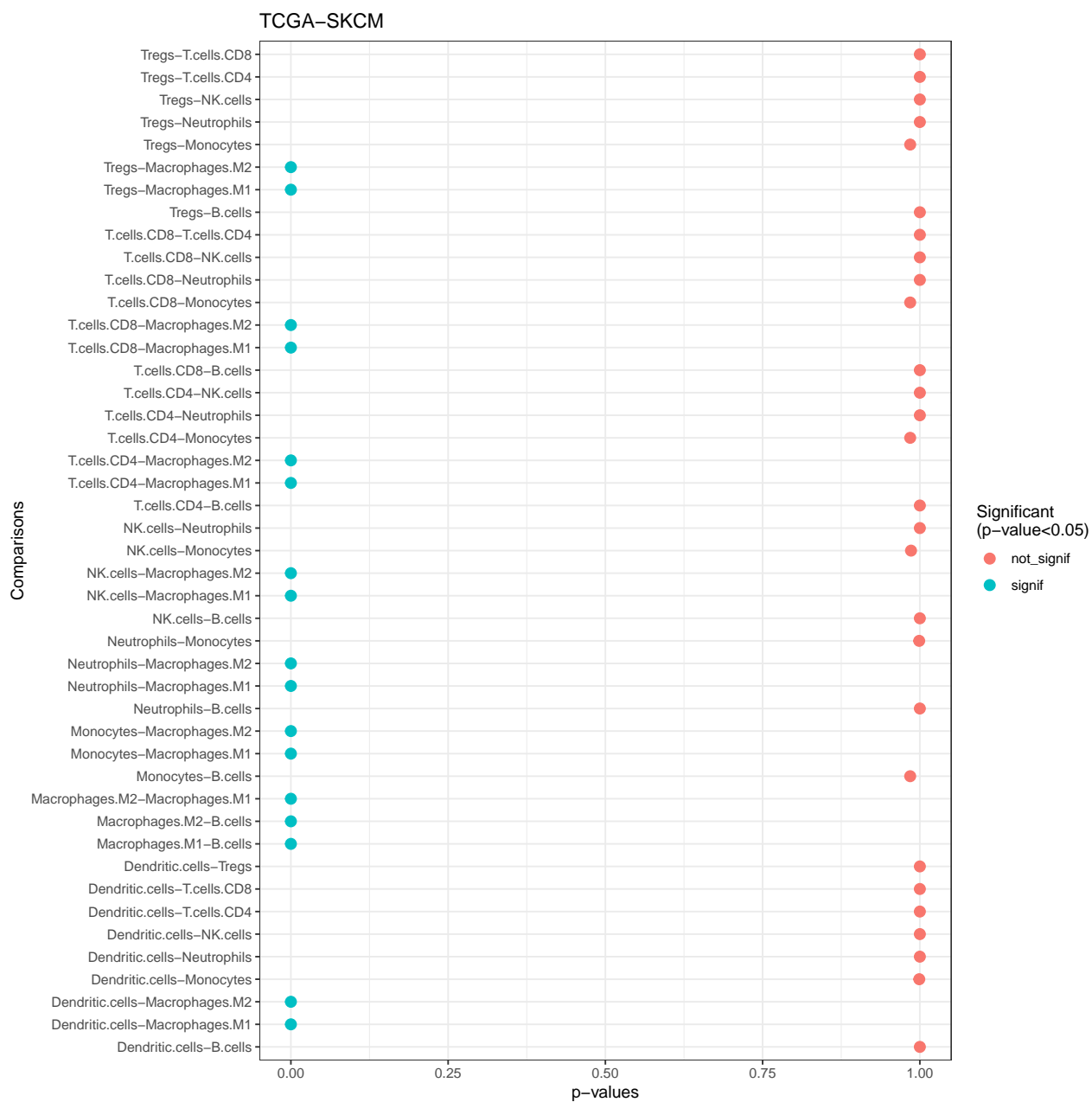
No id variables; using all as measure variables

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## cell_type    9  2.108  0.23420   369.6 <2e-16 ***
## Residuals 4710  2.985  0.00063
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

95% family-wise confidence level

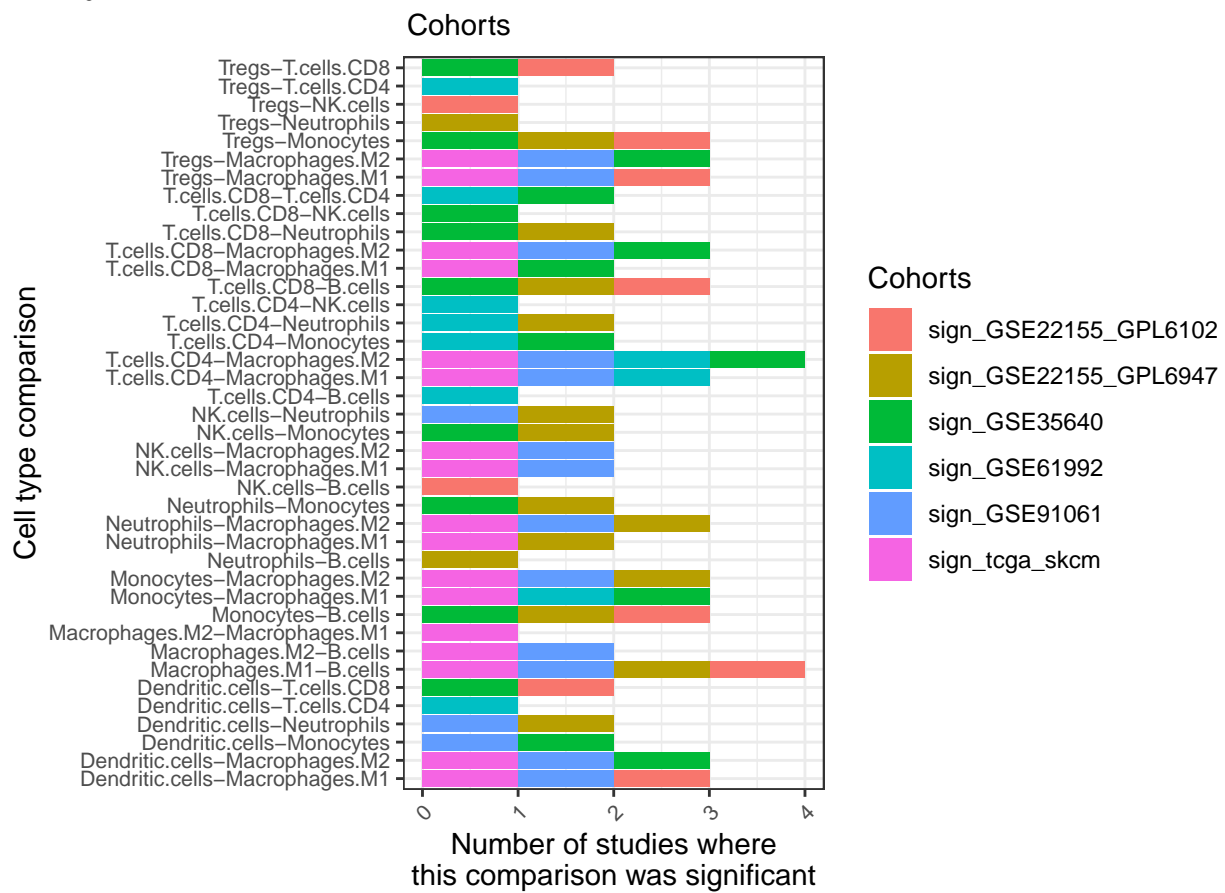
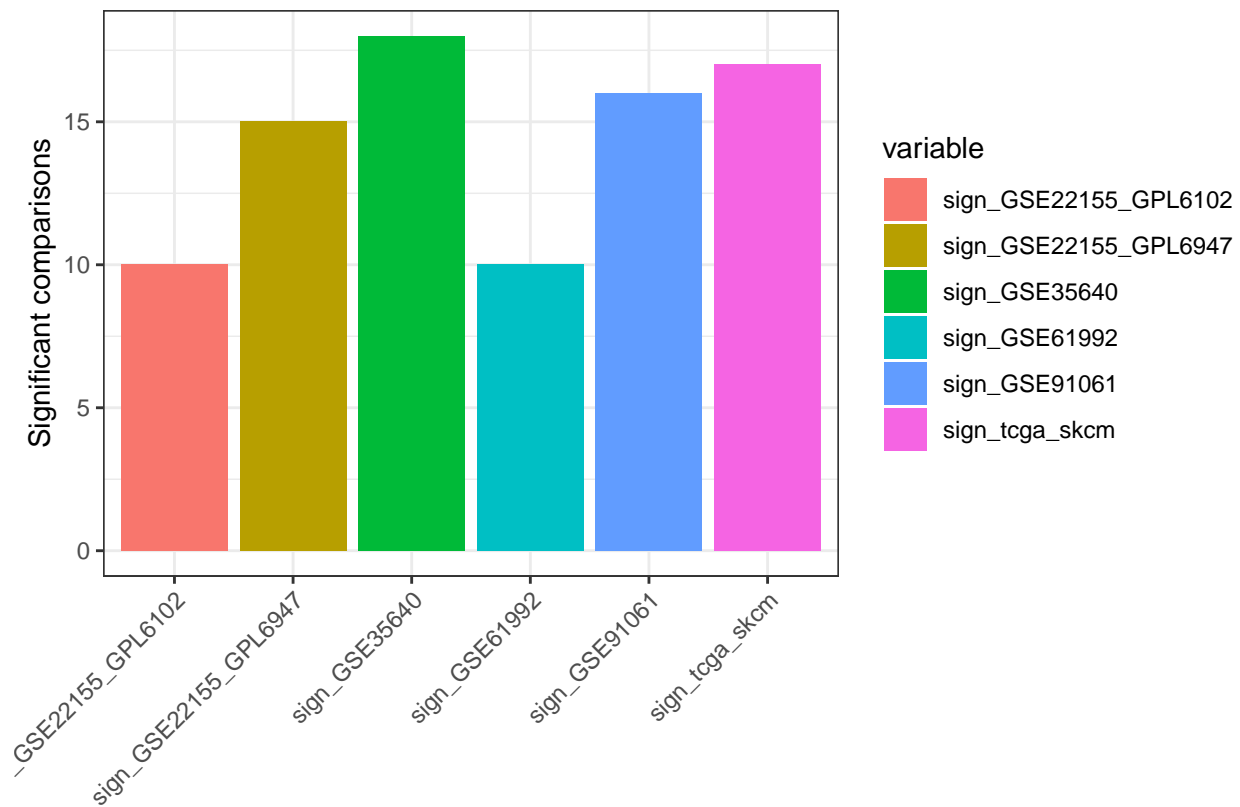


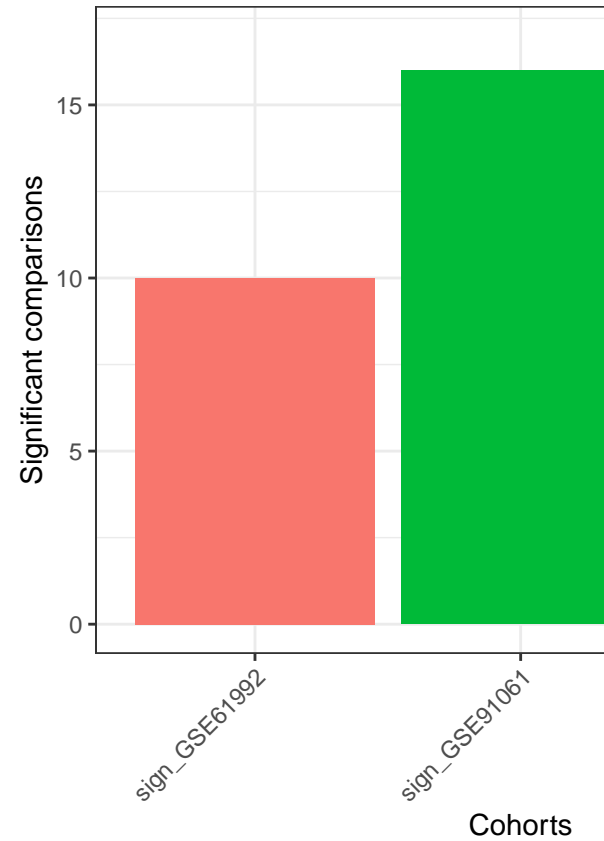
```
## Warning: Use of `` tk$p adj`` is discouraged.
## i Use `p adj` instead.
```



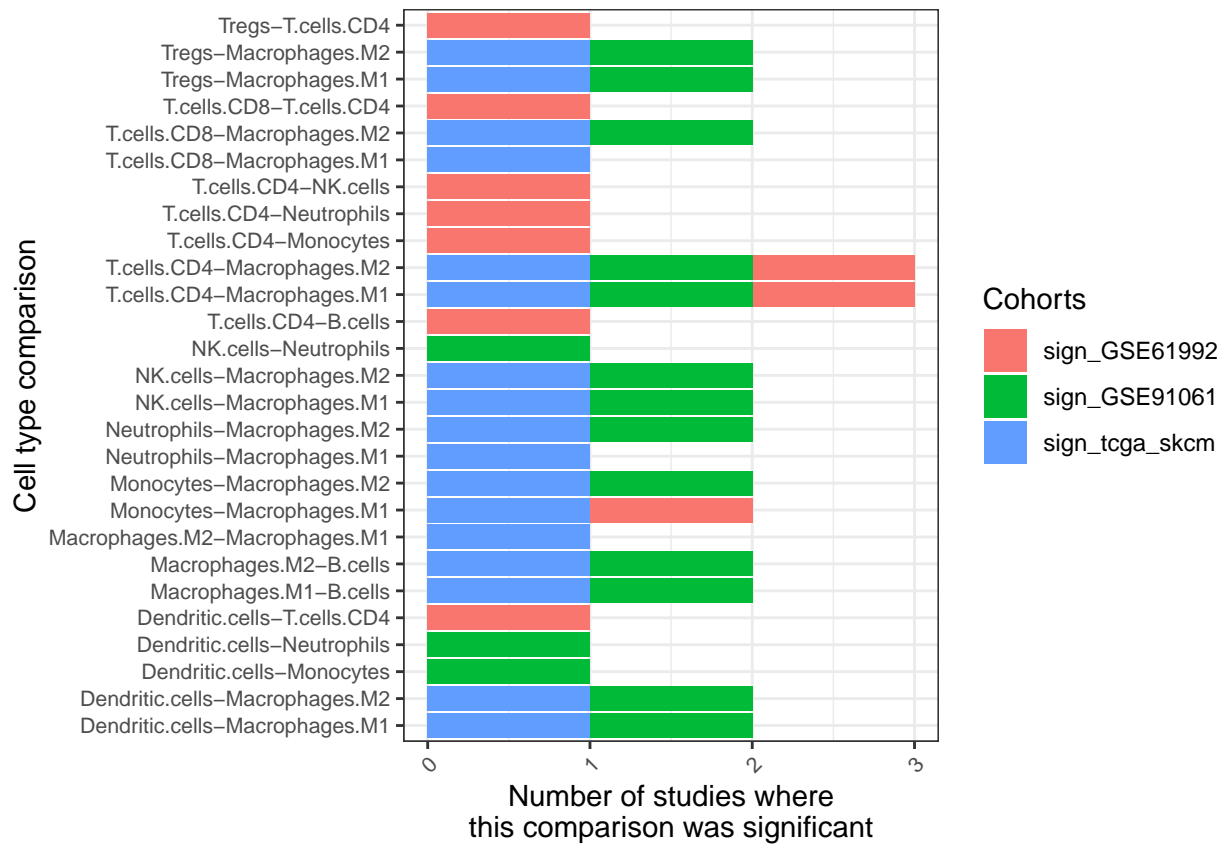
3.7 Todas las cohortes

Teniendo en cuenta el resultado del post-test de Tukey, podemos comparar las comparaciones significativas:





Si retiro los estudios de los que se extrajeron biopsias antes del tratamiento:



4 Heatmaps

Dado que disponemos de los datos de los análisis de deconvolución, también podemos hacer comparaciones de éstos mediante heatmaps:

5 scRNA-seq vs deconvoluted bulk RNA-seq

Para demostrar la validez de los análisis realizados de deconvolución, también comparo el cohorte analizado GSE91061 con datos bulk RNA-seq frente al cohorte scRNA-seq en el que también se había utilizado los mismos tratamientos.

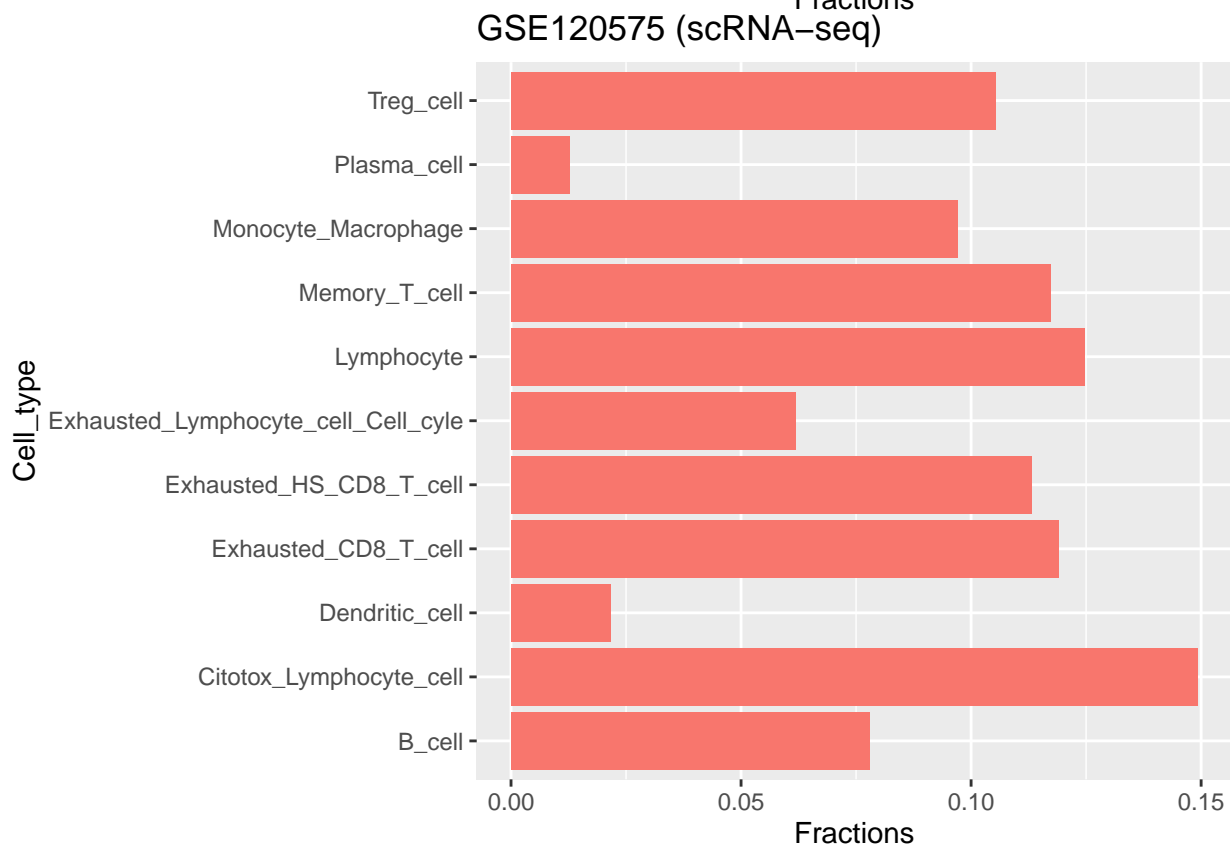
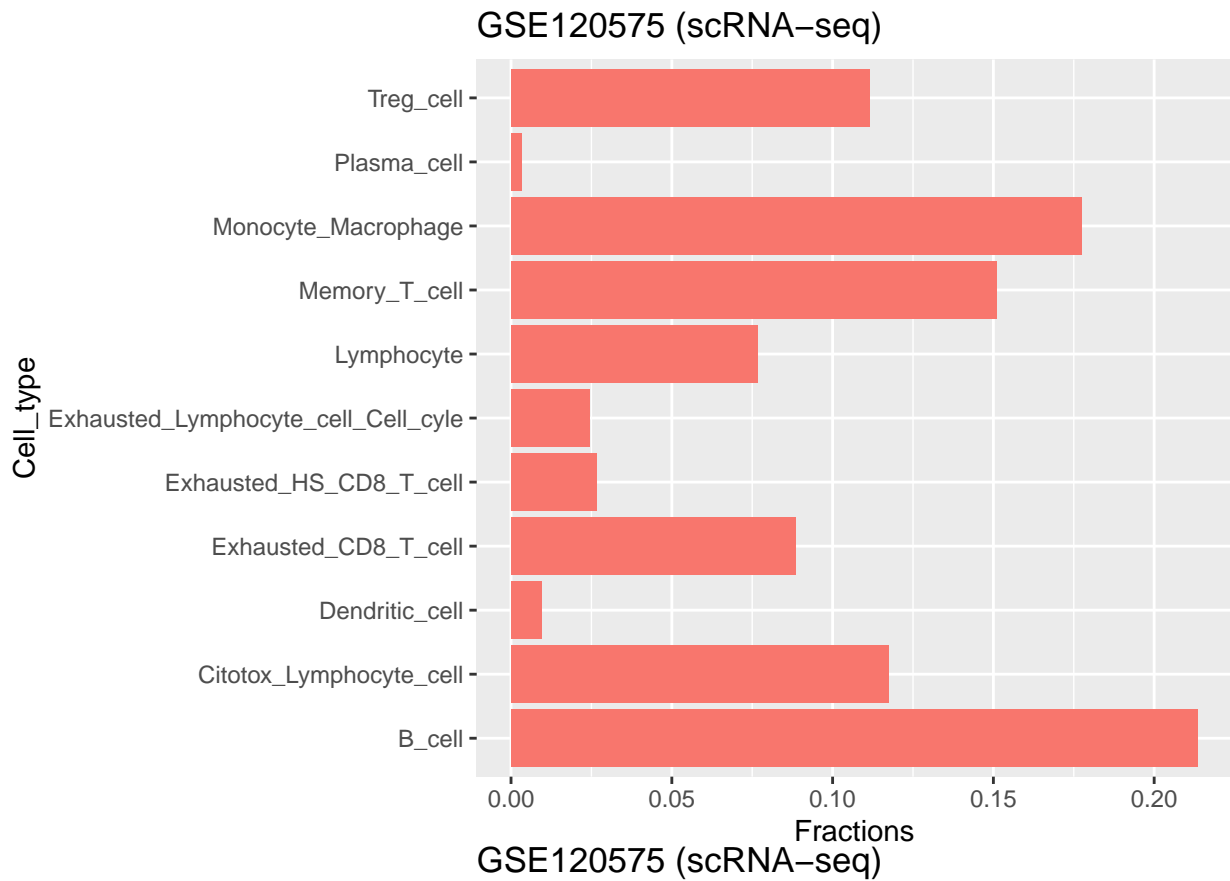
5.1 GSE120575

Dado que este cohorte tiene tratamientos combinados o separados, para poder comparar con el cohorte analizado en deconvolución, filtramos los que no sean tratados de la misma manera: nos quedamos con la combinación CTLA4 + PD1.

```
## New names:
## Rows: 16319 Columns: 35
## -- Column specification
## ----- Delimiter: "\t" chr
## (7): Sample name, title, source name, organism, characteristics: patine... lgl
## (28): molecule, description, processed data file, raw file, ...12, ...13...
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...14`
## * `` -> `...15`
## * `` -> `...16`
## * `` -> `...17`
## * `` -> `...18`
## * `` -> `...19`
## * `` -> `...20`
## * `` -> `...21`
## * `` -> `...22`
## * `` -> `...23`
## * `` -> `...24`
## * `` -> `...25`
## * `` -> `...26`
## * `` -> `...27`
## * `` -> `...28`
## * `` -> `...29`
## * `` -> `...30`
## * `` -> `...31`
## * `` -> `...32`
## * `` -> `...33`
## * `` -> `...34`
## * `` -> `...35`

## Warning in inner_join(metadata, metadata2, by = "title"): Detected an unexpected many-to-many relationship
## Detected an unexpected many-to-many relationship between `x` and `y`.
## i Row 16292 of `x` matches multiple rows in `y`.
## i Row 16292 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
```

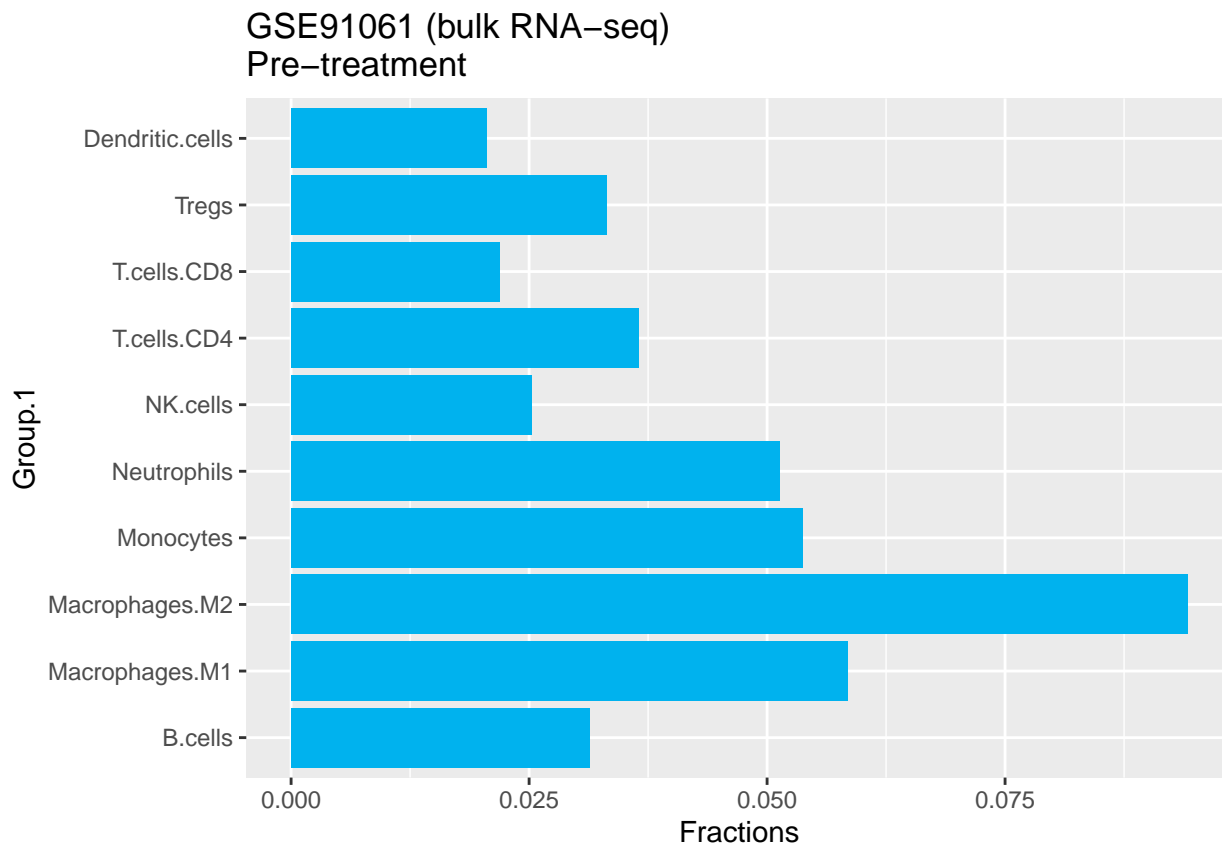
```
## "many-to-many" to silence this warning.
```



Dado que el tratamiento de este cohorte distingue entre los niveles previos al tratamiento y durante el tratamiento, se puede utilizar también para la comparación:

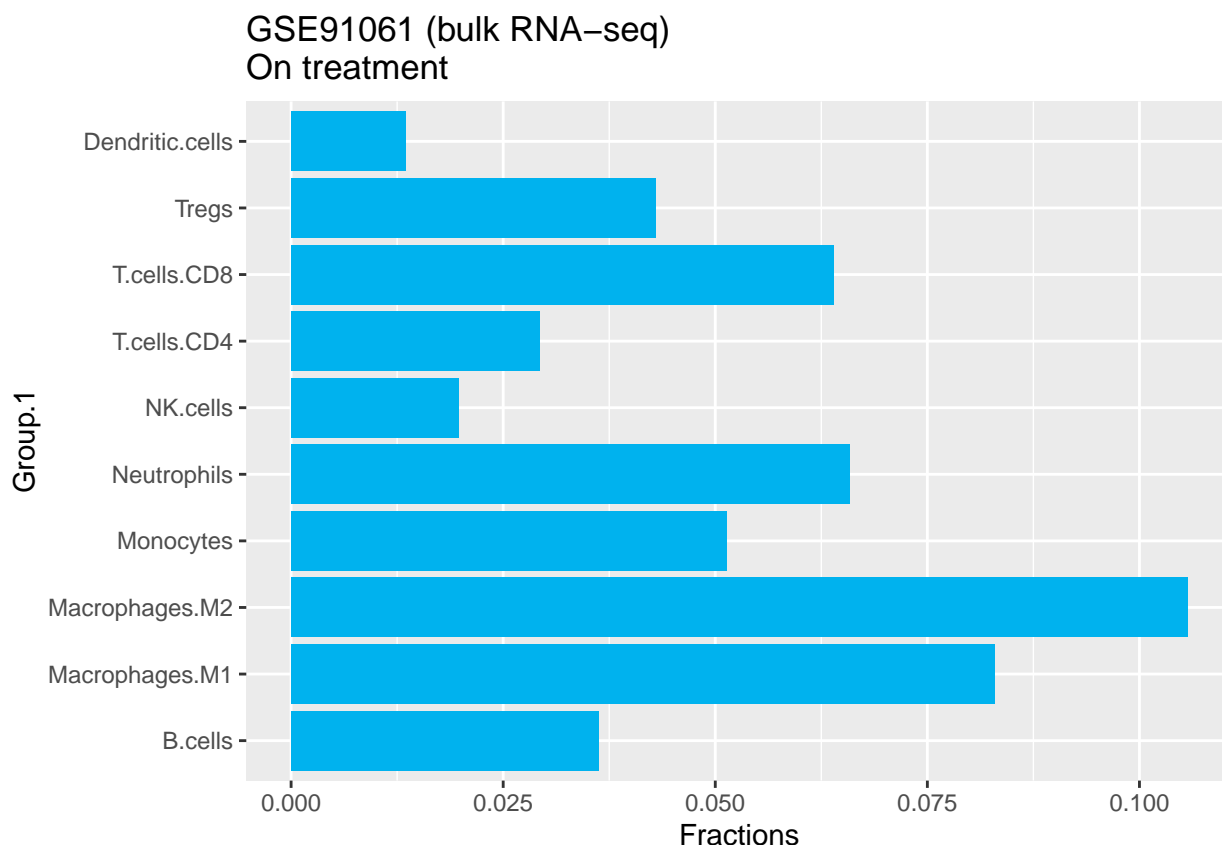
19

```
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
```



```
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
## Warning in mean.default(X[[i]], ...): argument is not numeric or logical:
## returning NA
```

[illegible]



6 DEGs

Resumen de los resultados de firmas génicas.

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
## Warning: Removed 14 rows containing missing values or values outside the scale range
## (`geom_col()`).
```

