

# Explorando Proyectos con ReCount3

García Carlos

2025-02-01

## Contents

ReCount3 . . . . .	1
Buscando Proyecto de Interes . . . . .	1

## ReCount3

**ReCount3** es un recurso en línea que consta de recuentos de genes, exones y uniones exón-exón de secuencias de ARN, así como archivos bigWig de cobertura para 8679 y 10 088 estudios diferentes para humanos y ratones respectivamente. Es la tercera generación del proyecto **ReCount** y parte de **recount.bio**.

## Buscando Proyecto de Interes

Lo primero que hay que hacer es cargar la librería de ReCount3.

```
library(recount3)

# Bloque para asegurarnos que se cargo de manera correcta recount3

if ("recount3" %in% loadedNamespaces()) {
  print("Se cargó de manera exitosa recount3")
} else {
  print("No se pudo cargar recount3")
}
```

```
## [1] "Se cargó de manera exitosa recount3"
```

```
# Obtener la lista de proyectos disponibles en recount3
human_projects <- available_projects()
```

```
## 2025-02-01 19:49:29.432916 caching file sra.recount_project.MD.gz.
```

```
## 2025-02-01 19:49:30.036422 caching file gtex.recount_project.MD.gz.
```

```
## 2025-02-01 19:49:30.564568 caching file tcga.recount_project.MD.gz.
```

```
# Mostrar los valores únicos de la columna "project_home" para ver de dónde provienen los datos
unique(human_projects$project_home)
```

```
## [1] "data_sources/sra" "data_sources/gtex" "data_sources/tcga"
```

```
# Filtrar solo los proyectos que pertenecen a TCGA (The Cancer Genome Atlas)
human_projects[human_projects$project_home == "data_sources/tcga", ]
```

```
##      project organism file_source      project_home project_type n_samples
## 8710     ACC      human      tcga data_sources/tcga data_sources      79
## 8711     BLCA     human      tcga data_sources/tcga data_sources     433
## 8712     BRCA     human      tcga data_sources/tcga data_sources    1256
## 8713     CESC     human      tcga data_sources/tcga data_sources     309
## 8714     CHOL     human      tcga data_sources/tcga data_sources      45
## 8715     COAD     human      tcga data_sources/tcga data_sources     546
## 8716     DLBC     human      tcga data_sources/tcga data_sources      48
## 8717     ESCA     human      tcga data_sources/tcga data_sources     198
## 8718      GBM     human      tcga data_sources/tcga data_sources     175
## 8719     HNSC     human      tcga data_sources/tcga data_sources     548
## 8720     KICH     human      tcga data_sources/tcga data_sources      91
## 8721     KIRC     human      tcga data_sources/tcga data_sources     618
## 8722     KIRP     human      tcga data_sources/tcga data_sources     323
## 8723     LAML     human      tcga data_sources/tcga data_sources     178
## 8724     LGG      human      tcga data_sources/tcga data_sources     532
## 8725     LIHC     human      tcga data_sources/tcga data_sources     424
## 8726     LUAD     human      tcga data_sources/tcga data_sources     601
## 8727     LUSC     human      tcga data_sources/tcga data_sources     555
## 8728     MESO     human      tcga data_sources/tcga data_sources      87
## 8729      OV      human      tcga data_sources/tcga data_sources     430
## 8730     PAAD     human      tcga data_sources/tcga data_sources     183
## 8731     PCPG     human      tcga data_sources/tcga data_sources     187
## 8732     PRAD     human      tcga data_sources/tcga data_sources     558
## 8733     READ     human      tcga data_sources/tcga data_sources     177
## 8734     SARC     human      tcga data_sources/tcga data_sources     265
## 8735     SKCM     human      tcga data_sources/tcga data_sources     473
## 8736     STAD     human      tcga data_sources/tcga data_sources     453
## 8737     TGCT     human      tcga data_sources/tcga data_sources     156
## 8738     THCA     human      tcga data_sources/tcga data_sources     572
## 8739     THYM     human      tcga data_sources/tcga data_sources     122
## 8740     UCEC     human      tcga data_sources/tcga data_sources     589
## 8741     UCS      human      tcga data_sources/tcga data_sources      57
## 8742     UVM      human      tcga data_sources/tcga data_sources      80
```

```
# Seleccionar la información del proyecto específico "LAML" (Leucemia Mieloide Aguda)
# y asegurarse de que proviene de la fuente de datos correcta
project_info <- subset(
  human_projects,
  project == "LAML" & project_type == "data_sources"
)
```

```
# Crear un objeto RangedSummarizedExperiment (RSE) con los datos de LAML
rse_LAML <- create_rse(project_info)
```

```
## 2025-02-01 19:49:34.680173 downloading and reading the metadata.

## 2025-02-01 19:49:35.245746 caching file tcga.tcga.LAML.MD.gz.

## 2025-02-01 19:49:35.794242 caching file tcga.recount_project.LAML.MD.gz.

## 2025-02-01 19:49:36.329555 caching file tcga.recount_qc.LAML.MD.gz.

## 2025-02-01 19:49:36.990752 caching file tcga.recount_seq_qc.LAML.MD.gz.

## 2025-02-01 19:49:37.683717 downloading and reading the feature information.

## 2025-02-01 19:49:38.144497 caching file human.gene_sums.G026.gtf.gz.

## 2025-02-01 19:49:38.669846 downloading and reading the counts: 178 samples across 63856 features.

## 2025-02-01 19:49:39.267999 caching file tcga.gene_sums.LAML.G026.gz.

## 2025-02-01 19:49:42.113479 constructing the RangedSummarizedExperiment (rse) object.
```

```
# Calcular los conteos de lectura (read counts) y almacenarlos en el assay "counts"
# dentro del objeto RSE
assay(rse_LAML, "counts") <- compute_read_counts(rse_LAML)

# Mostrar el objeto RSE con los datos de LAML
rse_LAML
```

```
## class: RangedSummarizedExperiment
## dim: 63856 178
## metadata(8): time_created recount3_version ... annotation recount3_url
## assays(2): raw_counts counts
## rownames(63856): ENSG00000278704.1 ENSG00000277400.1 ...
## ENSG00000182484.15_PAR_Y ENSG00000227159.8_PAR_Y
## rowData names(10): source type ... havana_gene tag
## colnames(178): 984f27ef-d4d7-4e68-bd64-776fd04d07
## 8ff9e94a-2ed2-4727-947f-d524d7ece815 ...
## 4c810ffa-ed07-4f4c-9f81-b8f1cf4956f7
## cebe9594-0f19-46b4-af7d-f8df33e00afb
## colData names(937): rail_id external_id ... recount_seq_qc.errq
## BigWigURL
```

Se seleccionó el proyecto **LAML** (Leucemia Mieloide Aguda) del prestigioso consorcio **TCGA** (The Cancer Genome Atlas), una de las iniciativas más importantes en oncogenómica. Este proyecto proporciona un recurso invaluable para el análisis del perfil molecular de la leucemia mieloide aguda, permitiendo una mejor comprensión de su biología y posibles estrategias terapéuticas.