This is a fantastic milestone! You have successfully built a consistent, reproducible R package.

Here is your complete "Recipe Book" for recreating SomaEnrich from scratch, followed by the guide on how to publish it to GitHub so the world (or just your lab) can use it.

Part 1: How to Build the Package (Step-by-Step)

Save this guide as BUILD_INSTRUCTIONS.txt inside your project folder for future reference.

1. Setup the Project

Open RStudio and run these commands in the Console.

```R
```

```R
# Install build tools if you haven't already
install.packages(c("devtools", "usethis", "roxygen2", "testthat"))
library(devtools)
library(usethis)

# Create the package directory (A new RStudio window will open)
create_package("~/SomaEnrich")
```

2. Setup the Data Pipeline (In the NEW window)

We need to embed the proteinatlas.tsv data into the package so users don't need to download it.

1. **Create the raw data folder:**
   ```R
   use_data_raw()
   ```

2. **Move the file:** Manually copy your proteinatlas.tsv file into the new data-raw/ folder that appeared in your "Files" pane.
3. **Create the processing script:**
   ```R
   use_r("data_processing") # Creates data-raw/data_processing.R
   ```

4. **Paste this code** into data-raw/data_processing.R and run it:
   ```R
   library(readr)
   ```

```r
library(dplyr)
library(tidyr)

# 1. Load Data
hpa_raw <- read_tsv("data-raw/proteinatlas.tsv", col_types = cols())

# 2. Filter & Process
# We use 'Gene' (Ensembl) as the ID, but keep it flexible
hpa_dataset <- hpa_raw %>%
  filter(`RNA tissue specificity` %in% c("Tissue enriched", "Group enriched", "Tissue
enhanced")) %>%
  select(Gene = Gene, Specificity = `RNA tissue specificity`, Tissue_String = `RNA tissue
specific nTPM`) %>%
  separate_rows(Tissue_String, sep = ";") %>%
  separate(Tissue_String, into = c("Tissue", "nTPM"), sep = ":") %>%
  mutate(Tissue = trimws(Tissue), nTPM = as.numeric(nTPM)) %>%
  select(Gene, Specificity, Tissue) %>%
  distinct()

# 3. Save to package
usethis::use_data(hpa_dataset, overwrite = TRUE)
```

3. Create the Main Function

1. **Create the function file:**
   R
   ```r
   use_r("enrichment_analysis") # Creates R/enrichment_analysis.R
   ```

2. **Paste the final code** (the one we verified matches your script) into that file.
3. **Define Dependencies:** Run these lines in the Console to tell the package what libraries it needs:
   R
   ```r
   use_package("dplyr")
   use_package("tidyr")
   use_package("readr")
   use_package("magrittr")
   use_package("stats")
   ```

4. Build and Install

1. **Document:** Generates the help files (man pages).
   R
   ```r
   document()
   ```

2. **Install:** Compiles the package and installs it to your local library.
   R
   install()

Part 2: How to Publish for Public Users (GitHub)

The standard way to share R packages today is via **GitHub**. This allows users to install your package with one command (devtools::install_github("YourName/SomaEnrich")).

Step 1: Initialize Git

In your RStudio terminal (or Console), run:

R

use_git()

- Select "Yes" to commit your files.
- Select "Yes" to restart RStudio if asked.

Step 2: Create a Repository on GitHub

1. Go to github.com and log in.
2. Click the **+** icon (top right) -> **New repository**.
3. Name it SomaEnrich.
4. **Important:** Do *not* check "Initialize with README", "Add .gitignore", or "Add license". Keep it empty.
5. Click **Create repository**.

Step 3: Link Your Local Package to GitHub

Back in RStudio, run this command (replace YourUsername with your actual GitHub username):

R

```
# This connects your local folder to the GitHub page you just made
use_github(protocol = "https", auth_token = gh::gh_token())
```

(Note: If use_github() complains about authentication, follow the instructions it prints to set up a Personal Access Token).

Alternatively, you can do this via the Terminal tab in RStudio:

Bash

```bash
git remote add origin https://github.com/YourUsername/SomaEnrich.git
git push -u origin master
```

Step 4: Add a README (The "Landing Page")

Users need to know how to install it. Create a README.md file:

R

```r
use_readme_md()
```

Paste this into the README.md file:

Markdown

```markdown
# SomaEnrich

**SomaEnrich** is an R package for performing tissue enrichment analysis specifically optimized for SomaScan data, using the Human Protein Atlas (HPA).

## Installation

You can install the development version from GitHub:

```r
```

```
# install.packages("devtools")
devtools::install_github("YourUsername/SomaEnrich")
```

Quick Start

R

```
library(SomaEnrich)

# 1. Load your data
my_de_genes <- c("ALB", "CRP", "INS") # Your DE list
my_background <- c("ALB", "CRP", "INS", "TP53", "EGFR", ...) # Your full SomaScan menu

# 2. Run Enrichment
results <- run_tissue_enrichment(
    input_genes = my_de_genes,
    background_genes = my_background
)

# 3. View Results
head(results)
```

Final Step: Push the README
1. Go to the **Git** tab in RStudio (top right pane).
2. Check the boxes next to `README.md` (Stage it).
3. Click **Commit**, type a message ("Added README"), and click Commit.
4. Click **Push** (Green arrow).

Done! Your package is now live. Anyone in the world can install it using `devtools::install_github("YourUsername/SomaEnrich")`.