ggpicrust2

AUTHOR
Jochum, Michael D.

Objective

The purpose of this script is to integrate the annotated KEGG pathways from the KO outputs from picrust2 and then compare differences observed between Type 2 diabetic and control cohorts using DESEQ2 wrapped inside ggpicrust2.

Analysis

1. Data Importation

2. Data Wrangling

3. Data Filtering

1. Data Import

Lets begin by loading our libraries and importing our metadata and KO predictions, then use ggpicrust2 to translate the KO to KEGG pathways

```
library(readr)
```

Warning: package 'readr' was built under R version 4.2.3

```
library(ggpicrust2)
library(tibble)
```

Warning: package 'tibble' was built under R version 4.2.3

```
library(tidyverse)
```

Warning: package 'tidyverse' was built under R version 4.2.3

Warning: package 'ggplot2' was built under R version 4.2.3

Warning: package 'tidyr' was built under R version 4.2.3

Warning: package 'purrr' was built under R version 4.2.3

Warning: package 'dplyr' was built under R version 4.2.3

Warning: package 'stringr' was built under R version 4.2.3

Warning: package 'forcats' was built under R version 4.2.3

Warning: package 'lubridate' was built under R version 4.2.3

```
── Attaching core tidyverse packages ──────────────────── tidyverse 2.0.0 ──
✓ dplyr     1.1.2     ✓ purrr     1.0.1
✓ forcats   1.0.0     ✓ stringr   1.5.0
✓ ggplot2   3.4.2     ✓ tidyr     1.3.0
✓ lubridate 1.9.2
── Conflicts ─────────────────────────────── tidyverse_conflicts() ──
✗ dplyr::filter() masks stats::filter()
✗ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
errors
```

```r
library(ggprism)
```

Warning: package 'ggprism' was built under R version 4.2.3

```r
library(patchwork)
library(mosaic)
```

```
Registered S3 method overwritten by 'mosaic':
  method                           from
  fortify.SpatialPolygonsDataFrame ggplot2

The 'mosaic' package masks several functions from core packages in order to add
additional features.  The original behavior of these functions should not be affected by this.

Attaching package: 'mosaic'

The following object is masked from 'package:Matrix':

    mean

The following objects are masked from 'package:dplyr':

    count, do, tally

The following object is masked from 'package:purrr':

    cross

The following object is masked from 'package:ggplot2':

    stat

The following objects are masked from 'package:stats':
```

```
    binom.test, cor, cor.test, cov, fivenum, IQR, median, prop.test,
    quantile, sd, t.test, var
```

The following objects are masked from 'package:base':

```
    max, mean, min, prod, range, sample, sum
```

```r
library(ggh4x)
```

Warning: package 'ggh4x' was built under R version 4.2.3

```r
getwd()
```

[1] "K:/github/lmatz"

```r
metadata <- read_delim("IRB_Human_Metadata.txt", delim = "\t", escape_double = FALSE, trim_ws = T
```

```
Rows: 12 Columns: 5
── Column specification ─────────────────────────────────────────────
Delimiter: "\t"
chr (3): SampleID, Diabetes_Status, Sex
dbl (2): BMI, Age

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.
```

```r
metadata
```

```
# A tibble: 12 × 5
   SampleID            Diabetes_Status Sex      BMI   Age
   <chr>               <chr>           <chr>  <dbl> <dbl>
 1 Buffington_201_001  T2D             Male    31.8    57
 2 Buffington_201_002  T2D             Male    39.5    59
 3 Buffington_201_003  HC              Female  25.2    50
 4 Buffington_201_005  HC              Female  34.5    53
 5 Buffington_201_006  T2D             Female  28.7    62
 6 Buffington_201_007  HC              Male    25.8    52
 7 Buffington_201_008  HC              Female  35.8    54
 8 Buffington_201_009  HC              Female  21.1    56
 9 Buffington_201_010  T2D             Female  33.5    64
10 Buffington_201_011  HC              Female  25      59
11 Buffington_201_013  T2D             Male    45.4    67
12 Buffington_201_014  T2D             Female  28.5    55
```

```r
# Load KEGG pathway abundance
kegg_abundance <- ko2kegg_abundance("picrust2/KO_metagenome_out/pred_metagenome_unstrat.tsv")
```

```
Loading data from file...
Rows: 5565 Columns: 13── Column specification
────────────────────────────────────────────────────

Delimiter: "\t"
chr  (1): function
dbl (12): Buffington_201_001, Buffington_201_002, Buffington_201_003, Buffin...
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.Loading KEGG
reference data. This might take a while...
Performing KO to KEGG conversion. Please be patient, this might take a while...
```

```
|                                                        |    0%
|
|                                                        |    1%
|
|=                                                       |    1%
|
|=                                                       |    2%
|
|==                                                      |    2%
|
|==                                                      |    3%
|
|===                                                     |    4%
|
|===                                                     |    5%
|
|====                                                    |    5%
|
|====                                                    |    6%
|
|=====                                                   |    7%
|
|=====                                                   |    8%
|
|======                                                  |    8%
|
|======                                                  |    9%
|
|=======                                                 |    9%
|
|=======                                                 |   10%
|
|========                                                |   11%
|
|========                                                |   12%
|
|=========                                               |   12%
```

|
|========= | 13%
|
|========== | 14%
|
|========= | 15%
|
|========== | 15%
|
|========== | 16%
|
|=========== | 17%
|
|=========== | 18%
|
|============ | 18%
|
|============ | 19%
|
|============= | 20%
|
|============= | 21%
|
|============== | 21%
|
|============== | 22%
|
|=============== | 22%
|
|=============== | 23%
|
|=============== | 24%
|
|================ | 24%
|
|================ | 25%
|
|================= | 25%
|
|================= | 26%
|
|================== | 26%
|
|================== | 27%
|
|================== | 28%
|
|=================== | 28%
|
|=================== | 29%

```
|
|=====================                                                    |   29%
|
|=====================                                                    |   30%
|
|=====================                                                    |   31%
|
|=====================                                                    |   32%
|
|======================                                                   |   32%
|
|======================                                                   |   33%
|
|======================                                                   |   34%
|
|=======================                                                  |   35%
|
|=======================                                                  |   35%
|
|=======================                                                  |   36%
|
|========================                                                 |   37%
|
|========================                                                 |   38%
|
|=========================                                                |   38%
|
|=========================                                                |   39%
|
|==========================                                               |   40%
|
|==========================                                               |   41%
|
|===========================                                              |   41%
|
|===========================                                              |   42%
|
|============================                                             |   42%
|
|============================                                             |   43%
|
|=============================                                            |   44%
|
|==============================                                           |   45%
|
|===============================                                          |   45%
|
|===============================                                          |   46%
|
|================================                                         |   47%
```
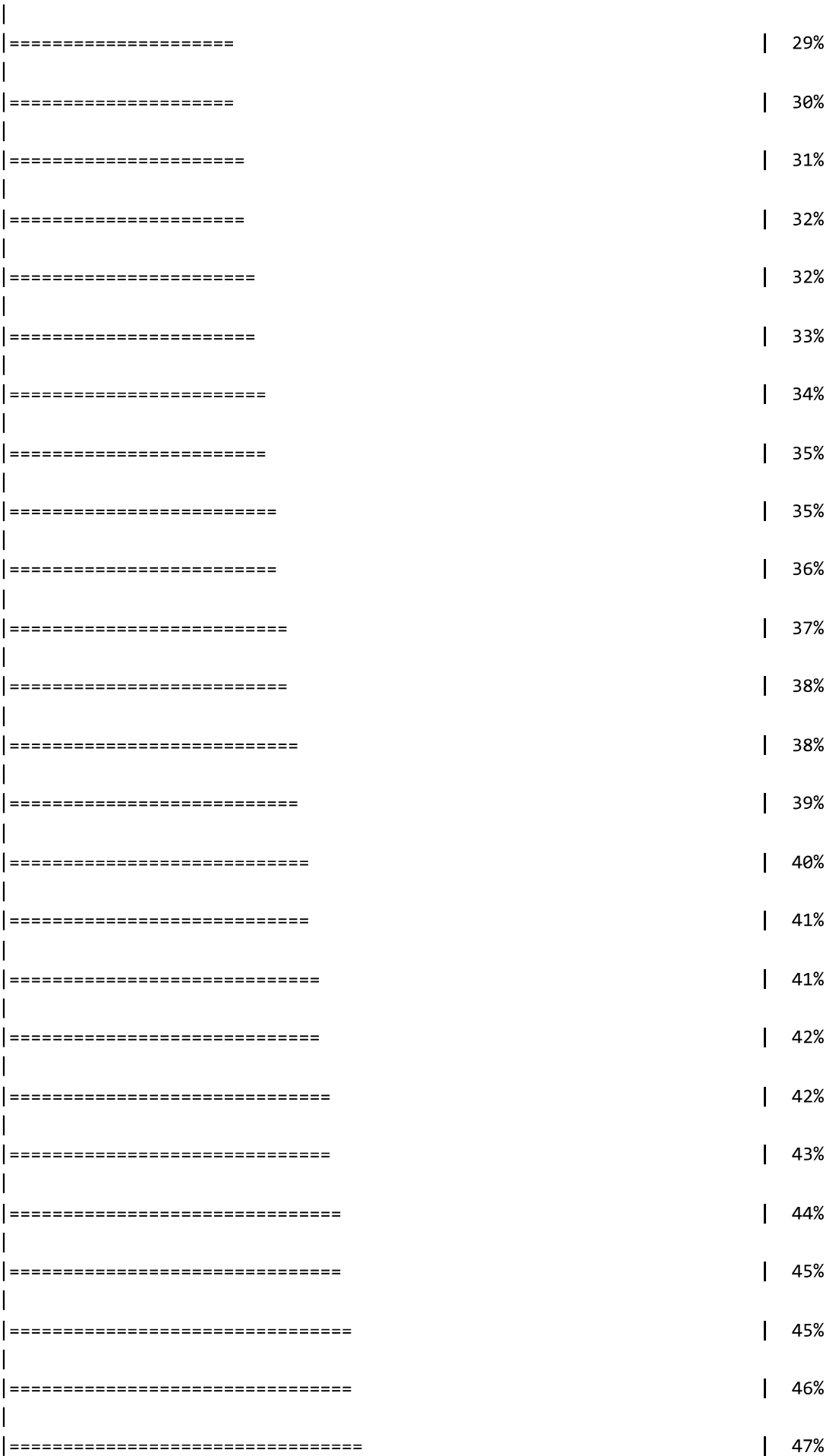
|=================================== | 48%

|================================== | 48%

|=================================== | 49%

|===================================== | 49%

|===================================== | 50%

|====================================== | 51%

|======================================= | 51%

|======================================= | 52%

|=== | 52%

|=== | 53%

|=== | 54%

|=== | 55%

|=== | 55%

|=== | 56%

|== | 57%

|=== | 58%

|== | 58%

|== | 59%

|== | 59%

|=== | 60%

|== | 61%

|== | 62%

|== | 62%

|== | 63%

|== | 64%

|=== | 65%

|== | 65%

|=== | 66%

|== | 67%

|== | 68%

|=== | 68%

|=== | 69%

|== | 70%

|== | 71%

|== | 71%

|== | 72%

|== | 72%

|== | 73%

|== | 74%

|=== | 74%

|=== | 75%

|== | 75%

|== | 76%

|== | 76%

|== | 77%

|== | 78%

|=== | 78%

|=== | 79%

|== | 79%

|== | 80%

```
|
|============================================================     |   81%
|
|============================================================     |   82%
|
|============================================================     |   82%
|
|============================================================     |   83%
|
|=============================================================    |   84%
|
|=============================================================    |   85%
|
|==============================================================   |   85%
|
|==============================================================   |   86%
|
|===============================================================  |   87%
|
|===============================================================  |   88%
|
|===============================================================  |   88%
|
|===============================================================  |   89%
|
|================================================================ |   90%
|
|================================================================ |   91%
|
|================================================================ |   91%
|
|================================================================ |   92%
|
|================================================================ |   92%
|
|================================================================ |   93%
|
|================================================================ |   94%
|
|================================================================ |   95%
|
|================================================================ |   95%
|
|================================================================ |   96%
|
|================================================================ |   97%
|
|================================================================ |   98%
|
|================================================================ |   98%
```

```
    |
    |======================================================================= |    99%
    |
    |=======================================================================|    99%
    |
    |=======================================================================|   100%
```

KO to KEGG conversion completed. Time elapsed: 16.24 seconds.

Removing KEGG pathways with zero abundance across all samples...

KEGG abundance calculation completed successfully.

differential pathway analysis

Using the kegg_abundance and metadata, run DESeq2 to obtain the sig. KO assoc. with Type 2 diabetes

```
deseq2_daa_results_df<- pathway_daa(abundance = kegg_abundance,
                                    metadata = metadata,
                                    group = "Diabetes_Status",
                                    p.adjust ="BH",
                                    daa_method = "DESeq2",
                                    select = NULL,
                                    reference = "HC")
```

Sample names extracted.

Identifying matching columns in metadata...

Matching columns identified: SampleID . This is important for ensuring data consistency.

Using all columns in abundance.

Converting abundance to a matrix...

Reordering metadata...

Converting metadata to a matrix and data frame...

Extracting group information...

Running DESeq2. Note: DESeq2 is only suitable for comparison between two groups.

Performing pairwise comparisons with DESeq2...

converting counts to integer mode

using pre-existing size factors

estimating dispersions

gene-wise dispersion estimates

mean-dispersion relationship

final dispersion estimates

fitting model and testing

Compiling DESeq2 results...

DESeq2 analysis complete.

```
deseq2_daa_results_df<-deseq2_daa_results_df%>%filter(!is.na(p_values)) #14

deseq2_daa_results_df%>%filter(p_values<0.05) #14
```

	feature	method	group1	group2	p_values	adj_method	p_adjust
1	ko00311	DESeq2	T2D	HC	0.0154634285344572	BH	0.243846373
2	ko05215	DESeq2	T2D	HC	0.00932532551594288	BH	0.173790157
3	ko05211	DESeq2	T2D	HC	0.0103302717505706	BH	0.176475476
4	ko05130	DESeq2	T2D	HC	0.00708270225776833	BH	0.173790157
5	ko04914	DESeq2	T2D	HC	0.00932532551594288	BH	0.173790157
6	ko05143	DESeq2	T2D	HC	0.0386912989347241	BH	0.520024327
7	ko00400	DESeq2	T2D	HC	0.0405872645834935	BH	0.520024327
8	ko05131	DESeq2	T2D	HC	0.00194346578891148	BH	0.079682097
9	ko00340	DESeq2	T2D	HC	0.00451037851803533	BH	0.154104599
10	ko04612	DESeq2	T2D	HC	0.00932532551594288	BH	0.173790157
11	ko04621	DESeq2	T2D	HC	0.00801484379440754	BH	0.173790157
12	ko05014	DESeq2	T2D	HC	0.0491542850008098	BH	0.555480616
13	ko05100	DESeq2	T2D	HC	0.000143460544632947	BH	0.007352353
14	ko04144	DESeq2	T2D	HC	0.032209896154432	BH	0.471644908

```
deseq2_daa_results_df%>%filter(p_adjust<0.05) #4
```

	feature	method	group1	group2	p_values	adj_method	p_adjust
1	ko00592	DESeq2	T2D	HC	9.66578157555687e-05	BH	0.0066049507
2	ko00624	DESeq2	T2D	HC	3.33156296678632e-05	BH	0.0034148520
3	ko05150	DESeq2	T2D	HC	5.88462099103448e-07	BH	0.0001206347
4	ko05100	DESeq2	T2D	HC	0.000143460544632947	BH	0.0073523529

Annotate the significant KO that have annotations

```
# Annotate pathway results using KO to KEGG conversion
deseq2_daa_annotated_sub_method_results_df <- pathway_annotation(pathway = "KO", daa_results_df =
```

Starting pathway annotation...

DAA results data frame is not null. Proceeding...

KO to KEGG is set to TRUE. Proceeding with KEGG pathway annotations...

We are connecting to the KEGG database to get the latest results, please wait patiently.
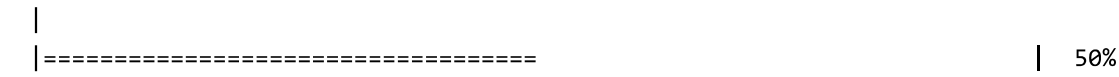
```
Processing pathways individually...




      |
      |                                                            |   0%

Beginning annotation for pathway 1 of 4...

Annotated pathway 1 of 4. Time taken: 1.79 seconds.




      |
      |==================                                          |  25%

Beginning annotation for pathway 2 of 4...

Annotated pathway 2 of 4. Time taken: 0.23 seconds.




      |
      |===================================                         |  50%

Beginning annotation for pathway 3 of 4...

Annotated pathway 3 of 4. Time taken: 0.24 seconds.




      |
      |====================================================        |  75%

Beginning annotation for pathway 4 of 4...

Annotated pathway 4 of 4. Time taken: 0.2 seconds.




      |
      |============================================================| 100%

Pathway annotation completed.

Returning DAA results filtered annotation data frame...
```
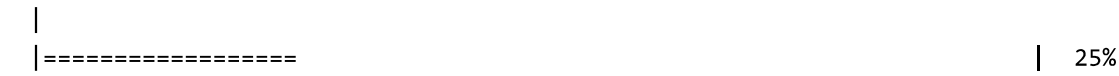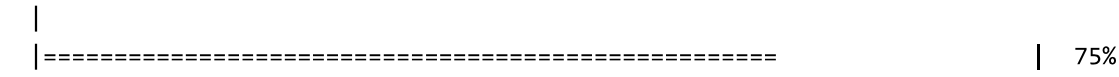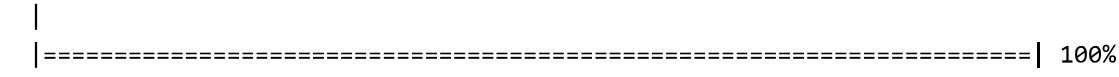
deseq2_daa_annotated_sub_method_results_df

```
    feature method group1 group2            p_values adj_method     p_adjust
45  ko00592 DESeq2    T2D     HC 9.66578157555687e-05         BH 0.0066049507
```

```
58   ko00624 DESeq2      T2D       HC 3.33156296678632e-05            BH 0.0034148520
140  ko05150 DESeq2      T2D       HC 5.88462099103448e-07            BH 0.0001206347
175  ko05100 DESeq2      T2D       HC 0.000143460544632947            BH 0.0073523529
```

```
                                pathway_name
45                   alpha-Linolenic acid metabolism
58   Polycyclic aromatic hydrocarbon degradation
140              Staphylococcus aureus infection
175      Bacterial invasion of epithelial cells
```

```
pathway_description
45
<NA>
58
<NA>
140
```
Staphylococcus aureus can cause multiple forms of infections ranging from superficial skin infections to food poisoning and life-threatening infections. The organism has several ways to divert the effectiveness of the immune system: secreting immune modulating proteins that inhibit complement activation and neutrophil chemotaxis or lysis, modulating the sensitivity to cationic antimicrobial peptides (such as defensin) by increasing the positive net charge of its cytoplasmic membrane, and expression of superantigens that prevent development of a normal immune response or cause an emetic response when ingested.

175 Many pathogenic bacteria can invade phagocytic and non-phagocytic cells and colonize them intracellularly, then become disseminated to other cells. Invasive bacteria induce their own uptake by non-phagocytic host cells (e.g. epithelial cells) using two mechanisms referred to as zipper model and trigger model. Listeria, Staphylococcus, Streptococcus, and Yersinia are examples of bacteria that enter using the zipper model. These bacteria express proteins on their surfaces that interact with cellular receptors, initiating signalling cascades that result in close apposition of the cellular membrane around the entering bacteria. Shigella and Salmonella are the examples of bacteria entering cells using the trigger model. These bacteria use type III secretion systems to inject protein effectors that interact with the actin cytoskeleton.

```
                                pathway_class
45                      Metabolism; Lipid metabolism
58   Metabolism; Xenobiotics biodegradation and metabolism
140          Human Diseases; Infectious disease: bacterial
175          Human Diseases; Infectious disease: bacterial
```

```
                                pathway_map
45                   alpha-Linolenic acid metabolism
58   Polycyclic aromatic hydrocarbon degradation
140              Staphylococcus aureus infection
175      Bacterial invasion of epithelial cells
```

Generate the pathway error plot

```
# Generate pathway error bar plot
# Please change Group to metadata$your_group_column if you are not using example dataset
p <- pathway_errorbar(abundance = kegg_abundance,
                 daa_results_df = deseq2_daa_annotated_sub_method_results_df,
                 Group = metadata$Diabetes_Status,
                 p_values_threshold = 0.25,
                 order = "pathway_class",
```

```
                       select = NULL,
                       ko_to_kegg = T,
                       p_value_bar = TRUE,
                       colors = NULL,
                       x_lab = "pathway_name")
```

```
Registered S3 method overwritten by 'GGally':
  method from
  +.gg   ggplot2
```

```
p
```