Portfolio

Signaling pathways caused by exogenous carcinogens in vitro associated with oncogenesis

2019250140 Jeon in a (BSMS222)

1. Introduction

1. 1. Background

Lung cancer remains the most common malignancy and the leading cause of cancer mortality worldwide. Lung cancer is known to be mainly caused by direct exposure to cigarettes. But LUAD (lung adenocarcinoma) in East Asia, especially in Taiwan, is characterized by a high rate of never-smokers, early onset, and predominant EGFR mutations (Chen et al., 2020). According to these analyses, APOBEC mutational signatures are frequently observed in younger females and enviromental carcinogen-like mutational signatures highly occur in older females. In addition, early onset is a distinct feature of LUAD in East Asia, especially among never-smokers. From this, we can think that LUAD is associated with genetic and environmental factors, especially in this paper, which analyzed that air pollution in Taiwan correlates with LUAD incidence in never-smokers. The carcinogen signals presented in the paper include (1) Nitrosamine-like, (2) Nitro-PAH, (3) radiation, (4) Alkylating agents, and (5) PAHs. Among these five carcinogen, I will focus on Nitrosamine, Nitro-PAH, and PAHs. Among the many components in tobacco smoke and outdoor and indoor air pollution are polycyclic aromatic hydrocarbons (PAHs), which are considered to be the most important carcinogens in these complex mixtures. Metabolism of PAHs leads to the formation of the active carcinogens. These reactive metabolites produce DNA adducts, resulting in DNA mutations, alteration of gene expression profiles, and tumorigenesis(Moorthy et al., 2015). Nitrosamines are formed by a reaction between nitrates or nitrites and certain amines. Nitrosamines and/or their precursors can be found in diverse consumer products such as processed meats, alcoholic beverages, cosmetics, cigarette smoke and also be formed in the mouth or stomach if the food contains nitrosamine precursors. Nitrosamines are considered to be strong carcinogens that may produce cancer in diverse organs and tissues including lung, brain, liver, kidney, bladder, stomach, esophagus, and nasal sinus(H. Robles, 2014). Nitrated polycyclic aromatic hydrocarbons (Nitro-PAHs) are derivatives of PAHs with at least one nitro-functional group (-NO2) on the aromatic ring. Nitro-PAHs are mainly generated by incomplete combustion and pyrolysis of fossil fuels and biomass. Nitro-PAHs are direct-acting mutagens and carcinogens. The mechanisms underlying some of these toxicological effects of nitro-PAHs include DNA damage, DNA adduct formation, aryl hydrocarbon receptor activation, changes in gene and protein expression, cell cycle alternations, increased levels of reactive oxygen species and pro-inflamation. Inhalation, oral ingestion and dermal contact are the main routes of nitro-PAH intake from the environment by humans and animals(Benjamin, 2017).

1. 2. Data Visualization Topic

The topic for data visualization is to plot the correlation between environmental carcinogen and enriched pathway. Particularly, focusing on the pathway that has a significant correlation with carcinogen, I will examine whether the carcinogens presented in the paper have a direct relationship with mutational metabolism.

2. Exploring Data

2. 1. Unboxing Dataset

Before drawing the plot, I loaded the packages needed to create a portfolio by using ‘library()’.

```
library(readxl)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
```

```
## Warning:    'tidyr'  R   4.1.2
```

```
## Warning:    'readr'  R   4.1.2
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(dplyr)
library(rio)
```

```
## Warning:    'rio'  R   4.1.2
```

```
library(ggridges)
library(cowplot)
```

I loaded the file by using ‘rio’ package. Plus, since the sheet to be used in this file is [4] “S5C_carcinogen 1Dpath_Fig5E”, the data set was loaded using ‘sheet’.

```
d <- rio::import('https://ars.els-cdn.com/content/image/1-s2.0-S0092867420307431-mmc5.xlsx', sheet = 4)
```

```
## New names:
## * `` -> ...1
## * `` -> ...2
## * `` -> ...3
## * `` -> ...5
## * `` -> ...6
## * ...
```

2. 2. Manipulating Data Frame

After checking the data using ‘head()’, let’s write ‘colnames()’ to check whether the column names are proper or not.

```r
head(d)
```

```
##                                ...1 ...2 ...3                PAHs
## 1       Carcinogen_enriched_pathway   NA   NA                mRNA
## 2           Chemical carcinogenesis    1   NA -0.31144931595774833
## 3 Drug metabolism - cytochrome P450    2   NA -0.15800318970588245
## 4             Vitamin B6 metabolism    3   NA -0.16170810810810804
## 5      Steroid hormone biosynthesis    4   NA -0.40752271666666662
## 6               Tyrosine metabolism    5   NA  0.64585152388059697
##                    ...5                  ...6             NitroPAHs
## 1               Protein               Phospho                  mRNA
## 2   0.12460930645465899  -0.43284930985915515  -0.14544231595774834
## 3     0.201279506087303  -0.45726570499999974 4.8945310294117556E-2
## 4   0.11493490636348699  -0.66166892765957441   0.24789189189189192
## 5 4.6774663031101199E-2                    NA  -0.37392271666666665
## 6 1.0219943709671501E-2 8.2334373770491687E-2  -0.16271947611940299
##                     ...8                  ...9                  Mixed
## 1                Protein               Phospho                   mRNA
## 2 -3.4253496676683398E-2  -0.40659030985915512  4.1076840422516747E-3
## 3    1.77990105003119E-2  -0.88807370499999971  -9.707118970588241E-2
## 4  4.6745907515287399E-2  -0.16967042765957446     -0.556708108108108
## 5   -6.39064386487007E-2                    NA -8.5266716666666645E-2
## 6 -2.4898756295442599E-2 4.8907373770491702E-2   -0.15017647611940299
##                   ...11                  ...12            Nitrosamine
## 1               Protein                Phospho                   mRNA
## 2 7.9630803316831547E-2 -2.4644309859155167E-2   -0.10393531595774835
## 3   0.17991451174020801 -9.0524204999999691E-2 -5.1056897058824324E-3
## 4 5.3167406469583553E-2  5.6881072340425554E-2  -6.130810810810805E-2
## 5  3.318106383085255E-2                     NA    0.23129978333333337
## 6   0.11672574281692499   -0.19960462622950831     -0.133062476119403
##                    ...14                 ...15
## 1                Protein               Phospho
## 2 -1.8061976879835476E-3   0.28762469014084485
## 3 -2.9448938556015498E-2    1.4518357950000003
## 4  4.3244070839136648E-3 1.9425572340425523E-2
## 5 -2.5487435515969999E-2                    NA
## 6   -1.82835562154651E-2   0.12265187377049169
##   Comparison of 6 carcinogen group (p-value)                 ...17
## 1                                       mRNA                  prot
## 2                          0.230818539240676 7.5559771573053998E-3
## 3                          0.256339756363334   1.15378813052363E-2
## 4                        0.65971092990573099 1.9998993868506301E-2
## 5                        0.28582841259592201 4.4840276861642998E-2
## 6                        0.67354691111237597 2.1550580777231002E-2
##                   ...18
## 1                  phos
## 2 9.1148064415487395E-2
## 3 8.5714344236999303E-3
## 4 3.3834974597418702E-2
## 5                     1
## 6 2.8656791444298901E-2
```

```
colnames(d)
```

```
##  [1] "...1"
##  [2] "...2"
##  [3] "...3"
##  [4] "PAHs"
##  [5] "...5"
##  [6] "...6"
##  [7] "NitroPAHs"
##  [8] "...8"
##  [9] "...9"
## [10] "Mixed"
## [11] "...11"
## [12] "...12"
## [13] "Nitrosamine"
## [14] "...14"
## [15] "...15"
## [16] "Comparison of 6 carcinogen group (p-value)"
## [17] "...17"
## [18] "...18"
```

Since the column names are not organized, I set them as I wanted. In addition, row 1 and column 2, 3 are unnecessary, so I deleted them. I changed the columns except for the first column to numeric for facilitate processing, and rearranged the columns into a 'type'.

```
d <- d[c(2:54), c(1, 4:18)]

d <- d %>%
  rename(pathway = ...1,
         PAHs_mRNA = PAHs,
         PAHs_prot = ...5,
         PAHs_phos = ...6,
         NitroPAHs_mRNA = NitroPAHs,
         NitroPAHs_prot = ...8,
         NitroPAHs_phos = ...9,
         Mixed_mRNA = Mixed,
         Mixed_prot = ...11,
         Mixed_phos = ...12,
         Nitrosamine_mRNA = Nitrosamine,
         Nitrosamine_prot = ...14,
         Nitrosamine_phos = ...15,
         mRNA = "Comparison of 6 carcinogen group (p-value)",
         prot = ...17,
         phos = ...18) %>%
  mutate(PAHs_mRNA = as.numeric(PAHs_mRNA),
         PAHs_prot = as.numeric(PAHs_prot),
         PAHs_phos = as.numeric(PAHs_phos),
         NitroPAHs_mRNA = as.numeric(NitroPAHs_mRNA),
         NitroPAHs_prot = as.numeric(NitroPAHs_prot),
         NitroPAHs_phos = as.numeric(NitroPAHs_phos),
         Mixed_mRNA = as.numeric(Mixed_mRNA),
         Mixed_prot = as.numeric(Mixed_prot),
         Mixed_phos = as.numeric(Mixed_phos),
         Nitrosamine_mRNA = as.numeric(Nitrosamine_mRNA),
         Nitrosamine_prot = as.numeric(Nitrosamine_prot),
         Nitrosamine_phos = as.numeric(Nitrosamine_phos),
         mRNA = as.numeric(mRNA),
         prot = as.numeric(prot),
         phos = as.numeric(phos)) %>%
  gather(key = "type", value = "log", -c("pathway", "mRNA", "prot", "phos")) %>%
  gather("mRNA", "prot", "phos", key = "p_value", value = "P")
```

```
## Warning in mask$eval_all_mutate(quo):          NA

## Warning in mask$eval_all_mutate(quo):          NA

## Warning in mask$eval_all_mutate(quo):          NA

## Warning in mask$eval_all_mutate(quo):          NA

## Warning in mask$eval_all_mutate(quo):          NA

## Warning in mask$eval_all_mutate(quo):          NA

## Warning in mask$eval_all_mutate(quo):          NA

## Warning in mask$eval_all_mutate(quo):          NA

## Warning in mask$eval_all_mutate(quo):          NA

## Warning in mask$eval_all_mutate(quo):          NA

## Warning in mask$eval_all_mutate(quo):          NA

## Warning in mask$eval_all_mutate(quo):          NA
```

Let’s check again to see whether it is changed properly.

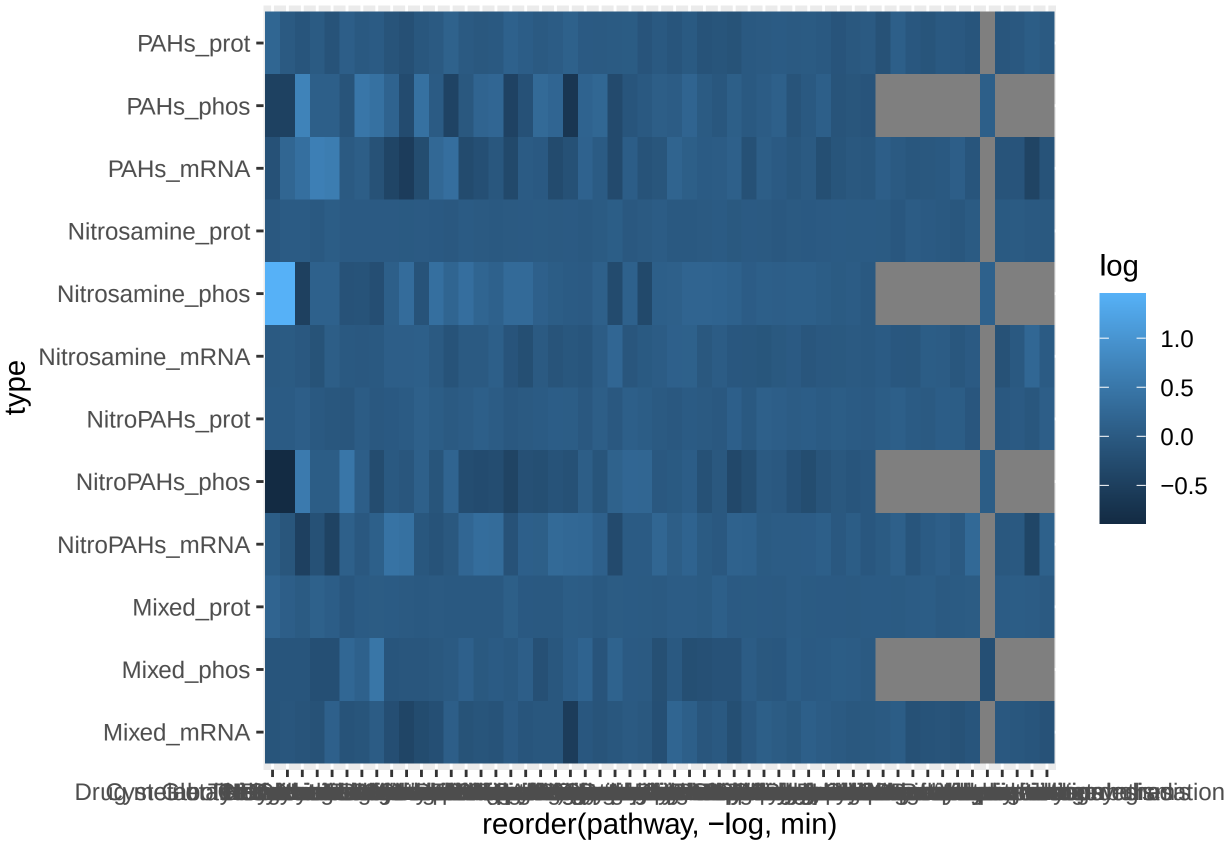
```
head(d)
```

```
##                              pathway      type        log p_value          P
## 1           Chemical carcinogenesis PAHs_mRNA -0.3114493    mRNA 0.23081854
## 2 Drug metabolism - cytochrome P450 PAHs_mRNA -0.1580032    mRNA 0.25633976
## 3             Vitamin B6 metabolism PAHs_mRNA -0.1617081    mRNA 0.65971093
## 4      Steroid hormone biosynthesis PAHs_mRNA -0.4075227    mRNA 0.28582841
## 5               Tyrosine metabolism PAHs_mRNA  0.6458515    mRNA 0.67354691
## 6          Renin-angiotensin system PAHs_mRNA  0.1463799    mRNA 0.06453428
```

3. Data Visualization

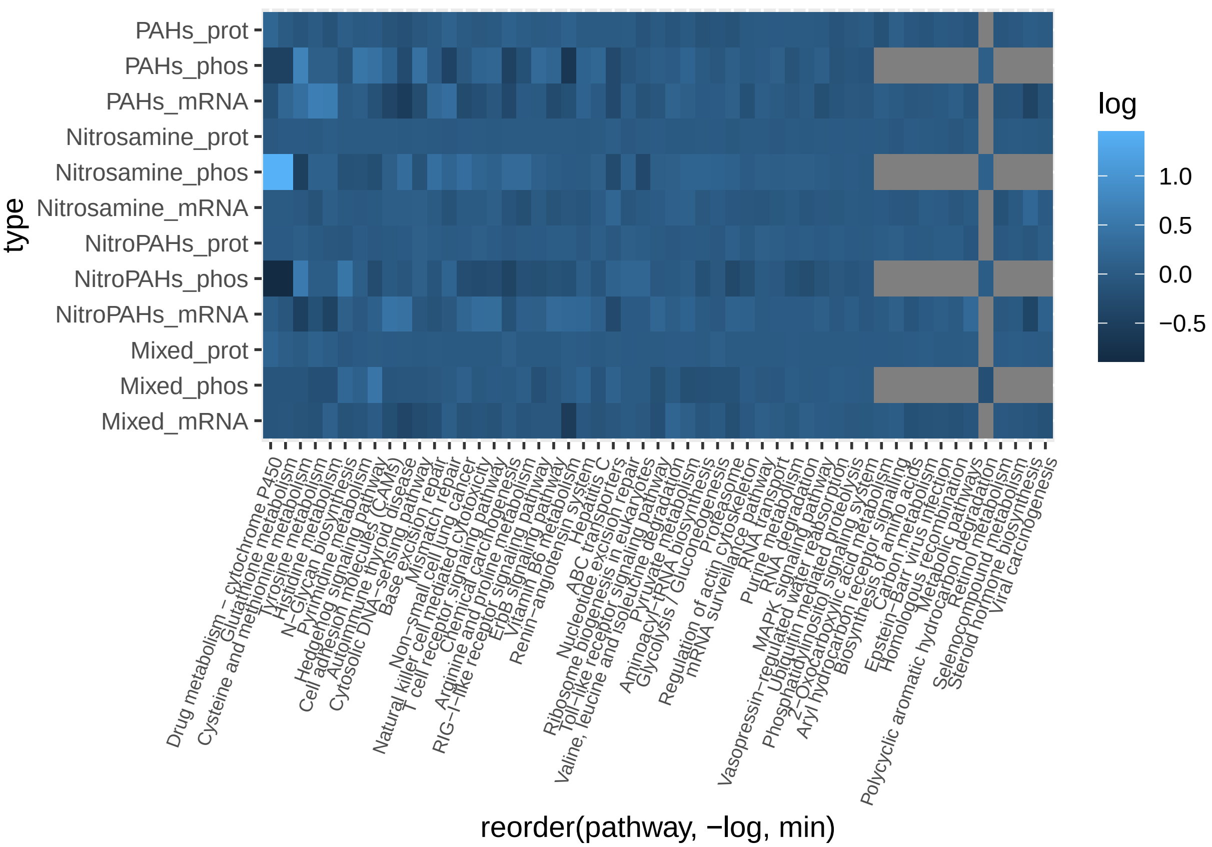
At first, I will draw Heatmap using the values of ‘relative log2T/N’, classified into mRNA, protein, and phosphate to see whether each carcinogens have significant effects on the enriched pathway. First, set the x-axis to ‘pathway’ and y-axis to ‘log2T/N’ values. Then write the code using ‘geom_tile()’ because I will draw a heatmap. At this time, the degree of log value will be compared, so write ‘aes(fill=log)’ in ’geom_tile().

```
d %>% ggplot(aes(x = reorder(pathway, -log, min), y = type)) +
  geom_tile(aes(fill = log))
```



Since ‘pathway’ letters on the x-axis overlap, let’s adjust the angle and size so that the letters don’t overlap.

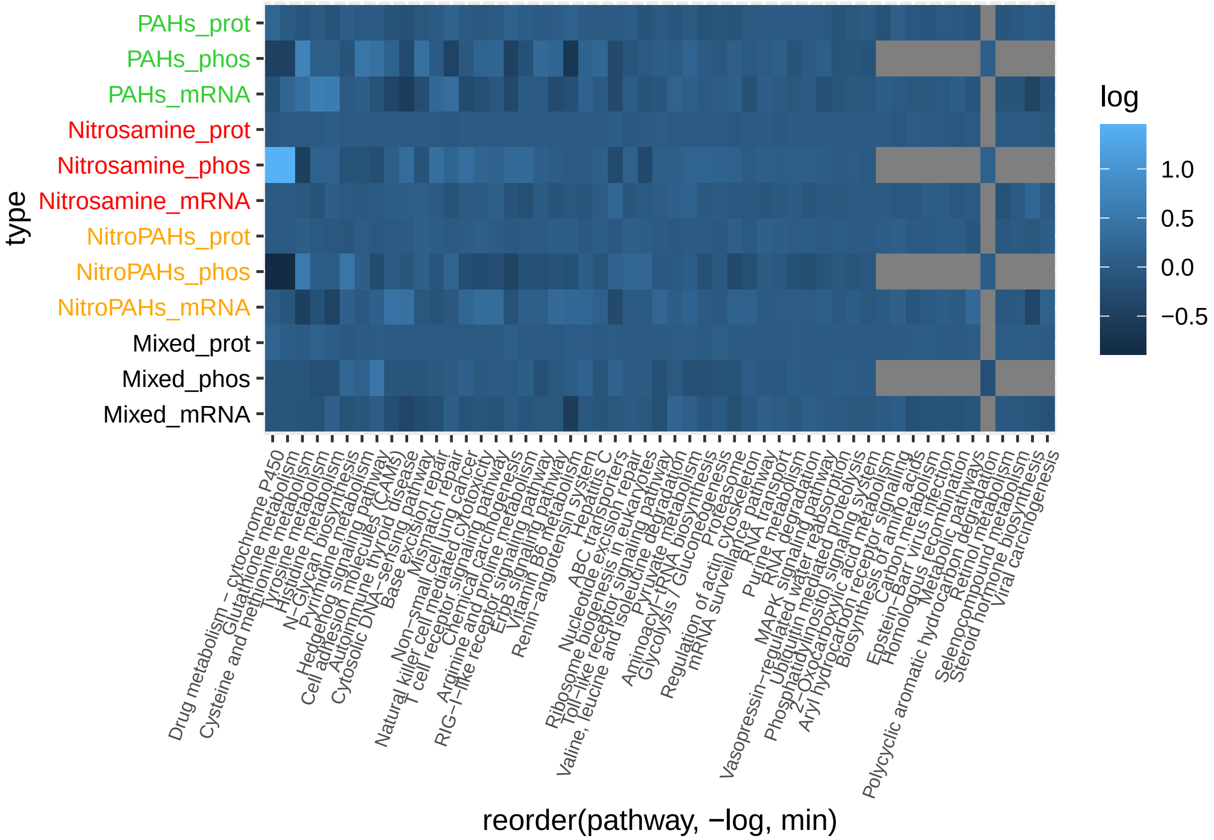
```
d %>% ggplot(aes(x = reorder(pathway, -log, min), y = type)) +
  geom_tile(aes(fill = log)) +
  theme(axis.text.x = element_text(angle = 70, hjust = 1, size = 7))
```



Assign the color for good visibility. So that the carcinogen on the y-axis can be easily seen for each type. Also adjust the font size.

```
d %>% ggplot(aes(x = reorder(pathway, -log, min), y = type)) +
  geom_tile(aes(fill = log)) +
  theme(axis.text.x = element_text(angle = 70, hjust = 1, size = 7),
        axis.text.y = element_text(hjust = 1, size = 9,
                                   colour = c("black", "black", "black",
                                              "orange", "orange", "orange",
                                              "red", "red", "red",
                                              "limegreen", "limegreen", "limegreen")))
```

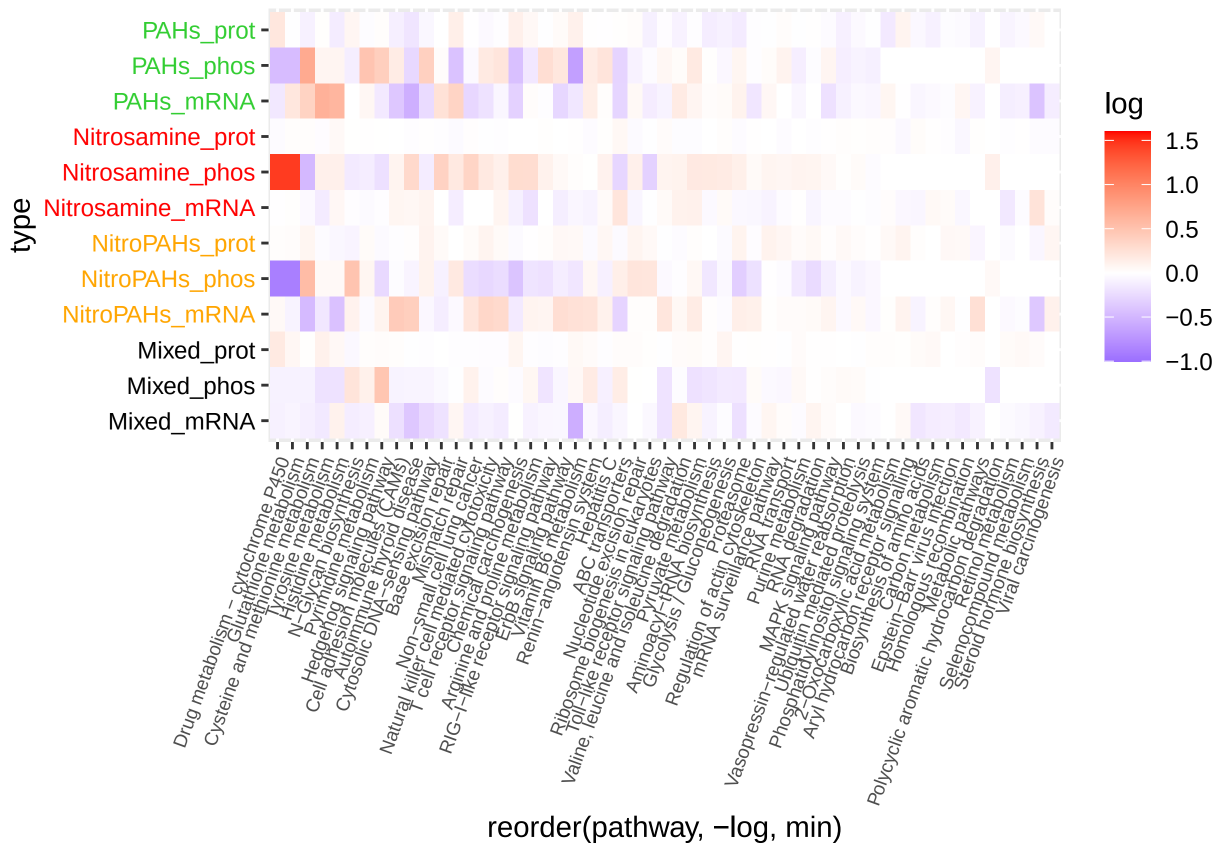
```
## Warning: Vectorized input to `element_text()` is not officially supported.
## Results may be unexpected or may change in future versions of ggplot2.
```



Change the color of heatmap to see the values easily, so the low-value was set to blue and the high-value to red. In this case, when the log value is 0, the value of Tumor and NAT is the same, and when the log value is negative, the NAT value is higher than Tumor, and vice versa. I set range using 'limits'. Values below 0 are marked in blue and over 0 are marked in red. The missing value 'NA' was marked in white.

```
d %>% ggplot(aes(x = reorder(pathway, -log, min), y = type)) +
  geom_tile(aes(fill = log)) +
  theme(axis.text.x = element_text(angle = 70, hjust = 1, size = 7),
        axis.text.y = element_text(hjust = 1, size = 9,
                                   colour = c("black", "black", "black",
                                              "orange", "orange", "orange",
                                              "red", "red", "red",
                                              "limegreen", "limegreen", "limegreen"))) +
  scale_fill_gradient2(midpoint = 0, low = "blue", high = "red",
                       limits = c(-1, 1.6), na.value = "white")
```

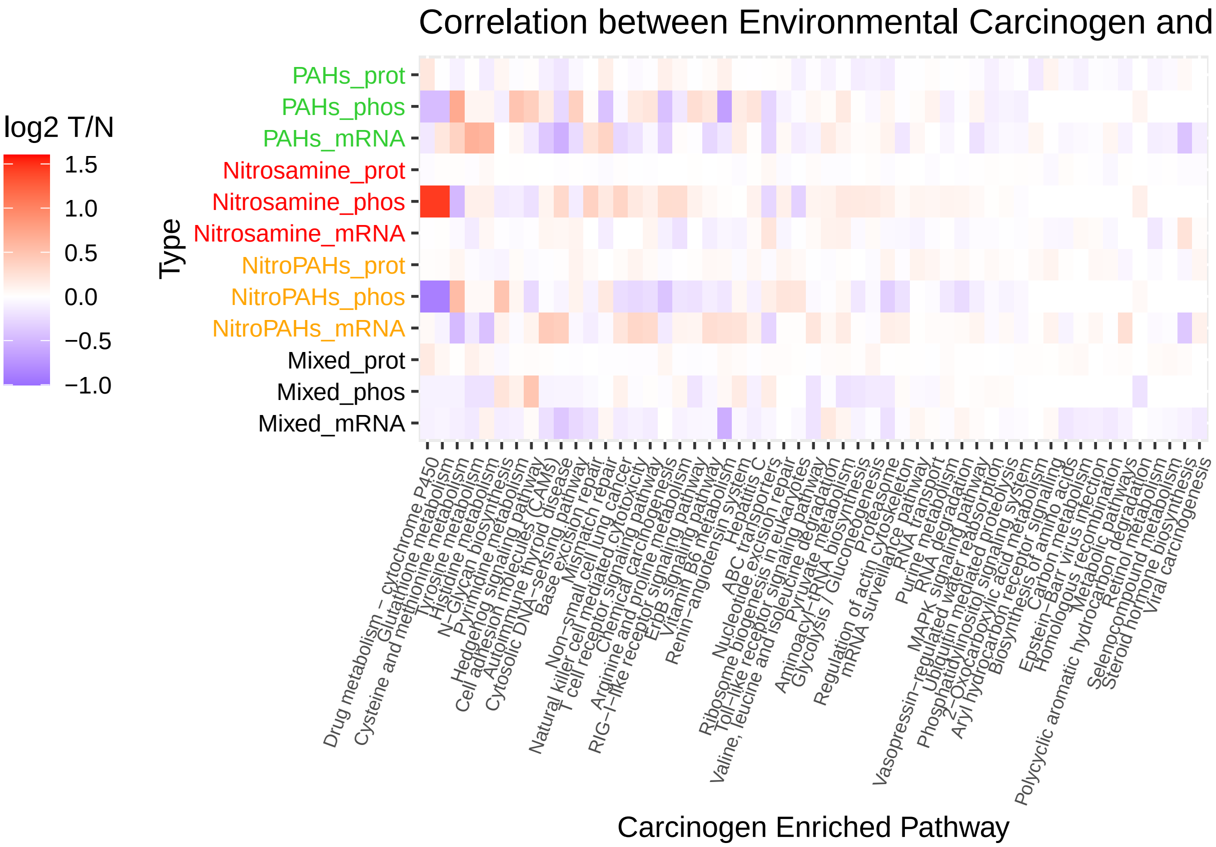
```
## Warning: Vectorized input to `element_text()` is not officially supported.
## Results may be unexpected or may change in future versions of ggplot2.
```



Name the title of the plot, the x- and y-axes, and the legend. Also change the position of legend to left.

```r
d %>% ggplot(aes(x = reorder(pathway, -log, min), y = type)) +
  geom_tile(aes(fill = log)) +
  theme(axis.text.x = element_text(angle = 70, hjust = 1, size = 7),
        axis.text.y = element_text(hjust = 1, size = 9,
                                   colour = c("black", "black", "black",
                                              "orange", "orange", "orange",
                                              "red", "red", "red",
                                              "limegreen", "limegreen", "limegreen"))) +
  scale_fill_gradient2(midpoint = 0, low = "blue", high = "red",
                       limits = c(-1, 1.6), na.value = "white") +
  labs(title = "Correlation between Environmental Carcinogen and Enriched Pathway",
       cex.main = 8,
       x = "Carcinogen Enriched Pathway",
       y = "Type",
       fill = "log2 T/N") +
  theme(legend.position="left")
```

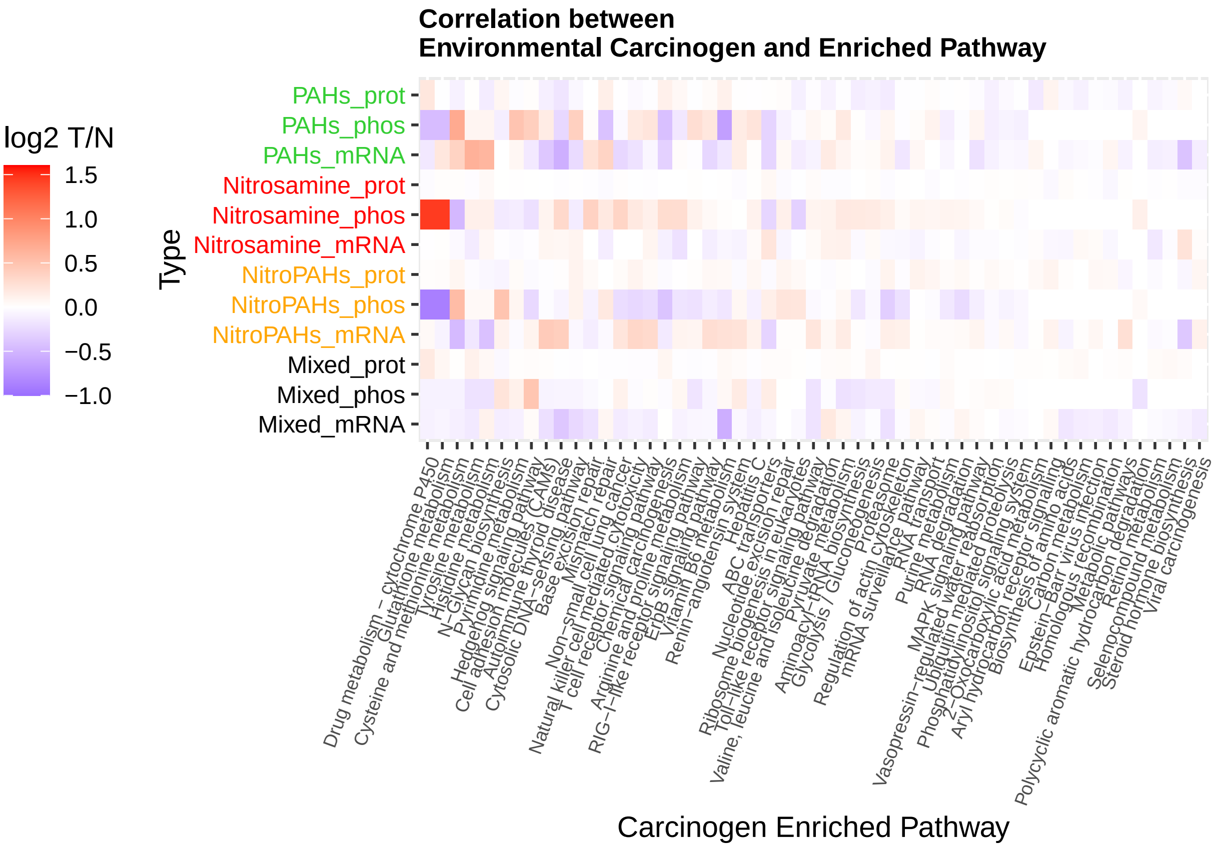
```
## Warning: Vectorized input to `element_text()` is not officially supported.
## Results may be unexpected or may change in future versions of ggplot2.
```



Since the title doesn't appear completely, adjust the size of the title.

```
d %>% ggplot(aes(x = reorder(pathway, -log, min), y = type)) +
  geom_tile(aes(fill = log)) +
  theme(axis.text.x = element_text(angle = 70, hjust = 1, size = 7),
        axis.text.y = element_text(hjust = 1, size = 9,
                                   colour = c("black", "black", "black",
                                              "orange", "orange", "orange",
                                              "red", "red", "red",
                                              "limegreen", "limegreen", "limegreen"))) +
  scale_fill_gradient2(midpoint = 0, low = "blue", high = "red",
                       limits = c(-1, 1.6), na.value = "white") +
  labs(title = "Correlation between \nEnvironmental Carcinogen and Enriched Pathway",
       cex.main = 8,
       x = "Carcinogen Enriched Pathway",
       y = "Type",
       fill = "log2 T/N") +
  theme(legend.position="left",
        plot.title = element_text(size=10, face="bold"))
```

```
## Warning: Vectorized input to `element_text()` is not officially supported.
## Results may be unexpected or may change in future versions of ggplot2.
```



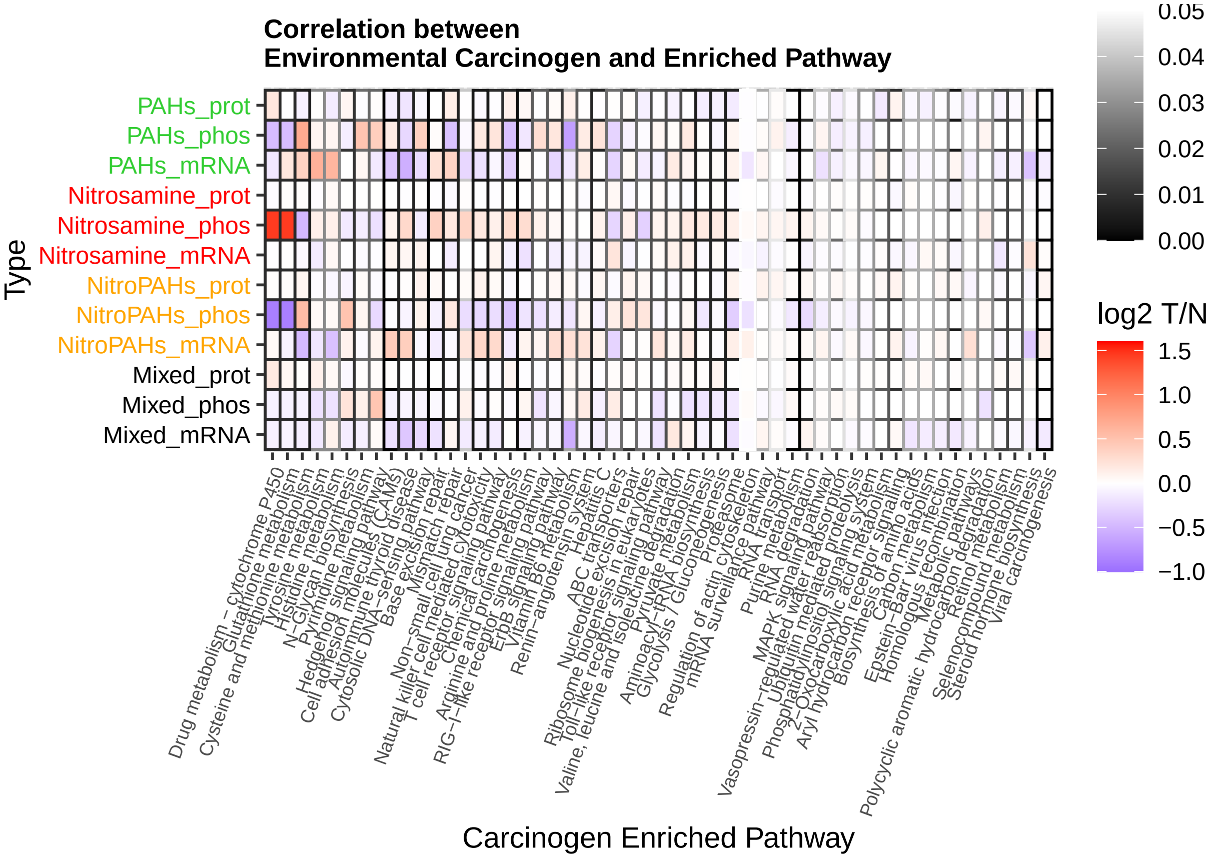
Next, I will mark oultines to compare p-values to see if the hypothesis that carcinogen affects to the enriched pathway is significant.

> As our statistical hypothesis will, by definition, state some property of the distribution, the null hypothesis is the default hypothesis under which that property does not exist. This hypothesis might specify the probability distribution of X precisely, or it might only specify that it belongs to some class of distributions. The p-value is used in the context of null hypothesis testing in order to quantify the statistical significance of a result, the result being the observed value of the chosen statistic T. The lower the p-value is, the lower the probability of getting that result if the null hypothesis were true. A result is said to be statistically significant if it allows us to reject the null hypothesis. All other things being equal, smaller p-values are taken as stronger evidence against the null hypothesis.

For typical analysis, using the standard = 0.05 cutoff, the null hypothesis is rejected when p <= 0.05 and not rejected when p > 0.05. So, I marked p = 0.05 as a reference point for significance.

```
d %>% filter(P <= 0.05) %>%
  ggplot(aes(x = reorder(pathway, -log, min), y = type)) +
  geom_tile(aes(fill = log, color = P), size = 0.5) +
  theme(axis.text.x = element_text(angle = 70, hjust = 1, size = 7),
        axis.text.y = element_text(hjust = 1, size = 9,
                                   colour = c("black", "black", "black",
                                              "orange", "orange", "orange",
                                              "red", "red", "red",
                                              "limegreen", "limegreen", "limegreen"))) +
  scale_fill_gradient2(midpoint = 0, low = "blue", high = "red",
                              limits = c(-1, 1.6), na.value = "white") +
  scale_color_gradient2(midpoint = 0.05, low = "black", high = "white",
                               limits = c(0, 0.05), na.value = "white") +
  labs(title = "Correlation between \nEnvironmental Carcinogen and Enriched Pathway",
       cex.main = 8,
       x = "Carcinogen Enriched Pathway",
       y = "Type",
       fill = "log2 T/N",
       color = "P-value") +
  theme(legend.position="right",
        plot.title = element_text(size=10, face="bold"))
```

```
## Warning: Vectorized input to `element_text()` is not officially supported.
## Results may be unexpected or may change in future versions of ggplot2.
```



```
ggsave('Figure 1.pdf', width = 7, height = 5, dpi = 300)
```

Next, I will visualize what proteins are enriched in the pathway that five carcinogen groups affecting. For this, I loaded two files by using 'rio' package. After loading, check the data whether it is proper or not.

```
e <- rio::import('https://ars.els-cdn.com/content/image/1-s2.0-S0092867420307431-mmc5.xlsx', sheet = 5)
f <- rio::import('https://ars.els-cdn.com/content/image/1-s2.0-S0092867420307431-mmc1.xlsx', sheet = 5)
head(e)
```

```
##   Uniprot     Gene KW.\r\np-value Rank Median s06 Rank Median s43
## 1  P27540     ARNT    0.467278280               6               1
## 2  O00170      AIP    0.148502871               5               6
## 3  P35869      AHR    0.643776691               5               1
## 4  P08238 HSP90AB1    0.030535120               6               5
## 5  Q15185   PTGES3    0.014916660               6               5
## 6  P11766     ADH5    0.003045171               6               3
##   Rank Median s52 Rank Median twmix Rank Median s36 Rank Median s33
## 1               2                 5               3               4
## 2               2                 1               4               3
## 3               6                 2               3               4
## 4               1                 2               4               3
## 5               1                 3               4               2
## 6               1                 2               4               5
##   Median T/N Ratio s06 Median T/N Ratio s43 Median T/N Ratio s52
## 1          -0.04949943            0.3592583           0.08636325
## 2           0.27833098            0.1824936           0.35820079
## 3           0.14877119            0.3157946           0.14041237
## 4           0.18876846            0.4339458           0.64531543
## 5           0.45426936            0.6055831           0.92123842
## 6          -0.32545915            0.3905531           0.48408212
##   Median T/N Ratio twmix Median T/N Ratio s36 Median T/N Ratio s33
## 1           -0.002325501           0.05150691          0.005782324
## 2            0.385569812           0.29798971          0.350368342
## 3            0.303291148           0.29989678          0.226947574
## 4            0.613702419           0.56963016          0.574279941
## 5            0.786082393           0.71275963          0.813761378
## 6            0.449827850           0.32422074          0.031469974
##
## 1
## 2
## 3
## 4 Antigen processing and presentation;Aryl hydrocarbon receptor signalling;Estrogen signaling pathway
## 5
## 6                                                   Carbon metabolism;Chemical carcinogenesis;Drug
##   Aryl hydrocarbon receptor signalling Chemical carcinogenesis
## 1                                    v                       v
## 2                                    v                    <NA>
## 3                                    v                    <NA>
## 4                                    v                    <NA>
## 5                                    v                    <NA>
## 6                                 <NA>                       v
##   Drug metabolism - cytochrome P450 NSCLC ErbB signaling pathway
## 1                              <NA>  <NA>                   <NA>
## 2                              <NA>  <NA>                   <NA>
## 3                              <NA>  <NA>                   <NA>
## 4                              <NA>  <NA>                   <NA>
## 5                              <NA>  <NA>                   <NA>
## 6                                 v  <NA>                   <NA>
##   MAPK signaling pathway
## 1                   <NA>
## 2                   <NA>
## 3                   <NA>
## 4                   <NA>
## 5                   <NA>
## 6                   <NA>
```

```
head(f)
```

```
##       gene ensembl_gene_id  Median   P002   P006   P007   P009   P010   P011
## 1   TSPAN6 ENSG00000000003  1.1720  1.175 -1.617  3.284  1.224  0.161 -0.426
## 2     TNMD ENSG00000000005 -0.5215 -0.477 -0.487 -0.068 -0.997 -0.473 -1.187
## 3     DPM1 ENSG00000000419 -0.1150  0.027 -0.082  0.102 -0.417 -0.751 -0.526
## 4    SCYL3 ENSG00000000457  0.4220  0.770  0.087  0.839  0.730  0.152  0.932
## 5 C1orf112 ENSG00000000460  0.4820  0.910  0.311  1.204  0.811  0.822  0.605
## 6      FGR ENSG00000000938 -1.5760 -0.242  2.329 -3.609 -3.323  0.643 -1.521
##     P012   P013   P015   P016   P017   P018   P019   P020   P021   P022   P023
## 1  1.402  2.031  0.320  1.398  1.840  0.651  0.186  1.380  1.330  0.828 -0.338
## 2 -0.631 -0.709 -0.077 -0.621  1.406  0.052 -0.556  0.276 -1.705 -1.484  0.538
## 3 -0.552  1.211 -0.388 -0.169  1.812 -0.573  0.090  0.477  0.284 -0.142 -1.168
## 4  1.253  0.508  0.715  0.144  0.916  0.808  0.173 -0.212 -0.303  0.893  0.813
## 5  2.007  0.461  0.145  0.221  0.434  0.279  0.038  0.089 -0.428  0.625  0.534
## 6 -2.328 -1.963 -1.295 -1.159 -4.960 -0.749 -1.244 -1.564 -1.288 -1.113 -1.343
##     P024   P025   P026   P027   P028   P029   P030   P031   P032   P033   P034
## 1  0.021  0.111  0.658 -0.782  1.278  1.620  1.644  1.762 -0.311  0.003 -0.788
## 2 -2.825 -1.019  0.706 -1.057 -1.561  4.394  0.990  3.040 -1.070  0.357 -1.110
## 3  0.045 -0.143 -0.314 -1.199 -0.017  0.513  0.117  0.293  0.053 -0.701 -0.438
## 4  0.474 -0.018  0.867  0.360  0.625  0.311 -0.086 -0.212  1.538  1.230 -0.439
## 5  0.260 -1.024  0.587 -0.122  0.594  1.463  0.391 -0.083  1.158  2.314  0.024
## 6 -2.605 -1.815 -1.062 -0.650 -1.585 -2.149 -1.975 -1.202 -4.374 -1.957 -3.333
##     P036   P037   P038   P039   P040   P042   P043   P044   P045   P048   P049
## 1 -0.802  1.261  0.765  1.496  1.526  2.829  0.875  1.710  1.043  0.981  1.363
## 2 -2.712  1.201 -1.119 -2.419 -1.901  3.241 -3.362 -1.671 -0.783  0.752  1.758
## 3 -0.124 -0.483  0.035  0.330 -0.367  0.477 -1.336  0.018  0.918 -0.287 -0.639
## 4  1.115  0.180 -0.239  0.453  0.516 -0.037  1.040  0.629  0.290  0.122  0.155
## 5  1.011 -0.314 -0.553  0.622  0.878 -0.471 -0.425  0.245  0.939  0.264 -0.062
## 6 -2.094 -2.534 -0.479 -1.805 -1.452 -5.345 -0.673 -0.864 -2.241 -2.046 -1.752
##     P050   P051   P052   P053   P054   P055   P056   P057   P058   P059   P060
## 1  1.408  1.597  3.427  1.463 -1.571  0.376  1.050  2.037  0.442  1.073  0.559
## 2  3.032  0.831  0.073  0.146 -2.712 -1.803  2.193 -0.928 -2.184 -2.009  1.824
## 3 -0.484  0.472  1.957  0.214 -0.574  0.386 -0.548  0.682  0.770  0.260 -0.138
## 4  0.490  0.265  0.378 -0.217  0.997  1.375  1.125 -0.114 -0.487  0.363 -0.441
## 5  1.099  0.154  0.736  0.166  1.111  1.678  0.092  0.308  1.961  0.227 -0.504
## 6 -3.417 -1.624 -2.037 -0.380 -3.196 -3.154 -2.566 -1.189 -1.537 -1.567 -0.130
##     P061   P062   P063   P064   P066   P067   P068   P070   P071   P072   P073
## 1  2.590  2.077  2.501  1.864  1.672 -0.753  1.496  0.997  0.938  1.702  0.840
## 2  1.512  0.207 -1.072  0.555 -1.807 -0.261 -0.571  1.587 -1.626  1.004 -1.599
## 3 -0.125 -0.274  0.676  0.262  0.455 -0.400 -0.341 -0.658 -0.233 -0.832  0.374
## 4 -0.643  0.783  0.718  0.672 -0.551 -0.027 -0.873  0.391  0.529  0.004  1.333
## 5  2.559  1.876  0.409  0.841 -0.447  0.040 -0.445 -0.495 -0.145  0.758  0.954
## 6 -2.268 -1.533 -1.195 -2.614  0.084  0.508 -1.246 -2.263 -0.053 -1.146 -3.279
##     P074   P075   P076   P077   P080   P081   P082   P085   P086   P088   P089
## 1  1.169  1.769  0.627  2.356  0.393  1.914  0.454  0.676  2.389  3.439  1.314
## 2  0.810 -0.403 -2.052 -0.901 -2.942 -1.084  0.315  0.000  1.134 -0.737  1.268
## 3  0.201  0.394 -0.745  1.010 -1.612 -0.515 -0.399 -0.876 -0.190 -0.187 -0.106
## 4  0.665  1.581  0.842 -0.616  1.155  0.093  0.470  0.301 -0.060  0.488  1.096
## 5  0.411  2.323  0.383 -0.057  0.881  0.531  0.159  0.985  0.203 -0.018  0.836
## 6 -1.484 -2.741 -2.430 -1.111 -3.216 -1.557 -0.677 -0.918 -2.640 -2.577 -2.804
##     P090   P091   P092   P093   P094   P095   P097   P098   P099   P100   P101
## 1  1.102 -1.457 -0.679  0.600  1.622  1.556 -0.040  1.231  1.931  0.878  0.830
## 2 -1.207 -0.238 -1.948 -2.488  1.090 -3.273  0.399 -0.634 -0.403 -1.756 -2.279
## 3  0.669 -0.163  0.422 -0.038  0.120 -0.190  0.013  0.435 -0.340  0.120  0.257
## 4  1.122  1.056 -0.491  0.390  0.309  0.151  0.004  1.303 -0.112  1.035  0.137
## 5  1.268  1.084 -0.191  0.655  0.788  1.216  0.391  1.258 -0.024  0.734  1.002
## 6 -3.769 -2.063 -0.821 -0.102 -0.460 -1.495 -1.023 -2.338 -2.385 -2.731 -1.301
##     P102   P103   P104   P109   P110   P111   P112
## 1  3.156  0.732  2.359 -0.852  1.046  2.443  2.222
## 2 -0.379  0.879  0.588 -2.345 -1.090  0.691  3.135
## 3  0.339 -0.532  0.049 -1.682  0.311 -0.794 -0.353
## 4  0.173  0.457 -0.475  0.867  0.917 -0.236  0.892
## 5  0.535  0.829  0.556  0.293  0.741  0.009  0.503
## 6 -1.122 -1.720 -0.846  2.172 -1.897 -1.643 -2.670
```

The column names are proper, so I will extract some data that could be of use. I will use enriched pathways in data 'e', and log values of patient in data 'f'.

```
e <- e[, c(2, 17:22)]
f <- f[, c(1, 4:93)]
head(e)
```

```
##       Gene Aryl hydrocarbon receptor signalling Chemical carcinogenesis
## 1     ARNT                                    v                       v
## 2      AIP                                    v                    <NA>
## 3      AHR                                    v                    <NA>
## 4 HSP90AB1                                    v                    <NA>
## 5   PTGES3                                    v                    <NA>
## 6     ADH5                                 <NA>                       v
##   Drug metabolism - cytochrome P450 NSCLC ErbB signaling pathway
## 1                              <NA>  <NA>                   <NA>
## 2                              <NA>  <NA>                   <NA>
## 3                              <NA>  <NA>                   <NA>
## 4                              <NA>  <NA>                   <NA>
## 5                              <NA>  <NA>                   <NA>
## 6                                 v  <NA>                   <NA>
##   MAPK signaling pathway
## 1                   <NA>
## 2                   <NA>
## 3                   <NA>
## 4                   <NA>
## 5                   <NA>
## 6                   <NA>
```

```
head(f)
```

```
##       gene   P002   P006   P007   P009   P010   P011   P012   P013   P015
## 1   TSPAN6  1.175 -1.617  3.284  1.224  0.161 -0.426  1.402  2.031  0.320
## 2     TNMD -0.477 -0.487 -0.068 -0.997 -0.473 -1.187 -0.631 -0.709 -0.077
## 3     DPM1  0.027 -0.082  0.102 -0.417 -0.751 -0.526 -0.552  1.211 -0.388
## 4    SCYL3  0.770  0.087  0.839  0.730  0.152  0.932  1.253  0.508  0.715
## 5 C1orf112  0.910  0.311  1.204  0.811  0.822  0.605  2.007  0.461  0.145
## 6      FGR -0.242  2.329 -3.609 -3.323  0.643 -1.521 -2.328 -1.963 -1.295
##      P016   P017   P018   P019   P020   P021   P022   P023   P024   P025   P026
## 1  1.398  1.840  0.651  0.186  1.380  1.330  0.828 -0.338  0.021  0.111  0.658
## 2 -0.621  1.406  0.052 -0.556  0.276 -1.705 -1.484  0.538 -2.825 -1.019  0.706
## 3 -0.169  1.812 -0.573  0.090  0.477  0.284 -0.142 -1.168  0.045 -0.143 -0.314
## 4  0.144  0.916  0.808  0.173 -0.212 -0.303  0.893  0.813  0.474 -0.018  0.867
## 5  0.221  0.434  0.279  0.038  0.089 -0.428  0.625  0.534  0.260 -1.024  0.587
## 6 -1.159 -4.960 -0.749 -1.244 -1.564 -1.288 -1.113 -1.343 -2.605 -1.815 -1.062
##      P027   P028   P029   P030   P031   P032   P033   P034   P036   P037   P038
## 1 -0.782  1.278  1.620  1.644  1.762 -0.311  0.003 -0.788 -0.802  1.261  0.765
## 2 -1.057 -1.561  4.394  0.990  3.040 -1.070  0.357 -1.110 -2.712  1.201 -1.119
## 3 -1.199 -0.017  0.513  0.117  0.293  0.053 -0.701 -0.438 -0.124 -0.483  0.035
## 4  0.360  0.625  0.311 -0.086 -0.212  1.538  1.230 -0.439  1.115  0.180 -0.239
## 5 -0.122  0.594  1.463  0.391 -0.083  1.158  2.314  0.024  1.011 -0.314 -0.553
## 6 -0.650 -1.585 -2.149 -1.975 -1.202 -4.374 -1.957 -3.333 -2.094 -2.534 -0.479
##      P039   P040   P042   P043   P044   P045   P048   P049   P050   P051   P052
## 1  1.496  1.526  2.829  0.875  1.710  1.043  0.981  1.363  1.408  1.597  3.427
## 2 -2.419 -1.901  3.241 -3.362 -1.671 -0.783  0.752  1.758  3.032  0.831  0.073
## 3  0.330 -0.367  0.477 -1.336  0.018  0.918 -0.287 -0.639 -0.484  0.472  1.957
## 4  0.453  0.516 -0.037  1.040  0.629  0.290  0.122  0.155  0.490  0.265  0.378
## 5  0.622  0.878 -0.471 -0.425  0.245  0.939  0.264 -0.062  1.099  0.154  0.736
## 6 -1.805 -1.452 -5.345 -0.673 -0.864 -2.241 -2.046 -1.752 -3.417 -1.624 -2.037
##      P053   P054   P055   P056   P057   P058   P059   P060   P061   P062   P063
## 1  1.463 -1.571  0.376  1.050  2.037  0.442  1.073  0.559  2.590  2.077  2.501
## 2  0.146 -2.712 -1.803  2.193 -0.928 -2.184 -2.009  1.824  1.512  0.207 -1.072
## 3  0.214 -0.574  0.386 -0.548  0.682  0.770  0.260 -0.138 -0.125 -0.274  0.676
## 4 -0.217  0.997  1.375  1.125 -0.114 -0.487  0.363 -0.441 -0.643  0.783  0.718
## 5  0.166  1.111  1.678  0.092  0.308  1.961  0.227 -0.504  2.559  1.876  0.409
## 6 -0.380 -3.196 -3.154 -2.566 -1.189 -1.537 -1.567 -0.130 -2.268 -1.533 -1.195
##      P064   P066   P067   P068   P070   P071   P072   P073   P074   P075   P076
## 1  1.864  1.672 -0.753  1.496  0.997  0.938  1.702  0.840  1.169  1.769  0.627
## 2  0.555 -1.807 -0.261 -0.571  1.587 -1.626  1.004 -1.599  0.810 -0.403 -2.052
## 3  0.262  0.455 -0.400 -0.341 -0.658 -0.233 -0.832  0.374  0.201  0.394 -0.745
## 4  0.672 -0.551 -0.027 -0.873  0.391  0.529  0.004  1.333  0.665  1.581  0.842
## 5  0.841 -0.447  0.040 -0.445 -0.495 -0.145  0.758  0.954  0.411  2.323  0.383
## 6 -2.614  0.084  0.508 -1.246 -2.263 -0.053 -1.146 -3.279 -1.484 -2.741 -2.430
##      P077   P080   P081   P082   P085   P086   P088   P089   P090   P091   P092
## 1  2.356  0.393  1.914  0.454  0.676  2.389  3.439  1.314  1.102 -1.457 -0.679
## 2 -0.901 -2.942 -1.084  0.315  0.000  1.134 -0.737  1.268 -1.207 -0.238 -1.948
## 3  1.010 -1.612 -0.515 -0.399 -0.876 -0.190 -0.187 -0.106  0.669 -0.163  0.422
## 4 -0.616  1.155  0.093  0.470  0.301 -0.060  0.488  1.096  1.122  1.056 -0.491
## 5 -0.057  0.881  0.531  0.159  0.985  0.203 -0.018  0.836  1.268  1.084 -0.191
## 6 -1.111 -3.216 -1.557 -0.677 -0.918 -2.640 -2.577 -2.804 -3.769 -2.063 -0.821
##      P093   P094   P095   P097   P098   P099   P100   P101   P102   P103   P104
## 1  0.600  1.622  1.556 -0.040  1.231  1.931  0.878  0.830  3.156  0.732  2.359
## 2 -2.488  1.090 -3.273  0.399 -0.634 -0.403 -1.756 -2.279 -0.379  0.879  0.588
## 3 -0.038  0.120 -0.190  0.013  0.435 -0.340  0.120  0.257  0.339 -0.532  0.049
## 4  0.390  0.309  0.151  0.004  1.303 -0.112  1.035  0.137  0.173  0.457 -0.475
## 5  0.655  0.788  1.216  0.391  1.258 -0.024  0.734  1.002  0.535  0.829  0.556
## 6 -0.102 -0.460 -1.495 -1.023 -2.338 -2.385 -2.731 -1.301 -1.122 -1.720 -0.846
##      P109   P110   P111   P112
## 1 -0.852  1.046  2.443  2.222
## 2 -2.345 -1.090  0.691  3.135
## 3 -1.682  0.311 -0.794 -0.353
## 4  0.867  0.917 -0.236  0.892
## 5  0.293  0.741  0.009  0.503
## 6  2.172 -1.897 -1.643 -2.670
```

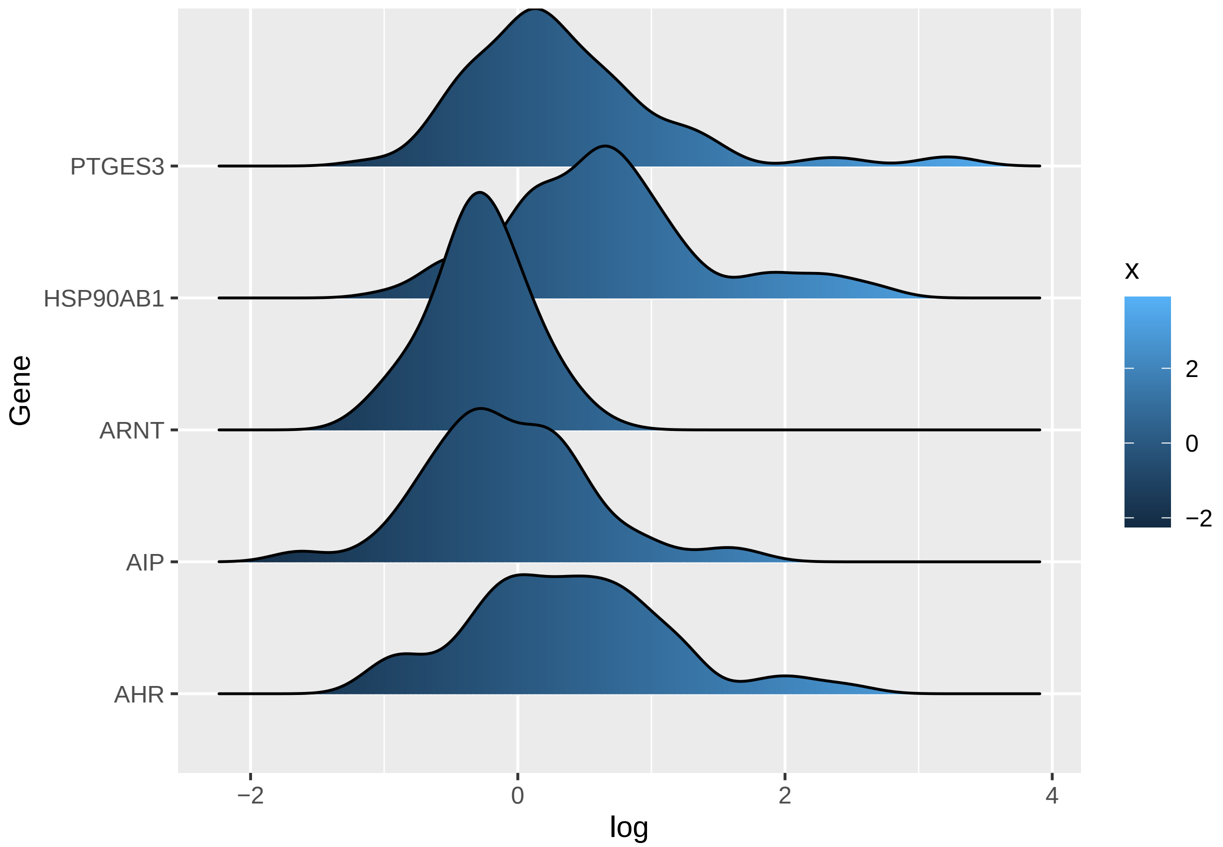
Two data sets will be merged to draw the plots. For this, there will have to be some standard. In this case, gene name can be the 'merge point'. But each column names are different, so I will rename the column name and merge them. And for handling data easily, other column names are also changed and columns are rearranged in data 'e' and 'f'.

```
e <- e %>% rename(Aryl = 'Aryl hydrocarbon receptor signalling',
                  Chem = 'Chemical carcinogenesis',
                  Drug = 'Drug metabolism - cytochrome P450',
                  ErbB = 'ErbB signaling pathway',
                  MAPK = 'MAPK signaling pathway')
f <- f %>% rename(Gene = gene) %>%
  gather(key = "patient", value = "log", -c("Gene"))

nd <- merge(e, f, by = "Gene")
```

Because I will use log values data to see genes' upregulation or downregulation, 'geom_density_ridges_gradient' function will be used. At first, use 'filter' function to extract 'Aryl Hydrocarbon Receptor Signalling'.

```
nd %>% filter(Aryl == 'v') %>%
  select(-c(2:7)) %>%
  ggplot(aes(log, Gene)) +
  geom_density_ridges_gradient(aes(fill = stat(x)))
```

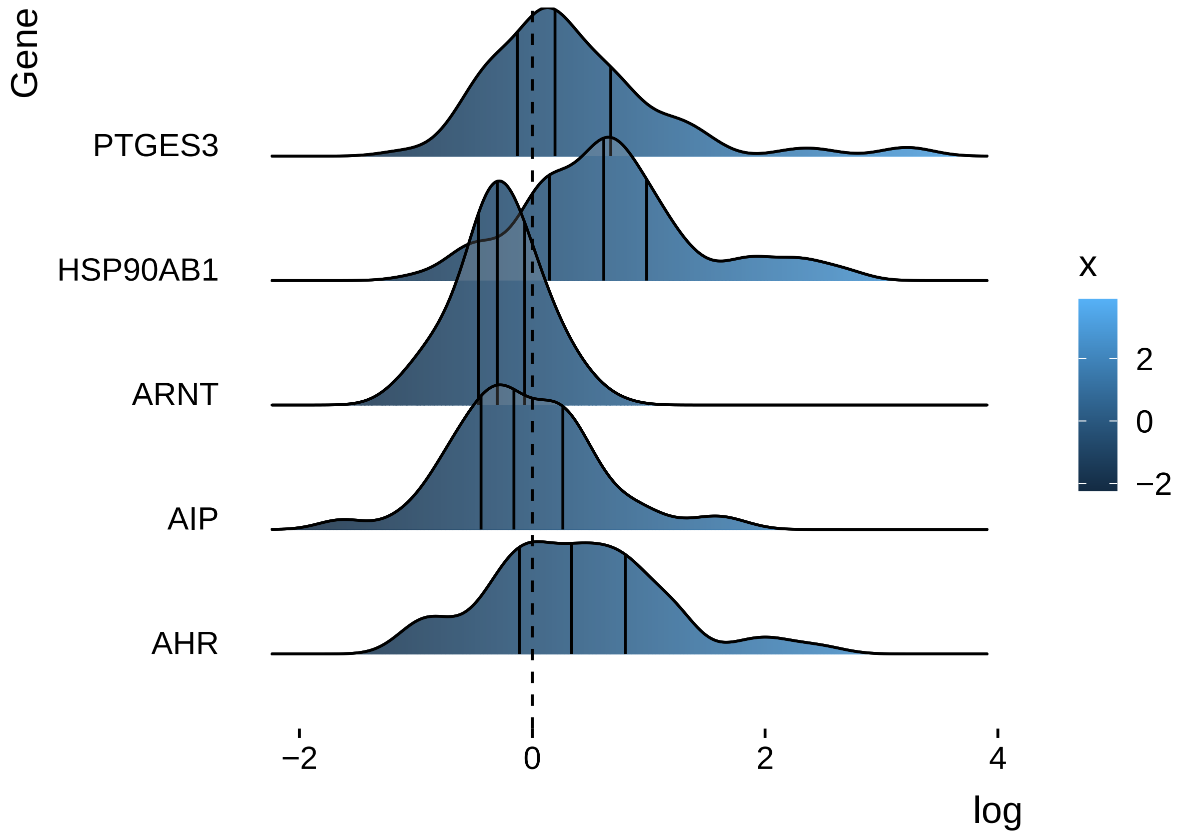
```
## Picking joint bandwidth of 0.199
```



Adjust theme and add quantile lines to see distribution of data, and also add vline to see the value 'zero'.

```
nd %>% filter(Aryl == 'v') %>%
  select(-c(2:7)) %>%
  ggplot(aes(log, Gene)) +
  geom_density_ridges_gradient(aes(fill = stat(x))) +
  theme_ridges(grid = F) +
  stat_density_ridges(quantile_lines = T, quantiles = c(0.25, 0.5, 0.75), alpha = 0.2) +
  geom_vline(xintercept = 0, col = "black", linetype = "dashed")
```

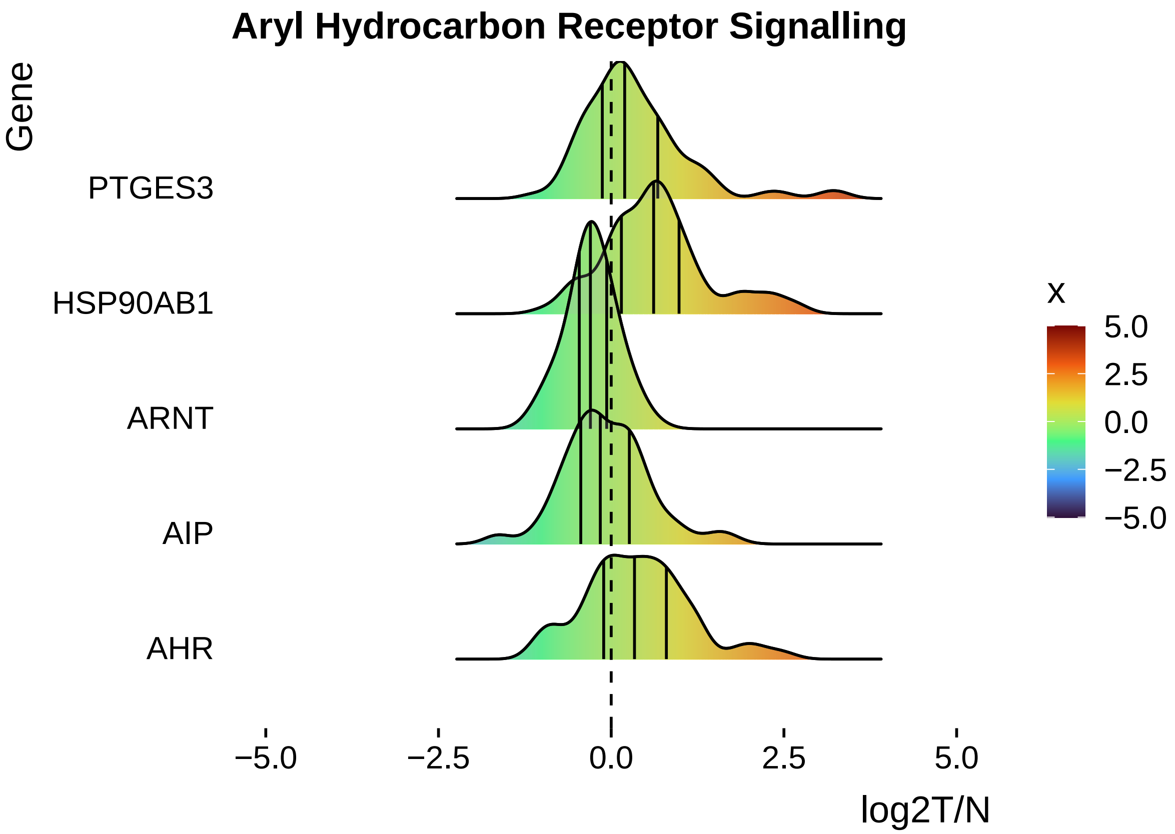
```
## Picking joint bandwidth of 0.199
## Picking joint bandwidth of 0.199
```



To improve readability, change the color and add titles.

```
nd %>% filter(Aryl == 'v') %>%
  select(-c(2:7)) %>%
  ggplot(aes(log, Gene)) +
  geom_density_ridges_gradient(aes(fill = stat(x))) +
  theme_ridges(grid = F) +
  stat_density_ridges(quantile_lines = T, quantiles = c(0.25, 0.5, 0.75), alpha = 0.2) +
  geom_vline(xintercept = 0, col = "black", linetype = "dashed") +
  scale_fill_viridis_c(option = "turbo", limits = c(-5, 5)) +
  xlim(-5, 5) +
  labs(x = "log2T/N",
       y = "Gene",
       title = "Aryl Hydrocarbon Receptor Signalling")
```

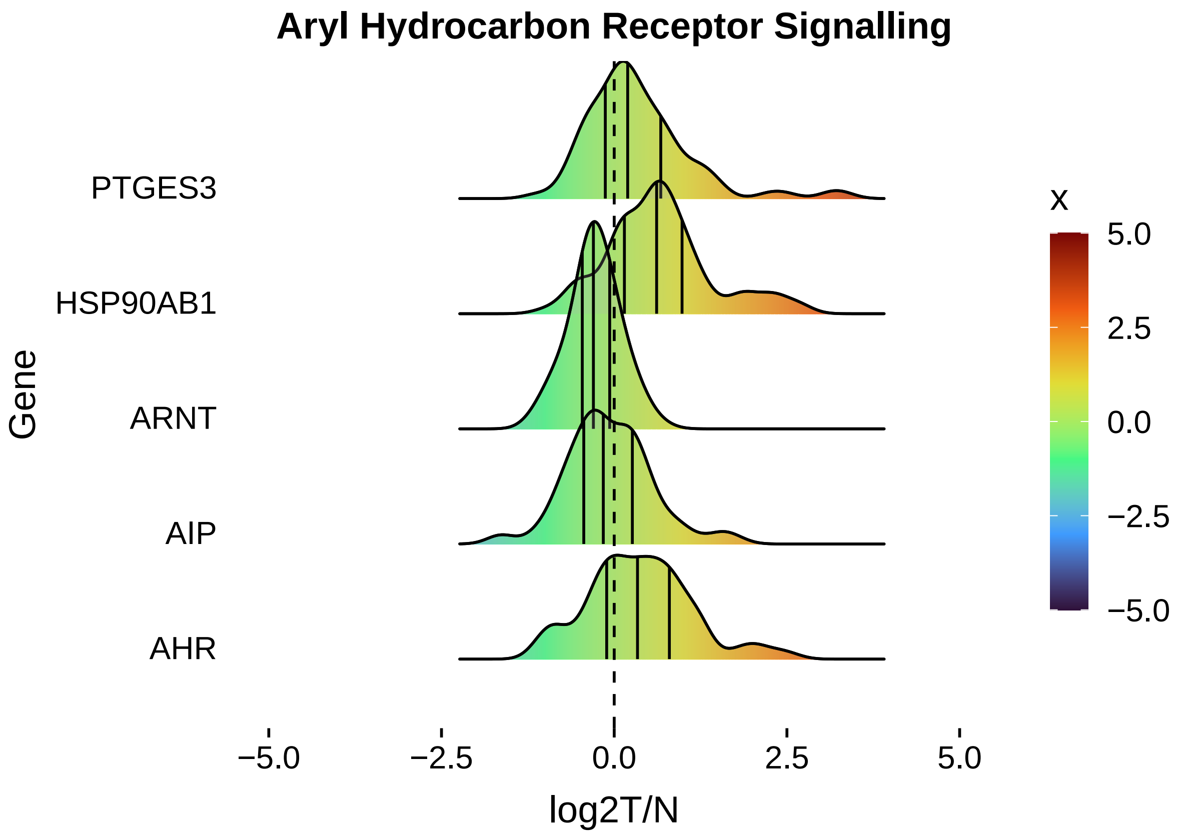
```
## Picking joint bandwidth of 0.199
## Picking joint bandwidth of 0.199
```



Finally, modify the positions and some details of plot.

```
nd %>% filter(Aryl == 'v') %>%
  select(-c(2:7)) %>%
  ggplot(aes(log, Gene)) +
  geom_density_ridges_gradient(aes(fill = stat(x))) +
  theme_ridges(grid = F) +
  stat_density_ridges(quantile_lines = T, quantiles = c(0.25, 0.5, 0.75), alpha = 0.2) +
  geom_vline(xintercept = 0, col = "black", linetype = "dashed") +
  scale_fill_viridis_c(option = "turbo", limits = c(-5, 5)) +
  xlim(-5, 5) +
  labs(x = "log2T/N",
       y = "Gene",
       title = "Aryl Hydrocarbon Receptor Signalling") +
  theme(axis.title.x = element_text(hjust = 0.5),
        axis.title.y = element_text(hjust = 0.5),
        plot.title = element_text(hjust = 0.5),
        legend.key.height = unit(1, 'cm'))
```

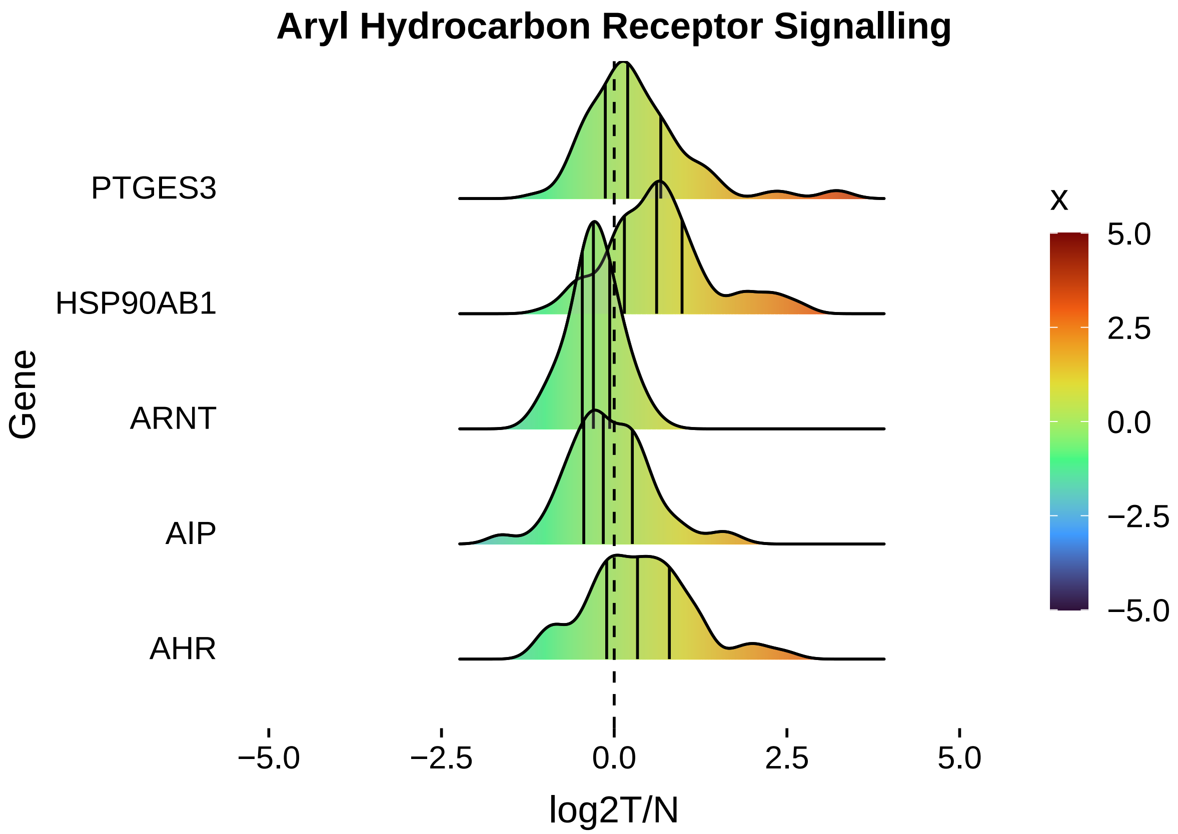
```
## Picking joint bandwidth of 0.199
## Picking joint bandwidth of 0.199
```



At the same way, draw other plots and save them respectively.

```
Aryl <- nd %>% filter(Aryl == 'v') %>%
  select(-c(2:7)) %>%
  ggplot(aes(log, Gene)) +
  geom_density_ridges_gradient(aes(fill = stat(x))) +
  theme_ridges(grid = F) +
  stat_density_ridges(quantile_lines = T, quantiles = c(0.25, 0.5, 0.75), alpha = 0.2) +
  geom_vline(xintercept = 0, col = "black", linetype = "dashed") +
  scale_fill_viridis_c(option = "turbo", limits = c(-5, 5)) +
  xlim(-5, 5) +
  labs(x = "log2T/N",
       y = "Gene",
       title = "Aryl Hydrocarbon Receptor Signalling") +
  theme(axis.title.x = element_text(hjust = 0.5),
        axis.title.y = element_text(hjust = 0.5),
        plot.title = element_text(hjust = 0.5),
        legend.key.height = unit(1, 'cm'))

Chem <- nd %>% filter(Chem == 'v') %>%
  select(-c(2:7)) %>%
  ggplot(aes(log, Gene)) +
  geom_density_ridges_gradient(aes(fill = stat(x))) +
  theme_ridges(grid = F) +
  stat_density_ridges(quantile_lines = T, quantiles = c(0.25, 0.5, 0.75), alpha = 0.2) +
  geom_vline(xintercept = 0, col = "black", linetype = "dashed") +
  scale_fill_viridis_c(option = "turbo", limits = c(-5, 5)) +
  xlim(-5, 5) +
  labs(x = "log2T/N",
       y = "Gene",
       title = "Chemical Carcinogenesis") +
  theme(axis.title.x = element_text(hjust = 0.5),
        axis.title.y = element_text(hjust = 0.5),
        plot.title = element_text(hjust = 0.5),
        legend.key.height = unit(1, 'cm'))

Drug <- nd %>% filter(Drug == 'v') %>%
  select(-c(2:7)) %>%
  ggplot(aes(log, Gene)) +
  geom_density_ridges_gradient(aes(fill = stat(x))) +
  theme_ridges(grid = F) +
  stat_density_ridges(quantile_lines = T, quantiles = c(0.25, 0.5, 0.75), alpha = 0.2) +
  geom_vline(xintercept = 0, col = "black", linetype = "dashed") +
  scale_fill_viridis_c(option = "turbo", limits = c(-5, 5)) +
  xlim(-5, 5) +
  labs(x = "log2T/N",
       y = "Gene",
       title = "Drug Metabolism \n- cytochrome 450") +
  theme(axis.title.x = element_text(hjust = 0.5),
        axis.title.y = element_text(hjust = 0.5),
        plot.title = element_text(hjust = 0.5),
        legend.key.height = unit(1, 'cm'))

NSCLC <- nd %>% filter(NSCLC == 'v') %>%
  select(-c(2:7)) %>%
  ggplot(aes(log, Gene)) +
  geom_density_ridges_gradient(aes(fill = stat(x))) +
  theme_ridges(grid = F) +
  stat_density_ridges(quantile_lines = T, quantiles = c(0.25, 0.5, 0.75), alpha = 0.2) +
  geom_vline(xintercept = 0, col = "black", linetype = "dashed") +
  scale_fill_viridis_c(option = "turbo", limits = c(-5, 5)) +
  xlim(-5, 5) +
  labs(x = "log2T/N",
       y = "Gene",
       title = "Non-Small Cell Lung Cancer") +
  theme(axis.title.x = element_text(hjust = 0.5),
        axis.title.y = element_text(hjust = 0.5),
        plot.title = element_text(hjust = 0.5),
        legend.key.height = unit(1, 'cm'))

ErbB <- nd %>% filter(ErbB == 'v') %>%
  select(-c(2:7)) %>%
  ggplot(aes(log, Gene)) +
  geom_density_ridges_gradient(aes(fill = stat(x))) +
  theme_ridges(grid = F) +
  stat_density_ridges(quantile_lines = T, quantiles = c(0.25, 0.5, 0.75), alpha = 0.2) +
  geom_vline(xintercept = 0, col = "black", linetype = "dashed") +
  scale_fill_viridis_c(option = "turbo", limits = c(-5, 5)) +
  xlim(-5, 5) +
  labs(x = "log2T/N",
       y = "Gene",
       title = "ErbB Signaling Pathway") +
  theme(axis.title.x = element_text(hjust = 0.5),
        axis.title.y = element_text(hjust = 0.5),
        plot.title = element_text(hjust = 0.5),
        legend.key.height = unit(1, 'cm'))

MAPK <- nd %>% filter(MAPK == 'v') %>%
  select(-c(2:7)) %>%
  ggplot(aes(log, Gene)) +
  geom_density_ridges_gradient(aes(fill = stat(x))) +
  theme_ridges(grid = F) +
  stat_density_ridges(quantile_lines = T, quantiles = c(0.25, 0.5, 0.75), alpha = 0.2) +
  geom_vline(xintercept = 0, col = "black", linetype = "dashed") +
  scale_fill_viridis_c(option = "turbo", limits = c(-5, 5)) +
  xlim(-5, 5) +
  labs(x = "log2T/N",
       y = "Gene",
       title = "MAPK Signaling pathway") +
  theme(axis.title.x = element_text(hjust = 0.5),
        axis.title.y = element_text(hjust = 0.5),
        plot.title = element_text(hjust = 0.5),
        axis.text.y = element_text(size = 8),
        legend.key.height = unit(1, 'cm'))

Aryl
```

```
## Picking joint bandwidth of 0.199
## Picking joint bandwidth of 0.199
```

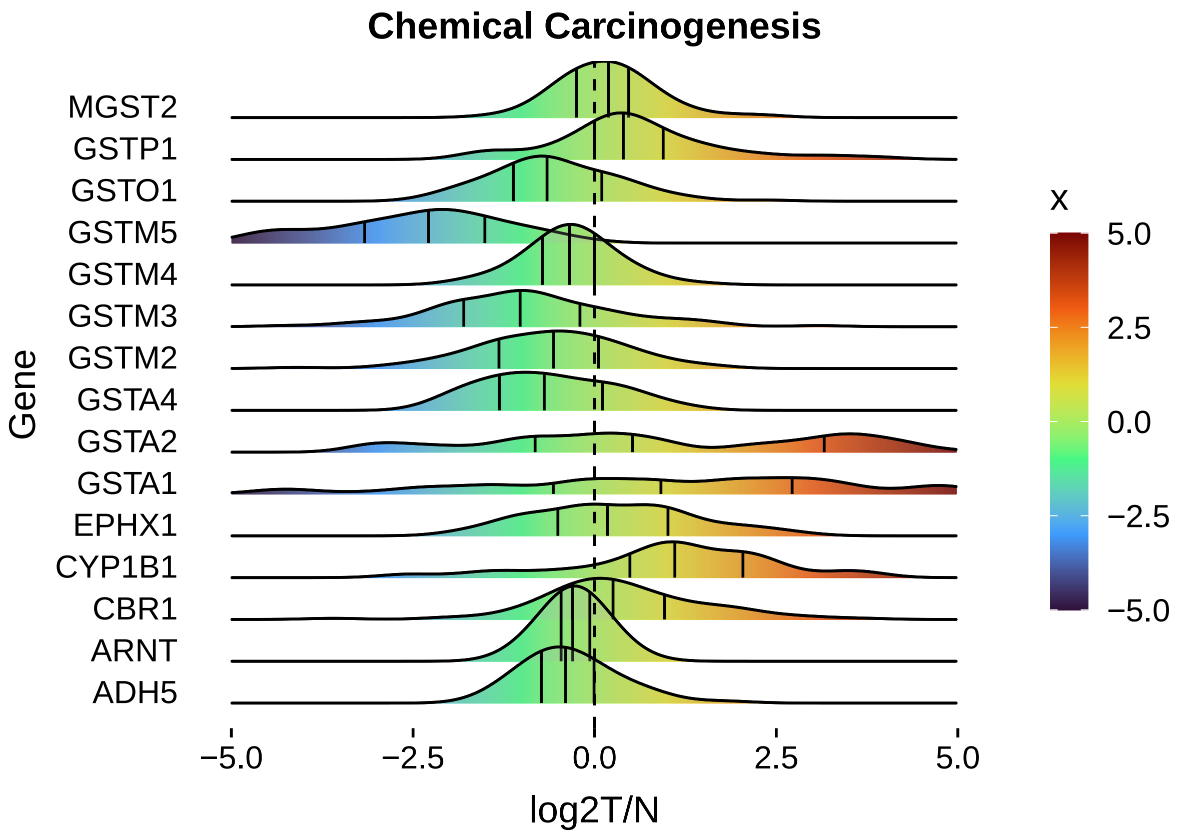


```
Chem
```

```
## Picking joint bandwidth of 0.387
```

```
## Picking joint bandwidth of 0.387
```

```
## Warning: Removed 40 rows containing non-finite values (stat_density_ridges).
```

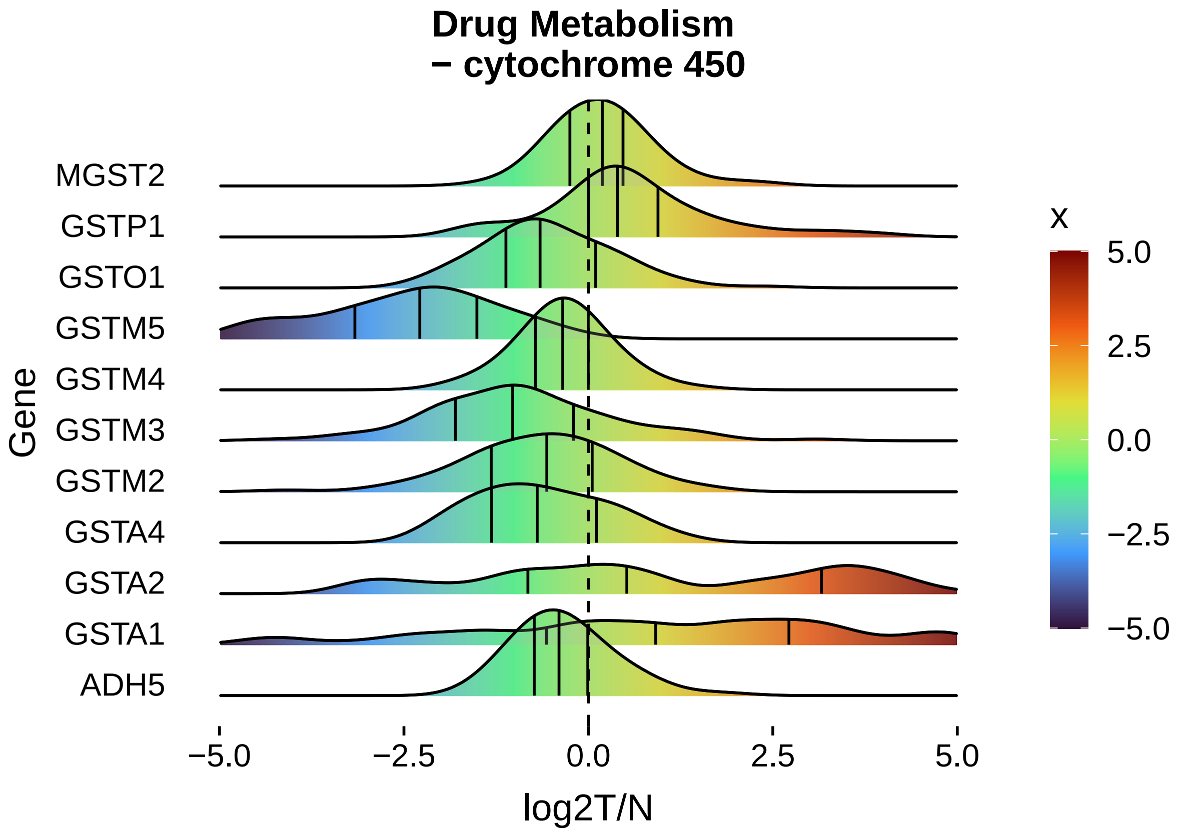


```
Drug
```

```
## Picking joint bandwidth of 0.411
```

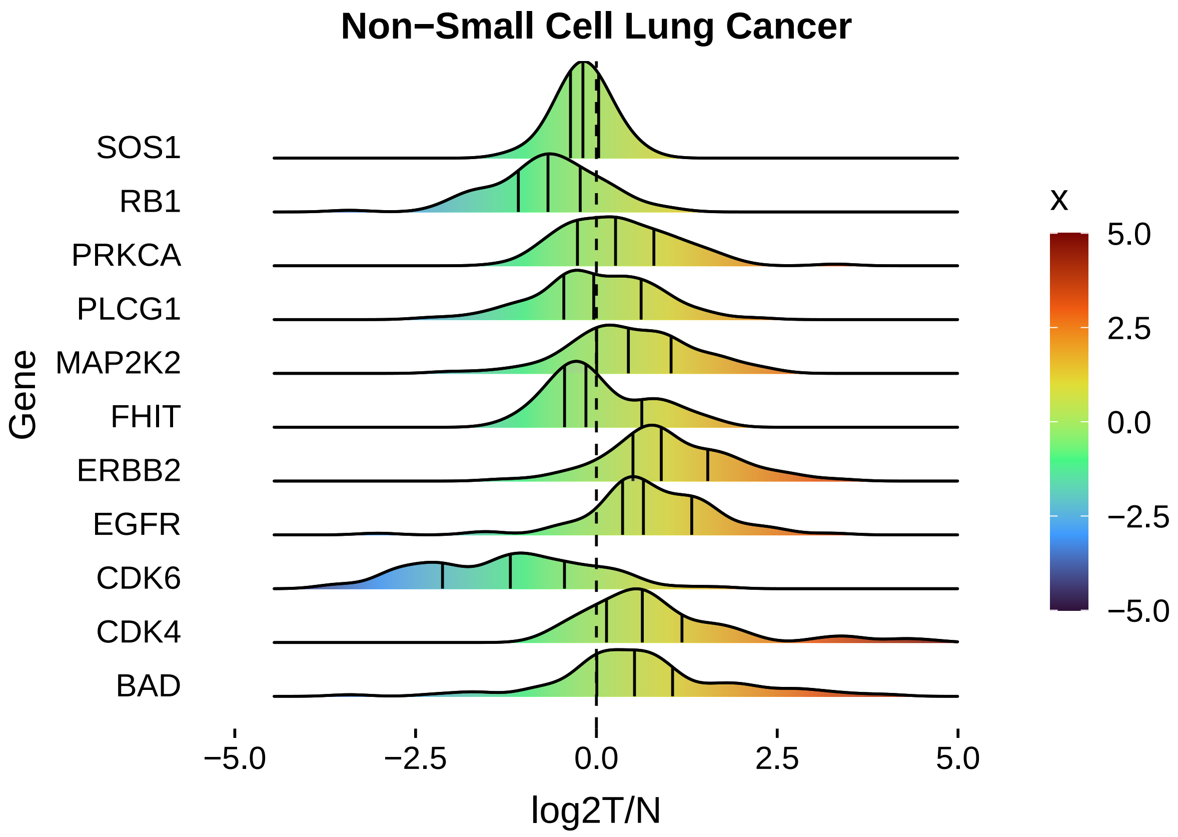
```
## Picking joint bandwidth of 0.411
```

```
## Warning: Removed 40 rows containing non-finite values (stat_density_ridges).
```



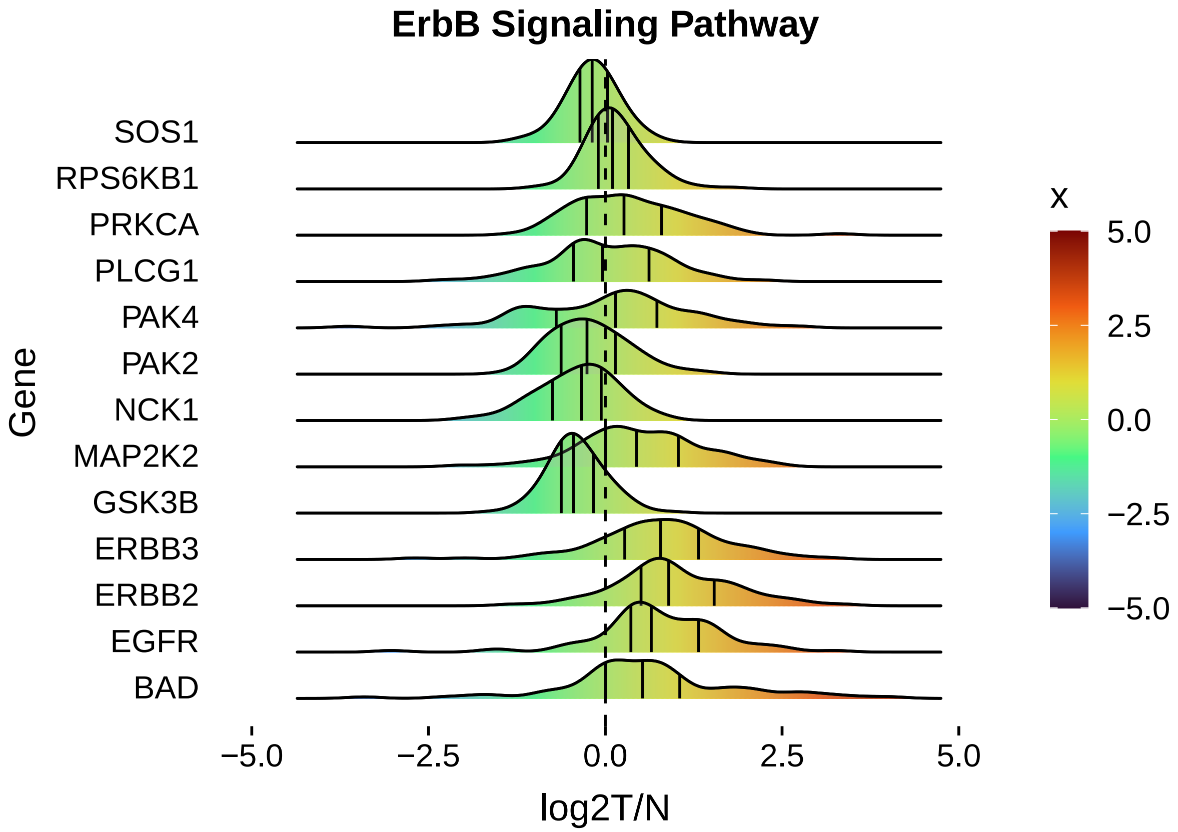
```
NSCLC
```

```
## Picking joint bandwidth of 0.271

## Picking joint bandwidth of 0.271

## Warning: Removed 1 rows containing non-finite values (stat_density_ridges).
```



```
ErbB
```

```
## Picking joint bandwidth of 0.238

## Picking joint bandwidth of 0.238

## Warning: Removed 1 rows containing non-finite values (stat_density_ridges).
```

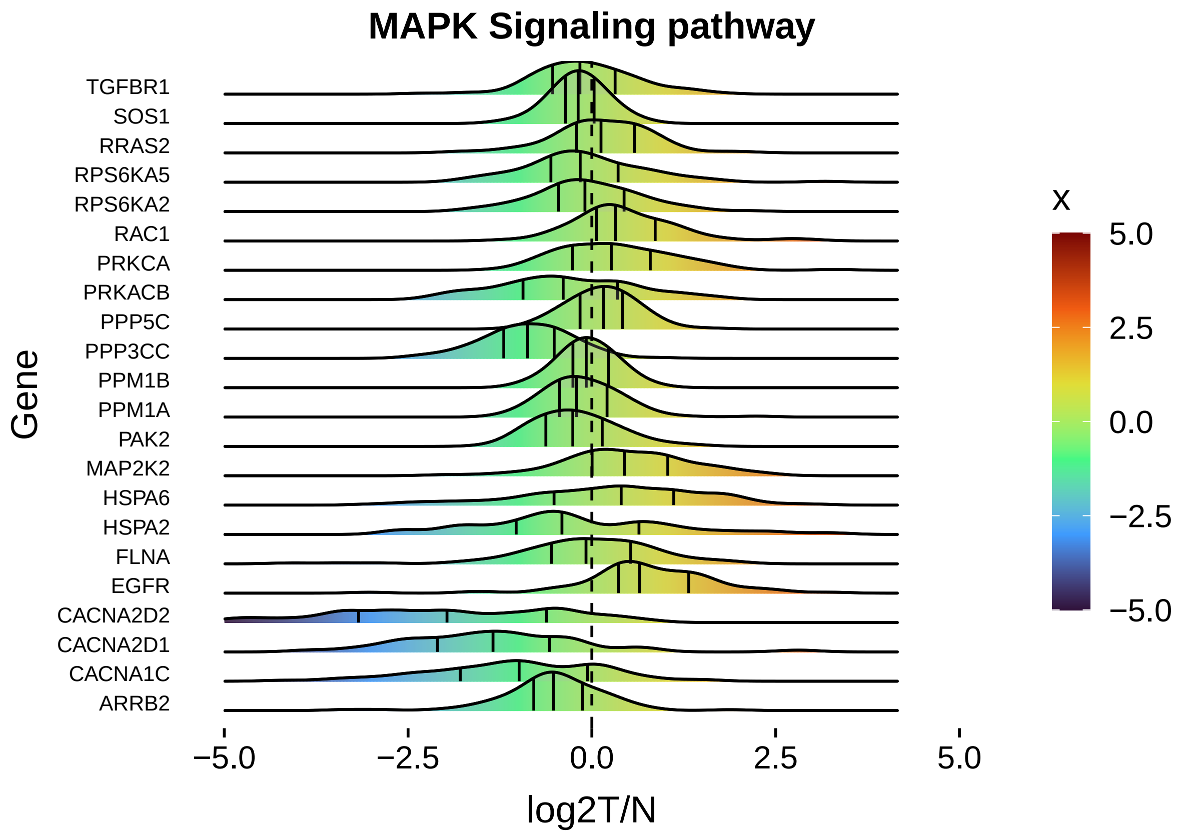


```
MAPK
```

```
## Picking joint bandwidth of 0.275
```

```
## Picking joint bandwidth of 0.275
```

```
## Warning: Removed 10 rows containing non-finite values (stat_density_ridges).
```



Lastly, let's merge pathway plots together. 'plot_grid' function will be used.

```
title <- ggdraw() +
  draw_label("Proteins Affecting to Enriched Pathway",
             fontface = "bold",
             size = 25,
             vjust = 0.5)

p <- plot_grid(Aryl + theme(legend.position = "none"),
               Chem + theme(legend.position = "none"),
               Drug + theme(legend.position = "none"),
               NSCLC + theme(legend.position = "none"),
               ErbB + theme(legend.position = "none"),
               MAPK + theme(legend.position = "none"),
               labels = c("A", "B", "C", "D", "E", "F"),
               label_size = 10)
```

```
## Picking joint bandwidth of 0.199
## Picking joint bandwidth of 0.199

## Picking joint bandwidth of 0.387
## Picking joint bandwidth of 0.387

## Warning: Removed 40 rows containing non-finite values (stat_density_ridges).

## Picking joint bandwidth of 0.411
## Picking joint bandwidth of 0.411
```

```
## Warning: Removed 40 rows containing non-finite values (stat_density_ridges).
```

```
## Picking joint bandwidth of 0.271
```

```
## Picking joint bandwidth of 0.271
```

```
## Warning: Removed 1 rows containing non-finite values (stat_density_ridges).
```

```
## Picking joint bandwidth of 0.238
```

```
## Picking joint bandwidth of 0.238
```

```
## Warning: Removed 1 rows containing non-finite values (stat_density_ridges).
```

```
## Picking joint bandwidth of 0.275
```

```
## Picking joint bandwidth of 0.275
```

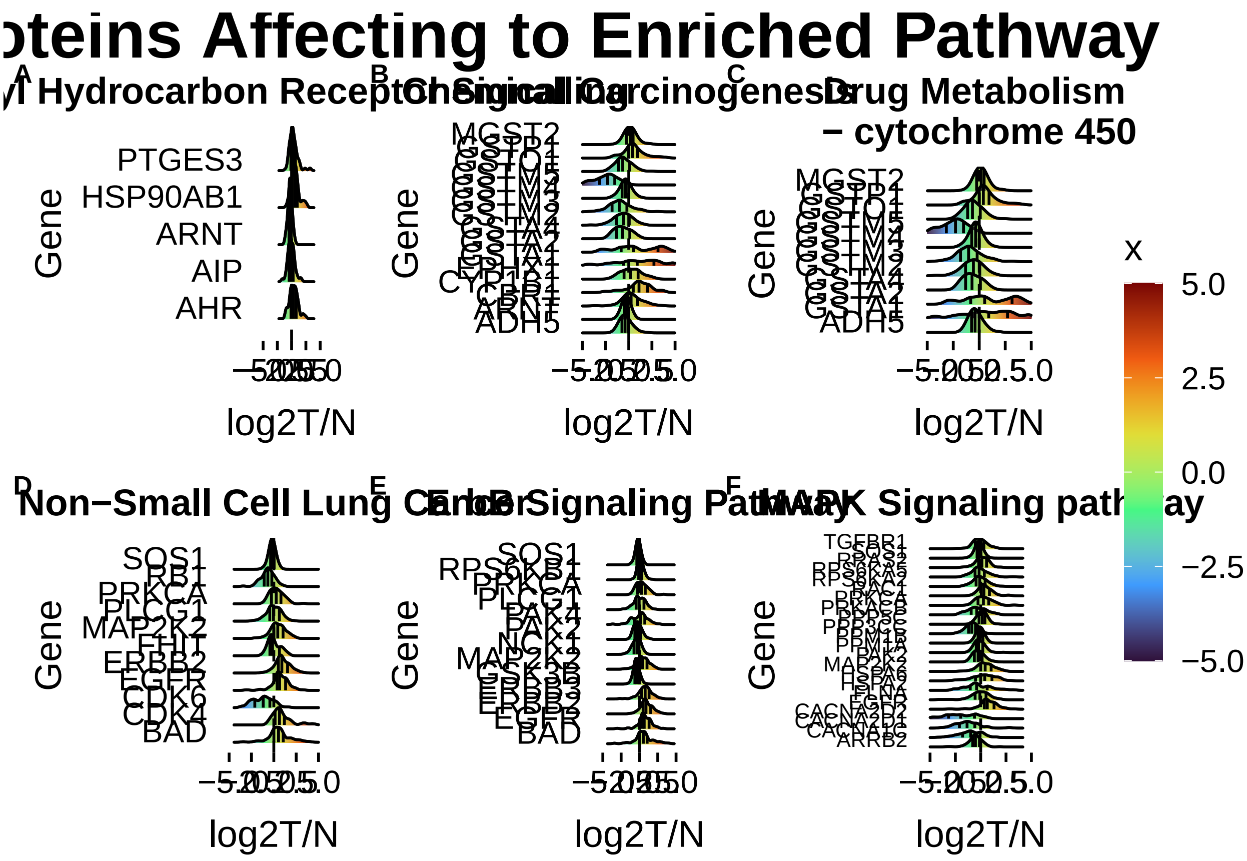
```
## Warning: Removed 10 rows containing non-finite values (stat_density_ridges).
```

```
ps <- plot_grid(title, p, ncol = 1, rel_heights = c(0.1, 2))

legend <- get_legend(Aryl + theme(legend.box.margin = margin(0, 0, 0, 12)))
```

```
## Picking joint bandwidth of 0.199
```

```
## Picking joint bandwidth of 0.199
```

```
plot_grid(ps, legend, rel_widths = c(3, 0.5))
```



```
ggsave('test.pdf', width = 17, height = 14)
```

4. Discussion

Figure A shows that PAHs and nitrosamine have a high mutational signature ratio in phosphate and NitroPAHs have a high proteinic mutation ratio. In addition, as we can see the p-value in first Figure, it was found that phosphate and protein affect pathway as carcinogen. According to the first and second Figure, tumors harboring PAH or nitro-PAH signatures showed significant enrichment for pathways associated with metabolism and detoxification of chemical carcinogens, including the AHR and Cytochrome P450 pathways, known to contribute to carcinogenesis by PAH. The nitro-PAH and nitrosamines-like groups were dominated by DNA repair, ERBB/MAPK pathway, and TLR/RIG-1 T-cell signaling, which potentially link to the tumor initiation, cell proliferation, EMT malignant progression, and immune modulation in early carcinogenesis(Chen et al., 2020). Through this plotting, it was possible to determine that various carcinogens (PAHs, nitroPAHs, nitrosamine, etc.) absorbed into the body through smoking or air pollution affect to metabolism. Especially, the fact that experimental group used in data is never-smoker makes us to think environmental pollution is main carcinogen in TW cohort.

5. Feedback

At first, I loaded the file from website by using 'rio' package instead of loading from folder in my laptop. So others now can also access the original data easily. And in the first plot, I stated before that the log value below zero is unnecessary information, but that opinion came from my ignorance. The less log value belows zero, the more downregulation it represents. Thus, I added negative log values in the plot. When the R file was knitted to pdf, some long codes diplayed beyond the code block, so others couldn't see whole code. I used ‘enter’ buttons properly to improve readability. And most importantly, I merged existing two plots together. I thought two Figures displayed same conclusion, so it wasn’t meaningful. Instead of merging them, I added outer border to show p-value so we can understand which value represents significance.

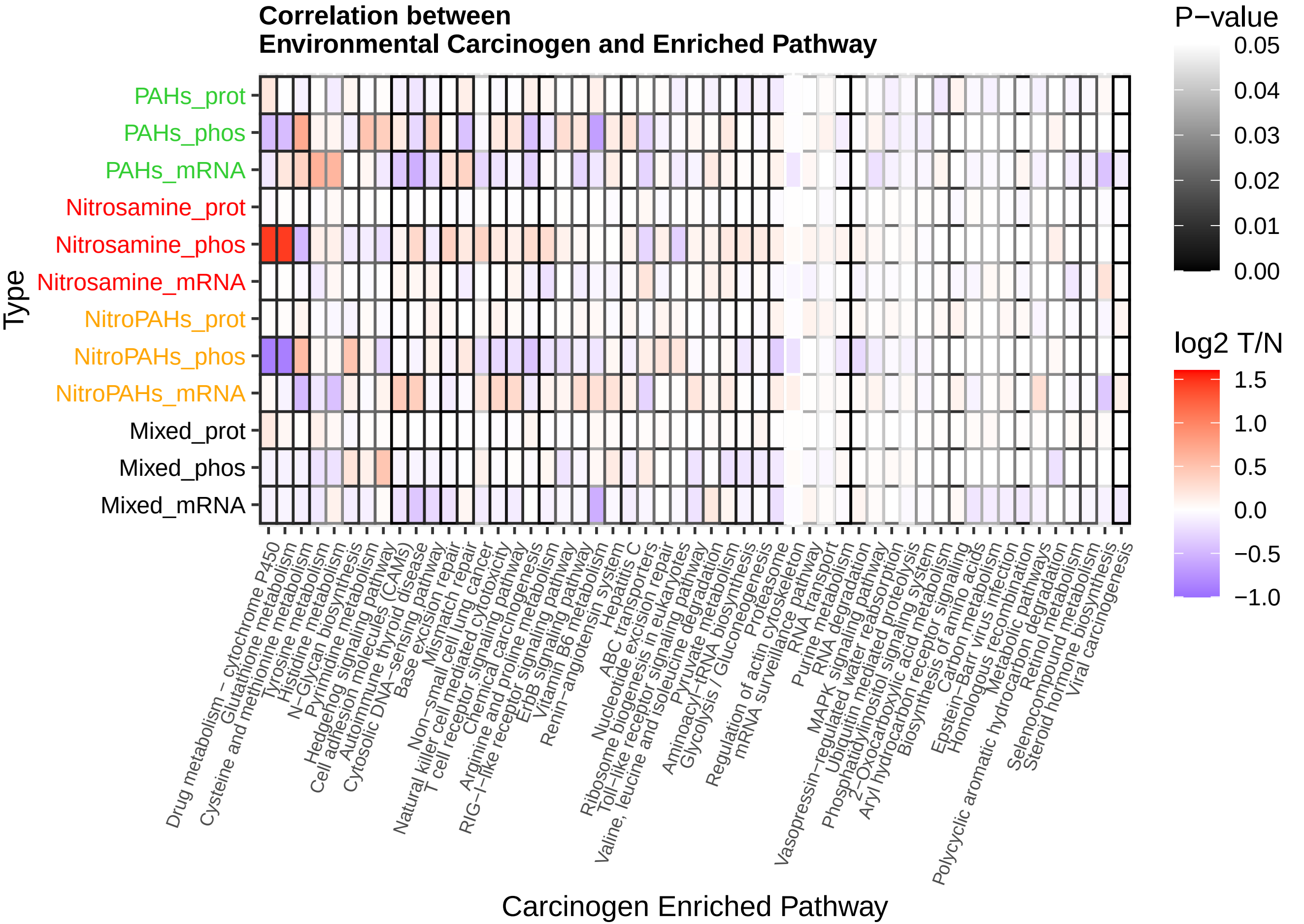
6. Reference

Chen et al. (2020), Cell, Proteogenomics of Non-smoking Lung Cancer in East Asia Delineates Molecular Signatures of Pathogenesis and Progression

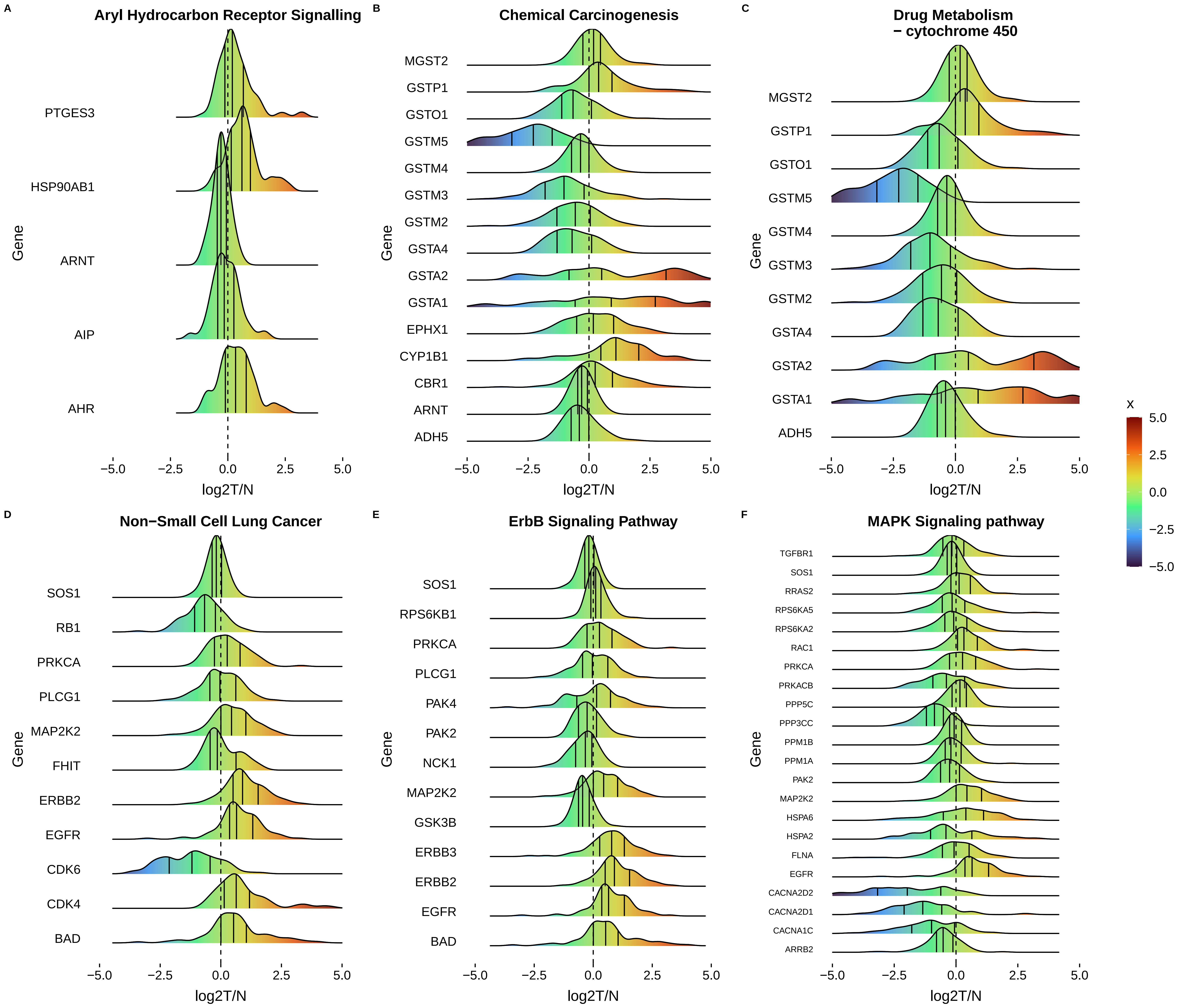
Moorthy et al. (2015), Toxicological Sciences, Polycyclic Aromatic Hydrocarbons: From Metabolism to Lung Cancer

H.Robles (2014), Encyclopedia of Toxicology

Benjamin A. Musa Bandowe et al. (2017), Science of The Total Environment, Nitrated polycyclic aromatic hydrocarbons (nitro-PAHs) in the environment – A review


Correlation between
Environmental Carcinogen and Enriched Pathway
Type
PAHs_prot
PAHs_phos
PAHs_mRNA
Nitrosamine_prot
Nitrosamine_phos
Nitrosamine_mRNA
NitroPAHs_prot
NitroPAHs_phos
NitroPAHs_mRNA
Mixed_prot
Mixed_phos
Mixed_mRNA
Drug metabolism – cytochrome P450
Glutathione metabolism
Cysteine and methionine metabolism
Tyrosine metabolism
Histidine metabolism
N−Glycan biosynthesis
Pyrimidine metabolism
Hedgehog signaling pathway
Cell adhesion molecules (CAMs)
Autoimmune thyroid disease
Cytosolic DNA−sensing pathway
Base excision repair
Mismatch repair
Non−small cell lung cancer
Natural killer cell mediated cytotoxicity
T cell receptor signaling pathway
Chemical carcinogenesis
Arginine and proline metabolism
RIG−I−like receptor signaling pathway
ErbB signaling pathway
Vitamin B6 metabolism
Renin−angiotensin system
Hepatitis C
ABC transporters
Nucleotide excision repair
Ribosome biogenesis in eukaryotes
Toll−like receptor signaling pathway
Valine, leucine and isoleucine degradation
Pyruvate metabolism
Aminoacyl−tRNA biosynthesis
Glycolysis / Gluconeogenesis
Proteasome
Regulation of actin cytoskeleton
mRNA surveillance pathway
RNA transport
Purine metabolism
RNA degradation
MAPK signaling pathway
Vasopressin−regulated water reabsorption
Ubiquitin mediated proteolysis
Phosphatidylinositol signaling system
2−Oxocarboxylic acid metabolism
Aryl hydrocarbon receptor signalling
Biosynthesis of amino acids
Carbon metabolism
Epstein−Barr virus infection
Homologous recombination
Metabolic pathways
Polycyclic aromatic hydrocarbon degradation
Retinol metabolism
Selenocompound metabolism
Steroid hormone biosynthesis
Viral carcinogenesis
Carcinogen Enriched Pathway
P−value
0.05
0.04
0.03
0.02
0.01
0.00
log2 T/N
1.5
1.0
0.5
0.0
−0.5
−1.0

Proteins Affecting to Enriched Pathway


A
Aryl Hydrocarbon Receptor Signalling
PTGES3
HSP90AB1
ARNT
AIP
AHR
Gene
log2T/N
B
Chemical Carcinogenesis
MGST2
GSTP1
GSTO1
GSTM5
GSTM4
GSTM3
GSTM2
GSTA4
GSTA2
GSTA1
EPHX1
CYP1B1
CBR1
ARNT
ADH5
Gene
log2T/N
C
Drug Metabolism
− cytochrome 450
MGST2
GSTP1
GSTO1
GSTM5
GSTM4
GSTM3
GSTM2
GSTA4
GSTA2
GSTA1
ADH5
Gene
log2T/N
x
D
Non−Small Cell Lung Cancer
SOS1
RB1
PRKCA
PLCG1
MAP2K2
FHIT
ERBB2
EGFR
CDK6
CDK4
BAD
Gene
log2T/N
E
ErbB Signaling Pathway
SOS1
RPS6KB1
PRKCA
PLCG1
PAK4
PAK2
NCK1
MAP2K2
GSK3B
ERBB3
ERBB2
EGFR
BAD
Gene
log2T/N
F
MAPK Signaling pathway
TGFBR1
SOS1
RRAS2
RPS6KA5
RPS6KA2
RAC1
PRKCA
PRKACB
PPP5C
PPP3CC
PPM1B
PPM1A
PAK2
MAP2K2
HSPA6
HSPA2
FLNA
EGFR
CACNA2D2
CACNA2D1
CACNA1C
ARRB2
Gene
log2T/N