

Uncovering Molecular Adaptations in the Addicted Brain: A Focus on Dopamine Genes and Sex Differences

Hoang Yen Nhi Nguyen

Introduction

This study aimed to evaluate how sex and prolonged cocaine withdrawal influence the expression of dopamine-related genes in the nucleus accumbens (NAc) of mice. Specifically, I reanalyzed a publicly available RNA-sequencing dataset focusing on genes involved in dopamine signalling, including *Cartpt* (Lohoff et al., 2008) and *Drd1a*, *Drd2*, *Drd3*, *Drd4*, and *Drd5* (Ozburn et al., 2015, Zhang et al., 2004), to identify patterns of sex-specific regulation and the lasting molecular impacts of stimulant exposure.

We focus on two core questions: 1. Are there significant gene expression differences between male and female mice? 2. Are there gene expression differences after 28 days of cocaine withdrawal?

We analyze this dataset using `ggplot2` to create visual comparisons and apply linear models to test for statistical significance.

Load Data & Data Wrangling

Source: R/remove.R `str_remove(string, pattern)`

Source: R/extract.R `str_extract(string, pattern, group = NULL)`

Source: vignettes/dplyr.Rmd

First, the dataset from the source paper is wide-format, includes gene and samples(experiment objects). We transform the data into a long-format, with samples into 1 column called (Sample) and the numeric expression values into Expression column.

To continue with our data analysis and organize our data, we need to extract the abbreviations from samples into separate columns named Condition (Naive or Cocaine), Sex (Female or Male), OVX_Group (Cocaine-treated group, ovariectomized female mice). Using `str_extract()` and Regular Expression (Regex) to extract specific code (e.g., OPS = Cocaine-treated, FS = Female,...). Additionally, the asterisk *, which indicated as data use for further gene analysis by authors, needs to be removed before string extraction as it might cause error when running. Using `str_remove()` and Regex to remove the asterisk.

After remove and creating a code object extracting sample abbreviation , we now organize our data into specific columns, including Condition/Sex/OVX_Group. Using conditional function `ifelse()` and `%in%` to check if a value is in the list containing specific sample abbreviations.

Sample Group Codes:

```
**Cocaine-treated groups (female OVX mice only):**
- **OPS**: OVX + Placebo + Saline (=male)
- **OES**: OVX + Estradiol + Saline
- **OPC**: OVX + Placebo + Cocaine (=male)
- **OEC**: OVX + Estradiol + Cocaine
```

```

**Naive groups (no cocaine):**
- **MS**: Male, Saline
- **FS**: Female, Saline
- **MC**: Male, Cortex
- **FC**: Female, Cortex

```

Using a logarithmic scale (log2) makes the data more symmetric and easier to analyze. The raw data include expression that can range from very small (0.0000001) to very big (100). This big variance between data point creates a large skewed distribution making it harder to interpret. Therefore, log transformation makes variance more stable across expression levels. Why log2(Expression + 1)? Because some expression values might be 0 which will cause undefined value and error later. Thus we adding 1 to avoid the issue.

#Load Library

```
library(tidyverse)
```

```

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr    1.5.1
## v ggplot2    3.5.1      v tibble     3.2.1
## v lubridate  1.9.4      v tidyr      1.3.1
## v purrr      1.0.4
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

```

```

library(readxl)
library(cowplot)

```

```

##
## Attaching package: 'cowplot'
##
## The following object is masked from 'package:lubridate':
##
##     stamp

```

```
library(ggplot2)
```

#Load data

```
data <- read_excel("FPKM_AllMice_AllTranscripts-S.Table_2.xlsx")
```

#pivot data

```

long_data <- data %>%
  pivot_longer(-Gene, names_to = "Sample", values_to = "Expression") %>%

```

#add columns using mutate

```

mutate(
  #Remove * and
  #set an object as Code, using Regex, for later extraction of Sample name
  Sample_clean = str_remove(Sample, "\\*"), #Source: R/remove.R
  Code = str_extract(Sample_clean, "[A-Z]+$"), #Source: R/extract.R

```

```

#add column Condition/Sex/OVX_Group to simplify all abbreviations
#Source: vignettes/dplyr.Rmd
Condition = ifelse(Code %in% c("OPS", "OES", "OPC", "OEC"), "Cocaine",
  ifelse(Code %in% c("FS", "MS", "FC", "MC"), "Naive",
    "Unknown")),

Sex = ifelse(Code %in% c("FS", "FC", "OES", "OEC"), "Female",
  ifelse(Code %in% c("MS", "MC", "OPS", "OPC"), "Male",
    "Unknown")),

OVX_Group = ifelse(Code == "OPS", "OVX_Placebo_Saline",
  ifelse(Code == "OES", "OVX_Estradiol_Saline",
  ifelse(Code == "OPC", "OVX_Placebo_Cocaine",
  ifelse(Code == "OEC", "OVX_Estradiol_Cocaine", NA))))

) %>%

mutate(LogExpression = log2(Expression + 1)) #using logExpression

```

Linear Models: Testing Sex and Condition Effects on Total Gene Expression

First, we explore the relationship of Sex and Condition on total gene expression. We applied three linear regression models to assess the effect of Sex and Condition on gene expression.

```

model_gender <- lm(LogExpression ~ Sex, data = long_data)
summary(model_gender)

##
## Call:
## lm(formula = LogExpression ~ Sex, data = long_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.749 -2.730 -0.365   2.093  14.543
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.749176   0.003840  715.958 < 2e-16 ***
## SexMale      -0.019272   0.005501  -3.503  0.00046 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.624 on 910609 degrees of freedom
## Multiple R-squared:  1.348e-05, Adjusted R-squared:  1.238e-05
## F-statistic: 12.27 on 1 and 910609 DF, p-value: 0.00046

model_condition <- lm(LogExpression ~ Condition, data = long_data)
summary(model_condition)

```

```
##
## Call:
## lm(formula = LogExpression ~ Condition, data = long_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8183 -2.6652 -0.3616  2.0877 14.6079
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.665167   0.003838  694.37  <2e-16 ***
## ConditionNaive 0.153168   0.005499   27.85  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.623 on 910609 degrees of freedom
## Multiple R-squared:  0.0008513, Adjusted R-squared:  0.0008502
## F-statistic: 775.8 on 1 and 910609 DF, p-value: < 2.2e-16
```

```
model_interactions <- lm(LogExpression ~ Sex + Condition, data = long_data)
summary(model_interactions)
```

```
##
## Call:
## lm(formula = LogExpression ~ Sex + Condition, data = long_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8256 -2.6575 -0.3617  2.0879 14.6155
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.672792   0.004722  566.031  < 2e-16 ***
## SexMale       -0.015251   0.005501  -2.772   0.00556 **
## ConditionNaive 0.152767   0.005501  27.771  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.623 on 910608 degrees of freedom
## Multiple R-squared:  0.0008597, Adjusted R-squared:  0.0008575
## F-statistic: 391.8 on 2 and 910608 DF, p-value: < 2.2e-16
```

```
library(interactions)
total_gene_plot <- interact_plot(model_interactions, pred = Condition, modx = Sex,
                                plot.points = TRUE, interval = TRUE)
```

```
## Warning: Condition and Sex are not included in an interaction with one another in
## the model.
```

```
## x Detected factor predictor.
## i Plotting with cat_plot() instead.
## i See '?interactions::cat_plot()' for full details on how to specify models
```

```
## with categorical predictors.
## i If you experience errors or unexpected results, try using cat_plot()
## directly.
```

```
#total_gene_plot
#ggsave("total_gene_plot.jpeg")
```

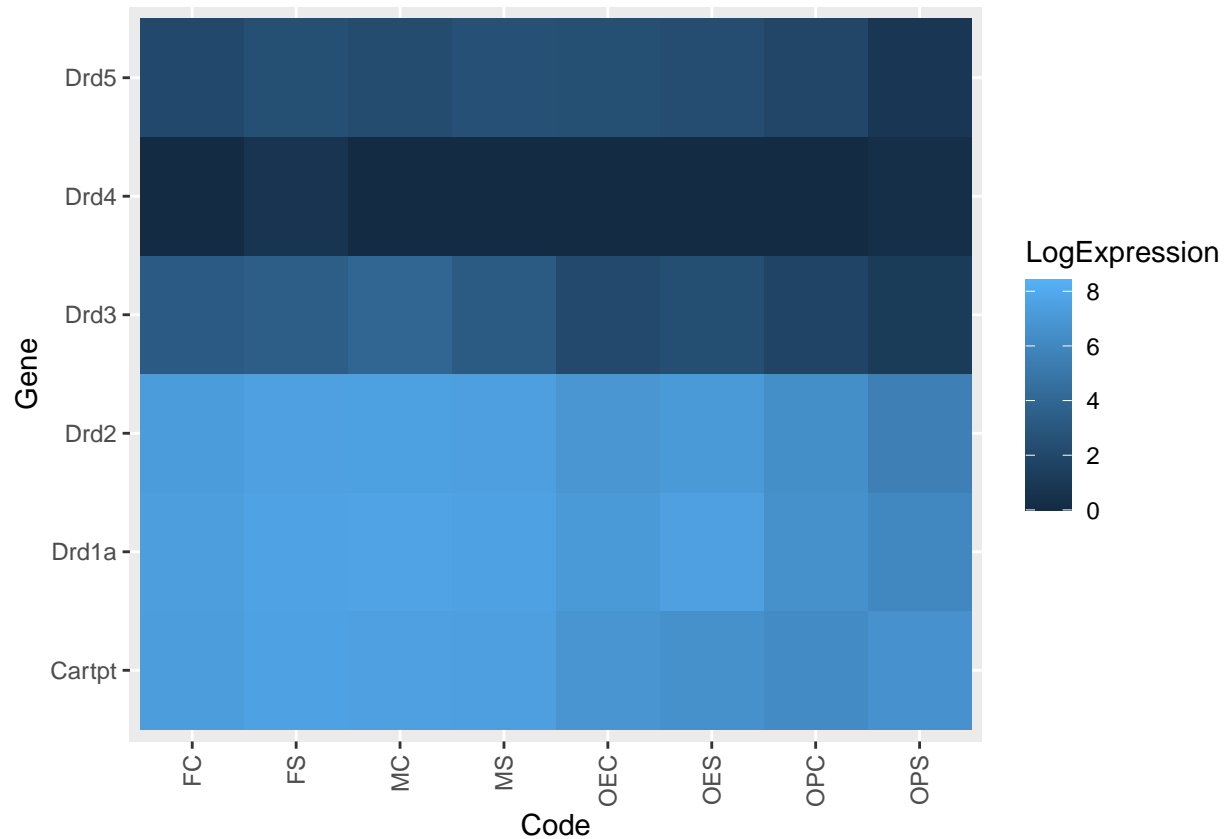
All models indicated statistically significant effects of Sex and/or Condition. Specifically, male mice had slightly lower expression than females, and Naive mice had slightly higher expression than those exposed to cocaine. However, the effect sizes were extremely small ($R^2 < 0.001$ in all models), suggesting that while statistically detectable, Sex and Condition do not explain meaningful variation in global gene expression levels. These findings may reflect subtle differences that become detectable only due to the large sample size, rather than biologically significant changes across the genome. The interaction plot shows no clear evidence of an interaction between Sex and Condition on gene expression. Both males and females exhibited similar expression levels across Cocaine and Naive conditions, suggesting no significant effect of sex or drug withdrawal on genes.

Select Dopamine-Associated Genes

Resoure <https://r-charts.com/>

We now filter out to select targeted dopanmine-associated genes. A heap map in ggplot2 can be created with `geom_tile()` to have a better visualization on how these genes expressed across samples.

```
dopamine_genes <- c("Cartpt", "Drd1a", "Drd2", "Drd3", "Drd4", "Drd5")
dopamine_data <- filter(long_data, Gene %in% dopamine_genes)
figure1 <- ggplot(dopamine_data, aes(x = Code, y = Gene, fill = LogExpression))+
  geom_tile()+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
figure1
```



```
#ggsave("figure1.jpeg")
```

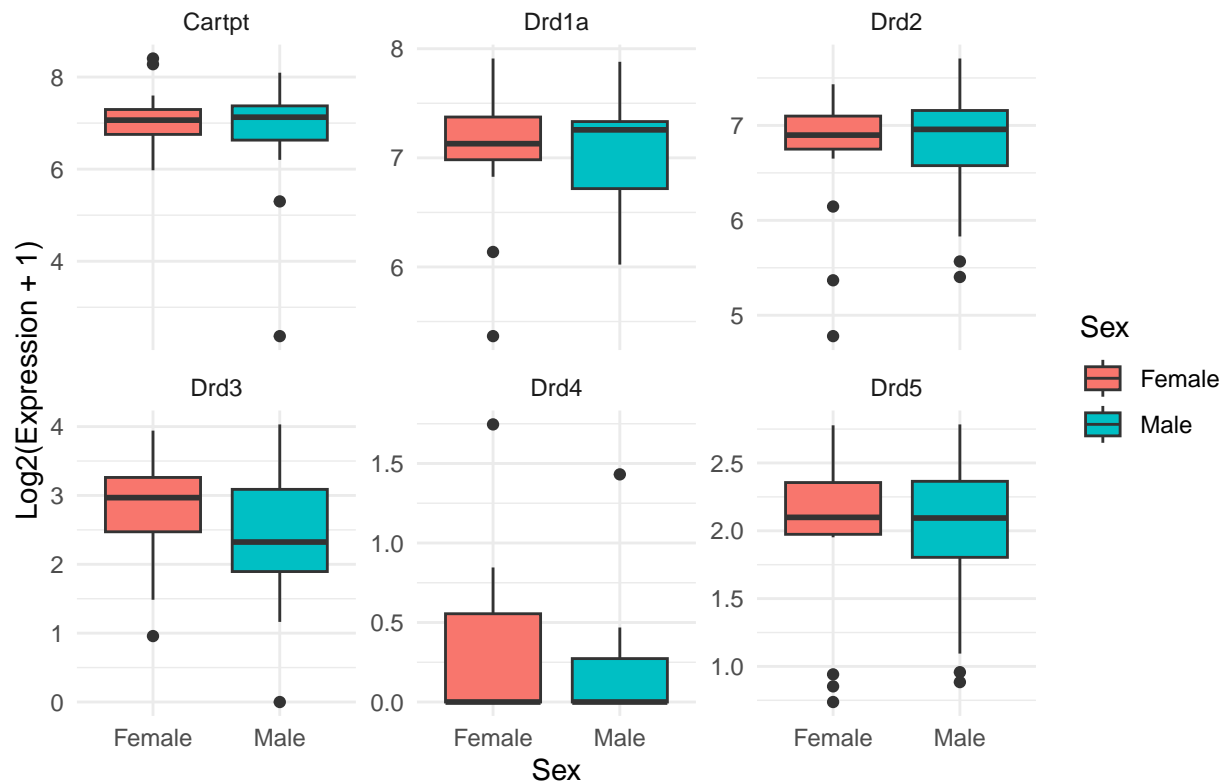
Drd2 and Drd1a show the highest overall expression. Drd3 and Drd4 exhibit lower expression, especially Drd4, which is nearly absent across all groups. No drastic expression shifts across treatment groups, but subtle differences may warrant further analysis

Boxplot: Expression by Sex

We then create boxplot to visualize the effect of Sex on these genes expression. Linear Model is used to detect statistic significance

```
figure2<- ggplot(dopamine_data, aes(x = Sex, y = LogExpression, fill = Sex)) +
  geom_boxplot() +
  facet_wrap(~Gene, scales = "free_y") +
  labs(title = "Gene Expression by Sex", y = "Log2(Expression + 1)") +
  theme_minimal()
figure2
```

Gene Expression by Sex



```
#ggsave("figure2.jpeg")
```

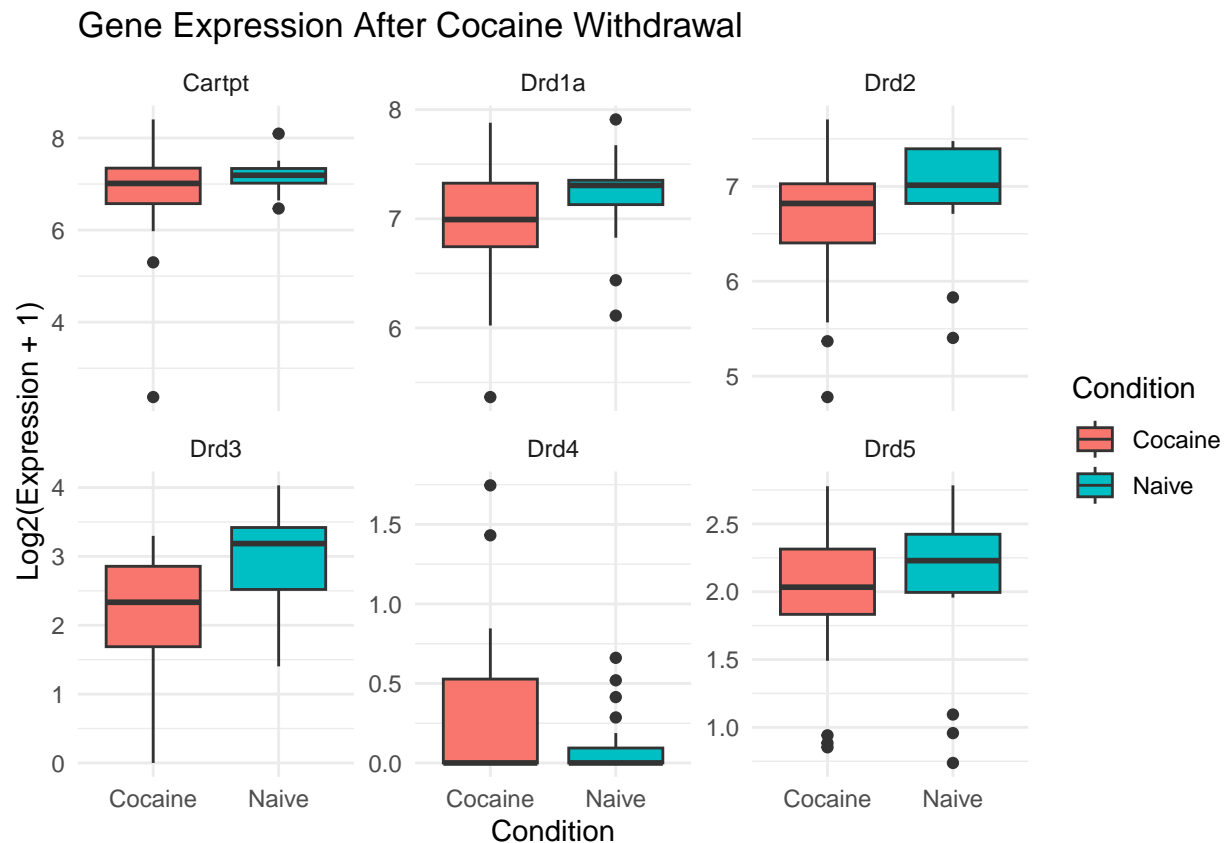
```
model_sex <- lm(LogExpression ~ Sex, data = dopamine_data)
summary(model_sex)
```

```
##
## Call:
## lm(formula = LogExpression ~ Sex, data = dopamine_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.3372 -2.3462 -0.2223  2.7571  4.0652
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.3372     0.2606  16.644  <2e-16 ***
## SexMale       -0.1324     0.3734  -0.355   0.723
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.855 on 232 degrees of freedom
## Multiple R-squared:  0.0005415, Adjusted R-squared: -0.003767
## F-statistic: 0.1257 on 1 and 232 DF, p-value: 0.7233
```

Boxplot: Expression After Cocaine Withdrawal

We then create boxplot to visualize the effect of Condition (cocaine vs naive) on these genes expression

```
figure3 <- ggplot(dopamine_data, aes(x = Condition, y = LogExpression, fill = Condition)) +  
  geom_boxplot() +  
  facet_wrap(~Gene, scales = "free_y") +  
  labs(title = "Gene Expression After Cocaine Withdrawal", y = "Log2(Expression + 1)") +  
  theme_minimal()  
figure3
```



```
#ggsave("figure3.jpeg")
```

```
# Linear model demonstrates condition effects on dopamine-related genes expression  
model_condition <- lm(LogExpression ~ Condition, data = dopamine_data)  
summary(model_condition)
```

```
##  
## Call:  
## lm(formula = LogExpression ~ Condition, data = dopamine_data)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -4.4202 -2.3264 -0.4334  2.7685  4.2698   
##
```



```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.1326    0.2603  15.875  <2e-16 ***
## ConditionNaive  0.2876    0.3730   0.771    0.441
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.852 on 232 degrees of freedom
## Multiple R-squared:  0.002556, Adjusted R-squared:  -0.001743
## F-statistic: 0.5946 on 1 and 232 DF, p-value: 0.4414
```

We combined both factor Sex and Condition and check for its linear model.

```
#Linear model demonstrate sex and condition effects on dopamine-related genes expression
model_sex_condition <- lm(LogExpression ~ Sex + Condition, data = dopamine_data)
summary(model_sex_condition)
```

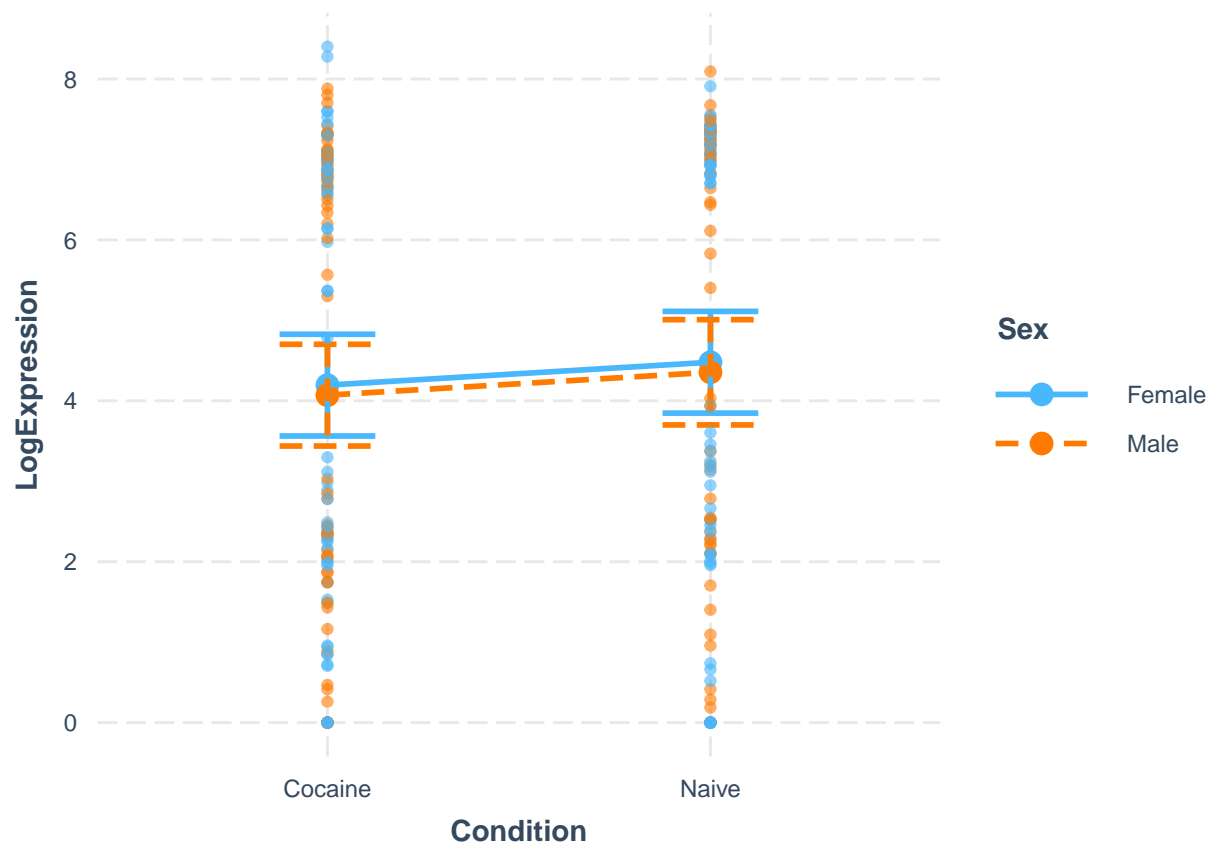
```
##
## Call:
## lm(formula = LogExpression ~ Sex + Condition, data = dopamine_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.479 -2.327 -0.372  2.767  4.207
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.1951    0.3209  13.073  <2e-16 ***
## SexMale        -0.1249    0.3738  -0.334    0.739
## ConditionNaive  0.2843    0.3738   0.761    0.448
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.857 on 231 degrees of freedom
## Multiple R-squared:  0.003038, Adjusted R-squared:  -0.005594
## F-statistic: 0.352 on 2 and 231 DF, p-value: 0.7037
```

```
figure4 <- interact_plot(model_sex_condition, pred = Condition, modx = Sex,
  plot.points = TRUE, interval = TRUE)
```

```
## Warning: Condition and Sex are not included in an interaction with one another in
## the model.
```

```
## x Detected factor predictor.
## i Plotting with cat_plot() instead.
## i See '?interactions::cat_plot()' for full details on how to specify models
## with categorical predictors.
## i If you experience errors or unexpected results, try using cat_plot()
## directly.
```

figure4



```
#ggsave("figure4.jpeg")
```

Although the effect of sex is statistically significant, it is very small and not biologically meaningful in the context of global gene expression ($p = 0.00046$, $R\text{-squared} = 0.000013$). The sample size is likely driving the statistical significance.

There is a statistically significant difference in expression between Naive and Cocaine groups, with Naive mice showing slightly higher gene expression overall ($p < 2e-16$). Again, this is statistically strong but biologically weak due to low effect size ($R\text{-squared} = 0.00085$).

Even when accounting for both predictors together, each contributes a statistically detectable but very minimal amount of variance in gene expression. This means the differences between sexes and drug conditions exist, but they are not strong enough to account for most of the variation in gene expression ($R\text{-squared} = 0.003038$). The model explained only 0.3% of the variation in gene expression, suggesting these predictors are not meaningful drivers of overall dopamine gene expression.

Linear Models: Testing Sex and Condition Effects for each Dopamine-related genes.

```

#dopamine_genes <- c("Cartpt", "Drd1a", "Drd2", "Drd3", "Drd4", "Drd5")

run_lm_for_gene <-function(data, gene){
  dataset <-filter(data, Gene == gene)
  model <- lm(LogExpression ~ Sex + Condition, data = dataset)
  summary(model)
}

for (x in dopamine_genes){
  print(x)
  print(run_lm_for_gene(dopamine_data,x))
}

```

```

## [1] "Cartpt"
##
## Call:
## lm(formula = LogExpression ~ Sex + Condition, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.2570 -0.3291 -0.0057  0.4472  1.4711
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.9313     0.2601  26.649  <2e-16 ***
## SexMale        -0.3006     0.3030  -0.992   0.328
## ConditionNaive  0.3641     0.3030   1.202   0.237
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9455 on 36 degrees of freedom
## Multiple R-squared:  0.06477,    Adjusted R-squared:  0.01281
## F-statistic: 1.247 on 2 and 36 DF,  p-value: 0.2996
##
## [1] "Drd1a"
##
## Call:
## lm(formula = LogExpression ~ Sex + Condition, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.58296 -0.14206  0.06575  0.32003  0.96133
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.94911     0.14256  48.746  <2e-16 ***
## SexMale        -0.03012     0.16607  -0.181   0.8571
## ConditionNaive  0.28183     0.16607   1.697   0.0983 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5182 on 36 degrees of freedom

```

```

## Multiple R-squared:  0.07529,    Adjusted R-squared:  0.02392
## F-statistic: 1.466 on 2 and 36 DF,  p-value: 0.2444
##
## [1] "Drd2"
##
## Call:
## lm(formula = LogExpression ~ Sex + Condition, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8191 -0.1512  0.1643  0.4123  1.0733
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.5989     0.1739  37.947  <2e-16 ***
## SexMale         0.0328     0.2026   0.162   0.8723
## ConditionNaive  0.3446     0.2026   1.701   0.0975 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6321 on 36 degrees of freedom
## Multiple R-squared:  0.07474,    Adjusted R-squared:  0.02333
## F-statistic: 1.454 on 2 and 36 DF,  p-value: 0.247
##
## [1] "Drd3"
##
## Call:
## lm(formula = LogExpression ~ Sex + Condition, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0153 -0.6317  0.1272  0.5394  1.2122
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.3611     0.2162  10.922 5.59e-13 ***
## SexMale        -0.3458     0.2518  -1.373  0.17822
## ConditionNaive  0.8046     0.2518   3.195  0.00291 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7858 on 36 degrees of freedom
## Multiple R-squared:  0.2551, Adjusted R-squared:  0.2138
## F-statistic: 6.165 on 2 and 36 DF,  p-value: 0.004982
##
## [1] "Drd4"
##
## Call:
## lm(formula = LogExpression ~ Sex + Condition, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.3715 -0.2879 -0.1485  0.1535  1.3743
##

```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.37148    0.11060   3.359  0.00186 **
## SexMale        -0.08362    0.12885  -0.649  0.52046
## ConditionNaive -0.22297    0.12885  -1.730  0.09211 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4021 on 36 degrees of freedom
## Multiple R-squared:  0.08534,    Adjusted R-squared:  0.03453
## F-statistic: 1.679 on 2 and 36 DF,  p-value: 0.2008
##
## [1] "Drd5"
##
## Call:
## lm(formula = LogExpression ~ Sex + Condition, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.35474 -0.09724  0.12391  0.36696  0.81924
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.95848    0.15358  12.752 6.46e-15 ***
## SexMale        -0.02191    0.17891  -0.122   0.903
## ConditionNaive  0.13369    0.17891   0.747   0.460
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5583 on 36 degrees of freedom
## Multiple R-squared:  0.01582,    Adjusted R-squared:  -0.03886
## F-statistic: 0.2893 on 2 and 36 DF,  p-value: 0.7505
```

Among the dopamine-associated genes analyzed, *Drd3* exhibited a significant increase in expression in naive mice compared to those exposed to cocaine withdrawal ($p = 0.0029$), suggesting a potential role in long-term neural adaptations following drug exposure. While other dopamine receptor genes (*Drd1a*, *Drd2*, *Drd4*) showed non-significant trends toward differential expression by condition, no gene demonstrated significant sex-based differences. These results support previous findings linking *Drd3* downregulation to drug-induced neuroplasticity, and highlight it as a candidate for further exploration in addiction-related gene expression changes.

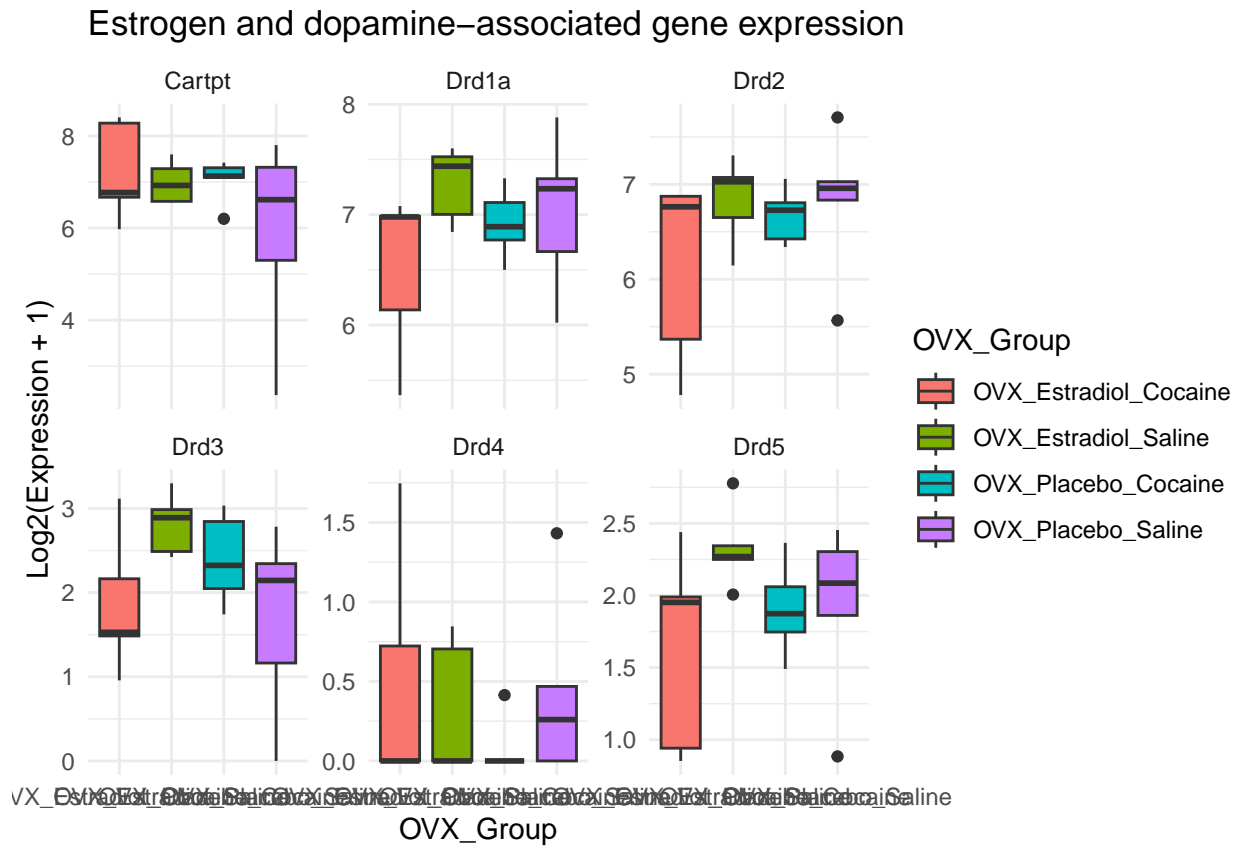
##Additional

Here, we tested whether estradiol treatment in ovariectomized female mice influenced dopamine-related gene expression after cocaine withdrawal.

```
#filtering out NA group in OVX_Group
OVX_dopamine_data <- dopamine_data %>%
  filter(!is.na(OVX_Group))

#plot dopamine genes in OVX_Group to check whether addition of estrogen influence gene expressions
figure5 <- ggplot(OVX_dopamine_data, aes(x = OVX_Group, y = LogExpression, fill = OVX_Group)) +
```

```
geom_boxplot() +
facet_wrap(~Gene, scales = "free_y") +
labs(title = "Estrogen and dopamine-associated gene expression", y = "Log2(Expression + 1)") +
theme_minimal()
figure5
```



```
#ggsave("figure5.jpeg")
```

```
model_OVX <- lm(LogExpression ~ OVX_Group, data = dopamine_data)
summary(model_OVX)
```

```
##
## Call:
## lm(formula = LogExpression ~ OVX_Group, data = dopamine_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.4281 -2.4218 -0.9932  2.7921  4.4291
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.97327    0.52170   7.616  7.7e-12 ***
## OVX_GroupOVX_Estradiol_Saline  0.45481    0.73779   0.616   0.539
## OVX_GroupOVX_Placebo_Cocaine  0.19534    0.73779   0.265   0.792
## OVX_GroupOVX_Placebo_Saline  -0.01269    0.73779  -0.017   0.986
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.857 on 116 degrees of freedom
##    (114 observations deleted due to missingness)
## Multiple R-squared:  0.004527,    Adjusted R-squared:  -0.02122
## F-statistic: 0.1759 on 3 and 116 DF,  p-value: 0.9126
```

Estrogen does not significantly alter dopamine gene expression in OVX mice after cocaine withdrawal.

While we observed trends in genes like *Cartpt* and *Drd3*, especially in the estradiol+cocaine group, the overall differences were not statistically significant.

Our linear model showed no significant effect of OVX treatment group on gene expression, and the R-squared was very low, indicating minimal explanatory power. This suggests that 28 days post-cocaine, estrogen may not exert a strong transcriptional effect on these dopamine genes — or that subtle differences require more targeted or gene-specific testing.

Conclusion

Using RNA-seq data from mouse brain samples, we compared gene expression between sexes and after long-term cocaine exposure. Some dopamine genes showed notable differences across conditions. Boxplots help visualize patterns and linear models test statistical significance.