Analytical Methods

Supplementary Information

Untargeted UHPLC-Q-Orbitrap Metabolomics Reveals Hormonal Suppression and Candidate Serum Markers as Evidence of Testosterone Misuse in Cattle

Diego G. Rocha[*1], Mary Ane G. Lana[2], Débora C. S. de Assis[3], Adriana F. Faria[1]

[1]Department of Chemistry, Federal University of Minas Gerais, Belo Horizonte/MG 31270-901, Brazil

[2]Federal Laboratory of Animal and Plant Health and Inspection - Ministry of Agriculture, Livestock and Food Supply, Pedro Leopoldo/MG 33600-000, Brazil

[3]Veterinary School, Federal University of Minas Gerais, Belo Horizonte/MG 30123-970, Brazil

*Corresponding Author: Tel.: +55 031 3660 5734. Fax: +55 031 3409 5720.

E-mail: diegogr@ufmg.br

Table S1 Integrated peak areas for M1, M2, M3, testosterone, androstenedione (AND), corticosterone (COR), progesterone (PRG) and the internal standard (17β-boldenone-d3) for serum samples analyzed by UHPLC-Q-Orbitrap

SAMPLE	M1	M2	M3	Testosterone	AND	COR	PRG	IS
Samples collected from the treated animals (early period) Peak Areas (x10^6 cps)								
T1-D1	0.00	1.25	5.60	2.19	5.80	0.32	2.80	23.2
T1-D2	0.00	5.70	12.4	1.88	0.80	0.06	1.00	14.9
T1-D5	2.75	10.7	0.877	2.83	0.88	0.10	0.13	21.4
T1-D8	3.04	2.45	0.957	1.36	0.61	0.99	0.12	24.0
T2-D1	0.00	1.44	2.96	2.76	4.50	1.00	1.60	28.1
T2-D2	0.00	6.24	2.81	3.40	1.80	0.18	0.47	21.4
T2-D5	0.593	10.7	1.38	4.62	0.70	0.22	0.24	27.0
T2-D8	0.551	13.6	1.37	1.28	0.38	0.25	0.58	24.6
T3-D1	0.00	2.59	4.15	2.75	4.60	0.42	1.10	21.5
T3-D2	0.371	31.0	3.23	3.19	2.60	0.11	0.34	21.1
T3-D5	0.478	33.4	0.629	3.22	0.56	0.17	0.18	23.9
T3-D8	0.533	9.50	0.268	1.34	0.33	0.076	0.10	23.0
T4-D1	0.00	2.25	20.7	3.58	4.50	1.20	0.62	26.6
T4-D2	1.01	6.73	3.83	3.90	2.00	0.22	0.60	21.2
T4-D5	2.02	5.58	1.57	4.77	0.24	0.66	0.34	24.3
T4-D8	1.56	11.6	1.24	1.30	0.53	0.47	0.36	25.2
QC-D2-A	0.629	6.49	11.8	2.94	2.70	0.17	0.97	22.4
QC-D2-B	0.597	6.90	12.0	3.14	3.00	0.21	0.36	24.0
QC-D2-C	0.561	6.39	11.8	3.00	2.80	0.19	0.30	25.3
QC-D2-D	0.563	6.95	12.1	3.07	2.90	0.19	0.40	24.4
QC-D5-A	1.54	1.72	15.0	2.92	0.96	0.43	0.18	21.1
QC-D5-B	1.44	1.73	15.5	3.33	0.56	0.46	0.18	22.4
QC-D5-C	1.42	2.36	15.6	3.38	1.00	0.42	0.21	24.6
QC-D5-D	1.37	1.97	14.9	3.36	0.71	0.42	0.19	2.41
Samples collected from the untreated animals Peak areas (x10^6 cps)								
T1-D-5	0.00	0.12	0.49	2.83	5.20	3.40	2.80	22.5
T1-D-4	0.00	0.00	0.38	1.16	6.10	3.10	2.30	23.6
T1-D0	0.00	0.00	0.32	1.31	5.60	2.90	2.20	23.0
T2-D-5	0.00	0.00	0.61	2.26	4.20	2.50	1.60	22.2
T2-D-4	0.00	0.00	0.14	1.18	3.70	2.80	1.70	28.5
T2-D0	0.00	0.00	0.08	1.62	5.20	2.30	1.90	21.2
T3-D-5	0.00	0.13	0.21	2.40	3.90	0.74	0.66	21.1
T3-D-4	0.00	0.00	0.02	0.84	4.70	0.83	1.40	20.3
T3-D0	0.00	0.00	0.00	1.00	4.90	0.81	1.60	20.3
T4-D-5	0.00	1.47	0.53	1.50	4.40	0.92	1.80	20.5
T4-D-4	0.00	0.00	0.07	0.86	4.30	1.60	2.10	22.0

T4-D0	0.00	0.07	0.10	1.21	4.30	0.93	2.00	24.5
C1-D1	0.00	0.00	0.13	3.78	3.40	3.50	1.50	21.7
C1-D8	0.00	0.00	0.38	1.68	4.60	0.60	1.50	23.4
C1-D18	0.00	3.60	0.16	0.98	4.80	1.30	1.80	23.2
C2-D1	0.10	0.00	0.13	5.35	4.60	4.10	2.00	22.5
C2-D8	0.00	0.00	0.14	1.29	1.10	2.10	1.80	23.0
C2-D18	0.00	0.00	0.23	1.07	3.20	3.80	1.60	22.8
C3-D1	0.10	0.00	1.00	4.67	3.10	6.60	1.50	20.1
C3-D8	0.00	0.00	0.00	0.92	3.00	1.20	1.50	17.7
C3-D18	0.00	0.07	0.10	1.04	3.90	0.82	1.70	26.1
C4-D1	0.12	0.63	0.10	1.23	3.00	1.30	1.00	21.3
C4-D8	0.00	0.00	0.01	0.87	2.60	2.00	0.45	26.5
C4-D18	0.00	0.00	0.02	0.95	2.50	0.91	1.40	23.2
Samples collected from the treated animals (later period)								
Peak areas (x10^6 cps)								
T1-D12	1.57	8.70	0.80	0.22	2.70	1.30	0.17	25.8
T1-D15	1.75	4.96	1.36	0.08	4.60	2.20	2.50	28.9
T1-D18	0.59	1.05	0.31	0.07	5.70	2.30	2.60	21.6
T2-D12	0.45	11.20	1.19	0.15	0.36	0.59	0.17	24.9
T2-D15	0.26	2.22	0.49	0.27	1.20	1.40	1.60	22.4
T2-D18	0.18	1.80	0.02	0.38	4.10	1.60	2.60	24.3
T3-D12	0.26	3.13	0.18	0.98	2.90	0.66	0.12	22.2
T3-D15	0.34	0.80	0.03	0.16	3.30	0.42	1.20	16.0
T3-D18	0.12	0.90	0.01	0.11	2.80	0.50	1.50	21.1
T4-D12	0.59	2.44	2.18	0.98	1.80	0.68	3.10	25.7
T4-D15	0.61	1.43	0.30	0.09	5.10	1.30	2.60	23.4
T4-D18	0.33	1.06	0.08	0.24	4.30	1.70	1.90	25.8

Sample identification: T1-D1 (sample collected 1 day after drug administration from animal 1, C1-D-5 (sample collected 5 days before drug administration from a control animal 1); IS: internal standard.

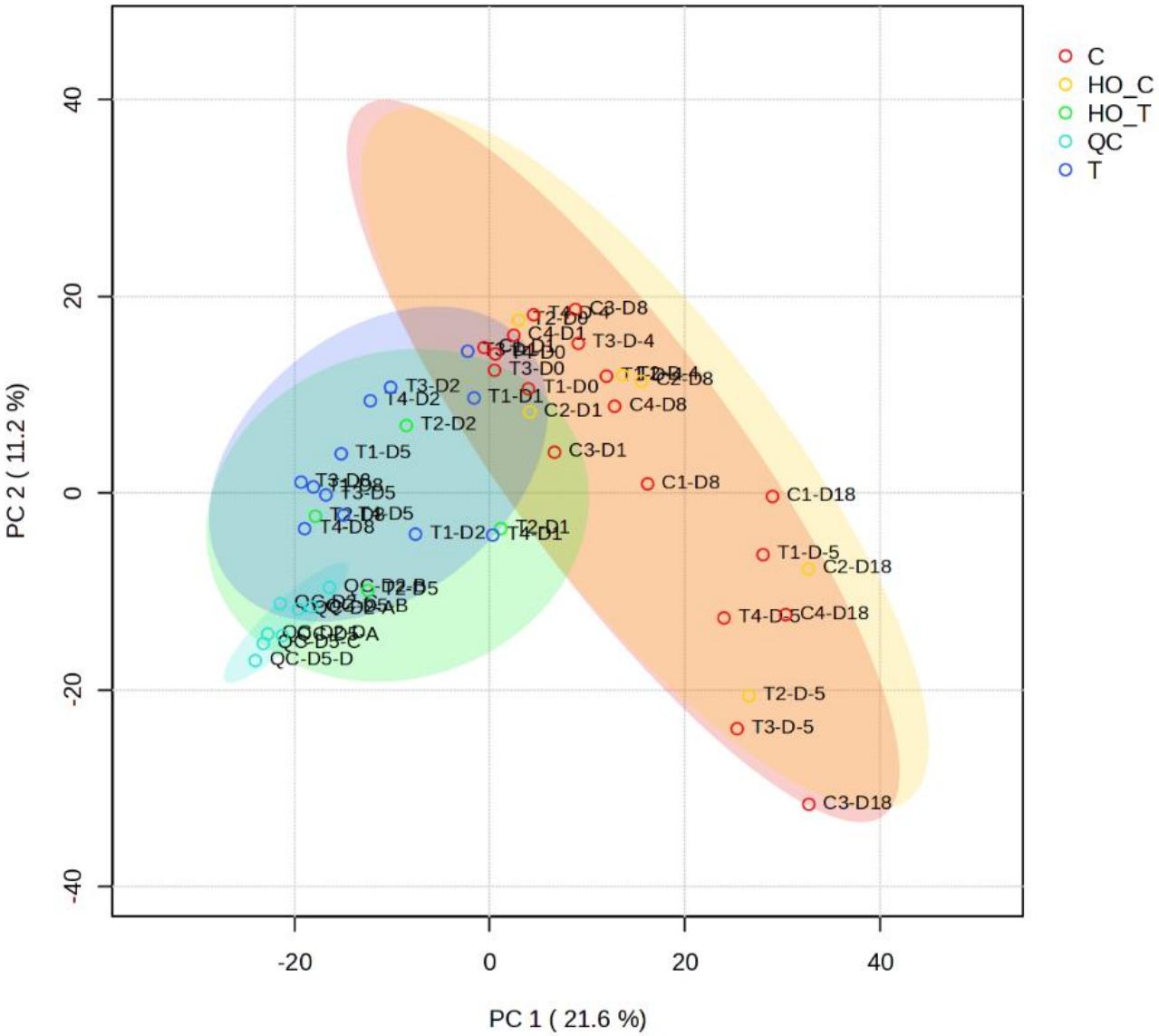

Figure S1: PCA scores plot obtained from serum samples: QCs (n = 8) are colored light blue, treated samples (n = 12) dark blue, control samples (n = 18) red, control hold-outs (n = 6) orange and treated hold-outs (n = 4) green.

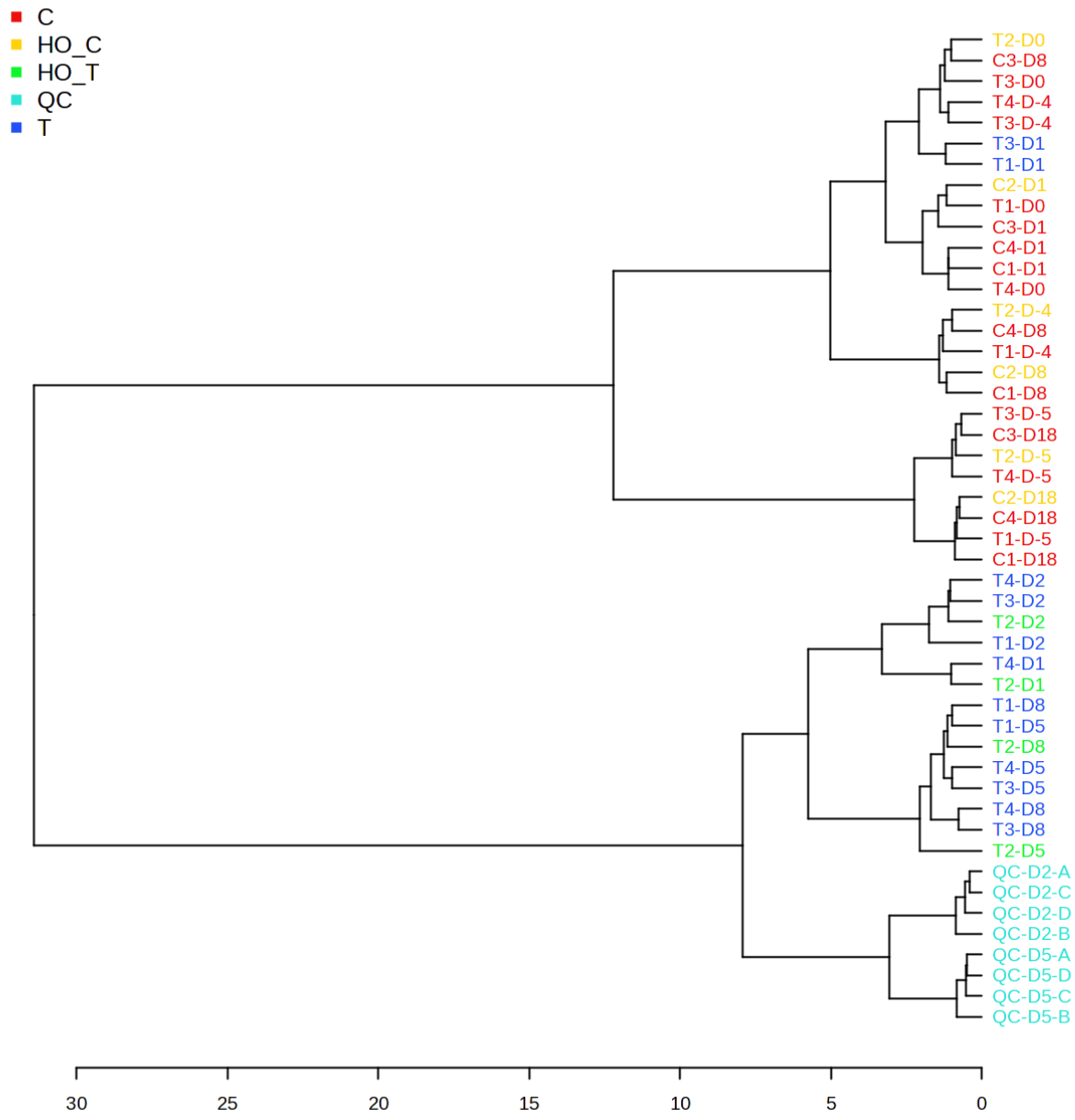

Figure S2: Hierarchical Cluster Analysis (HCA) generated by applying the Pearson Distance Measure and Ward clustering algorithm for the following serum samples: control (red), QCs collected at D2 (light blue), QCs collected at D5 (pink), treated (yellow), control hold-outs (green), treated hold-outs (dark blue).

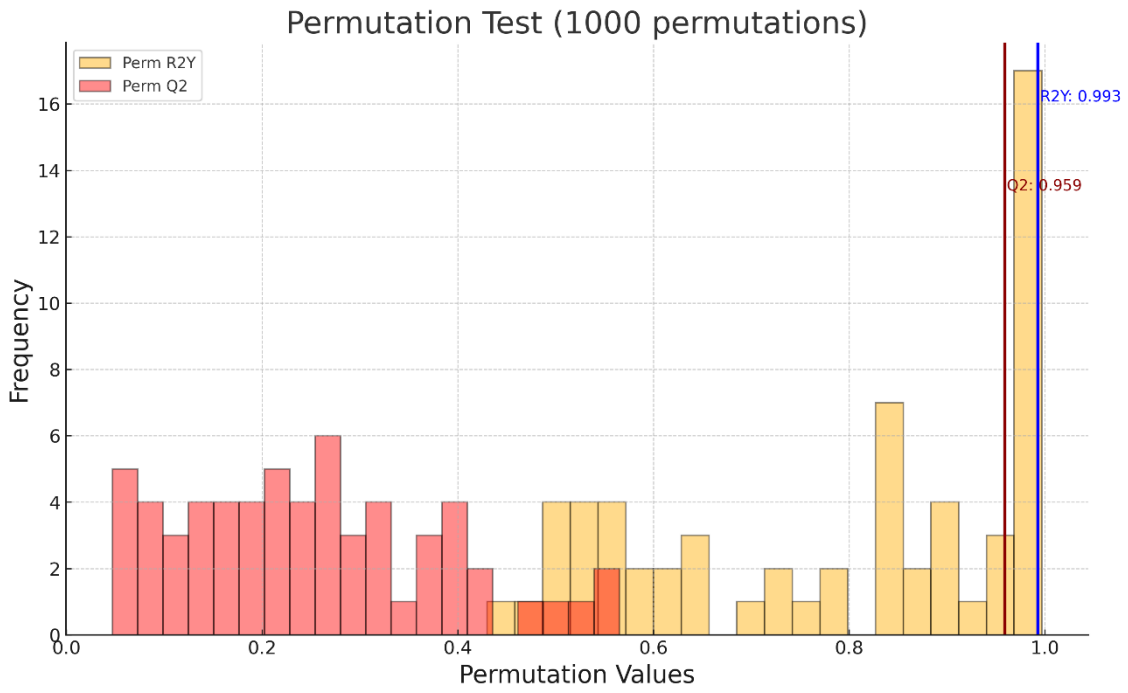

Figure S3: Permutation test (1000 permutations) assessing the robustness of the OPLS-DA model based on empirical p-values for R^2Y and Q^2.

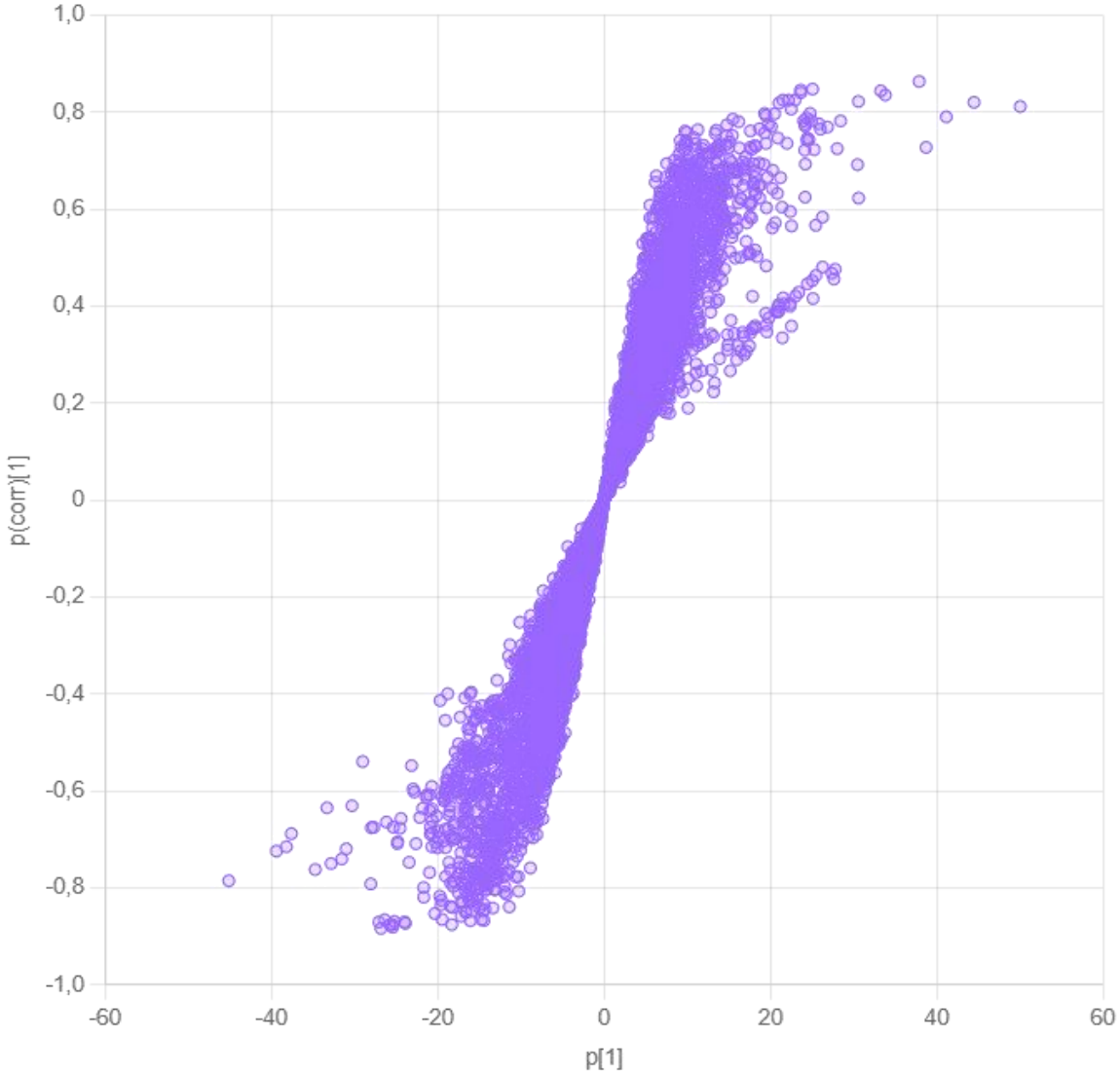

Figure S4: Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) S-plot generated in MetaboAnalyst for biomarker search in serum samples.

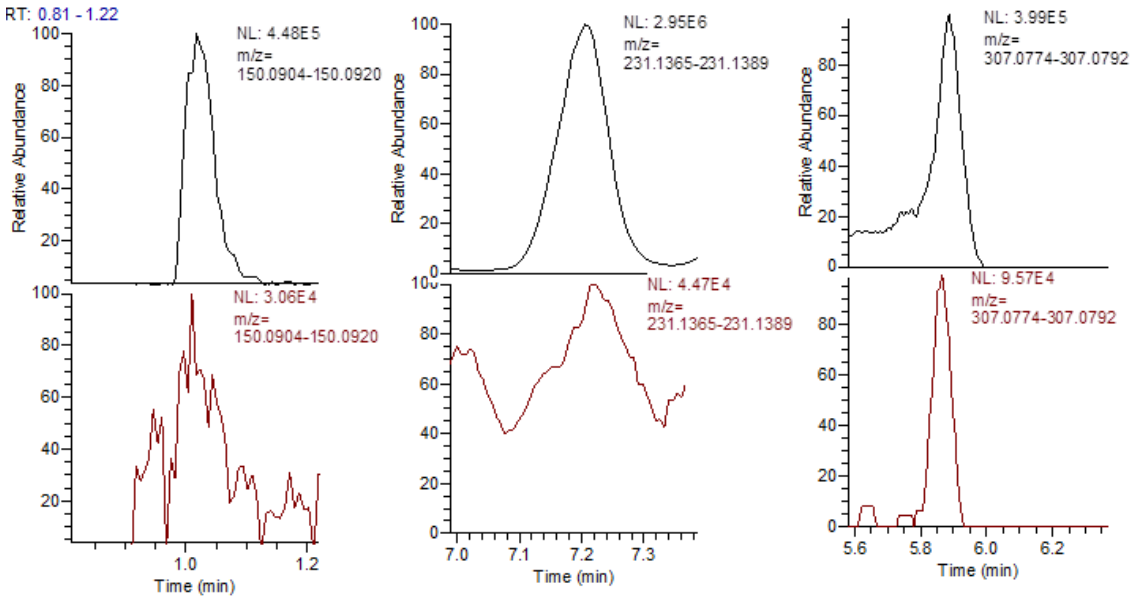

Figure S5: Extracted ion chromatograms for the proposed biomarkers M1, M2 and M3, obtained from UHPLC-Q-Orbitrap for one of the studied steers with a sample collected from day 0 shown in the lower panels and day 4 in the upper panels.

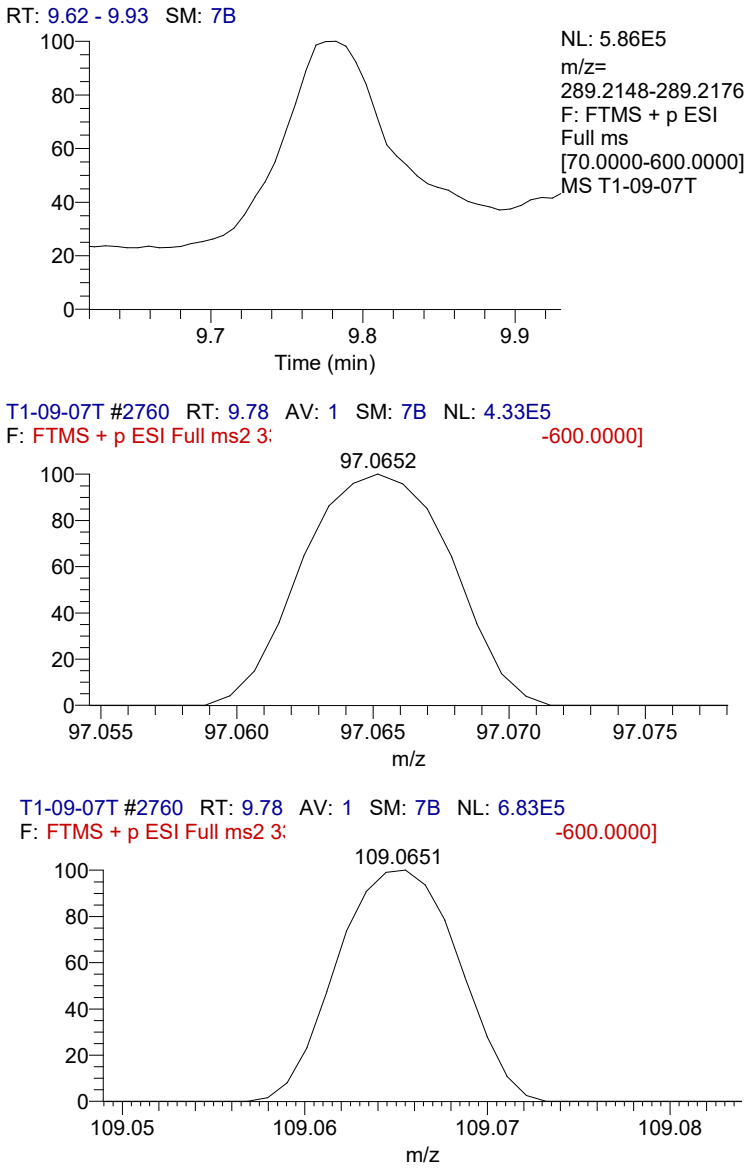

Figure S6: Extracted ion chromatograms for testosterone: parent (*m/z* = 289.2162) and the product ions (*m/*z = 109.0650 and 97.0652) obtained from a steer from the treated group at day 1 by UHPLC-Q-Orbitrap.

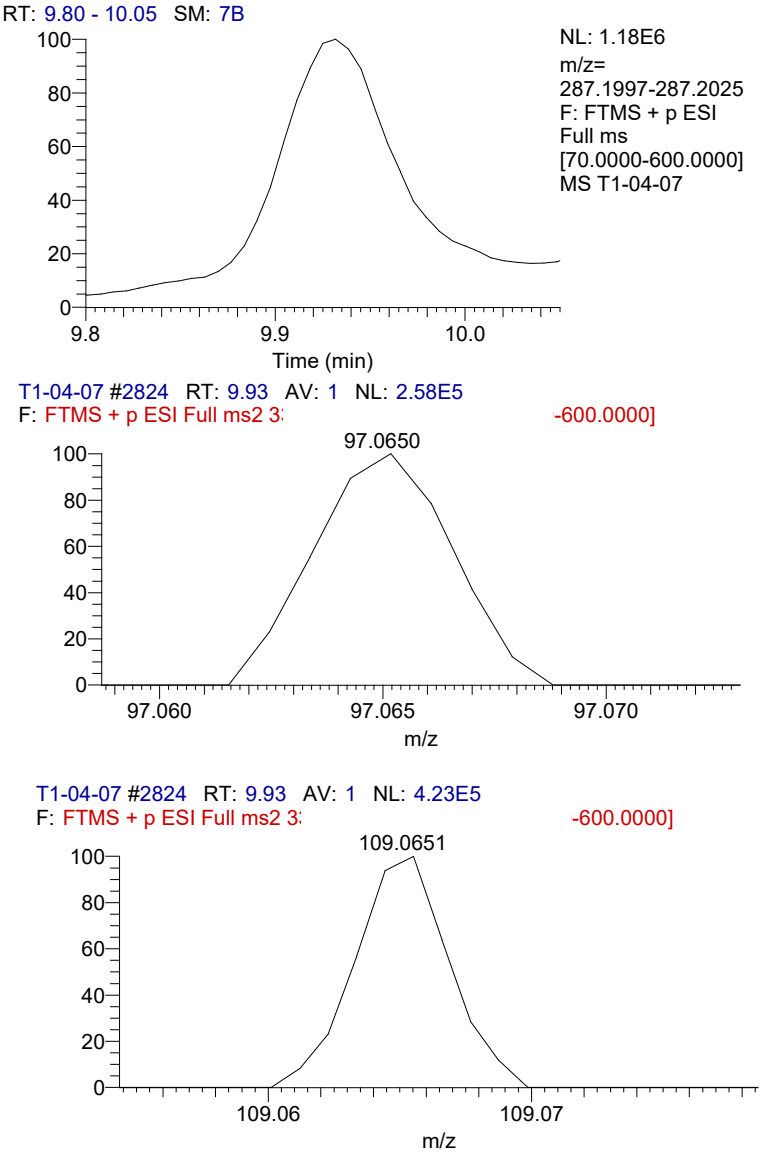

Figure S7 a: Extracted ion chromatogram of androstenedione and the obtained product ions monitored in the all-ions fragmentation mode in a bovine serum sample collected at day -4.

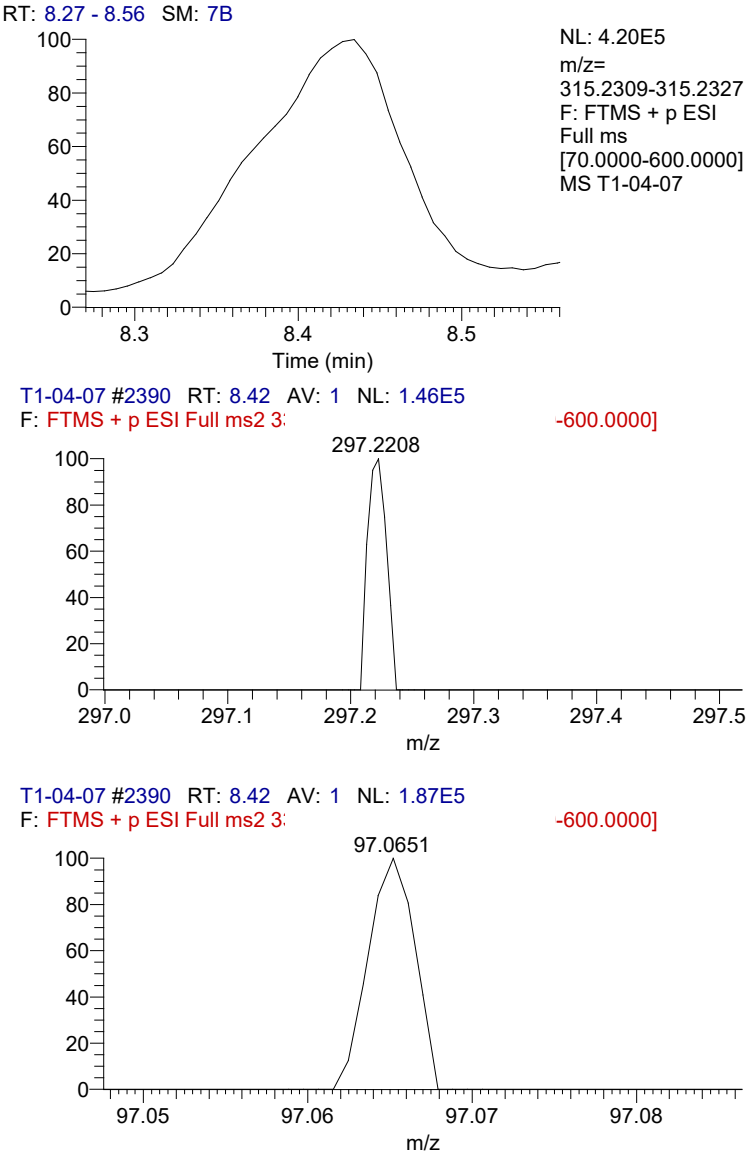

Figure S7 b: Extracted ion chromatogram of progesterone and the obtained product ions monitored in the all-ions fragmentation mode in a bovine serum sample collected at day -4.

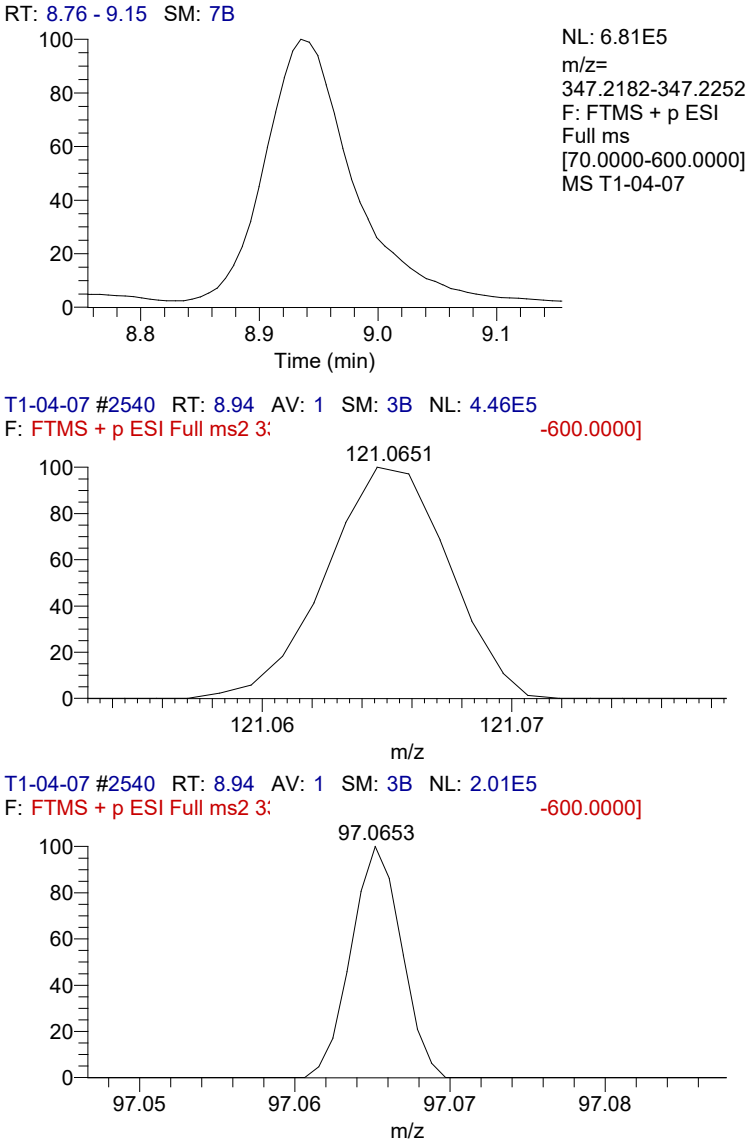

Figure S7 c: Extracted ion chromatogram of corticosterone and the obtained product ions monitored in the all-ions fragmentation mode in a bovine serum sample collected at day -4.

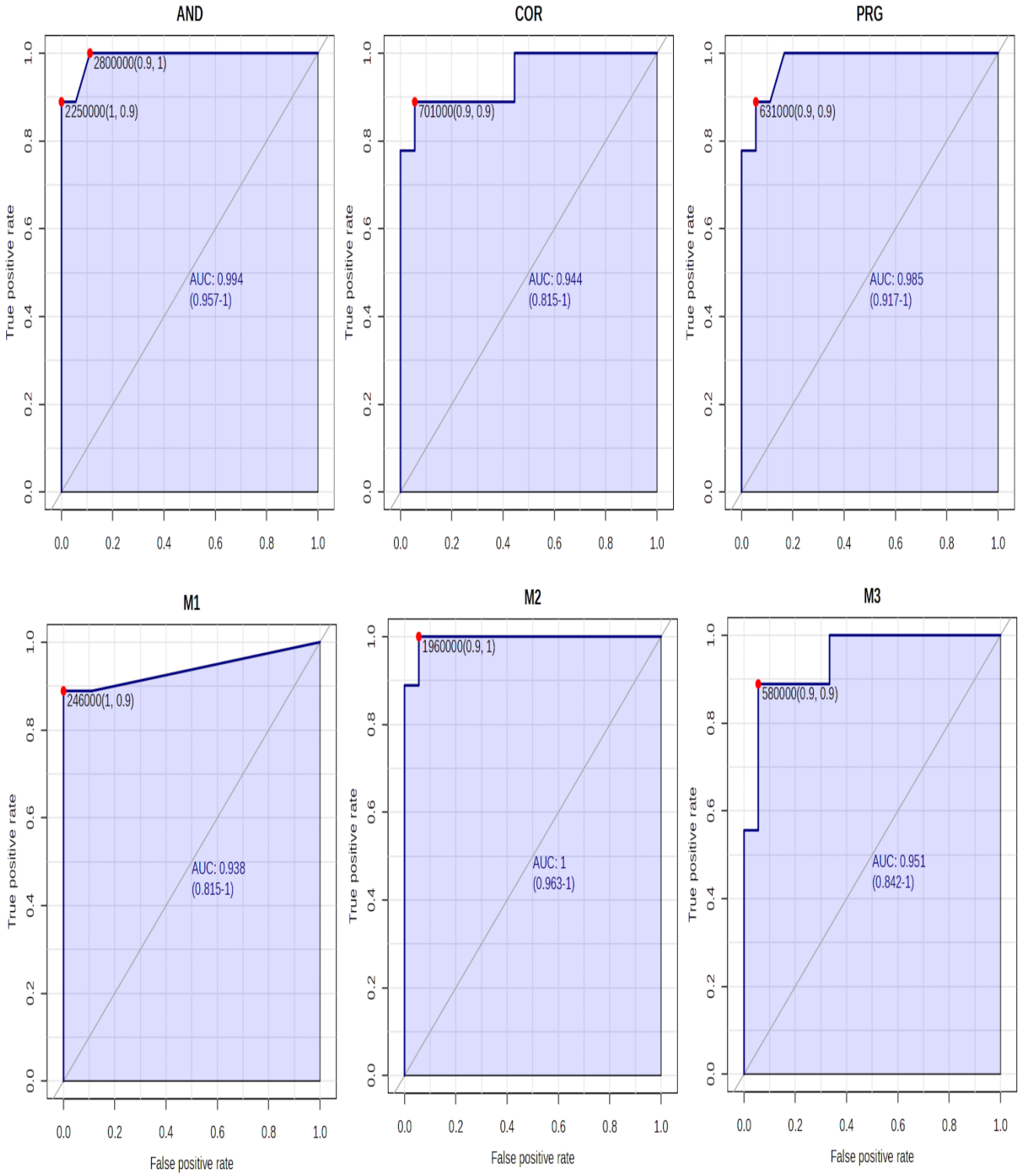

Figure S8: Receiver Operating Characteristic (ROC) curve analysis for AND, COR, PRG and testosterone, where the 95% confidence interval was calculated using 500 bootstrappings in MetaboAnalyst - the solid gray lines represent the ROC for which the AUC value was calculated.

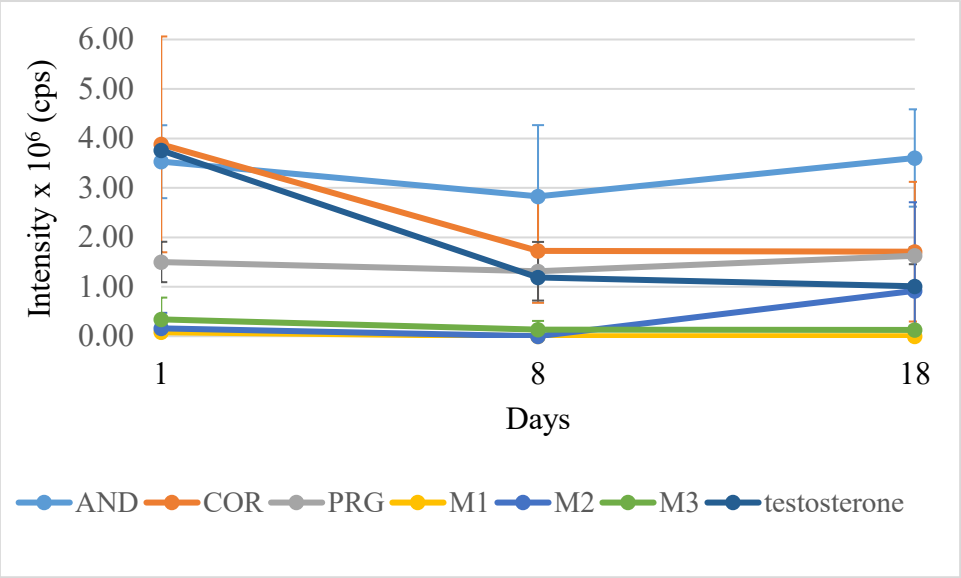

Figure S9: Monitored peak areas in bovine serum for AND, COR, PRG, testosterone, M1, M2 and M3 over 18 days after drug administration for the steers from the control group.

```
########################################################################
# IPO-OPTIMIZED XCMS PROCESSING PIPELINE FOR LC-HRMS METABOLOMICS #
########################################################################

# This script provides a complete workflow for untargeted metabolomics data processing
# using IPO-optimized XCMS parameters. To use with your own data:
# 1. Replace all user-defined variables
# 2. Run script sequentially
# 3. Check output files in specified directory

# ----------------------------------
# SECTION 1: USER-DEFINED VARIABLES
# ----------------------------------

# Working directory (containing raw files in open format: .mzML/.mzXML)
data_directory <- "PATH/TO/YOUR/RAW/DATA"

# Experimental design (case-sensitive!)
sample_groups <- list(
  "GROUP_A" = c("Sample1", "Sample2"), # e.g. "Control"
  "GROUP_B" = c("Sample3", "Sample4")  # e.g. "Treated"
)

# Peak processing parameters (optimized by IPO)
peak_params <- list(
  method = "centWave",
  peakwidth = c(7.2, 46.5),
  ppm = 12.5,
  snthresh = 10,
  mzdiff = 0.0045,
  prefilter = c(3, 100)
)

# Retention time correction parameters
rtcor_params <- list(
  method = "obiwarp",
  profStep = 0.5848,
  center = 3
)

# Grouping parameters
group_params <- list(
  method = "density",
  bw = 0.25,
  mzwid = 0.00538,
  minfrac = 0.931
)

# ----------------------------------
# SECTION 2: PACKAGE INSTALLATION
# ----------------------------------
if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")

required_packages <- c("IPO", "xcms", "RColorBrewer")
```

```
for (pkg in required_packages) {
  if (!require(pkg, character.only = TRUE)) {
    BiocManager::install(pkg)
    library(pkg, character.only = TRUE)
  }
}

# ----------------------------------
# SECTION 3: DATA PROCESSING
# ----------------------------------

# Set working directory
setwd(data_directory)

# Step 1: Peak picking
xset <- xcmsSet(
  method = peak_params$method,
  peakwidth = peak_params$peakwidth,
  ppm = peak_params$ppm,
  snthresh = peak_params$snthresh,
  mzdiff = peak_params$mzdiff,
  prefilter = peak_params$prefilter,
  verbose.columns = FALSE
)

# Step 2: Retention time correction
xset_rt <- do.call("retcor", c(list(object = xset), rtcor_params))

# Step 3: Peak grouping
xset_grouped <- do.call("group", c(list(object = xset_rt), group_params))

# Step 4: Fill missing peaks
xset_filled <- fillPeaks(xset_grouped)

# ----------------------------------
# SECTION 4: DIFFERENTIAL ANALYSIS
# ----------------------------------

# Generate differential report
report <- diffreport(
  xset_filled,
  class1 = names(sample_groups)[1],
  class2 = names(sample_groups)[2],
  filebase = "Metabolomics_Report",
  eicmax = 100
)

# ----------------------------------
# SECTION 5: OUTPUT VALIDATION
# ----------------------------------

# Diagnostic plots
pdf("QC_Diagnostic_Plots.pdf")
  plotrt(xset_rt) # Retention time correction
  plotQC(xset_filled) # Peak quality
```

```
dev.off()

# Save workspace
save.image("Processing_Workspace.RData")

cat("\nPROCESSING COMPLETE!\n",
    "Check output files in:", data_directory)
```

```
######################################################################
# Supplementary Script – Untargeted Metabolomics Analysis
# OPLS-DA Model Construction, VIP Feature Selection,
# Hold-Out Validation and Score Plot Generation
# This script reproduces all analyses used in the manuscript.
######################################################################

#################################
# 1) Load Required Packages
#################################
install.packages(c("data.table", "dplyr", "ggplot2"))
library(data.table)
library(dplyr)
library(ggplot2)
library(ropls)


################################################
# 2) Import Data (generic path)
################################################
file_path <- "path/to/Features.csv"   # <-- replace with your local path

data_raw <- fread(file_path)


################################################
# 3) Transpose Data and Prepare Metadata
################################################

# Transpose: features become columns
data_t <- t(data_raw)
data_t <- as.data.frame(data_t)
```

```
# First row becomes column names
colnames(data_t) <- data_t[1, ]
data_t <- data_t[-1, ]

# Add sample identifier column
data_t$Sample <- rownames(data_t)

#####################################################
# Define experimental groups (Control vs Treated)
#####################################################
control_samples <- c(
  "C1-D1","C1-D8","C1-D18","C2-D1","C2-D8","C2-D18","C3-D1","C3-D8","C3-D18",
  "C4-D1","C4-D8","C4-D18","T1-D0","T1-D-4","T1-D-5","T2-D0","T2-D-4","T2-D-5",
  "T3-D0","T3-D-4","T3-D-5","T4-D0","T4-D-4","T4-D-5"
)

treated_samples <- c(
  "T1-D1","T1-D2","T1-D5","T1-D8",
  "T2-D1","T2-D2","T2-D5","T2-D8",
  "T3-D1","T3-D2","T3-D5","T3-D8",
  "T4-D1","T4-D2","T4-D5","T4-D8"
)

# Assign group labels
data_t$Group <- ifelse(
  data_t$Sample %in% treated_samples, "Treated",
  ifelse(data_t$Sample %in% control_samples, "Control", NA)
)

# Convert all feature columns to numeric
feature_cols <- setdiff(colnames(data_t), c("Sample", "Group"))
```

```
data_t[feature_cols] <- lapply(
  data_t[feature_cols],
  function(x) as.numeric(as.character(x))
)

# Remove rows without valid group assignment
data_t <- data_t %>% filter(!is.na(Group))


################################################
# 4) Define Hold-Out Samples for Validation
################################################

holdout_samples <- c(
  "T4-D1","T4-D2","T4-D5","T4-D8",            # treated hold-outs
  "C4-D1","C4-D8","C4-D18","T4-D0","T4-D-4","T4-D-5"  # control hold-outs
)

data_holdout <- data_t %>% filter(Sample %in% holdout_samples)
data_train   <- data_t %>% filter(!Sample %in% holdout_samples)

X_train <- data_train %>% select(-Sample, -Group)
X_test  <- data_holdout %>% select(-Sample, -Group)
Y_train <- factor(data_train$Group)


################################################
# 5) Full OPLS-DA Model (All Features)
################################################

set.seed(123)
```

```
full_model <- opls(
  X_train, Y_train,
  predI = 1,
  orthoI = NA,
  permI = 50
)

vip_values  <- getVipVn(full_model)
vip_sorted  <- sort(vip_values, decreasing = TRUE)



################################################
# 6) OPLS-DA Using Top VIP Features
################################################

vip_test_sizes <- c(20, 50, 100, 500)
model_quality  <- data.frame()
model_results  <- list()

for (n in vip_test_sizes) {

  selected_vars <- names(vip_sorted)[1:n]
  X_n <- X_train[, selected_vars, drop = FALSE]

  model_n <- opls(
    X_n, Y_train,
    predI = 1,
    orthoI = NA,
    permI = 50
  )

  R2Y_val <- model_n@summaryDF$`R2Y(cum)`
```

```
  Q2_val  <- model_n@summaryDF$`Q2(cum)`

  model_quality <- rbind(
    model_quality,
    data.frame(Model = paste0("VIP_", n), R2Y = R2Y_val, Q2 = Q2_val)
  )

  model_results[[paste0("VIP_", n)]] <- model_n
}


################################################
# 7) Select Best VIP Model
################################################

best_model_name <- model_quality %>%
  arrange(desc(Q2)) %>%
  slice(1) %>%
  pull(Model)

best_model <- model_results[[best_model_name]]
n_selected <- as.numeric(gsub("VIP_", "", best_model_name))
vip_final  <- names(vip_sorted)[1:n_selected]


################################################
# 8) Predict Hold-Out Samples
################################################

prediction <- predict(best_model, X_test)
```

```
#################################################################
# 9) Project Hold-Out Samples into OPLS-DA Space (p1 × o1)
#################################################################

# Extract necessary model values
W_final   <- best_model@weightMN[vip_final, "p1"]
X_center  <- best_model@xMeanVn[vip_final]
X_scale   <- best_model@xSdVn[vip_final]

# Apply model scaling
X_test_scaled <- sweep(X_test[, vip_final], 2, X_center, "-")
X_test_scaled <- sweep(X_test_scaled, 2, X_scale, "/")

# Project into predictive component
p1_holdout <- as.numeric(as.matrix(X_test_scaled) %*% W_final)

# Orthogonal component cannot be computed externally → set to zero
o1_holdout <- rep(0, length(p1_holdout))



################################################
# 10) Build Score Matrices (Training + Hold-Out)
################################################

scores_holdout <- data.frame(
  Sample = data_holdout$Sample,
  Group  = data_holdout$Group,
  p1     = p1_holdout,
  o1     = o1_holdout,
  Type   = "Hold-Out"
)
```

```
scores_train <- data.frame(
  Sample = data_train$Sample,
  Group  = data_train$Group,
  p1     = best_model@scoreMN[, "p1"],
  o1     = best_model@orthoScoreMN[, "o1"],
  Type   = "Training"
)

scores_all <- rbind(scores_train, scores_holdout)
scores_train_no_na <- scores_train %>% filter(!is.na(p1), !is.na(o1))



################################################
# 11) Generate OPLS-DA Score Plot (p1 × o1)
################################################

ggplot(scores_all, aes(x = p1, y = o1)) +
  geom_point(aes(color = Group, shape = Type), size = 4, alpha = 0.9) +
  geom_text(aes(label = Sample), vjust = -0.6, size = 3) +
  stat_ellipse(
    data = scores_train_no_na,
    aes(group = Group, color = Group),
    level = 0.95, linewidth = 0.8,
    linetype = "dashed"
  ) +
  theme_bw(base_size = 13) +
  theme(
    panel.grid = element_blank(),
    plot.title = element_text(hjust = 0.5, face = "bold")
  ) +
  labs(
    title = paste("OPLS-DA Score Plot (p1 × o1) –", best_model_name),
```

```
  x = "Predictive Component (p1)",
  y = "Orthogonal Component (o1)"
)
```