Supporting Information

Metabolome Profiling of Fish Muscle Tissue Exposed to Benzo[a]pyrene Using *In Vivo* Solid-Phase Microextraction

Anna Roszkowska[1#], Miao Yu[1#], Vincent Bessonneau[1], Leslie Bragg[2], Mark Servos[2], Janusz Pawliszyn[1*]

[1]Department of Chemistry, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

[2]Department of Biology, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

[*]Corresponding author; e-mail address: janusz@uwaterloo.ca (J. Pawliszyn); phone: +1-519-888-4567 ext: 35123; fax: +1-519-888-4348

[#]These authors contributed equally to the manuscript.

Table S1 Unique and multiple features detected with high confidence level in the three groups of fish. Not detected features means that the annotation algorithm could not assign a high confidence level to this compounds in certain groups.

Compound	Precursor m/z	Precursor adducts	RT (sec)	Control group		Low-dose group		High-dose group	
				Day 1	Day 14	Day 1	Day 14	Day 1	Day 14
L-Acetylcarnitine	205.12 204.12 226.1	M+H_[+1] M+H 2M+2Na	105.5	•					
Cysteic acid	170.01 126.53	M+H M+2ACN+2H	54.6		•				
L-Proline	117.07 116.07	M+H_[+1] M+H	56.3		•		•		•
L-Tryptophan	206.1 205.09	M+H_[+1] M+H	180.6						•
Tricosanoylglycine	413.38	M+H_[+1] M+H	567.9					•	
Pristanoylglycine	356.31	M+H	504.6						•
N1-Methyl-4-pyridone-3-carboxamide N1-Methyl-2-pyridone-5-carboxamide	153.06	M+H	79.3	•			•		
5-Aminopentanoic acid N-Methyl-a-aminoisobutyric acid L-Valine Betaine	119.08 118.08	M+H_[+1] M+H	54	•					
3-Phenylpropionylglycine Phenylpropionylglycine N-Acetyl-L-phenylalanine	209.1 208.09	M+H_[+1] M+H	275.5	•					
Threonic acid Erythronic acid	138.04 137.04	M+H_[+1] M+H	54	•			•		
Taurine	148.0 128.01 126.02	M+Na M+H_[+2] M+H	48.4				•		
L-Leucine L-Isoleucine	132.1	M+H	89.7				•		
L-Glutamate L-4-Hydroxyglutamate semialdehyde O-Acetylserine N-Acetylserine N-Methyl-D-aspartic acid	149.06 148.06	M+H_[+1] M+H	58.1		•				
5-Oxoprolinate	131.04 131.05 130.04 171.07	M+H_[+1] M+H_[+1] M+H M+ACN+H	65.2		•				
Leucyl-Proline Isoleucyl-Proline	230.15 229.15	M+H_[+1] M+H	101.9		•				
1-Pyrroline-4-hydroxy-2-carboxylate N-Acryloylglycine Pyrroline hydroxycarboxylic acid Pyrrolidonecarboxylic acid Pyroglutamic acid	131.05 130.04 171.07	M+H_[+1] M+H M+ACN+H	65.2 56.3		•				
PC(38:8)	802.53 813.52	M+H 2M+H+Na	593 588.4		•				

Table S2 Unique features with medium confidence level detected in the three groups of fish. Not detected features means that the annotation algorithm could not assign a medium confidence level to this compounds in certain groups.

Compound	Precursor m/z	Precursor adducts	RT (sec)	Control group		Low-dose group		High-dose group	
				Day 1	Day 14	Day 1	Day 14	Day 1	Day 14
2,3-cyclic CMP	306.04	M+H	49.1	•					
3-Hydroxy-2-methylpyridine-4,5-dicarboxylate	198.04	M+H	91.1	•					
3-Dehydroxycarnitine	146.11	M+H	78.5	•					
Hypotaurine	110.02	M+H	49.1	•					
Taurine	126.02	M+H	49.1	•	•				
Tricosanoylglycine	412.37	M+H	567.6		•				
Pristanoylglycine	356.31	M+H	505.1		•	•			
Beta-Citryl-L-glutamic acid	322.07	M+H	695.4		•				
PC(44:12)	878.56	M+H	612		•				
3-Dehydroxycarnitine	146.11	M+H	79.5		•		•		
Glucose 6-phosphate	261.03	M+H	48.4		•				
Citrulline	176.1	M+H	57.2		•			•	
L-Acetylcarnitine	204.12	M+H	106.4		•				
L-Aspartic acid	134.04	M+H	55.5		•				
L-Asparagine	133.06	M+H	55.4		•				
L-Carnitine	162.11	M+H	73.1		•		•	•	
SM(32:1)	675.54408	M+H	559.6			•			
LysoPC (22:6)	568.34	M+H	407.3			•			
Propionylcarnitine	218.13	M+H	157.2			•			
5-phosphonooxy-L-lysine	243.07	M+H	50.2				•		
3-Cyclohexyldodecane	253.28	M+H	580.2				•		
5-Methylthioadenosine	298.09	M+H	141				•		
L-Tyrosine	182.08	M+H	65.2				•		
Arachidonoyl Serinol	378.29	M+H	505.3					•	•
L-Aspartic acid	134.04	M+H	52.3					•	
L-Threonine	120.06	M+H	53.3					•	
L-Glutamic acid	148.06	M+H	54.5					•	

R script for data analysis

```
# you need install enviGCMS(version>0.5.0) from Github(
devtools::install_github("yufree/enviGCMS") ) for plot function
library(IPO)
library(xcms)
library(xMSannotator)
library(enviGCMS)

# find the parameters for xcms
peakpickingParameters <- getDefaultXcmsSetStartingParams('centWave')
path <- list.files('data/QC/',full.names = T,recursive = T)
peakpickingParameters$ppm <- 5
resultPeakpicking <-
  optimizeXcmsSet(files = path,
                  params = peakpickingParameters,
                  subdir = NULL)
optimizedXcmsSetObject <- resultPeakpicking$best_settings$xset
retcorGroupParameters <- getDefaultRetGroupStartingParams()
resultRetcorGroup <-
  optimizeRetGroup(xset = optimizedXcmsSetObject,
                   params = retcorGroupParameters,
                   subdir = NULL)
writeRScript(resultPeakpicking$best_settings$parameters,
             resultRetcorGroup$best_settings)
sessionInfo()
# get the data
getopqedata <- function(path,
                        index = F,
                        xsmethod = "centWave",
                        peakwidth = c(14, 25),
```

```
ppm = 2.5,
noise = 0,
snthresh = 10,
mzdiff = -0.00395,
prefilter = c(3, 100),
mzCenterFun = "wMean",
integrate = 1,
fitgauss = FALSE,
verbose.columns = FALSE,
BPPARAM = BiocParallel::SnowParam(workers = 4),
rmethod = "obiwarp",
plottype = "none",
distFunc = "cor_opt",
profStep = 1,
center = 2,
response = 1,
gapInit = 0.6176,
gapExtend = 2.4,
factorDiag = 2,
factorGap = 1,
localAlignment = 0,
gmethod = "density",
bw = 0.25,
mzwid = 0.0021748,
minfrac = 1,
minsamp = 1,
gmax = 50,
...) {
```

```
cdffiles <- list.files(path, recursive = TRUE, full.names = TRUE)
if (index) {
        cdffiles <- cdffiles[index]
}
xset <- xcms::xcmsSet(
        cdffiles,
        method = xsmethod,
        snthresh = snthresh,
        mzdiff = mzdiff,
        BPPARAM = BPPARAM,
        peakwidth = peakwidth,
        ppm = ppm,
        noise = noise,
        prefilter = prefilter,
        mzCenterFun = mzCenterFun,
        integrate = integrate,
        fitgauss = fitgauss,
        verbose.columns = verbose.columns,
        ...
)
if (index & length(index) == 1) {
        xset3 <- xset
} else{
        xset <- xcms::group(
                xset,
                method = gmethod,
                bw = bw,
                mzwid = mzwid,
```

```
        minfrac = minfrac,
        minsamp = minsamp,
        max = gmax
)
xset2 <- xcms::retcor(
        xset,
        method = rmethod,
        plottype = plottype,
        distFunc = distFunc,
        profStep = profStep,
        center = center,
        response = response,
        gapInit = gapInit,
        gapExtend = gapExtend,
        factorDiag = factorDiag,
        factorGap = factorGap,
        localAlignment = localAlignment
)
# you need group the peaks again for this corrected data
xset2 <- xcms::group(
        xset2,
        method = gmethod,
        bw = bw,
        mzwid = mzwid,
        minfrac = minfrac,
        minsamp = minsamp,
        max = gmax
)
```

```
        xset3 <- xcms::fillPeaks(xset2, BPPARAM = BPPARAM)
    }
    return(xset3)
}
# make annotation

fanno <-
    function(xset,
             outloc = "./result/",
             mode = 'pos',
             db_name = 'HMDB', num_nodes = 10,...) {
        data <- xcms::groupval(xset, 'medret', "into")
        adduct_weights = cbind.data.frame(Adduct = c('M+H','M-H'),Weight = c(5,5))
        mz <- xcms::groups(xset)[, 1]
        time <- xcms::groups(xset)[, 4]
        data <- as.data.frame(cbind(mz, time, data))
        data <- unique(data)
        if ( mode == 'neg') {
            annotres <-
                xMSannotator::multilevelannotation(
                    dataA = data,
                    mode = mode,
                    outloc = outloc,
                    db_name = db_name,
                    adduct_weights = adduct_weights,
                    filter.by = c("M-H"),
                    mass_defect_mode = mode,
                    num_nodes = num_nodes,
```

```
                ...
            )
        }else{
            annotres <-
                xMSannotator::multilevelannotation(
                    dataA = data,
                    mode = mode,
                    outloc = outloc,
                    db_name = db_name,
                    adduct_weights = adduct_weights,
                    filter.by = c("M+H"),
                    mass_defect_mode = mode,
                    num_nodes = num_nodes,
                    ...
                )
        }
        return(annotres)
    }

# make pathway analysis
mumdata <-
    function(xset,
             lv = NULL,
             name = 'test',
             method = "medret",
             intensity = 'inio') {
        data <- xcms::groupval(xset, method, intensity)
        if (intensity == "intb") {
```

```
        data[is.na(data)] = 0
    }
    if (is.null(lv)) {
        lv <- xset@phenoData[, 1]
    }
    mz <- xset@groups[, 1]
    rt <- xset@groups[, 4]
    mod <- stats::model.matrix(~ lv)
    mod0 <- as.matrix(c(rep(1, ncol(data))))
    fstats <- sva::fstats(data, mod, mod0)
    pvalue <- sva::f.pvalue(data, mod, mod0)
    df <- cbind.data.frame(mz, rt, pvalue, fstats)
    filename <- paste0(name, '.txt')
    utils::write.table(df,
        file = filename,
        sep = "\t",
        row.names = F)
    return(df)
}

# get the data
path <- 'dayvivo/day1/control/'
day1control <- getopqedata(path)
path <- 'dayvivo/day1/high/'
day1high <- getopqedata(path)
path <- 'dayvivo/day1/low/'
day1low <- getopqedata(path)
```

```
path <- 'dayvivo/day14/control/'
day14control <- getopqedata(path)
path <- 'dayvivo/day14/high/'
day14high <- getopqedata(path)
path <- 'dayvivo/day14/low/'
day14low <- getopqedata(path)

path <- 'dayvivo/day1/'
day1 <- getopqedata(path)
path <- 'dayvivo/day14/'
day14 <- getopqedata(path)

# make annotation

fanno(day1,outlot = 'data/day1/')
fanno(day14,outlot = 'data/day14/')

# get the files for annotation
mumdata(day1,name = 'day1')
mumdata(day14,name = 'day14')
# use ‘mummichog -f day1.txt -o myResult’ and ‘mummichog -f day14.txt -o myResult’ to make pathway analysis

# plot the mzrt profile

day1control <- getmzrt(day1control)
day1high <- getmzrt(day1high)
day1low <- getmzrt(day1low)
day14control <- getmzrt(day14control)
```

```
day14high <- getmzrt(day14high)
day14low <- getmzrt(day14low)

pdf(file = 'day1control.pdf',width = 6,height = 5)
plotmr(day1control,inscf = 4,rsdcf = 100)
dev.off()
pdf(file = 'day1high.pdf',width = 6,height = 5)
plotmr(day1high,inscf = 4,rsdcf = 100)
dev.off()
pdf(file = 'day1low.pdf',width = 6,height = 5)
plotmr(day1low,inscf = 4,rsdcf = 100)
dev.off()
pdf(file = 'day14control.pdf',width = 6,height = 5)
plotmr(day14control,inscf = 4,rsdcf = 100)
dev.off()
pdf(file = 'day14high.pdf',width = 6,height = 5)
plotmr(day14high,inscf = 4,rsdcf = 100)
dev.off()
pdf(file = 'day14low.pdf',width = 6,height = 5)
plotmr(day14low,inscf = 4,rsdcf = 100)
dev.off()
```