

Copyright

by

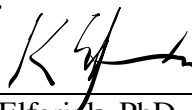
DIANNE ASTER YUNQUE - YAP

2021

**The Dissertation Committee for Dianne Aster Yunque - Yap
certifies that this is the approved version of the following dissertation:**

**BIOMARKERS OF HUMAN EXPOSURE AND EFFECT TO
PETROGENIC POLYCYCLIC AROMATIC HYDROCARBONS**

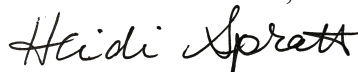
Committee:



Cornelis J. Elferink, PhD,
Mentor and Chair



G.A. Shakeel Ansari, PhD



Heidi Spratt, PhD



[George Golovko \(Aug 24, 2021 19:17 CDT\)](#)

George Golovko, PhD



[Enrique Fuentes Mattei \(Aug 24, 2021 18:29 CDT\)](#)

Enrique Fuentes-Mattei, PhD

**BIOMARKERS OF HUMAN EXPOSURE AND EFFECT TO
PETROGENIC POLYCYCLIC AROMATIC HYDROCARBONS**

by

Dianne Aster Yunque – Yap, BS

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas Medical Branch

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas Medical Branch

August 2021

Dedication

To Paul, my partner in everything, who constantly encourages me to do better every day,

To Astrid, our little ball of sunshine, who has brought so much joy in our lives,

To Pa, Ma, and Van, whose passion have inspired me to be who I am today,

To my lolos and lolas, whose humble beginnings have continued to borne fruit,

To my in-laws, whose strength have always been an inspiration to me,

To Perry, whose life was a testament to what unconditional love can do,

this dissertation is dedicated to all of you.

Acknowledgements

This dissertation would not have been possible if not for the support of so many people.

To Paul, with whom I equally share this huge milestone with, for having infinite patience and for stepping up especially when I was writing this dissertation, for being my constant source of strength, for always believing in me, and for encouraging us to dream;

To Astrid, our shining star, who has been our constant source of happiness especially during this pandemic;

To Pa, Ma, and Van, for the prayers and encouragement;

To my grandparents, for the hard work and dedication to our family that have put us all in places we once just dreamt of;

To my in-laws, for the endless stream of support and Delia's tamales, and for always being there when we need you;

To Kees, my mentor, for all the guidance, patience, and support especially as I navigated my way through graduate school as a new parent, for believing in this project, and for the recommendation letters that always put a smile on my face;

To Li, Adi, Nate, Eric, and Karen, my labmates who practically became my second family, for all the scientific and non-scientific inputs in my life;

To Tracie, for the encouraging talks, endless treats, and for always having our backs;

To Dr. Lisa Elferink and Dr. Ivana Gazovia Stevenson, for the encouragement and the invaluable contributions to my project;

To Dr. Casey Wright and his present lab members Barun, Amy, Madison, and past members Israel, Neli, and Luke, for the valuable insights during lab meetings and for letting us use the qPCR;

To Dr. Bill Ameredes and the T32 group, for providing us with the specialized toxicology training and for always encouraging us to be better scientists;

To the PharmTox graduate program, especially to Dr. Fernanda Laezza, Nicole, and Brenda, for the encouragement and support as we finish strong;

To Dr. Kenneth Johnson and Dr. Wayne Snodgrass, for the words of encouragement that made me stay in graduate school;

To Dr. George Golovko and Dr. Kamil Khanipov for helping me navigate through the intricacies of NGS, and for always being available;

To Dr. Heidi Spratt, for being very patient in helping me with the statistical analysis;

To Dr. Harshica Fernando and Dr. Shakeel Ansari, for generously providing the GC-MS data that is a major part of this project;

To Dr. Widen, Debbie, Jill, and Angie of the Molecular Genetics and NGS core facilities, for being so patient and responsive to my queries, and for running my samples and always generating excellent results;

To Dr. Michael Denison for generously providing the H1L7.5c3 CALUX cell line;

To my dissertation committee members, Dr. Cornelis “Kees” Elferink, Dr. Shakeel Ansari, Dr. Heidi Spratt, Dr. George Golovko, and Dr. Enrique Fuentes-Mattei, for providing very valuable inputs into my project and for always being so supportive and responsive when I send emails especially as I scheduled my committee meeting and the final oral defense;

To Laura and Esme, for the endless laughter and sprinkles of encouragement especially when we were making our way through the first two years of graduate school;

And to all the members of the UTMB GSBS administration, especially to Dr. Janice Endsley for going over the dissertation very thoroughly, and Laura Teed for guiding us until we receive our degree;

THANK YOU.

Biomarkers of Human Exposure and Effect to Petrogenic Polycyclic Aromatic Hydrocarbons

Publication No. _____

Dianne Aster Yunque – Yap, PhD
The University of Texas Medical Branch, 2021

Supervisor: Cornelis J. Elferink

Safety of seafood harvested from the Gulf of Mexico has been a primary concern since the 2010 Deepwater Horizon (DWH) oil spill. Petrogenic polycyclic aromatic hydrocarbons (PAHs) found in crude oil have the potential to bioaccumulate in seafood and render consumers vulnerable to adverse health effects. Despite the knowledge that alkylated PAHs abundant in petrogenic PAH mixtures can be potent activators of the Aryl Hydrocarbon Receptor (AhR) which mediates the formation of PAH toxic metabolites, there has been very limited research done on the human health impacts of petrogenic PAHs. Moreover, biomonitoring protocols applied after the DWH oil spill were predicated on pyrogenic PAHs commonly produced through combustion processes that are not representative of the petrogenic PAHs found in oil. Using an optimized Chemically Activated LUCiferase gene eXpression (CALUX) bioassay, highly bioactive PAHs in longitudinally collected plasma samples of DWH oil spill-affected human populations were identified as biomarkers of petrogenic PAH exposure. Additionally, Next Generation Sequencing (NGS) and quantitative Real Time – Polymerase Chain Reaction (qRT-PCR) were used to identify and verify differentially expressed circulating plasma microRNAs that may serve as novel minimally invasive biomarkers of effect following exposures to

bioactive PAHs including those found in contaminated seafood. The development of novel PAH biomonitoring methods will lend scientific support to toxicological risk assessments and management of human exposures to harmful PAHs.

TABLE OF CONTENTS

List of Tables	xi
List of Figures.....	xiii
List of Abbreviations	xvi
CHAPTER 1: INTRODUCTION.....	1
Deepwater Horizon (DWH) Oil Spill	1
Polycyclic Aromatic Hydrocarbons (PAHs).....	3
Pyrogenic vs. Petrogenic PAHs	3
Aryl Hydrocarbon Receptor (AhR)	5
PAH Metabolism and Toxicity	7
Biomonitoring of Petrogenic PAH Exposure	10
Gaps in Knowledge in Human Petrogenic PAH Exposure.....	12
CALUX Bioassay To Detect Highly Bioactive Petrogenic PAHs	15
MicroRNA and PAH Biomonitoring	17
GC – HARMS.....	20
Specific Aims.....	23
Specific Aim 1: Potent Petrogenic PAHs as Biomarkers of Exposure	24
Specific Aim 2: Plasma MicroRNAs as Biomarkers of Effect.....	24
Significance of the Study	24
CHAPTER 2: BIOACTIVE PAH AS BIOMARKER OF EXPOSURE	26
Introduction.....	26
Hypothesis.....	28
Significance.....	29
Research Strategy	30
Study volunteers and human plasma sample collection	30
Gas Chromatography – Mass Spectrometry (GC-MS).....	30
Chemically Activated LUCiferase gene eXpression (CALUX) Bioassay	31
Maintenance and Seeding of H1L7.5c3 Cells	31
Treatment with Diluted Human Plasma Samples	32

Luciferase Assay	33
Normalization to double stranded DNA Concentration	33
Calculations and Statistical Analysis	34
Optimization of Plasma Dilution	36
Optimal CALUX Exposure Duration to Detect PAHs	37
Results.....	38
Human Plasma samples can be directly used in the CALUX Bioassay	38
Robust CALUX induction is observed at 24-hr BaP exposure of H1L7.5C3	39
dsDNA as a Novel Normalization Factor in CALUX Bioassay	41
Correlation between total PAH body burden and CALUX AhR Bioactivity.....	43
Discussion.....	53
Chapter Summary	64
CHAPTER 3: PLASMA MICRORNA AS BIOMARKER OF EFFECT	66
Introduction.....	66
Hypothesis.....	68
Significance.....	68
Research Strategy	69
Small RNA Extraction	69
Plasma Sample Quality Control – Hemolysis QC	70
Discovery Phase: Next Generation Sequencing (NGS).....	71
Verification Phase: Quantitative Real Time – Polymerase Chain Reaction (qRT-PCR).....	72
Verification of <i>miR-199a-1</i> and <i>let-7a-1</i> precursors.....	74
Target Gene Prediction and Functional Enrichment Analysis.....	75
Statistical Analyses	76
Results.....	77
Phase I: Discovery with Small RNA Next Generation Sequencing	77
Phase II: Verification with qRT – PCR	78
Target Gene Prediction, Gene Ontology Functional Analyses	80
Supplemental Bioinformatics Analysis.....	81

Discussion.....	82
Chapter Summary	88
CHAPTER 4: DISSERTATION SUMMARY	90
APPENDICES.....	95
APPENDIX A: H1L7.5c3 Luciferase reporter gene construct.....	95
APPENDIX B: CALUX Bioassay Plate Layout	96
APPENDIX C: Quant-iT™ Picogreen™ Standard Calculation.....	97
APPENDIX D: Total PAH vs CALUX AhR Bioactivity of all 30 volunteers in each cohort.....	98
APPENDIX D (CONTINUED).....	99
APPENDIX E: Pearson Correlations of all detected PAHs	100
APPENDIX E (CONTINUED)	101
APPENDIX E (CONTINUED)	102
APPENDIX E (CONTINUED)	103
APPENDIX E (CONTINUED)	104
APPENDIX F: Hemolysis Ratios of Plasma Samples for NGS and qRT-PCR	105
APPENDIX G: TaqMan miRNA Assay Target Sequences	106
APPENDIX H: PCR and qRT-PCR Protocols	107
APPENDIX I: Upregulated miRNAs in High CALUX Bioactivity Samples ..	108
APPENDIX J: Downregulated miRNAs in High CALUX Bioactivity Samples	109
APPENDIX K: <i>hsa-miR-17-5p</i> Putative Target Genes	110
APPENDIX L: Second NGS Analysis – Top 30 Differentially Expressed miRNAs	118
APPENDIX M: Ingenuity Pathway Analyses (IPA)	119
APPENDIX N: Ingenuity Pathway Analysis Network.....	124
APPENDIX O: Institutional Review Board Letter	125
REFERENCES.....	127
VITA.....	142

List of Tables

Table 1. Optimized human plasma dilution for the CALUX bioassay.....	39
Table 2. GC-MS Total PAH (ng/mL of plasma) and CALUX Bioactivities (BaPTEQ ng/mL of plasma) median in human plasma samples from four partner communities.	44
Table 3. Total PAH body burden (ng/mL of plasma) and AhR bioactivities (BaPTEQ ng/mL of plasma) of human plasma samples in every wave (annual collection period).	47
Table 4. Summary of PAHs with significant two-tailed Pearson correlations between their concentrations and CALUX AhR bioactivities.	50
Table 5. Discovery Phase: Next Generation Sequencing (NGS).....	77
Table 6. Biological pathways that involve <i>hsa-miR-17-5p</i> putative target genes.....	80
Table 7. CALUX Bioassay 96-well Plate Layout.....	96
Table 8. Example calculation for Lambda DNA standards.	97
Table 9. Pearson correlation results generated from analyzing ALL locations.	100
Table 10. Pearson correlation results for MVC Gulfport, MS.....	101
Table 11. Pearson correlation results of CEEJ Biloxi, MS.....	102
Table 12. Pearson correlation results for UHN Houma, LA.....	103

Table 13. Pearson correlation results for Galveston, TX.....	104
--	-----

List of Figures

Figure 1. Map of the Deepwater Horizon (DWH) oil spilled in the Gulf of Mexico.	1
Figure 2. Parent and Substituted C2-Naphthalene.	4
Figure 3. Petrogenic vs. Pyrogenic PAH mixture composition.	4
Figure 4. Aryl Hydrocarbon Receptor (AhR)-mediated formation of PAH toxic metabolites.	7
Figure 5. CYP450-mediated BaP biotransformation that leads to BPdG Bulky DNA adducts.	9
Figure 6. Crude oil PAH composition vs. EPA's 16 priority PAH pollutant list.	13
Figure 7. Chemically Activated LUCiferase gene eXpression (CALUX) bioassay..	16
Figure 8. MicroRNA biogenesis, function, and distribution.	18
Figure 9. MicroRNAs are stable in frozen human plasma.....	19
Figure 10. Locations of partner communities in the Gulf of Mexico.	21
Figure 11. Seafood quality perception and consumption rose post DWH oil spill....	22
Figure 12. Schematic of extended exposure of H1L7.5c3 cells to BaP or TCDD. ...	37
Figure 13. Luciferase expression (RLU/ng of DNA) of H1L7.5c3 cells upon treatment with BaP and TCDD.	40

Figure 14. Modified CALUX Bioassay with novel dsDNA normalization method generated reproducible data. Independent experiments achieved acceptable RSD values (<30%).	41
Figure 15. Treatment of the H1L7.5c3 CALUX cell line with human plasma samples generated acceptable and reproducible data.	42
Figure 16. Total PAH body burden measured by GC-MS for each location do not correlate with AhR bioactivities measured by the CALUX bioassay.	44
Figure 17. Total PAH body burden from each wave do not correspond with the AhR bioactivities.	47
Figure 18. Total PAH body burden of individual samples do not follow the same trend as their mean AhR bioactivity.	48
Figure 19. Two-tailed Pearson Correlations of C3-Naphthalene plasma concentration and CALUX AhR bioactivity in all four cohorts.	52
Figure 20. Firefly luciferase emission do not overlap with Quant-iT™ Picogreen™ excitation and emission wavelengths.	57
Figure 21. Small RNA Next Generation Sequencing identified 91 unique differentially expressed miRNAs.	78
Figure 22. Verification Phase: quantitative Real-Time PCR.	79
Figure 23. Predicted diseases associated with the candidate biomarker <i>hsa-mir-17-5p</i>	81

Figure 24. Luciferase reporter gene vector backbone used to construct the pGudLuc7.5.	95
Figure 25. MVC Gulfport, MS CALUX AhR bioactivities and GC-MS Total PAH body burdens.....	98
Figure 26. CEEJ Biloxi, MS CALUX AhR bioactivities and GC-MS Total PAH body burdens.....	98
Figure 27. UHN Houma, LA CALUX AhR bioactivities and GC-MS Total PAH body burdens.....	99
Figure 28. Galveston, TX CALUX AhR bioactivities and GC-MS Total PAH body burdens.....	99
Figure 29. GC-HARMS human plasma samples passed hemolysis test.	105

List of Abbreviations

AhR	Aryl Hydrocarbon Receptor
AhRR	Aryl Hydrocarbon Receptor Repressor
AIB1	Amplified in Breast Cancer 1
ANOVA	Analysis of Variance
ARE	Antioxidant Response Element
ARNT	Aryl Hydrocarbon Receptor Nuclear Translocator
BaP	Benzo[a]pyrene
BaPDE	Benzo[a]pyrene-7,8-diol-9,10-epoxide
BaPTEQ	Benzo[a]pyrene Toxic Equivalent
bHLH/PAS	Basic helix-loop-helix/Per-ARNT-Single minded family of transcription factors
BMI	Body Mass Index
BP	British Petroleum
CALUX	Chemically Activated LUCiferase gene eXpression Bioassay
CBPR	Community-based Participatory Research
CEEJ	Center for Environmental and Economic Justice in Biloxi, Mississippi
ChIP	Chromatin Immunoprecipitation
Cq	Quantitation Cycle
CT	Quantitation Cycle
CYP	Cytochrome p450 monooxygenase
CYP1A1	Cytochrome P450 monooxygenase family 1 subfamily A polypeptide 1
df	Degrees of freedom
DGCR8	DiGeorge syndrome chromosomal region 8
DMBA	7,12-dimethylbenz[a]anthracene
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
dsDNA	Double stranded deoxyribonucleic acid
DWH	Deepwater Horizon
E2F1	E2F Transcription Factor 1
EH	Epoxide hydroxylase
ER	Estrogen receptor
ERMA	Emergency Response Management Application
FBS	Fetal Bovine Serum
FC	Fold change
FDR	False Discovery Rate

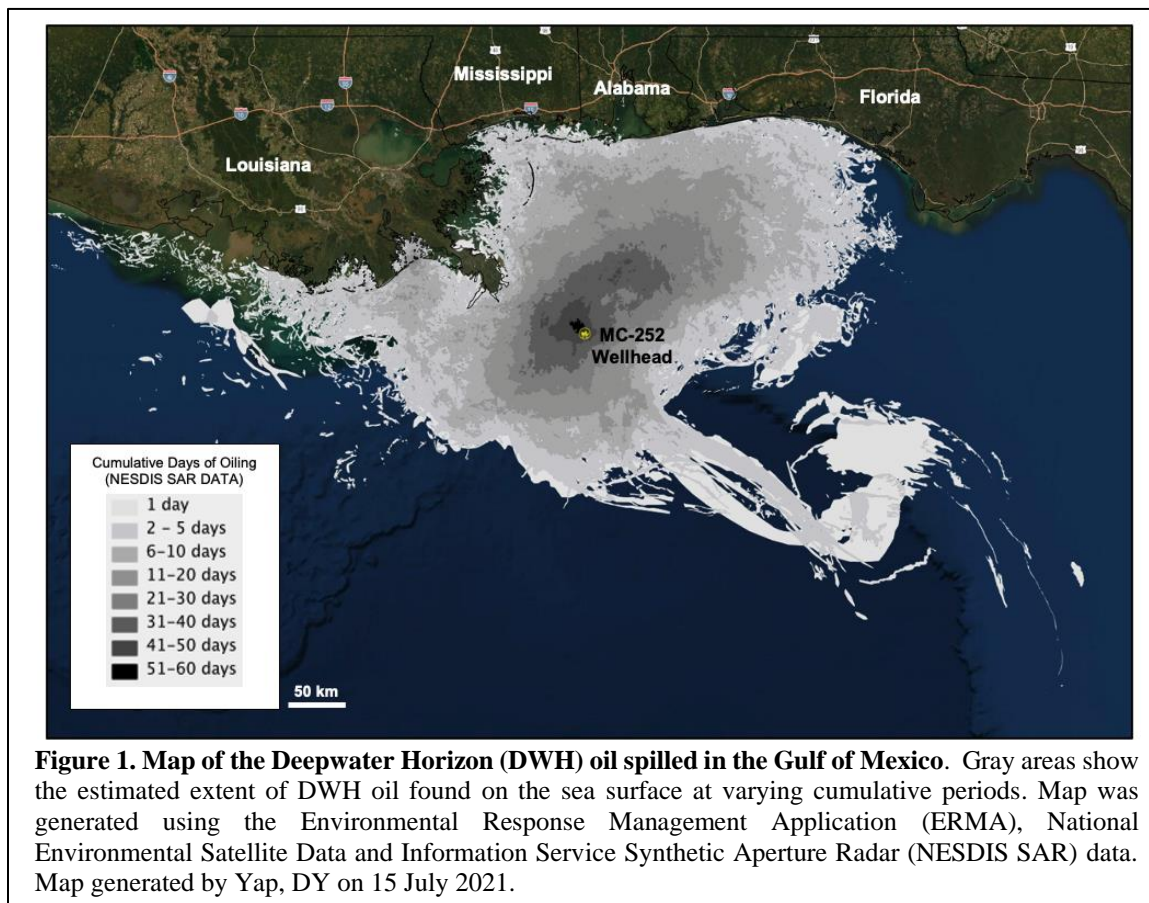
G	Guanine base
GB	Galveston Bay
GC-HARMS	Gulf Coast Health Alliance: health Risks associated with the Macondo Spill
GC-MS	Gas Chromatography – Mass Spectrometry
GJIC	Gap junction intercellular communication
GO	Gene Ontology
GRE	Glucocorticoid Response Element
HMW	High molecular weight
HSP90	Heatshock protein 90
IPA	Ingenuity Pathway Analysis
IQR	Interquartile Range
IRB	Institutional Review Board
KLF	Krüppel-like Factor
LMW	Low molecular weight
LOD	Limit of Detection
LOQ	Limit of Quantitation
LSU	Louisiana State University
MCF-7	Michigan Cancer Foundation – 7 breast cancer cell line
miEAA	microRNA Enrichment Analysis and Annotation
mir	Precursor microRNA
miR	Mature microRNA
miRNA	MicroRNA
MN	micronuclei
MRE	Metal Response Element
miRNA	MicroRNA
mRNA	Messenger RNA
MSDS	Material Safety Data Sheet
MVC	Mississippi Vietnamese Community in Gulfport, Mississippi
NCOA3	Nuclear Receptor Coactivator 3
NESDIS	National Environmental Satellite Data and Information Service
NGS	Next Generation Sequencing
NIC	National Incident Command
NOAA	National Oceanic and Atmospheric Administration
PAH	Polycyclic Aromatic Hydrocarbon
PLB	Passive Lysis Buffer
PCR	Polymerase Chain Reaction
Pre-miRNA	Precursor miRNA
Pri-miRNA	Primary miRNA
qRT-PCR	Quantitative Real Time – Polymerase Chain Reaction

RISC	RNA-induced Silencing Complex
RNA	Ribonucleic acid
ROS	Reactive Oxygen Species
SAR	Synthetic Aperture Radar
SD	Standard Deviation
SEM	Standard Error of the Mean
snRNA	Small nuclear ribonucleic acid
T	Thymidine base
TCDD	2,3,7,8-Tetrachlorodibenzo-p-dioxin
TEF	Toxic Equivalent Factor
TEQ	Toxic Equivalency
TKTD	Toxicokinetic toxicodynamic
<i>TP53</i>	Gene encoding the tumor suppressor protein p53
UHN	United Houma Nation in Houma, Louisiana
US EPA	US Environmental Protection Agency
US FDA	US Food and Drug Administration
UTMB	University of Texas Medical Branch
UTR	Untranslated Region
W	Wave, one annual sample collection
WHO	World Health Organization
XAP2	X-associated Protein 2
XRE	Xenobiotic Response Element

CHAPTER 1: INTRODUCTION

DEEPWATER HORIZON (DWH) OIL SPILL

On April 20, 2010, gaseous hydrocarbons that seeped into the Deepwater Horizon (DWH) semi-submersible oil rig ignited causing a massive explosion that killed 11 workers and injured 17 others¹. Located at the British Petroleum (BP) – operated Macondo Prospect in the Mississippi Canyon 252 (MC-252) 40 miles southeast of the coast of Louisiana, Gulf of Mexico, USA, the offshore operation was going through a series of final pressure tests in preparation for a smaller and less expensive oil rig replacement when the accident occurred¹⁻³. The blowout preventer designed to stop crude oil and gas from travelling upward from the well to the rig platform failed causing the explosion and significant unabated release of crude oil into the Gulf of Mexico^{1,3,4} (Figure 1).



Designated as the largest oil spill in the history of the United States of America to date, the damaged DWH oil rig system dumped approximately 4.9 million barrels of crude oil into the gulf in just a span of 87 days until the wellhead was successfully sealed on July 15, 2010^{1,5,6}. Commercial and recreational activities that were primary economic drivers in the area were then gradually allowed to resume by NOAA after a battery of tests for toxic crude oil chemical levels were satisfied⁷. However, twenty-two months after the capping of the well, fresh oil slicks that matched the chemical characteristics of the DWH crude oil were found in the northern parts of the Gulf of Mexico². This suggests that crude oil from the damaged well continued to leak and possibly contaminated water and food sources already deemed to be safe for consumption. Aside from the massive economic loss by the Gulf Coast communities, which was estimated at \$8.7 billion from 2012-2019 by Sumaila et al. (2012), short- and long-term health effects of consuming crude oil-contaminated seafood was of utmost concern^{1,5,8-10}.

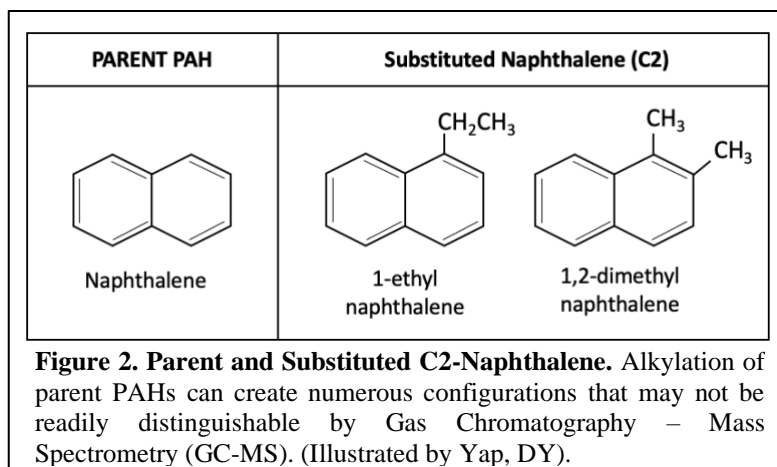
According to the November 2010 Oil Budget Tool technical report by the National Incident Command (NIC), only 25% of the estimated 4.9 million barrels of spilled crude oil was recovered through direct wellhead collection, burning, or skimming¹¹. This leaves 3.675 million barrels of crude oil remaining in the environment either through chemical dispersal (16%), natural dispersal (13%), evaporated or dissolved (23%), or unaccounted for (23%)^{3,11}. To facilitate the breakdown of crude oil into smaller particles and hasten the clean-up process, BP sprayed approximately 1.8 million gallons of Corexit 9527A and Corexit 9500A dispersants onto surface oil, and directly injected into oil plumes escaping the wellhead 5,000 ft below sea level¹²⁻¹⁴. Aside from their controversial inherent biological toxicities, the use of dispersants transformed the crude oil into ultrafine particles that could enter the respiratory system, and to smaller globular particles that were easily digested by marine organisms. This breakdown process is thought to increase the propensity of chemicals of concern found in crude oil, specifically polycyclic aromatic hydrocarbons (PAHs), to be ingested and bioaccumulated by marine organisms^{5,9,12,13}.

POLYCYCLIC AROMATIC HYDROCARBONS (PAHS)

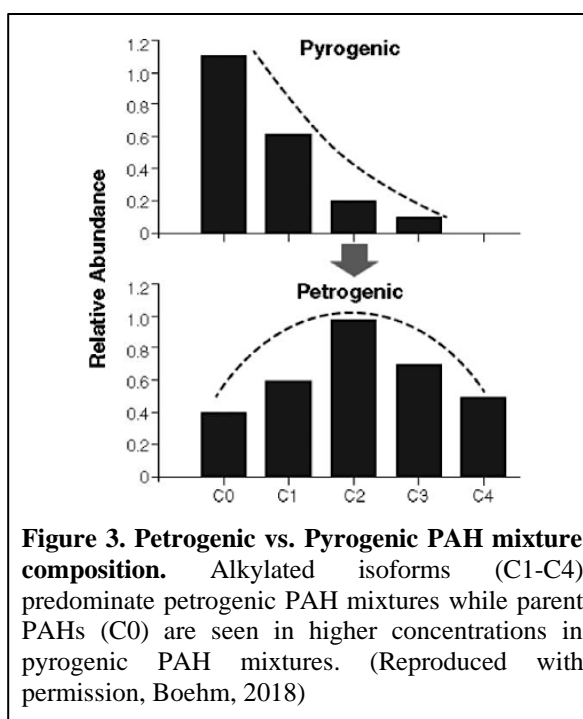
Crude oil contains a wide variety of chemicals, most of which are considered environmental toxicants. While crude oil is of utmost value to the energy sector, the oil contains polycyclic aromatic hydrocarbon (PAHs) that constitute human health concerns^{2,5,15}. PAHs are created through incomplete combustion of organic compounds and are composed of at least two fused aromatic ring structures. These are widely spread environmental pollutants that can be produced both by nature (e.g. crude oil, volcanic eruptions, forest fire, plant products) and anthropogenically (e.g. automobile exhaust, industrial waste, fossil fuel burning)¹⁶⁻¹⁹. There are hundreds of different types of PAHs with the list continuously expanding as different configurations of alkylated, nitrated, and oxygenated conjugates are identified^{16,20-22}. One way of classifying this diverse group, which also naturally exists as complex mixtures, is through their process of formation: petrogenic or pyrogenic^{17,20,23}.

Pyrogenic vs. Petrogenic PAHs

The nature of how PAH mixtures are formed can greatly affect their structures, chemical stability, mixture composition, and environmental persistence. Pyrogenic PAH mixtures are comprised of 5 or more fused aromatic rings and are formed through quick pyrolysis at high temperatures (350°C to 1200°C). These pyrogenic PAHs are commonly found in smoked food, automobile exhaust products, soot, and coal. Conversely, petrogenic PAH mixtures found in crude oil consist of low molecular weight PAHs, usually 2 to 4 fused ring structures, that are formed through slow combustion of organic material at lower temperatures (100°C to 150°C)¹⁷. This process can span millions of years, as in



the case of crude oil, and allows this group of PAHs to possess alkyl groups in many different substitution configurations (Figure 2). The predominance of alkylated PAH structures compared to the unsubstituted parent PAH has also been used to identify the contamination source^{15,24,25}. Petrogenic PAHs are composed largely of alkylated isomers, whereas unsubstituted parent PAHs predominate in pyrogenic PAH mixtures^{20,25,26} (Figure 3). The varying levels of these substituted PAHs, together with other distinct crude oil components, have also been used as chemical fingerprints by environmental regulators and petroleum industries to determine the specific oil reservoir source^{15,26,27}.

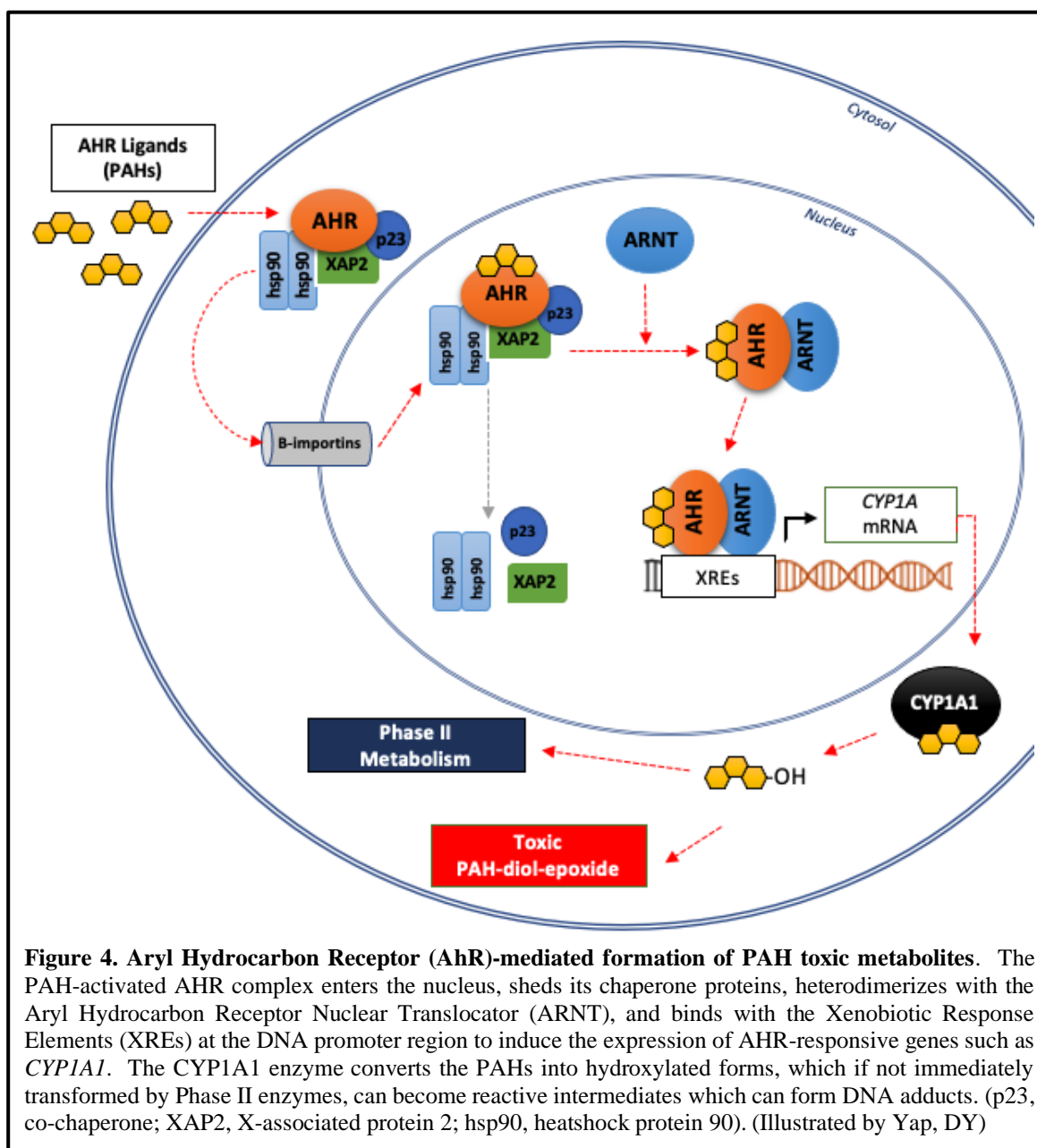


As PAHs become bigger and more angular, they become more lipophilic and more electrochemically stable and potentially more persistent in the environment^{17,20,28}. Thus, studies in the past have heavily focused on the environmental persistence and health impacts of high molecular weight (HMW) pyrogenic PAH exposure²⁰. Numerous studies have already been done to characterize the prototypical pyrogenic PAH benzo[a]pyrene (BaP), and these have uncovered a myriad of biological toxicities associated to high molecular weight PAHs. However, the toxicological effects of alkylation of PAHs at different positions, predominant in the petrogenic PAH isoforms, have not been fully explored. In 2008, Marvanova et al. highlighted that alkylation can increase the bioactivity of PAHs. Their research showed that mono-alkylation of the four ring benz[a]anthracene can make it more potent than unsubstituted benz[a]anthracene, 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), or 7,12-dimethylbenz[a]anthracene (DMBA) to increase the expression of cytochrome P450 1A1 (Cyp1A1) in rat liver cells²⁹. CYP1A1, the primary enzyme responsible for metabolizing PAHs, is controlled by the ligand activation of the aryl hydrocarbon receptor (AhR)^{30,31}.

ARYL HYDROCARBON RECEPTOR (AHR)

The metabolism of xenobiotics such as drugs and PAHs into water-soluble compounds that can be easily eliminated from the body is primarily mediated through the aryl hydrocarbon receptor (AhR) pathway. The AhR is a ubiquitously expressed ligand-activated transcription factor and a member of the basic helix-loop-helix/Per-ARNT-Single minded (bHLH/PAS) family. In its inactive form, the AhR can be found in the cytoplasm where it can be activated by endogenous ligands such as kynurenine or lipophilic exogenous ligands such as PAHs that can enter the cell through passive diffusion³²⁻³⁴. In the canonical AhR pathway, the ligand-bound AhR enters the nucleus through β -importins, sheds its chaperone proteins heatshock protein 90 (hsp90), p23, and X-associated protein

2 (XAP2), then binds with the Aryl Hydrocarbon Receptor Nuclear Translocator (ARNT) to form the active AhR:ARNT dimer. This transcription complex then binds to the xenobiotic response elements (XREs) with the core recognition sequence 5'GCGTG3' in the promoter regions of AhR-responsive genes such as *CYP1A1* and induce their expression. The *CYP1A1* mRNA then gets translated into the CYP1A1 enzyme which hydroxylates the AHR ligands, transforming them into hydrophilic compounds ready for phase II metabolism where the addition of acyl, glucuronic acid, or sulfate functional groups can render the metabolite more hydrophilic, inactive, and ready for elimination from the body³⁵⁻³⁷ (Figure 4). Unfortunately, the AhR pathway responsible for eliminating potentially toxic aromatic compounds can also lead to the bioactivation of PAHs into highly reactive intermediates^{22,38} (Figure 4 and Figure 5).



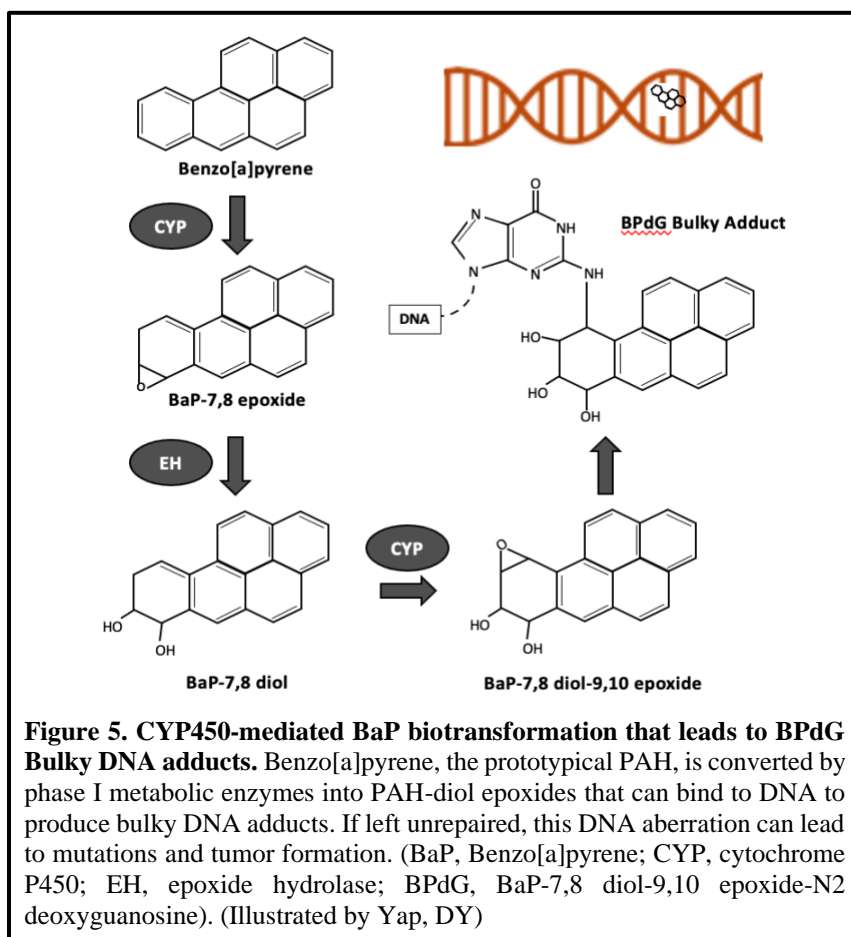
PAH METABOLISM AND TOXICITY

Soot wart, or chimney sweep's carcinoma is one of the earliest occupational exposures recorded that associate PAH exposure to scrotal cancer development. It was described by Dr. Percivall Pott in 1775 as a prevalent disease in London chimney sweeps who were young boys aged 4 to 7 years that fit London's narrow chimney flues^{39,40}. The

boys' exposure to soot was exacerbated by poor hygiene and direct skin contact as they oftentimes worked naked⁴⁰. Since then, more occupational, and environmental exposures to PAH mixtures have been linked to the development of various cancers. Yamagiwa and Ichikawa in 1917 described the formation of skin cancer when coal tar was directly applied on rabbit skin⁴¹. In a comprehensive report in 2012 by the International Agency for Research and Cancer (IARC), cigarette smoking has been strongly associated with lip, larynx, esophagus, and lung cancer; coal and soot exposure to skin, blood, and lung cancer; aluminum smelting to lung and bladder cancer; and coal tar distillation to skin cancer⁴².

Although the primary function of CYP-mediated transformation of PAHs is for detoxification⁴³, several studies have shown that this mechanism can also transform PAHs into bioactive compounds capable of producing genotoxic and non-genotoxic effects with cancer as one of the ultimate endpoints^{20,44,45}. Of the two effects, genotoxic effects of PAHs, particularly of BaP, has been rigorously studied. BaP is a 5-ring PAH that is categorized under the IARC Group 1 – carcinogenic to humans class⁴². It has a high molecular weight, is very lipophilic, and is commonly found in smoked food, automobile exhaust products, coal, and tobacco smoke. Exposure to BaP has been observed to cause adverse biological effects such as immunotoxicity, teratogenicity, epigenetic reprogramming, and tumor formation in animal studies^{16,46}. In humans, BaP exposure is mainly associated with an increase in incidence of cancer development^{17,22,30,42,47}.

Benzo[a]pyrene-7,8 diol-9,10 epoxide (BPDE) is an electrophilic reactive intermediate of CYP-mediated BaP metabolism that can bind to DNA, particularly to guanine (G) bases, to form the adduct BaP-7,8 diol-9,10 epoxide-N2 deoxyguanosine (BPdG) (Figure 5). Fortunately, biology has a way of detecting and repairing this aberration through DNA damage repair mechanisms. However, if the DNA damage repair system is inefficient and the BPdG kink on the DNA remains unchecked, G → T transversion mutations can be introduced in the next round of DNA replication^{17,38,48}. This can be



particularly disconcerting if the damage occurs in tumor suppressor genes or DNA damage repair protein genes rendering them ineffective leading to tumorigenesis and cancer development. Pfeifer et al. (2003) outlined that an increase in occurrence and preference of G → T transversions in *TP53*, the gene encoding the tumor suppressor p53, is seen in cigarette smoking-related lung cancer development⁴⁹.

Compared to the well-studied genotoxicity of PAH exposure, researchers are just starting to uncover their equally concerning non-genotoxic effects. Some PAHs, such as fluorene, phenanthrene, and pyrene, which are not classifiable carcinogens according to the US EPA⁵⁰ have been seen to produce non-genotoxic effects that are also related to cancer. The low molecular weight (LMW) PAHs fluorene, phenanthrene, pyrene, and their alkylated isomers can inhibit gap junction intercellular communication (GJIC), removing

the cell from growth inhibition, and consequently promoting tumorigenesis in rat liver epithelial systems^{51,52}.

The crosstalk between AhR and estrogen receptor (ER) pathways also plays a major role in mediating the non-genotoxic effects of PAHs. Hýzďalová et al. (2018) showed that the metabolite 3-OH-BaP produced through CYP-mediated biotransformation, and not the parent compound BaP, activated the ER pathway in MCF-7 breast cancer cells and promoted ER-dependent progression of cell cycle and cell proliferation. Biotransformed PAHs can thus become ER ligand mimics and interfere with normal estrogen receptor pathways. This can exacerbate cell proliferation, an important mechanism in tumor development^{53,54}. As endocrine disruptors, PAHs particularly BaP have also been observed to affect female reproductive health including embryo implantation^{48,54-56}.

Since PAHs naturally occur as mixtures composed of varying levels of both low and high molecular weight PAHs and their conjugated isomers, the opportunity to promote synergistic toxicities is inevitable^{17,57}. Bioactivated petrogenic PAHs that may have nongenotoxic effects, in mixture with genotoxic PAHs such as BaP, can potentially promote adverse human health effects.

BIOMONITORING OF PETROGENIC PAH EXPOSURE

A biological marker, or biomarker, is a measurable indicator of biological state that may establish exposure to a xenobiotic, a biological response to this exposure, and the inherent susceptibility that may affect the organism's response to the toxicant. There are three different types of biomarkers: biomarkers of exposure, biomarkers of effect, and biomarkers of susceptibility^{45,58,59}.

Biomarkers of exposure directly indicate the biological presence of a toxicant⁵⁹. Examples are the internal doses of the parent compound (e.g. pyrene) or biotransformed compounds (e.g. 1-OH-pyrene, metabolite of pyrene) that may reach biological targets and

potentially elicit toxic effects^{60,61}. Other examples of biomarkers of exposure are BPDE-DNA and protein adducts⁴². The Gas Chromatography – Mass Spectrometry (GC-MS) is an established method that has been used to detect and measure the presence of the original compound and their metabolites in biological samples. Although this method is reliable and very sensitive in detecting the xenobiotics, it is dependent on the existence of known standards and established toxic equivalent factors (TEFs) to allow risk evaluation for complex mixtures^{9,62,63}. It also requires expensive equipment and sample clean-up reagents which may not be readily available⁶⁴⁻⁶⁶.

Biomarkers of effect are measurable indicators of biological alteration in the exposed organism as a consequence to adequate xenobiotic exposure^{45,59,67}. These can be particularly useful in determining preclinical status and help inform decisions to prevent further adverse health outcomes⁶⁸. Some of the biomarkers of effect currently used in PAH toxicology are detection of micronuclei frequency, chromosomal aberrations, DNA damage, and 8-oxo-deoxyguanosine^{42,69,70}. In toxicology, the use of plasma microRNAs has recently been explored as a potential molecular indicator for ample biological exposure to PAHs^{69,71-74}.

Lastly, the biomarker of susceptibility provides an indicator of variability that may affect how an organism responds to a xenobiotic exposure⁵⁹. These may be mutations on genes of enzymes involved in xenobiotic metabolism such as *CYP1A1* or *GST* (glutathione-S-transferase), or in DNA damage repair enzymes that can alter the organism's ability to prevent potential toxicity^{43,45,70,75}.

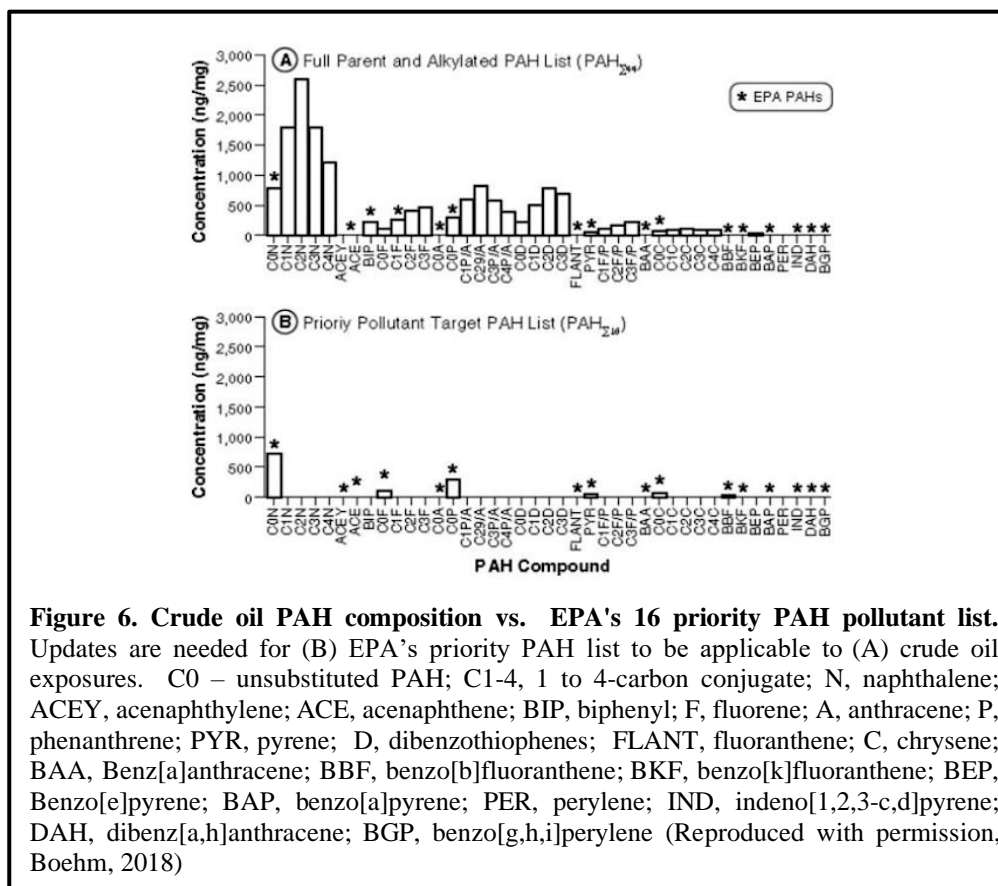
The discovery and verification of different types of biomarkers for petrogenic PAH exposure can yield valuable data that when integrated may better inform risk assessments and uncover important biological pathways implicated in the development of adverse human health effects.

GAPS IN KNOWLEDGE IN HUMAN PETROGENIC PAH EXPOSURE

Despite the recent developments in PAH toxicology research, robust scientific data on human exposures to petrogenic PAHs that can help inform risk assessments and support updates on policies are still lacking. In the last 50 years, oil spills of varying degrees have occurred and yet a scarcity in studies on the biomonitoring and biological ramifications of human exposure to crude oil remains^{5,76}. As a matter of fact, only 7 oil spills prior to the DWH catastrophe have launched human health studies and most of which are cross-sectional in design with a larger focus on acute physical and psychological effects⁷⁶.

The use of *in vitro* or animal models in uncovering PAH toxicities is convenient and provides valuable data. However, *in vitro* models do not provide bioavailability data and some mechanisms are species-specific which may hinder translatability to human biological effects^{28,77}. Dietary exposures can be tricky as the toxicokinetics (i.e. absorption, distribution, metabolism or biotransformation, elimination) of PAHs may not be the same in different species⁷⁸. The latest Toxicological Review of Benzo[a]pyrene document released by the US EPA (2017) provides multiple exposure criteria but are mostly calculated from animal experiments using interspecies uncertainty factors. The authors highlighted that although mice and rat models are currently being used in BaP studies, it is unknown which of the two models is more comparable to humans in terms of toxicokinetics. This limits the utility of Physiologically Based Pharmacokinetic (PBPK) mathematical modeling in predicting exposure criteria values^{77,79}. Thus, there is a need for human exposure studies that can provide species-specific information to update toxicity values used in risk assessments particularly in oil spill events.

Most studies on PAH exposures have also focused on the toxic effects of pyrogenic PAHs such as BaP^{28,42,44,79}. These studies have been the basis of guidelines established by the US Environmental Protection Agency (US EPA) and the US Food and Drug



Administration (US FDA) on PAH exposure. However, petrogenic and pyrogenic PAH mixtures have different chemical compositions (i.e. prevalence of alkylated and LMW PAHs vs. unsubstituted HMW PAHs) and may utilize different mechanisms to induce adverse health effects. The US EPA 16 priority PAHs penned in the 1970's relied on the availability of method standards, and toxicity studies which were mostly on pyrogenic PAHs during that time^{80,81}. Therefore, alkylated PAHs that are predominant in petrogenic PAH mixtures have been excluded from the priority list (Figure 6)^{25,26}. Unfortunately, this priority list was still applied during the DWH oil spill response by various institutions such as the US FDA. As highlighted by Rotkin-Ellman et al. (2012), the FDA excluded alkylated conjugates of naphthalene in calculating risk criteria and allowable PAH threshold levels. The FDA's response was heavily criticized for underestimating the human toxicity risks of consuming crude oil-contaminated seafood^{14,82–84}. US regulations on PAH exposure, albeit

historically served their purpose, undoubtedly need to be updated using more recent human toxicological studies.

The main route of human exposure to PAHs trapped in sediments, particularly LMW petrogenic PAHs, is through the consumption of contaminated marine organisms^{17,28}. Filter feeders such as oysters and mussels that can bioaccumulate these PAHs from the sediments serve as food sources for higher predators which are ultimately consumed by humans⁸⁵. Moreover, crude oil that has been broken down into smaller droplets by dispersants are more readily available to marine organisms^{8,10,28,64}. These pose increased risk for PAH exposure, particularly in populations that rely heavily on seafood for sustenance. Hence, there is a need for dietary PAH exposure studies to establish appropriate levels of concern. Unfortunately, according to the World Health Organization (WHO), data currently available for human PAH exposures are mostly derived from inhalation and dermal exposures in occupational settings. Furthermore, current studies on PAHs rely on the presence of pyrogenic PAHs (e.g. BaP) or metabolites of unsubstituted PAHs (e.g. 1-hydroxypyrene) in human biological samples to monitor exposure^{60,61,86,87}. Therefore, there is a need to identify biomarkers that encompass exposures to petrogenic PAHs associated with the consumption of crude oil-contaminated seafood. These may also help establish more appropriate threshold levels for dietary PAH exposure and provide insights on biological processes specifically affected by petrogenic PAH mixture exposure.

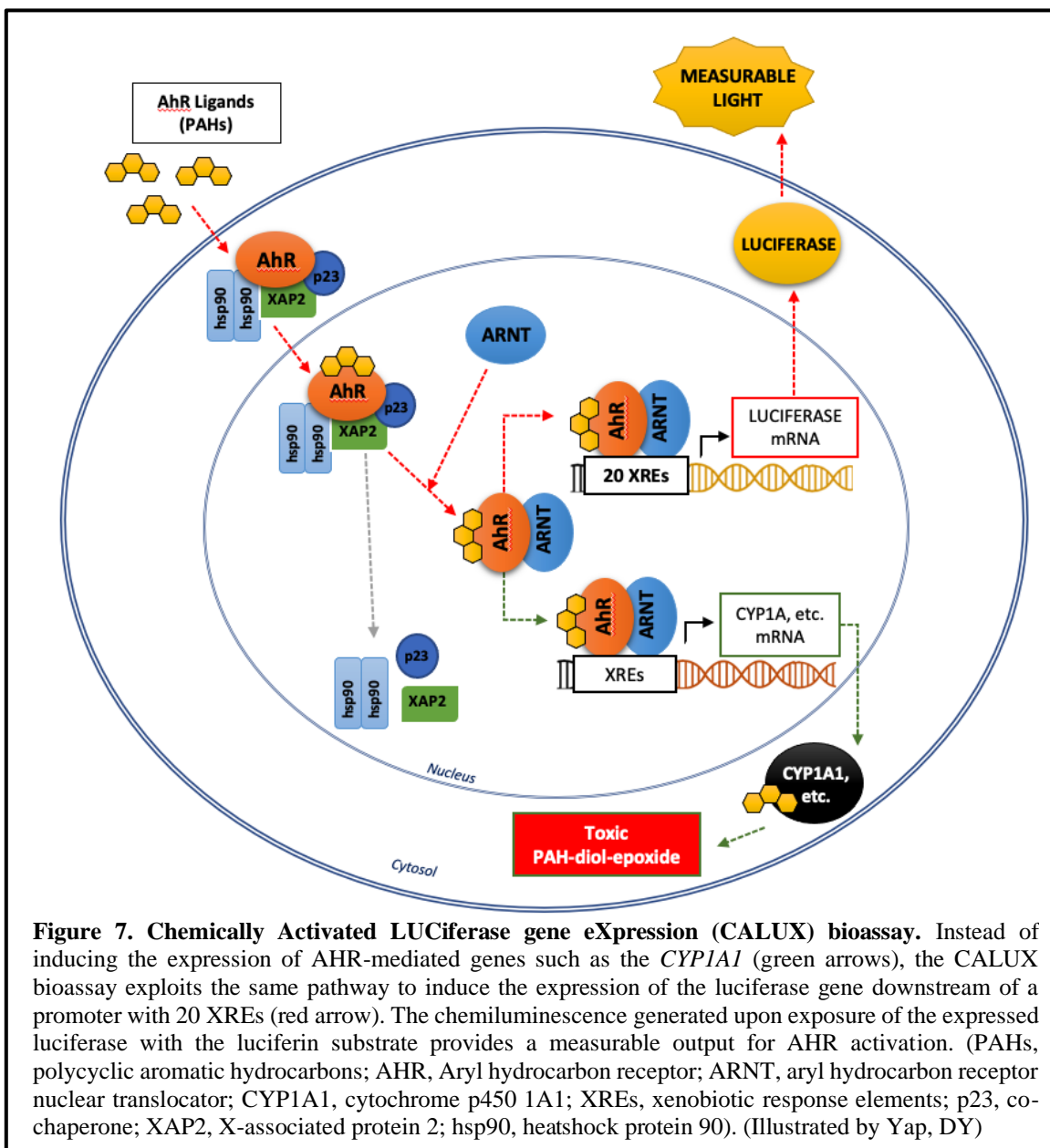
Several methods have been developed to identify biomarkers for monitoring PAH mixture exposure. The use of GC-MS to detect the presence and levels of PAHs has been the gold standard for years^{65,80,88,89}. It uses the toxic equivalency (TEQ) method to estimate the overall toxicity of a given exposure based on toxic equivalency factors (TEFs) of PAHs with established comparisons to benzo[a]pyrene toxicity^{19,80,90}. However, this method relies heavily on selected known PAH standards and the additivity of their TEQs, if available, which may lead to a gross underestimation of mixture toxicity^{9,62,63}. It is also impractical to establish standards for all the different types of PAHs including their

conjugated isomers that may exist in hundreds of configurations. And as human exposure to these toxicants never involves just a single PAH^{16,17}, there is a need to develop and verify other types of biomonitoring tools that may better estimate overall mixture toxicity.

CALUX BIOASSAY TO DETECT HIGHLY BIOACTIVE PETROGENIC PAHS

The Chemically Activated LUCiferase gene eXpression (CALUX) bioassay is a US EPA certified relatively low cost and rapid screening method developed to detect the presence of bioactive toxicants in a variety of sample types⁹¹⁻⁹⁴. This method involves a recombinant luciferase reporter gene system that generates a quantifiable output for AhR-mediated gene expression, the primary biological pathway activated that can induce the CYP-mediated biotransformation of PAHs into toxic metabolites. In this cell-based assay, ligand-activated AhR binds to the xenobiotic response elements (XREs) present on the promoter upstream of the luciferase gene, thereby inducing the expression of the luciferase enzyme.⁹⁵⁻⁹⁷ Figure 7 shows this mechanism and depicts the reporter system harboring a cassette encompassing 20 XREs upstream of a luciferase reporter gene stably expressed in the H1L7.5c3 cell line. This system is very sensitive down to femtomolar concentrations of bioactive AhR ligands. The chemiluminescence generated after the addition of the luciferase substrate provides a measurable output that represents degree of pathway activation by AhR ligands present in the sample. These light intensities are compared to luciferase expression levels obtained following induction with the prototypical PAH benzo[a]pyrene and are ultimately expressed in benzo[a]pyrene toxic equivalents (BaPTEQ/mL of plasma sample)^{92,98}. Exposure of this cell-based assay to bioactive PAHs in the sample can lead to robust CALUX readouts. This underscores the importance of the CALUX bioassay as a biomonitoring tool in identifying samples harboring functional AhR agonist activities with potential biological toxicities.

Compared to analytical methods such as the GC-MS, the CALUX bioassay is not limited to a selected—and likely incomplete—list of known PAHs to derive overall sample potential toxicity^{19,63,99,100}. For years, the biomonitoring of PAH exposure using GC-MS has been limited to the identification of 16 PAHs in the EPA 1972 priority list for practical reasons^{18,80,101}. This list only includes parent PAHs and mostly of high molecular weight^{80,102}. On the contrary, petrogenic PAH mixtures found in crude oil have higher

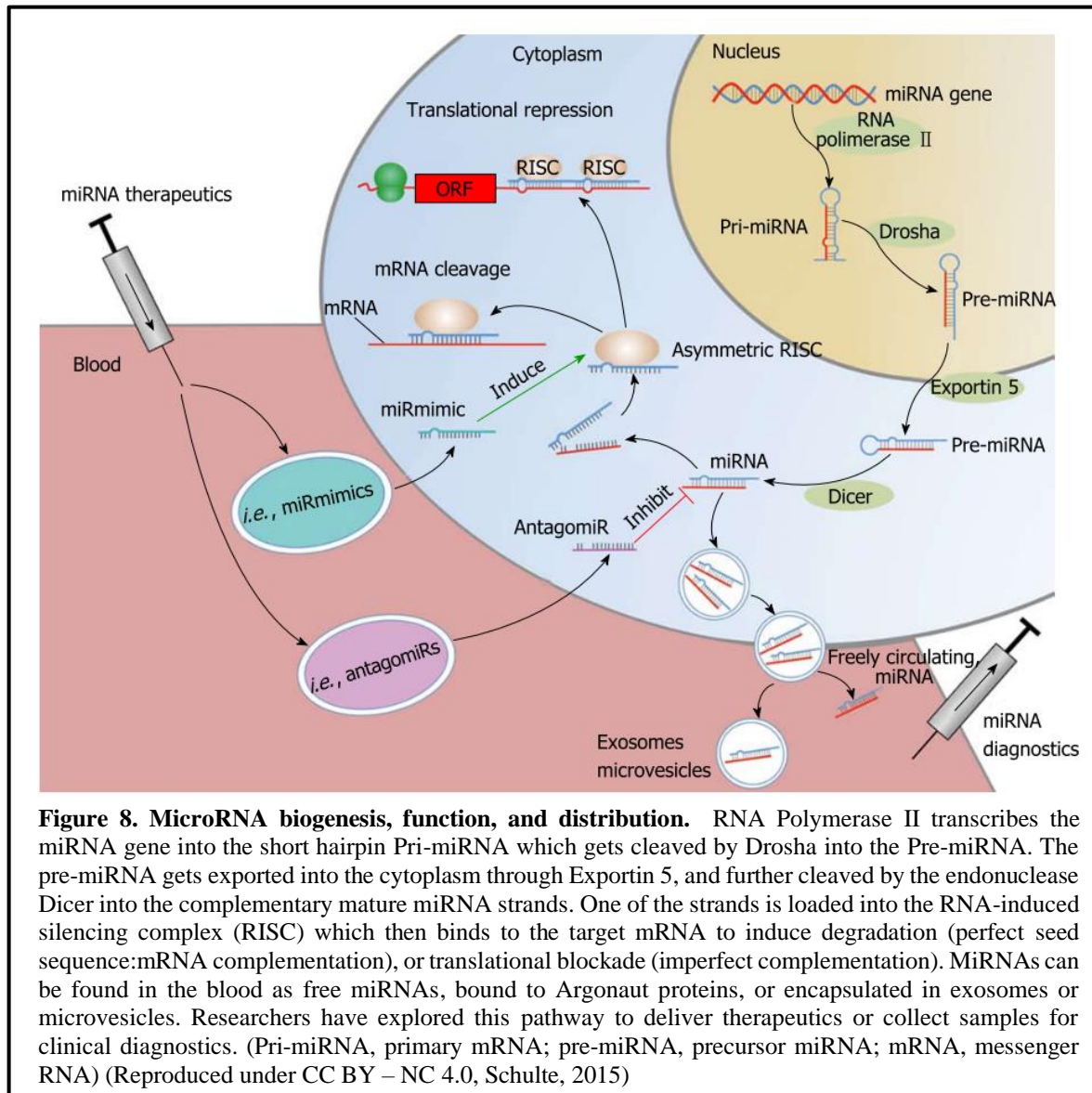


concentrations of alkylated and low molecular weight PAHs, most of which do not have available analytical standards^{15,26,103}.

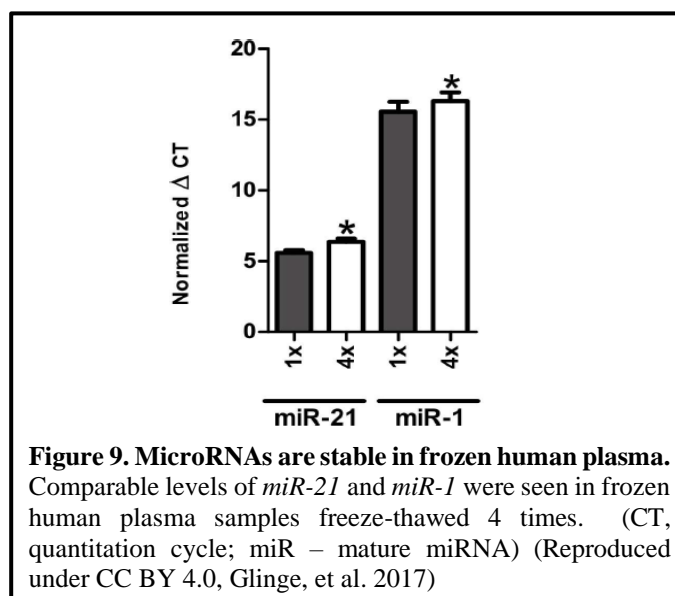
Integrating the overall toxicity potential data generated by the CALUX bioassay with the identities and concentrations of an expanded list of individual PAHs that include low molecular weight alkylated isomers can help identify highly bioactive petrogenic PAHs present in plasma that may serve as appropriate biomarkers of human exposure.

MICRORNA AND PAH BIOMONITORING

Recent developments in the biomarker discovery field have highlighted the use of plasma microRNAs (miRNAs) in detecting exposures to xenobiotics^{104,105}. MiRNAs are short 18-22 nucleotide (nt) non-coding RNAs that can post-transcriptionally regulate gene expression. These biomolecules are derived from a primary miRNA (pri-miRNA) transcribed from the miRNA gene by RNA polymerase II. The pri-miRNA is cleaved by Drosha:DGCR8 complex into the 60-70 nt stem-loop precursor miRNA (pre-miRNA), which gets transported to the cytoplasm through Exportin 5. The pre-miRNA is further processed by the endonuclease Dicer into the mature miRNA complementary strands. One strand is loaded into the RNA-induced silencing complex (RISC) and the 2-7nt sequence on its 5' end hybridizes to the 3' untranslated region (3'UTR) of perfectly complementary messenger RNAs (mRNAs) sequences to cause mRNA destabilization and degradation, or forms imperfect duplexes to physically block the ribosome complex from completing translation (Figure 8)^{104,106–111}.



The mechanism underlying these events were first described by the separate research groups of Ruvkun and Ambros in 1993. They observed that in *C. elegans*, *lin-4* RNA is capable of binding to the 3' UTR of the *lin-14* RNA and repressing the expression of LIN-14 protein without changing *lin-14* transcript levels^{112,113}. In subsequent years, the discovery of the highly conserved *let-7* miRNA family in different species triggered an explosion of research on miRNAs¹⁰⁷. To date, there are over 2000 miRNAs identified in the human genome and are predicted to control the expression of at least 60% of protein-coding genes^{108,114}.



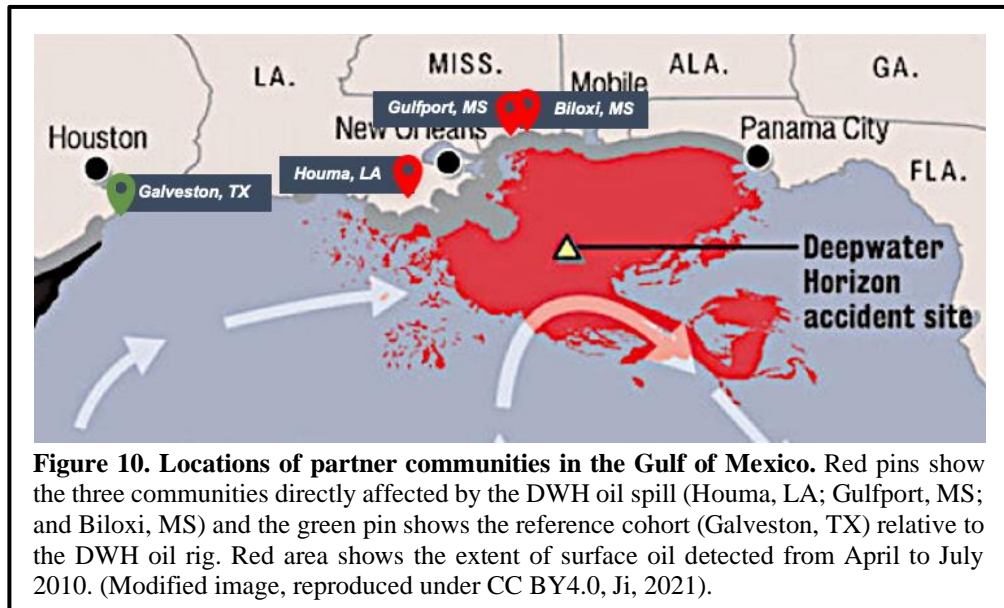
With its ability to control over thousands of biological pathways, miRNAs were originally examined for their therapeutic potential such as shown in Figure 8^{104,107,109}. However, these regulatory biomolecules became even more interesting when they were reported to be excellent biomarker candidates for diseases such as cancer^{115–117}. Dubbed as a “liquid biopsy”^{116–118}, these biomolecules possess the characteristics of an ideal biomarker – present and measurable in easily obtainable biological specimens such as blood, biologically and clinically related to the degree of disease or exposure in an organism, require small sample quantities, and provide a highly sensitive measure using readily performed techniques^{58,68}. MiRNAs are resistant to degradation owing to their short length and integration into protein complexes in the RISC or being encapsulated in exosomes or microvesicles. Glinge et al. (2017) have also observed that miRNAs levels are stable in human biofluids even after multiple freeze-thaws (Figure 9). In 2009, Choudhuri et al. alluded to the possibility of xenobiotics mediating their biological effects through the transcriptional regulation of miRNAs. Their group mentioned that miRNA biogenesis largely involves RNA polymerase II promoters such as glucocorticoid response elements (GREs), antioxidant response elements (AREs), and metal response elements (MREs)¹⁰⁴. Therefore, xenobiotics that can act through these promoter elements can also

regulate the expression of the miRNA genes^{68,104}. Using Chromatin Immunoprecipitation (ChIP), Hanieh (2015) demonstrated that activated AhR binds to XRE's upstream of the miRNA 212/132 cluster in T47D human breast cancer¹¹⁹. Deng et al. (2014) explored the utility of circulating miRNAs in detecting the degree of exposure to pyrogenic PAHs in the plasma of coal coke oven workers in China. Five candidate miRNAs were correlated with an increase in micronuclei frequency, as biomarkers of effect in the exposed group⁶⁹. To date, there are no publications on the discovery and verification of circulating miRNAs as biomarkers for human exposure to petrogenic PAHs.

GC – HARMS

The Gulf Coast Health Alliance: health risks related to the Macondo Spill (GC-HARMS) was established in 2011 in response to the DWH oil spill. Deep concerns for potential short and long-term human health impacts of consuming contaminated seafood were conveyed by Gulf Coast stakeholder communities that had a longstanding relationship with the University of Texas Medical Branch (UTMB). Spearheaded by the UTMB P30 Center in Environmental Toxicology, a partnership with research groups from the University of Pennsylvania P30 Center for Excellence in Environmental Toxicology, Texas A&M University at Galveston, Louisiana State University (LSU), and Gulf Coast stakeholder communities hailing from different backgrounds, was formed. To allay the fears of the Gulf communities as it pertains to seafood quality and human health impacts of the oil spill, the consortium developed a U19 inter-institutional project that utilized a Community-Based Participatory Research (CBPR) approach. This strategy required the active involvement of the Gulf Coast partner communities in all aspects of the project—from planning to implementation of methodologies that supported the consortium's goals. The GC-HARMS consortium pursued four main goals divided amongst the research groups: 1) assess the petrogenic PAH contamination of seafood collected from different

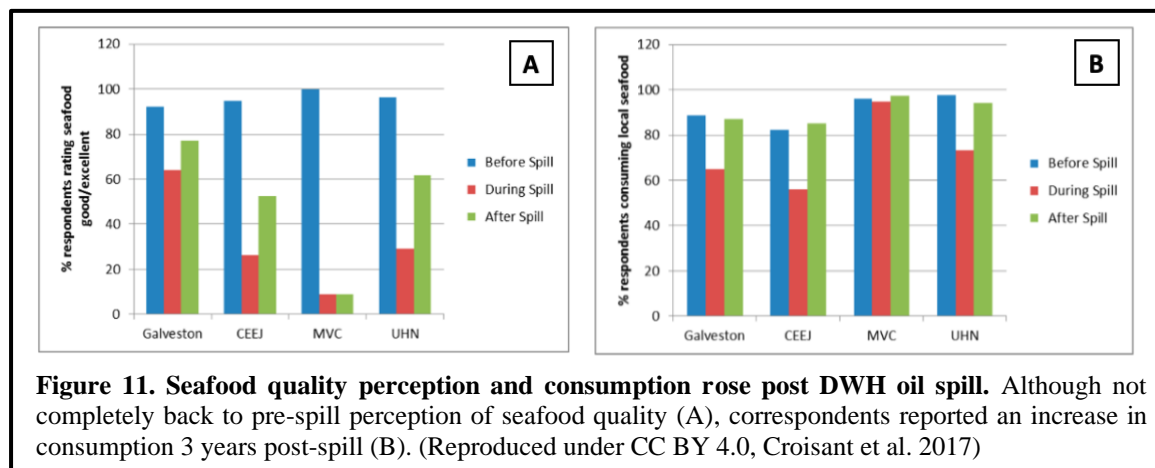
areas in the Gulf of Mexico, 2) determine the toxicity of the petrogenic PAHs, 3) evaluate long term health outcomes of petrogenic PAH exposure using a longitudinal cohort study, and 4) communicate the findings to the concerned stakeholders particularly the fisherfolks and population along the coast that rely on the Gulf for food and income^{5,9,120}.



One main strength of this project is the implementation of a longitudinal sampling framework that allowed researchers to identify attributes in seafood quality, and mental and physical human health that were altered as years passed by. Due the unpredictability of the occurrence of these massive oil spills, baseline data from seafood and human populations are normally not available. The GC-HARMS annual collection of seafood from affected areas in the Gulf of Mexico and human biological samples (i.e. plasma and urine) from approximately 100 volunteers encompassed four distinct communities (Figure 10)¹²¹. These were MVC (Mississippi Vietnamese Fishing Community, Gulfport, MS, USA); CEEJ (Center for Environmental and Economic Justice, Biloxi, MS, USA); UHN (United Houma Nation, Houma, LA, USA); and Galveston, TX, USA which served as the comparison unaffected cohort. Sampling spanned a 3-year period post-oil spill and provided an opportunity to see changes in trends of selected parameters in each site and

allowed for the use of the initial sampling collection as a surrogate baseline^{5,9}. Each human sample collection was also accompanied by a robust array of information captured using an extensive survey that included questions on seafood consumption patterns, overall health assessment performed by medical professionals, and thorough blood and urine clinical analyses. To date, numerous publications have been generated and continue to be developed by the consortium which addresses the different goals outlined earlier^{5,9,120,122}.

Pertaining to behavioral or mental health effects of the DWH oil spill, Croisant et al. (2017) evaluated the wave 1 survey data (collected 3 years after DWH oil spill) and showed that there were no significant differences in anxiety and depression among the four partner communities who participated. However, there was an increase in post-traumatic stress disorder (PTSD) in the three communities that were directly affected by the oil spill. Interestingly, the researchers noted a mark deterioration of self-reported health status in the three cohorts directly affected by the oil spill which was not observed in control cohort in Galveston, TX. And although they reported a decrease in seafood quality during and post-oil spill, MVC Gulfport, MS volunteers continued to consume seafood from the Gulf at pre-oil spill frequencies (Figure 11). The percentage of correspondents who consumed seafood returned to almost pre-spill levels for all the other cohorts despite a decrease perception of Gulf seafood quality⁵.



To aid the general population in determining the amount of Gulf seafood that can be safely consumed while limiting dietary exposures to carcinogenic PAHs, the group developed a seafood consumption calculator (<https://www.utmb.edu/scg/>)¹²³. This calculator uses current risk criteria in line with current US regulations, and other user input parameters such as weight, self-reported risk level, and seafood consumption frequency⁹. Jackson et al. (2019) also evaluated the current risk assessment methods applied to major oil spill incidents and highlighted the need for updates to include alkylated petrogenic PAHs.

Fernando et al. (2019) reported that the seafood (i.e. shrimp, crab, and oysters) collected from DWH-affected marine sites have PAH levels that are below FDA levels of concern. However, they emphasized that the biological and potential toxicological effects of alkylated petrogenic PAHs have not been fully elucidated and may thus need more comprehensive research. They also referred to the immense value that discovering novel biomarkers of exposure and effect may contribute to fully realizing the human health impacts of the DWH oil spill¹²⁰.

SPECIFIC AIMS

Equipped with the human plasma samples and the accompanying rich data sets collected annually over three years by the GC-HARMS consortium from 400 volunteers in four Gulf Coast communities after the DWH oil spill, and our current understanding of AhR biology and xenobiotic biomonitoring, this dissertation will bridge knowledge gaps in human health risk assessments pertaining to PAH exposures including those with petrogenic origins, in two specific aims:

Specific Aim 1: Potent Petrogenic PAHs as Biomarkers of Exposure

We aim to identify petrogenic PAHs in the human plasma samples with the highest potencies to activate the AhR pathway. We hypothesize that the petrogenic PAHs detected and measured in the GC-HARMS human plasma samples will exhibit varying AhR activation potencies using the Chemically Activated LUciferase gene eXpression (CALUX) bioassay. Integrating the individual concentrations of 36 PAHs (parent and their alkylated analogues) quantified by GC-MS with the bioactivation potencies measured using the CALUX bioassay will help identify the highly bioactive petrogenic PAHs that can serve as our biomarkers of exposure.

Specific Aim 2: Plasma MicroRNAs as Biomarkers of Effect

We seek to identify circulating miRNAs in plasma samples that can serve as novel biomarkers of effect following exposure to PAHs, including petrogenic PAHs found in crude oil. We hypothesize that human plasma samples containing highly bioactive petrogenic PAHs that can activate the AhR pathway will yield differentially expressed plasma miRNAs valuable in identifying petrogenic PAH exposures.

In essence, this dissertation proposed that human exposure to highly bioactive petrogenic PAHs (biomarkers of exposure) will generate a unique microRNA expression profile that can be used as our biomarkers of effect.

SIGNIFICANCE OF THE STUDY

This dissertation will address knowledge gaps in human exposure to petrogenic PAHs and will help support updates on human health risk assessments as previously expressed in the gaps in knowledge section of this chapter (page 10).

Currently available biomonitoring tools for PAH exposure were established based on pyrogenic PAH studies and may not translate well to petrogenic PAH exposures. To address this need, this dissertation assessed biomarkers of exposure using human plasma samples that were collected in the wake of the DWH disaster with concomitant exposures to petrogenic PAHs found in crude oil through the consumption of tainted seafood. Results of this study will provide valuable petrogenic PAH exposure-specific data that may ultimately better inform human health risk assessments applied to oil spill events.

The direct identification of biomarkers of exposure and effect using the human biological samples in lieu of animal models lends strength to the translational value of this dissertation's results.

Lastly, identifying differentially expressed miRNAs in human plasma associated with exposure to highly bioactive PAHs will provide a novel and rapid biomonitoring tool that may aid risk assessments in petrogenic PAH exposure events resulting from inevitable future oil spills.

CHAPTER 2: BIOACTIVE PAH AS BIOMARKER OF EXPOSURE

INTRODUCTION

The catastrophic explosion of the Deepwater Horizon oil spill led to the release of approximately 4.9 million barrels of crude oil into the Gulf of Mexico^{5,11}. Aside from the massive economic loss estimated to amount to \$8.7 billion dollars from 2012-2019^{5,10}, communities surrounding the Gulf that rely on locally sourced seafood for their economic livelihood and sustenance were deeply concerned with seafood quality and adverse health outcomes from consuming contaminated seafood^{5,9,120}. One of the chemicals of concern in crude oil are the polycyclic aromatic hydrocarbons (PAHs)^{1,8,11}.

PAHs are products of incomplete combustion of organic compounds and are composed of at least 2 fused aromatic rings^{18,47,64}. These ubiquitous xenobiotics can be produced by nature (e.g. volcanic eruption, fossil fuel), or through anthropogenic processes (e.g. industrial products, automobile exhaust, cigarette smoke)^{16,18,124}. Owing to their electrochemically stable structure and increase in hydrophobicity as more rings are fused^{16,64}, PAHs have been found to be environmentally persistent and can be bioaccumulated in organisms^{16,19,125}. These also exist as mixtures and as such, exposure to PAHs is never to a single compound^{17,22}.

PAH mixtures can be classified based on the process of their formation – pyrogenic or petrogenic. Pyrogenic PAH mixtures are formed through quick pyrolysis at high temperatures. These mixtures such as those found in cigarette smoke and coal are characterized by the predominance of high molecular weight unsubstituted PAHs. On the other hand, petrogenic PAH mixtures found in crude oil are formed at low temperatures over a long period of time. Due to this slow formation, alkylated forms of the PAHs become a predominant characteristic of this type of PAH mixture^{17,23,64}.

Our current knowledge on the human health impacts of PAHs has been developed largely through studying the biological effects of high molecular weight (HMW) PAHs such as benzo[a]pyrene (BaP)^{79,126}. BaP is a carcinogenic compound predominant in cigarette smoke, coal, and forest fires. This xenobiotic has been established to mediate its genotoxic effects through activating the Aryl Hydrocarbon Receptor (AhR) pathway^{43,124,127}. Activation of the AhR by PAHs induces the expression of PAH-metabolizing enzymes such as cytochrome P450 1A1 (CYP1A1) which transforms these xenobiotics into hydrophilic and more easily eliminated compounds. However, although humans are capable of metabolizing and eliminating PAHs through this pathway, the same process can lead to the formation of reactive intermediates that can form DNA adducts^{20,45,124}. If the bulky PAH intermediate binds to critical nucleotides in gene sequences and the DNA damage repair system falters in the removal of the DNA adduct, tumorigenesis can occur^{17,38,44}. Aside from this well-known genotoxic effect, PAH metabolites can also have nongenotoxic effects such as becoming estrogen mimics and affecting the estrogen receptor (ER) pathway^{30,54}.

Current studies have shown that alkylation of PAHs can increase their bioactivity particularly via the AhR pathway^{29,128}. Petrogenic PAH mixtures contain a larger proportion of the alkylated isomers^{15,24}. Although genotoxic effects of these alkylated isomers have not been fully elucidated yet, their presence in crude oil in mixture with high molecular weight PAHs with established carcinogenic effects is concerning.

Despite the number of massive oil spill events in the last 50 years, only a few studies have followed the human health impacts of exposure to crude oil^{5,76}. Biomonitoring of these events has been challenging due to gaps in knowledge on human health risk assessments on petrogenic PAH exposures. Although cell and animal models are valuable in PAH toxicity studies, interspecies variability can introduce issues on translatability⁷⁷⁻⁷⁹. Most of the toxicity values have also been established for exposures to individual and pyrogenic HMW PAH compounds which may not be representative of petrogenic PAH

mixture exposures^{79,82,83}. Our current understanding of PAH toxicity has also been anchored to the 16 priority PAHs listed by the US EPA since the 1970's which only include unsubstituted and mostly HMW PAHs^{80,102}.

Hence, there is a need to develop sensitive and rapid biomonitoring tools tailored to human exposures to mixtures of petrogenic PAHs. Direct identification and measurement of highly bioactive petrogenic PAHs in human plasma samples that can serve as biomarkers of exposure can help fill this gap. These will be valuable in our understanding of the biological ramifications of petrogenic PAH exposures and will provide the needed update to human health risk assessments and guidelines that are vital in our preparation and response to massive petrogenic PAH exposure events such as the DWH oil spill.

Hypothesis

We hypothesize that the petrogenic PAHs detected and quantified by Gas Chromatography – Mass Spectrometry (GC-MS) in the human plasma samples that were longitudinally collected after the DWH oil spill will exhibit distinct AhR activation potencies in the Chemically Activated Luciferase gene eXpression (CALUX) bioassay. Consolidating the AhR bioactivation potencies measured by the CALUX bioassay and concentrations of 36 different parent and alkylated PAHs measured by GC-MS in the human plasma samples will identify potent PAHs that can serve as biomarkers of petrogenic PAH exposures.

Significance

Upon completion of this aim, this dissertation will be the first to identify biomarkers of petrogenic PAH exposure in human plasma samples. These highly bioactive petrogenic PAHs can be used in monitoring human exposure to petrogenic PAHs such as those found in crude oil-contaminated seafood. Results of this study are particularly valuable in the formulation of our preparedness and early response to large population petrogenic PAH exposure events such as oil spills. The longitudinal collection of human plasma samples that come with robust survey and clinical data provide a unique opportunity to follow changes in petrogenic PAH levels and AhR bioactivity as other factors change through time. This can provide human exposure information that are critical in the derivation of risk assessment values that inform guidelines and policies set by regulatory bodies.

Furthermore, the study will provide valuable baseline information for future exposure assessments as the populations in the Gulf of Mexico will always be vulnerable to petrogenic PAH exposure events, particularly as oil exploration and extraction continue in the Gulf, and other associated disasters.

In conducting this study, improvements made to the CALUX bioassay and adapting it for rapid biomonitoring of petrogenic PAH exposures using human plasma samples significantly enhanced our capacity to quickly screen vulnerable individuals. The validation of an optimized rapid high-throughput screening method tailored for crude oil exposures that does not require elaborate sample clean up can hasten our emergency responses in the future. The ability to test petrogenic PAH exposures through biomarkers and bioassays will help allay the fear of deleterious exposures through seafood consumption in our partner communities in the Gulf as they continue to rely on the waters for their livelihood and sustenance.

RESEARCH STRATEGY

Study volunteers and human plasma sample collection

Using a community-based participatory research approach (CBPR), each GC-HARMS partner community (MVC – Mississippi Vietnamese Fishing Community, Gulfport, MS, USA; CEEJ – Center for Environmental and Economic Justice, Biloxi, MS, USA; UHN – United Houma Nation, Houma, LA, USA; and Galveston, TX, USA which served as the comparison site) randomly selected their volunteer cohorts^{5,129}. For each community, the 100 volunteers selected from the community enrolment list with a stratified random sampling consisted of approximately the same number of males and females, with 25 children aged 5 to 19, 50 adults aged 20 to 55, and 25 adults above the age of 55. Licensed medical professionals collected whole blood from each volunteer in EDTA-coated tubes. Plasma samples were aliquoted into four properly labeled 1.5mL screw top cryotubes and stored at -80°C. These collections were performed annually for 3 years from 2013 to 2015 for MVC, CEEJ, and UHN groups, and 2014-2015 for the Galveston group. Each annual sample collection is referred to as a “wave” in this dissertation. All collected samples were properly deidentified and chain of custody transfers were documented as required by the Institutional Review Board (IRB). To comply with the human study IRB requirements, I completed Collaborative Institutional Training Initiative (CITI) courses which I continued to maintain while working on the project. I was added to the IRB (#11-194) after successfully completing the initial required trainings (Appendix O).

Gas Chromatography – Mass Spectrometry (GC-MS)

The group of Dr. Harshica Fernando and Dr. Shakeel Ansari analyzed and quantified 36 different PAHs [naphthalene and alkylated analogues (C1, C2, C3, C4), acenaphthene, acenaphthylene, fluorene and alkylated analogues (C1, C2, C3),

phenanthrene/anthracene and alkylated analogues (C1, C2, C3, C4), fluoranthene/pyrene and alkylated analogues (C1, C2, C3, C4), benz[a]anthracene, chrysene and alkylated analogues (C1, C2, C3, C4), benzo[b]fluoranthene, benzo[k]fluoranthene, benzo[a]pyrene, indeno[1,2,3-c,d]pyrene, dibenz[a,h]anthracene, perylene, and benzo[g,h,i]perylene] in the human plasma samples using Gas Chromatography – Mass Spectrometry (unpublished results). Retention time and selective ion monitoring (SIM) which focuses on a desired range in the spectrum were employed in this study. Specific details of the GC-MS procedure are described in a GC-HARMS manuscript currently in preparation by Dr. Fernando.

Chemically Activated LUciferase gene eXpression (CALUX) Bioassay

MAINTENANCE AND SEEDING OF H1L7.5C3 CELLS

A Hepalcl7 mouse hepatoma cell line stably transfected with pGudLuc7.5 luciferase reporter plasmid (H1L7.5c3, Appendix A) was generously gifted to the laboratory by Dr. Michael Denison (University of California, Davis). It contains 20 XREs in the promoter region of the luciferase gene construct and imparts high sensitivity for the presence of ligands that can induce activation of the AhR pathway (Figure 7)¹⁰⁰. The cells were maintained in 75-cm² flasks with maintenance media: filter-sterilized Minimum Essential Medium Alpha (MEM α , Gibco Cat #12000-063), supplemented with 10% Fetal Bovine Serum (FBS Premium, Atlanta Biologicals, Cat #S11150), and 400mg/L active G418 Sulfate (Gentamicin, Corning Cat #30-234-CR) at 37°C and 5.0% carbon dioxide (CO₂) for 1-2 days until 70-90% confluent. Parallel cell cultures were established when FBS lot numbers were changed. The cells were washed with 5.0mL Dulbecco Phosphate Buffered Saline (D-PBS) without calcium chloride and magnesium chloride (SIGMA, Cat#D-5652), and allowed to completely detach from the flask using 1.0mL Trypsin

without EDTA (Gibco, Cat#T4549) in 37°C for 2-4 minutes. The reaction was stopped by adding 9.0mL of seeding media (MEM α supplemented with 1X Penicillin/Streptomycin (P/S) and 10% FBS). The detached cells were collected into a 10mL tube and centrifuged at 180xG for 3 minutes. The cell pellet was resuspended in 10mL fresh seeding media and was passed through a 23-gauge needle to allow optimal cell dissociation. Cell numbers were quantified using a hemocytometer. The cells were diluted in maintenance media and seeded at 10,000 cells/well onto 96-well white-wall clear flat bottom plates (Corning, Cat #07-200-566) following the plate layout in Appendix B. The cells were then allowed to reach 80% confluence at 37°C, 5.0% CO₂ for 24 hours. The next day, the tray was flicked to remove the seeding media, blotted on clean tissue, and the wells were gently washed twice with 200 μ L of D-PBS per well in preparation for treatment.

TREATMENT WITH DILUTED HUMAN PLASMA SAMPLES

Blank control, BaP standard controls, and plasma samples were diluted as follows: blank control - 5.0 μ L DMSO: 995 μ L MEM α + 1X P/S; benzo[a]pyrene standards (Sigma-Aldrich, CAS #50-32-8) reconstituted in dimethyl sulfoxide (DMSO, Sigma-Aldrich, Cat#D8414) – 5.0 μ L BaP: 995 μ L MEM α with 1X P/S, to 1.0x10⁻¹¹ to 1.0x10⁻⁶ M BaP final concentrations; and plasma samples - 10 μ L plasma: 5 μ L DMSO: 985 μ L MEM α with 1X P/S (1:100 plasma dilution). The pipette tips were pre-wet with the plasma samples to account for viscosity effects prior to dispensing and diluting in the media. The treatments were also performed without FBS to prevent serum components and lot variations from interfering with the assay. New DMSO was also used every 2 months to prevent DMSO photodegradation products from affecting the bioassay.

After D-PBS washing, the cells were treated with 100 μ L of the blank control, BaP standards, or the diluted plasma samples. All exposures were done in triplicate serum-free treatments and incubated at 33°C, 5.0% CO₂ for 24 hours.

LUCIFERASE ASSAY

The Promega Luciferase Assay System (Cat#E4450) was used to measure the relative expression of the induced luciferase enzyme¹³⁰. After incubation with the designated controls and plasma samples, the 96-well plate was carefully flicked, blotted dry, and kept upside down on a clean tissue paper to allow for efficient removal of liquid residues. The treated cells were lysed using 20 μ L 1X Passive Lysis Buffer (PLB, Promega, Cat# E1941) per well, and placed on a plate shaker at room temperature for 20 minutes. The plate was then placed in the GloMax Explorer Plate Reader (Promega, Cat #3500) equipped with dual automated injectors which delivered 50 μ L of Luciferase Assay Reagent (LAR, Promega, Cat#E151C, Cat#E152B, reconstituted according to the manufacturer's recommendation) into each well. To protect the luciferase substrate from photodegradation, the substrate reservoir was fully covered with aluminum foil and the injector lines were encapsulated in a dark cardboard box. Two second reading delay and 10 seconds integration time were employed before the luminescence was read at 550-570 nm. This whole sequence of LAR delivery and luminescence reading was done in succession for all 96 wells. The CALUX assay was performed on plasma samples from all three waves representing 30 volunteers from each of the four partner communities (total of 360 plasma samples). Luciferase activity was allowed to decay on the clean benchtop for an hour before proceeding to the dsDNA quantitation step. A clear lid was placed on the 96-well plates to prevent contamination.

NORMALIZATION TO DOUBLE STRANDED DNA CONCENTRATION

Quant-iT™ Picogreen™ double stranded DNA (dsDNA) fluorescent assay (Invitrogen, Cat#P11496) was adapted for use in a 96-well plate to measure dsDNA concentration in our CALUX bioassay. This is a novel normalization procedure introduced into the work-flow described in this dissertation that allowed dsDNA to serve as a surrogate

for cell number. The assay remains linear from 25pg/mL to 1000ng/mL in the presence of detergents and other preparation reagents making it ideal for use with cell-based assays. Lambda DNA stock provided with the kit (100µg/mL, Component C, Quant-iT™ Picogreen™ kit, Invitrogen) was diluted to working DNA standard solutions with molecular grade, DNase-free 1X TE (diluted from 20X TE Component B, Quant-iT™ dsDNA kit, Invitrogen) and 1X passive lysis buffer (PLB). The volume of 1X PLB added to the working solution was calculated by multiplying 20µL 1X PLB/well with the total number of wells for the given standard plus an extra well equivalent for dead volume. The Lambda DNA stock solution and 1X TE volumes were adjusted accordingly (Appendix C). To each designated DNA standard well that already contained 50µL of luciferase substrate, 50µL of working Lambda DNA standard solutions that would result to 0 ng/mL, 25 ng/mL, 50 ng/mL, 100 ng/mL, 250 ng/mL, 500 ng/mL, 750 ng/mL, and 1000 ng/mL in a final reaction volume of 200 µL per well were added. Thirty microliters (30 µL) were added to the other wells to adjust the total volume to 100 µL. The 96-well plate was returned to the plate reader where 100 µL of diluted Quant-iT™ Picogreen™ reagent was dispensed to all the wells using the automated injector. To protect the Quant-iT™ Picogreen™ from photodegradation, the Picogreen reagent reservoir was also covered with aluminum foil. The plate was briefly visually inspected for bubbles and returned to the plate reader to incubate in the dark at room temperature for a total of 5 minutes, with shaking in the initial 30 seconds. Fluorescence readings were then taken using 475 nm excitation and 500-550 nm emission wavelengths in the GloMax Explorer Plate Reader.

CALCULATIONS AND STATISTICAL ANALYSIS

The dsDNA concentrations were calculated based on the generated linear equation of the Lambda DNA standard curve for each plate. Normalized CALUX bioactivity values were expressed as relative light units (RLUs) per dsDNA concentration for each well

(RLU/ng of dsDNA). To account for media and diluent effects in each plate, the average normalized CALUX bioactivity of the 0.5% DMSO blank control triplicate wells was deducted from the standard and plasma treatment normalized CALUX values. Using GraphPad Prism ver. 9.1 for MacOS, BaP standard calibration curves were generated, and logEC₅₀ value for each plate was determined using three parameters least square best fit constrained at 1.0 for the hill slope, 0 for minimum or bottom value, and 10⁻⁶ M BaP standard normalized CALUX bioactivity (RLU/ng DNA) for maximum or top value. The resulting equations (Eq. 1 and 2) were then used to calculate for benzo[a]pyrene toxic equivalents (ng BaPTEQ/mL of plasma sample) taking into consideration the 100 dilution factor used in the CALUX bioassay.

$$\text{Log [BaP, M]} = \text{logEC}_{50} - \frac{\log \left[\left(\frac{\text{Top} - \text{Bottom}}{y - \text{Bottom}} \right) - 1 \right]}{\text{Hill Slope}} \quad (\text{Eq. 1})$$

$$\frac{\text{ng BaPTEQ}}{\text{mL of plasma}} = 10^{\log[\text{BaP}]} \times \text{BaP MW} \times \frac{1 \times 10^9 \text{ ng/g}}{1 \times 10^3 \text{ mL/L}} \times \text{dilution factor} \quad (\text{Eq. 2})$$

Values are reported as average BaPTEQ ng/mL \pm standard deviation of the treatment triplicates. CALUX runs were repeated as necessary when coefficient of variation (CV) values went above 30%.

Shapiro-Wilke test was used to assess normality of total and individual PAH concentrations, and the CALUX BaPTEQ levels of the plasma samples. Non-parametric Kruskal-Wallis one-way analysis of variance (ANOVA) was used to determine if significant differences in total PAH body burden and CALUX AhR bioactivities exist between the four cohorts (n=90 per cohort). Dunn's multiple comparison post-hoc test was used to determine which location comparisons are statistically significant. Friedman repeated measures ANOVA was employed to detect if there are significant differences between the different annual sampling points (i.e. waves) in each location for both CALUX

bioactivity and total PAH body burden (n=30 plasma samples per wave). Dunn's multiple comparison post-hoc test determined which wave comparisons were significantly different. The analyses were illustrated using Tukey boxplots with the middle horizontal bar indicating the median, and the distance between the 25th lower and 75th upper quartile box boundaries representing the interquartile range (IQR). Lower whisker indicates 25th lower quartile minus 1.5 times IQR, and upper whisker shows 75th upper quartile plus 1.5 times IQR. Outlier values beyond these ranges were plotted as individual data points.

Two-tailed Pearson correlation was used to describe the relationship between the CALUX bioactivities, and the individual PAHs identified and quantified by GC-MS. Since not all the 36 PAHs were reliably detected in all the plasma samples, the number of CALUX-PAH pairs vary for each PAH detected.

OPTIMIZATION OF PLASMA DILUTION

Before the CALUX bioassay could be applied to measure AhR activation, it was important to develop a methodology that would allow the use of plasma samples on the cell-based bioassay. Excess EDTA and presence of clotting factors can impede the use of plasma samples in cell culture^{131,132}. To address this issue, we used diluted plasma samples to treat the CALUX cell line.

A single donor matched set of K2 EDTA-anticoagulated plasma and serum samples was ordered from Innovative Research (Lot#2393789D). Upon receipt, the plasma and serum samples were aliquoted in properly labelled 1.5mL tubes and stored at -80°C.

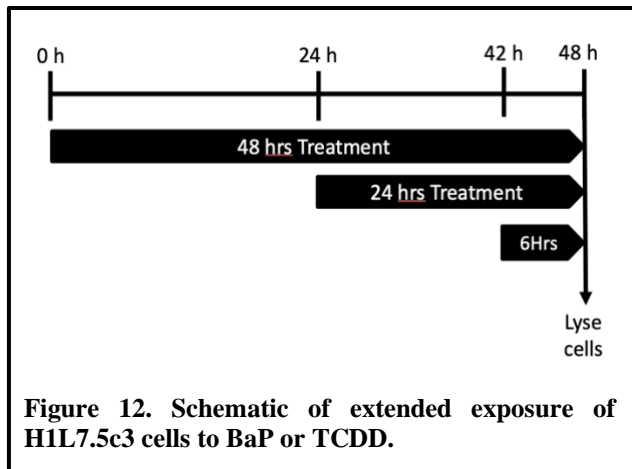
Three different plasma dilutions in treatment media (0.5% DMSO, MEM α with 1X P/S) 1:10, 1:50, and 1:100, with treatment media as control, and corresponding serum sample dilutions were used as treatments and performed in triplicates. Cell seeding and the CALUX bioassay were performed as described in this chapter with 100 μ L as treatment volumes. At this stage we have not yet introduced the use of dsDNA concentration as

normalization procedure. To normalize our CALUX luciferase activity values with a cell number surrogate, the lysis buffer compatible colorimetric Pierce DC Protein Assay (Thermo Scientific, Cat#23246) was performed according to the manufacturer's protocol. Briefly, 10 μ L of lysate was removed from the wells that contain cells lysed with 30 μ L 1X Passive Lysis Buffer and transferred to a 96 well clear absorbance plate in duplicates. To each treatment and standard well, 300 μ L of Pierce Detergent Compatible Bradford Assay Reagent were added and gently mixed five times with a multi-channel pipette. The plate was left in room temperature for 10 minutes to allow color development and the absorbance was read at 600 nm using the GloMax Explorer Plate Reader.

Cell adherence and growth were inspected using an inverted microscope and the presence of clots were gauged by normal visual inspection prior to conducting the luciferase and protein concentration assays. Clotting was also gauged during the initial tray flick to rid the wells of treatments.

OPTIMAL CALUX EXPOSURE DURATION TO DETECT PAHS

The optimal exposure duration that would allow the H1L7.5c3 cell line to produce a robust and quantifiable luciferase signal upon exposure to labile agonists such as PAHs was investigated. The H1L7.5c3 cells were seeded in 96-well plates as previously



described. The cells were treated with 100 μ L of 10^{-8} M, 10^{-7} M, or 10^{-6} M benzo[a]pyrene diluted in MEM α supplemented with 1X P/S, with 0.5% DMSO MEM α supplemented with 1X P/S as blank control, and 6 nM TCDD as positive non-labile control at exposure durations of 6 hours, 24 hours, or 48 hours (Figure 12). Treatments were performed in triplicate and the luciferase assay and dsDNA concentration measurements were performed as previously described. CALUX luciferase induction values were reported as RLU/ng of DNA.

RESULTS

HUMAN PLASMA SAMPLES CAN BE DIRECTLY USED IN THE CALUX BIOASSAY

EDTA-anticoagulated plasma samples pose a difficult challenge when directly used in cell culture-based methodologies such as the CALUX bioassay. Excess EDTA can chelate divalent cations needed for cell-cell and cell to plate surface adhesion^{131,133,134}. On the other hand, depletion of EDTA can cause the activation of clotting factors which may solidify the media and cause cells to be removed from the plate during washes^{131,132}. Having a limited volume of the GC-HARMS plasma samples also discouraged the use of sample cleanup procedures. To address these issues, we set out to determine the best plasma dilution that would prevent clotting and ensure comparable results with a matching serum sample.

Table 1 shows a summary of the CALUX bioactivities of single donor matched plasma and serum samples in different dilutions as normalized to protein concentration. At this point of this study, the novel dsDNA concentration assay was not yet introduced to the methodology. Higher volumes of plasma at 1:20 (50 μ L) and 1:50 (20 μ L plasma) in 0.5% DMSO MEM α + 1X P/S media produced clotting in the wells. These clots, when removed through flicking and blotting the tray upside down, created bald spots where cells detached from the well. Presence of clots also drastically affected the reproducibility of the results

as seen by the high coefficient of variations of the CALUX results for 1:50 and 1:20 plasma dilutions. On the other hand, 1:100 (10 μ L) plasma dilution showed comparable CALUX bioactivity to the corresponding serum sample without the formation of clots. The 1:100 plasma dilution also achieved better %CV (2.5%) than the corresponding serum sample (3.9%). In the diluted serum samples, we observed a decrease in CALUX bioactivity as dilutions decreased. This may be attributed to biofluid matrix effects, a phenomenon also observed in other CALUX bioassay experiments that use serum samples⁹⁸. We did not go lower than the 10 μ L plasma sample volume as viscosity of human plasma samples may introduce pipetting errors that may affect assay reproducibility.

Table 1. Optimized human plasma dilution for the CALUX bioassay.

Sample	1:100 ^a		1:50 ^b		1:20 ^c	
	RLU/ng	%CV	RLU/ng	%CV	RLU/ng	%CV
Plasma	3,867,785.38 ± 95,654.91	2.5%	*4,733,787.80 ± 1,163,667.53	24%	*1,536,798.67 ± 753,057.59	49%
Serum	3,452,722.06 ± 135,114.54	3.9%	3,068,972.00 ± 204,708.18	6.7%	1,766,517.19 ± 30,496.04	1.7%
Difference	415,063.31		1,664,815.79		-229,718.52	

*clotting. ^a 10 μ L plasma: 990 μ L media. ^b 20 μ L plasma: 980 μ L media. ^c 50 μ L plasma: 950 μ L media

ROBUST CALUX INDUCTION IS OBSERVED AT 24-HR BAP EXPOSURE OF H1L7.5C3

CALUX bioassay cell lines have greatly improved in sensitivity and reproducibility^{92,97,100,135}. The H1L7.5c3 cell line used in this study is a third-generation mouse hepatoma cell line that contains a modified luciferase gene construct with 20 XREs¹⁰⁰. Compared to its predecessors, this stably transfected cell line can detect the presence of dioxin-like AhR activators at a minimum detection limit (MDL) of 10 femtomolars, 10 times more sensitive than other cell lines such as the DR-Ecoscreen developed by Takeuchi et al. in 2008¹⁰⁰. It uses a pGudluc7.0 luciferase reporter plasmid with a mouse mammary tumor virus (MMTV) Δ 94 promoter that has no glucocorticoid

responsive elements (GRE), and ligated with 5 dioxin response domains (DRD, contains 4 XREs each) at the BglIII site upstream of the MMTV promoter^{97,100}. PAH bioactivity measurements in previous iterations of the CALUX AhR cell lines required shorter exposure durations of 4 to 6 hours to capture the maximal luciferase signal due to decreased signal strengths at longer time points associated with degradation of the luciferase enzyme^{97,136}. However, newly developed luciferase reporter gene vectors harbor a modified luciferase enzyme that is more stable and is directly targeted to the cytosol instead of localizing to the peroxisome^{137–139}. As the utility of the H1L7.5c3 cell line to detect labile AhR agonists such as PAHs has not been thoroughly tested, we investigated the appropriate exposure duration to use that would result in robust and differentiable luciferase signals between different concentrations of PAHs.

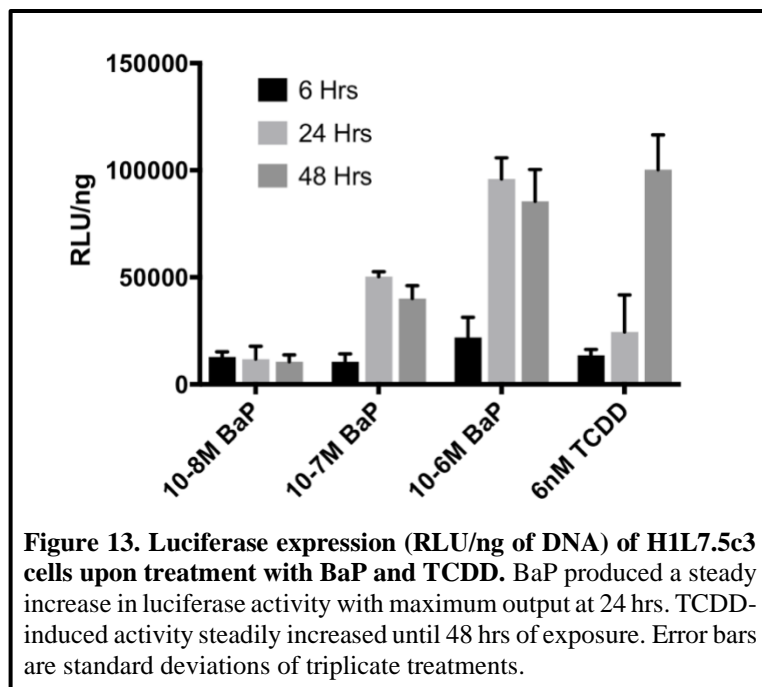
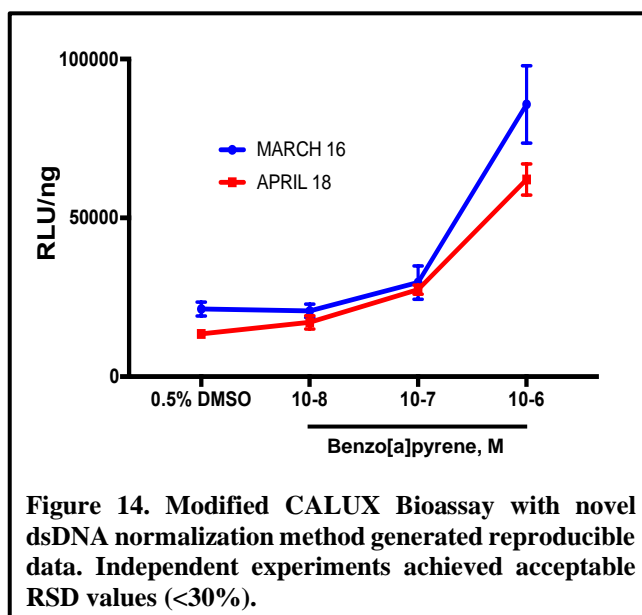


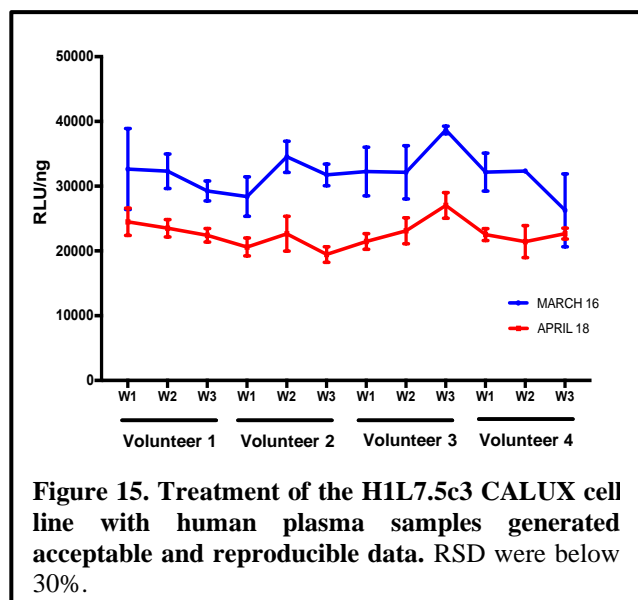
Figure 13 shows the normalized luciferase activity due to the exposure of H1L7.5c3 to 10^{-8} M, 10^{-7} M, 10^{-6} M BaP, and 6 nM TCDD for 6 hrs, 24 hrs, and 48 hrs. Results show that at 6 hours, the luciferase signal was not dramatically induced and differentiable between the different BaP concentrations used. At 24 hours, the H1L7.5c3

cell line produced robust luciferase activity in all BaP concentrations with an evident luciferase signal increase as BaP concentration was increased. At 48 hours, luciferase activity decreased for all the BaP treatments which was expected since PAHs get metabolized and luciferase activity diminishes accordingly. This contrasts with the steady luciferase induction observed in the exposure to TCDD which is a persistent non-labile AhR agonist. Results of this experiment clearly indicated that the 24-hour exposure duration and not the previously recommended 6-hr time point should be used with the H1L7.5c3 cell line in estimating potential toxicity of the GC-HARMS human plasma samples.

dsDNA AS A NOVEL NORMALIZATION FACTOR IN CALUX BIOASSAY

The colorimetric protein concentration assay is currently the method used to provide a normalization factor for the cell-based CALUX bioassay^{98,100,140,141}. However, this method could not be multiplexed with current luciferase reporter gene assays as it requires the transfer of lysates to a different plate for color development and absorbance reading. This transfer step introduces extra manipulations and potential pipetting errors that





risk introducing profound variability into the bioactivity measurements. To address this issue and simultaneously maintain the high throughput ability of the bioassay, we introduced the novel use of the Quant-iT™ Picogreen™ double stranded DNA (dsDNA) fluorescent assay^{142,143}. This method allows sequential measurement of the luciferase reporter system and dsDNA content that can be used as a surrogate for cell number. Figure 14 shows the results generated in two modified CALUX bioassay runs performed on separate days using the blank control (0.5% DMSO with media) and BaP standards as treatment with dsDNA concentration as normalization factor. Relative standard deviations (RSD) of 10-17% for the March 16 run and 7-12% RSD for the April 18 run were obtained. These are well below the 30% RSD required by the European Community guidelines for validation of screening methods^{95,144}. Figure 15 shows the consistency in normalized luciferase activity trends when the method was applied to the GC-HARMS human plasma samples. Acceptable 2-21% RSD for March 16 CALUX run and 4-12% RSD for the April 18 run were obtained. These data indicate that using this novel normalization method ensures reproducible and acceptable CALUX bioactivity measurements.

CORRELATION BETWEEN TOTAL PAH BODY BURDEN AND CALUX AHR BIOACTIVITY

COMPARISON BETWEEN COHORTS

To identify petrogenic PAHs present in the post-spill human plasma samples that are potent activators of the AhR pathway, we first evaluated the total PAH concentrations measured by GC-MS and the CALUX AhR bioactivities in complete sample sets of 30 volunteers (1 sample per wave for each volunteer) from each partner community (n=90 per community).

Table 2 presents the medians, 25th and 75th quartile boundaries depicted in Figure 16 and as well as the lowest and highest data for total PAH body burden and CALUX AhR bioactivity in each location. Kruskal-Wallis one-way analysis of variance (ANOVA) indicated that there are significant differences in total PAH body burdens ($H(3)=68.93$, $P<0.0001$) and the CALUX AhR bioactivities ($H(3)=47.75$, $P<0.0001$) between the four cohorts.

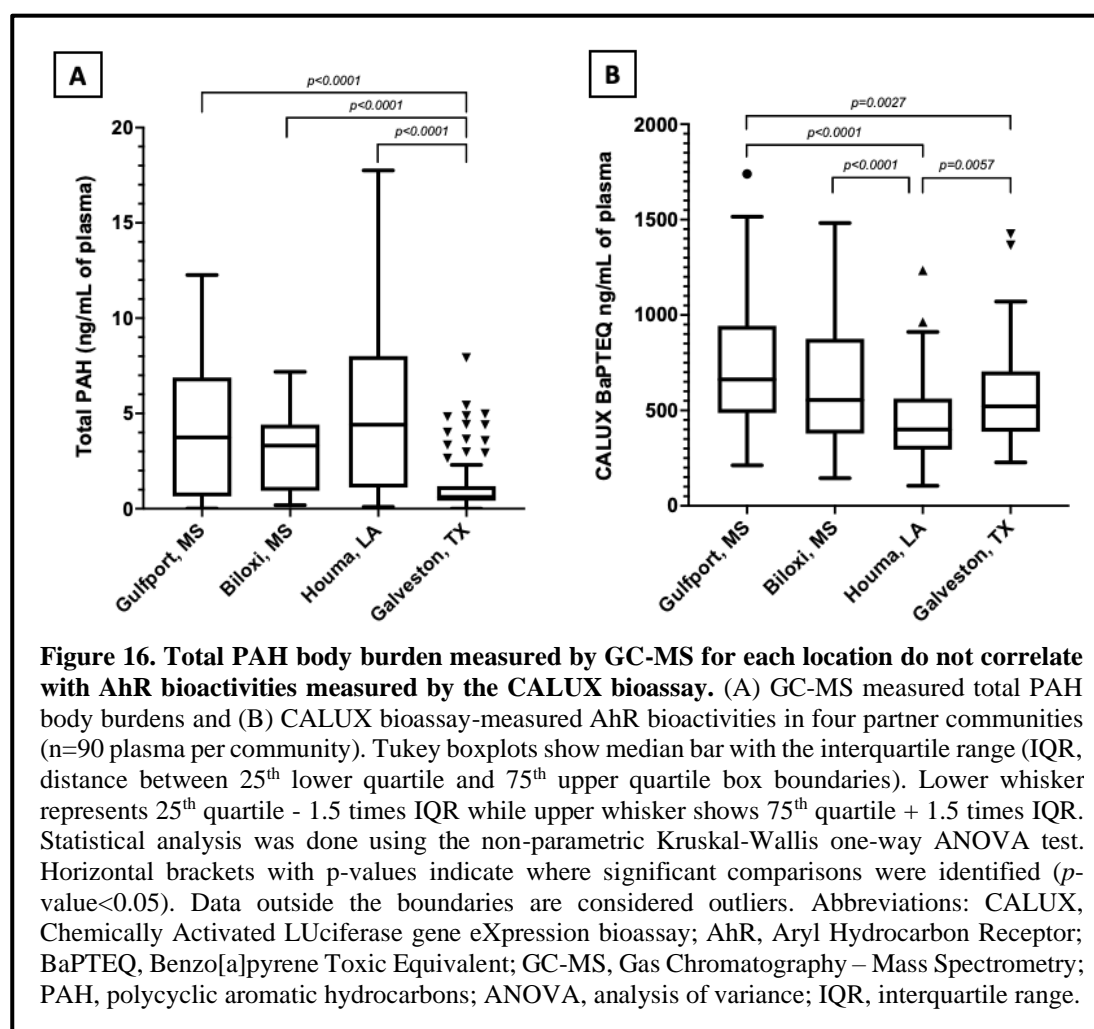
Pairwise comparisons using the Dunn's post-hoc test revealed that Galveston, TX total PAH body burden was highly statistically significant compared to the other cohorts ($p<0.0001$). It also showed the lowest total PAH body burden median (0.620) which indicated that Galveston, TX was indeed a good choice as a control group for total PAH exposure after the DWH oil spill. No other statistically significant comparisons were seen between the other cohorts.

Table 2. GC-MS Total PAH (ng/mL of plasma) and CALUX Bioactivities (BaPTEQ ng/mL of plasma) median in human plasma samples from four partner communities.

Partner Community ^a	Total PAH (ng/mL of plasma)			CALUX Bioactivity (BaPTEQ ng/mL of plasma)		
	Median ^b	Low	High	Median ^b	Low ^c	High ^c
Gulfport, MS (MVC)	3.735 (0.643, 6.878)	0.00127	12.26	662.20 (486.52, 943.75)	211.96 ± 25.06	1739.67 ± 114.53
Biloxi, MS (CEEJ)	3.310 (0.939, 4.415)	0.187	7.18	554.51 (378.39, 875.39)	144.93 ± 19.04	1481.80 ± 115.13
Houma, LA (UHN)	4.405 (1.114, 8.013)	0.0989	17.75	399.96 (294.06, 561.72)	105.13 ± 17.64	1233.58 ± 113.05
Galveston, TX	0.620 (0.423, 1.189)	BDL ^d	7.92	521.66 (387.92, 704.36)	227.21 ± 5.33	1424.83 ± 54.13

^a n=90 per location, 30 volunteers with 1 sample per year for 3 years. ^b Median (25th percentile, 75th percentile). ^c Mean ± SEM, n=3

^d Below detection limit



As seen on Figure 16, the GC-MS measured total PAH body burdens do not follow the same trend as the CALUX AhR bioactivities. Houma, LA cohort has the highest median (4.405 ng/mL of plasma) for total PAH body burden but showed the lowest CALUX AhR bioactivity median (399.96 BaPTEQ ng/mL of plasma). Galveston, TX which showed the

lowest total PAH body burden median gave a slightly higher but statistically significant ($p=0.0057$) AhR bioactivity than the Houma, LA group. Taken together, these results indicated although the PAH body burden in the Houma, LA plasma samples was markedly elevated, these PAHs did not appear to be biologically active in AhR-mediated pathways compared to the PAHs in the Galveston, TX cohort plasma samples.

MVC Gulfport, MS cohort which has the highest AhR bioactivity median (662.20 BaPTEQ ng/mL of plasma) showed statistically significant differences compared to Houma, LA ($p<0.0001$) and Galveston, TX ($p=0.0027$) cohorts. Comparing the MVC Gulfport, MS and the CEEJ Biloxi, MS cohorts that occupy the same geographical region, no statistically significant differences were seen in their total PAH body burdens and AhR bioactivities. However, the medians for both total PAH body burden and AhR bioactivity of Biloxi, MS were lower than the corresponding data on the MVC Gulfport, MS. This suggests that both populations were exposed to similar PAH profiles but at slightly higher concentrations in the MVC Gulfport, MS group, possibly due to seafood consumption differences.

COMPARISON WITHIN EACH COHORT

Next, we investigated the differences in total PAH body burden and CALUX AhR bioactivity trends between each wave or annual sampling period within each cohort. Non-parametric Friedman repeated measures ANOVA test indicated that there are highly statistically significant differences in total PAH body burden between waves in MVC Gulfport, MS ($\chi^2(2) = 24.47, p<0.0001$), Biloxi, MS ($\chi^2(2) = 18.60, p<0.0001$), Houma, LA ($\chi^2(2) = 44.60, p<0.0001$), and Galveston, TX ($\chi^2(2) = 34.07, p<0.0001$). These significant differences were further identified using Dunn's multiple comparison post-hoc test. On the other hand, the Friedman test did not detect significant differences between waves of the CALUX AhR bioactivity data.

As seen on Table 3 and Figure 17A, wave 1 total PAH body burdens in each cohort were considerably lower compared to subsequent waves and a general increase in total PAH body burden was observed in the period following the oil spill in all locations except for a small insignificant decrease in MVC Gulfport, MS wave 3, and a significant dip in wave 3 ($p<0.0001$) in the Galveston, TX cohort. Although MVC Gulfport, MS total PAH body burden slightly dropped in wave 3, it was still statistically significantly higher compared to the first year of sample collection ($p<0.0001$).

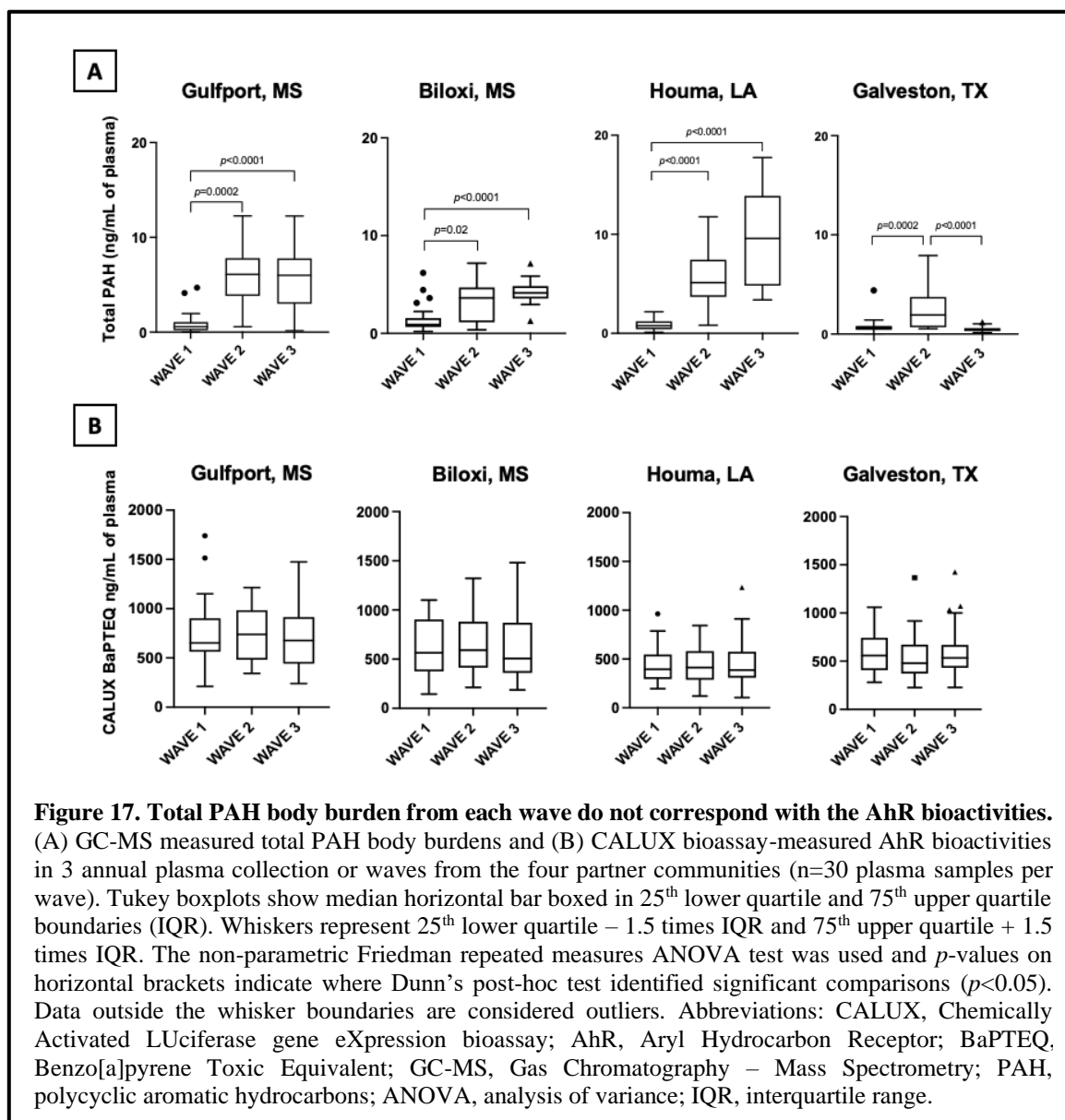
When we examined the data for Galveston, TX, we saw a significant increase in total PAH concentration in wave 2 but the corresponding CALUX AhR bioactivity decreased. This indicated that the detectable increase in PAHs present in the Galveston, TX plasma sample collected for wave 2 did not contribute substantively to the bioactivity of the PAHs detected in wave 1 and wave 3. A similar observation applies to the plasma samples collected in wave 3 from the Houma, LA cohort.

The total PAH body burden for each wave within each cohort does not follow the same trend as their CALUX AhR bioactivities. Again, these results indicate the presence of more potent PAHs that could potentially be developed as key biomarkers of petrogenic PAH exposure, and these need to be identified.

Table 3. Total PAH body burden (ng/mL of plasma) and AhR bioactivities (BaPTEQ ng/mL of plasma) of human plasma samples in every wave (annual collection period).

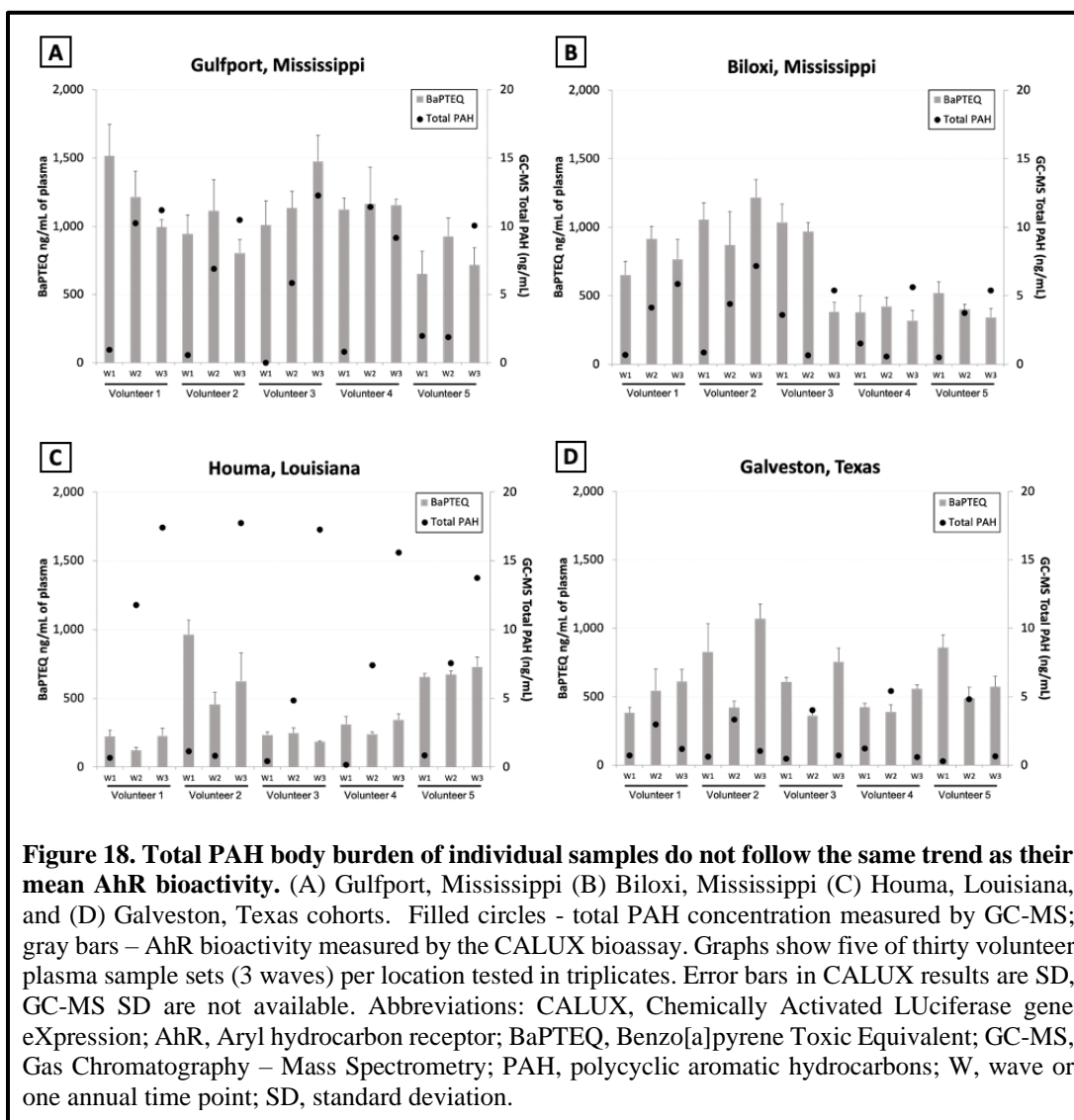
Partner Community ^a	Total PAH (ng/mL of plasma) ^b			CALUX Bioactivity (BaPTEQ ng/mL of plasma) ^b		
	Wave 1	Wave 2	Wave 3	Wave 1	Wave 2	Wave 3
Gulfport, MS (MVC)	0.577 (0.203,1.072)	6.09 (3.823,7.808)	5.995 (2.998,7.780)	652.6 (562.5,902.0)	738.0 (483.2,983.7)	675.9 (441.5,914.8)
Biloxi, MS (CEEJ)	0.899 (0.651,1.560)	3.631 (1.131,4.700)	4.135 (3.555,4.843)	563.9 (375.7,903.9)	591.7 (412.4,881.2)	506.3 (360.6,870.1)
Houma, LA (UHN)	0.772 (0.427,1.216)	5.113 (3.69,7.428)	9.582 (4.808,13.88)	395.2 (294.1,546.4)	413.1 (286.6,580.8)	386.7 (308.6,574.3)
Galveston, TX	0.618 (0.420,0.823)	1.921 (0.677,3.74)	0.417 (0.297,0.598)	558.2 (406.6,743.1)	479.4 (371.9,671.9)	534.6 (431.7,668.5)

^a n=30 per location ^b Median (25th percentile, 75th percentile).



COMPARISON IN INDIVIDUAL PLASMA SAMPLES

To further understand these trend differences observed between and within cohorts, we looked at the total PAH body burden and CALUX AhR bioactivities of individual plasma samples. Figure 18 shows total PAH concentration (filled circles) and CALUX bioactivity (gray bars) for 5 representative complete sample sets—representing 5 individuals—in each location. Graphs of the total PAH concentration and CALUX AhR bioactivities for all plasma samples of the 30 volunteers from each location are found in Appendix D.



As seen on the Figure 18 graphs, total PAH concentration detected in individual samples do not correlate with their CALUX AhR bioactivities. For example, plasma samples collected from volunteer 1 in MVC Gulfport, MS comprising waves 1 and 2 (Figure 18A), revealed that the total PAH burden was very low for wave 1 compared with wave 2, yet exhibited a higher CALUX signal than obtained with the wave 2 sample. These data indicate that certain PAHs present in the wave 1 plasma sample of volunteer 1 in MVC Gulfport, MS were proportionately more bioactive compared to those that contributed to the elevated PAH burden detected in the wave 2 sample.

The UHN Houma, LA graph (Figure 18C) revealed the striking differences between the levels of its total PAH concentrations and the CALUX AhR bioactivities also seen in Figure 16. Total PAH concentrations were markedly elevated, whereas the CALUX AhR bioactivities were relatively low across the Houma, LA volunteers. The generally higher total PAH and CALUX bioactivity of samples from the MVC Gulfport, MS population (Figure 18A) compared to CEEJ Biloxi, MS cohort is also noted (Figure 18B). Overall, these data reinforce our hypothesis that the bioactivity of PAHs present in samples can vary enormously, and their characterization as potential key biomarkers of exposure to crude oil PAHs deserves further investigation.

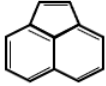
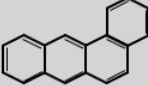
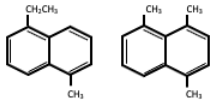
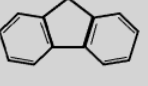
Pearson Correlation reveals significant associations between individual petrogenic PAHs and AhR potencies

To identify highly bioactive petrogenic PAHs that can serve as candidate biomarkers of human exposure to potentially harmful petrogenic PAHs, we examined associations between individual PAH concentrations identified by GC-MS and the BaPTEQ ng/mL of plasma CALUX AhR bioactivity potencies of the plasma samples by using the two-tailed Pearson correlation statistical analysis. We examined the data generated from the 30 complete plasma sample sets from each location (per location = 90 samples, all locations = 360 samples). Table 4 summarizes the correlation results of PAHs

that showed statistically significant associations ($p<0.05$). Associations for all PAHs detected are found in Appendix E. Note that not all PAHs were detected and quantified by GC-MS which resulted in fewer PAH-CALUX correlations for some PAHs.

Acenaphthylene was detected and quantified in the greatest number of samples (118 of 360 samples). Its presence was also significantly associated with an increase in AhR bioactivity in Houma, LA plasma samples ($r(19)=0.44$, $p=0.046$). However, when samples from all locations were analyzed, acenaphthylene showed a very weak positive association and was not statistically significant ($r(116)=0.033$, $p=0.72$).

Table 4. Summary of PAHs with significant two-tailed Pearson correlations between their concentrations and CALUX AhR bioactivities.

PAH	Location	df (n-2) ^a	<i>r</i> , Pearson coefficient ^b	<i>p</i> -value ^c
Acenaphthylene TEF: 0.001 IARC Class 3 CAS# 208-96-8 	Gulfport, MS	26	-0.076	0.70
	Biloxi, MS	38	0.012	0.94
	Houma, LA	19	0.44	*0.046
	Galveston, TX	27	-0.19	0.32
	ALL locations	116	0.033	0.72
Benz[a]anthracene TEF: 0.1 IARC Class 2B CAS# 56-55-3 	Gulfport, MS	18	-0.26	0.27
	Biloxi, MS	6	-0.0021	>0.99
	Houma, LA	16	0.72	***0.0008
	Galveston, TX	17	-0.52	*0.023
	ALL locations	63	-0.076	0.55
C3-Naphthalenes ^d TEF: 0.001 ^d IARC: 2B  (Two of many possible configurations)	Gulfport, MS	16	0.48	*0.045
	Biloxi, MS	1	-0.99	0.070
	Houma, LA	2	0.91	0.095
	Galveston, TX	1	0.42	0.72
	ALL locations	26	0.38	*0.049
Fluorene TEF: 0.001 IARC Class 3 CAS# 86-73-7 	Gulfport, MS	4	-0.69	0.13
	Biloxi, MS	Too few pairs (n=1)	-	-
	Houma, LA	Too few pairs (n=1)	-	-
	Galveston, TX	Too few pairs (n=1)	-	-
	ALL locations	7	-0.76	*0.018

^a degrees of freedom, n-2, n = number of PAH-CALUX pairs ^b Positive *r* indicates positive linear association while negative *r* indicates inverse linear association. ^c *r* is considered significant at * $p<0.05$, ** $p<0.01$, *** $p<0.001$. ^d Data only available for parent compound, naphthalene. TEF, Toxic Equivalency Factor (Nisbet, et al. 1992). IARC Class 2B – possibly carcinogenic to humans, IARC Class 3 – carcinogenicity is unclassifiable for humans. CAS, Chemical Abstract Service registry number.

Benz[a]anthracene, a class 2B possibly carcinogenic compound, also showed a very significant positive linear association with AhR bioactivity in Houma, LA cohort ($r(16)=0.72$, $p=0.0008$). However, its presence was negatively associated with CALUX assay activity in the Galveston, TX cohort ($r(17)= -0.52$, $p=0.023$). Moreover, when samples

from all locations were analyzed, presence of benz[a]anthracene showed a very weak negative association with AhR bioactivity ($r(63) = -0.076$, $p = 0.55$).

C3-naphthalene, a highly alkylated isomer of naphthalene, showed positive linear association in samples from the MVC Gulfport, MS cohort ($r(16) = 0.48$, $p < 0.045$). Although not statistically significant, C3-naphthalene also showed positive associations in other locations except for Biloxi, MS where it exhibited a strong negative association ($r(1) = -0.99$, $p = 0.07$) albeit only in 3 PAH-CALUX pairs. Of the four petrogenic PAHs identified, only C3-naphthalene showed a positive association when samples from all locations were analyzed.

Fluorene was rarely detected by GC-MS in our samples as evidenced by an association in only 9 PAH-CALUX pairs. Of these associations, 6 were detected in MVC Gulfport, MS samples, potentially skewing the data when information from all locations were integrated ($r(7) = -0.76$, $p = 0.018$). Although statistically significant, the few datapoints obtained for fluorene may render the association unreliable.

C3-Naphthalene as a potential biomarker of human exposure to petrogenic PAHs

As mentioned previously, petrogenic PAH mixtures are predominantly composed of low molecular weight 2-4 ring PAHs and their alkylated isomers^{17,23,64}. As seen in Table 4, the four PAHs that produced significant associations fulfill these criteria. Of particular interest is C3-naphthalene, an alkylated analogue of naphthalene. It is only one of the four identified PAHs in this study that is not included in the 16 PAH priority list by the EPA. Naphthalene, which is the parent compound of C3-naphthalene, is considered an IARC Class 2B possible carcinogenic to humans compound. Despite reports of alkylation increasing the bioactivity of parent PAHs, information in the published literature specific for C3-naphthalene's human biological effects is sparse.

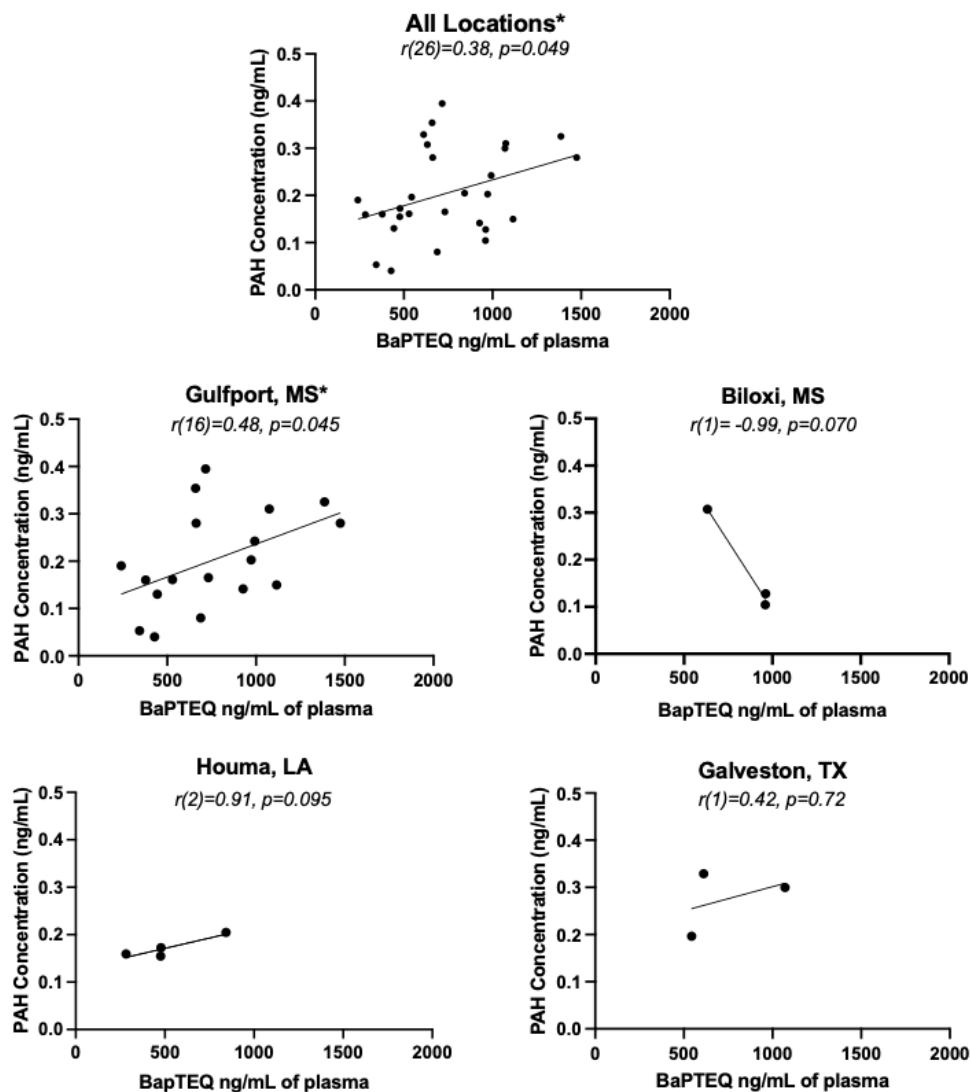


Figure 19. Two-tailed Pearson Correlations of C3-Naphthalene plasma concentration and CALUX AhR bioactivity in all four cohorts. Positive r indicates positive linear association whereas negative r indicates inverse association. Number of datapoints shown reflect the PAH-AhR bioactivity pairs for each location or all locations. Abbreviations: CALUX, Chemically Activated LUCiferase gene eXpression; AhR, Aryl hydrocarbon receptor; BaPTEQ, Benzo[a]pyrene Toxic Equivalent; PAH, polycyclic aromatic hydrocarbons; r , Pearson correlation coefficient, ranges from -1 to 1; (df), degrees of freedom $n-2$. Correlation is significant at $*p<0.05$.

As seen on Figure 19 and listed in Table 4, the presence of C3-naphthalene elicited positive associations with its CALUX AhR bioactivity except in the Biloxi, MS derived samples. Although statistically significant, the CEEJ Biloxi, MS data provided only 3 PAH-CALUX AhR bioactivity pairs, which may not be sufficient to draw meaningful conclusions. UHN Houma, LA and Galveston, TX cohorts also provided only 3 data points

each. However, when we examine the data for all locations, thus encompassing a more general population, we see a modest albeit significant association between the C3-naphthalene concentration and CALUX AhR bioactivity ($r(26)=0.38$, $p=0.049$).

DISCUSSION

This chapter describes the results obtained in the studies associated with the first aim of this dissertation, which was to identify highly bioactive petrogenic PAHs that can be used as biomarkers of human exposure to crude oil, particularly because of the consumption of contaminated seafood. To realize this aim, we first had to modify the recommended methodology to enable assessment of human plasma, a biofluid not normally used in cell-based assays, in the CALUX bioassay. Validation of our optimized methodology imparts confidence in the reproducibility of our results which were correlated with the GC-MS PAH analytical data to identify the highly bioactive candidate biomarker(s) of petrogenic PAH exposure.

CALUX Bioassay Optimization

Optimal Plasma Dilution

Previous studies have shown that serum samples can be used directly on the CALUX bioassay to generate AhR bioactivity data^{98,140,145}. However, only plasma samples were collected in the GC-HARMS project which posed several challenges pertaining to its direct use on the cell-based CALUX bioassay. To reiterate, excess EDTA in the plasma samples can chelate divalent cations such as calcium and magnesium, components in culture media that are essential for stable cell-cell or cell-matrix adhesions^{131,133,134}. Consequently, detachment of cells during culture would undermine the reliability of the CALUX assay. In contrast, complete elimination of the EDTA from the plasma samples,

activated coagulation factors and trigger clot formation^{146,147}. These clots congeal the media and were observed to cause the loss of cells during washing steps. Clotting the plasma samples prior to use in the CALUX assays or using elaborate PAH extraction processes require larger starting volumes than were available. Thus, it was imperative to develop a method that eliminated these confounders without affecting the AhR bioactivity measurements.

Table 1 shows that by using 1:100 plasma to treatment media dilution, the issues can be mitigated. This low plasma concentration resulted in a CALUX AhR bioactivity measurement that was comparable to the data generated by a matching 1:100 serum dilution. The coefficient of variation of the luciferase induction values for the 1:100 plasma dilution remained low at 2.5% which was even better than the matching serum dilution (3.9% CV). No trace of clotting was observed at this dilution but were detected at 1:50 and 1:20 plasma dilutions. Diluting the plasma samples served two distinct endpoints: 1) EDTA concentrations were reduced thus preventing excess titration of the divalent cations needed for cell adhesion, and 2) sufficient dilution of the coagulation factors contained in plasma to negate their ability to form clots¹⁴⁸. The use of the high plasma dilution treatment also diminished sample matrix effects, a common phenomenon in which other components of the biofluid may interfere with the ability of the agonist (i.e. PAHs) to induce luciferase activity. Dilution of biofluids has routinely been used in clinical diagnostics to maintain the linearity of an assay and to eliminate matrix effects. This dissertation, to this author's knowledge, is the first to identify a sample preparation procedure that validated the direct use of small volumes of human plasma samples in the CALUX bioassay, with sufficient sensitivity to analyze low levels of PAHs.

Optimal Exposure Duration

To detect labile AhR agonists such as PAHs, the CALUX bioassay was historically performed using shorter exposure durations of 4 to 6 hours^{97,136}. This allowed detection of maximal luciferase induction prior to luciferase enzyme degradation. Recent modifications in the sequence encoding the luciferase reporter has generated expression vectors with vastly improved stability of the expressed luciferase and its ability to accumulate in the cytosol over a longer period^{137,138}. The H1L7.5c3 cells used in this dissertation harbors the pGudLuc7.0 vector construct that expresses a firefly luciferase enzyme where the peroxisomal targeting tripeptide sequence (Ser-Lys-Leu) at the C-terminus has been substituted with the Ile-Ala-Val to prevent its compartmentalization to peroxisomes^{138,149,150}. This peroxisome localization has been reported to affect normal cellular physiology, sequesters expressed luciferase, and possibly enhances luciferase turnover which may have profound effects on the properties of the CALUX assay^{97,137,139}. As the response of the H1L7.5c3 cell line to PAHs was not previously characterized, we sought to determine the optimal exposure duration that would give a maximum detectable luciferase induction upon exposure to PAHs.

The induction of luciferase signals shown in Figure 13 demonstrates that the H1L7.5c3 cell line could detect AhR activation in response to the prototypical PAH benzo[a]pyrene, and the positive control 6nM TCDD, at 6-, 24-, and 48-hours post-exposure. However, despite showing above baseline activity induction in the reporter assay, the 6-hr exposure duration did not exhibit a reliable dose response at the lower concentrations of BaP. It was not until the cell line was exposed to 10^{-6} M BaP that a robust luciferase induction was generated. In contrast, a 24-hour exposure of the cell line to the different BaP concentrations generated robust signals with a clear dose response in luciferase activity. Our results corroborated findings reported by Promega—a commercial supplier of CALUX assay cell lines—by showing that the engineered protein stability conferred by the modifications on the luciferase enzyme lengthened the time required for

the CALUX cell line to detect distinguishable differences between different BaP concentrations compared to earlier iterations of cell lines developed for the CALUX assay^{139,151}.

From a large-scale rapid PAH biomonitoring point of view, a shorter CALUX bioassay exposure duration would have been preferred to enhance sample processing and hasten the turnaround time. However, the H1L7.5c3 cell line does not give discernable results between the increasing BaP standard concentrations at the shorter exposure duration (6 hrs) making it harder to interpret. However, the increased stability of the expressed luciferase greatly enhanced the sensitivity of the CALUX assay to detect even minute concentrations of bioactive PAHs^{138,139}. Thus, our study using the H1L7.5c3 cell line has characterized and optimized the conditions required to measure the presence of bioactive and potentially harmful PAHs in readily attainable human plasma biospecimens.

dsDNA concentration as Novel Normalization Method for the CALUX bioassay

Protein concentration assays constitute the most common normalization method for luciferase gene expression bioassays. However, this method was not recommended by the developers of the CALUX bioassay as it can introduce substantial variations in luciferase activity¹⁴⁰. Detergents from the lysis buffer may interfere with the protein assay and the transfer of lysates to a different plate for colorimetric protein assay development may introduce technical (i.e., pipetting) errors^{140,152}. Thus, we introduced the use of the fluorescent Quant-iT™ Picogreen™ dsDNA concentration assay as a compatible normalization method that can be multiplexed with the CALUX bioassay^{142,143,152}.

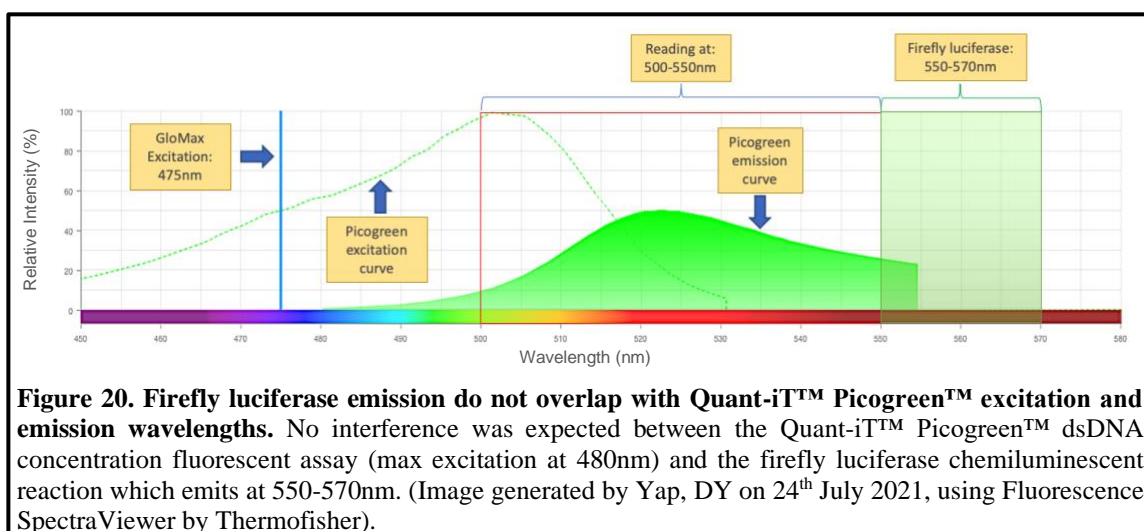


Figure 20 shows the excitation and emission details for the CALUX firefly luciferase and Quant-iT™ Picogreen™ fluorescence assays. Firefly luciferase emits in the 550-570 nm region of the visible spectrum^{130,153,154}. This emission is far from the excitation curve of the Quant-iT™ Picogreen™ which is maximal at 480 nm (green dash) and ends at 530 nm¹⁵⁵. We therefore expected no interference from the luciferase chemiluminescent reaction during dsDNA concentration fluorescent assay reading which was performed using a 475 nm excitation wavelength and emission reading at the 500-550 nm range. To further minimize the possibility of interference, we allowed the luciferase activity to deplete for 60 minutes prior to performing the dsDNA concentration fluorescent assay. This allowed for near complete decay of luciferase signal since the estimated half-life for the assay reaction is 10 minutes¹³⁰. Simulating this scenario using the Fluorescence SpectraViewer of Thermofisher¹⁵⁶, we saw that there is an expected 50% reduction in relative intensity of the Quant-iT™ Picogreen™ emission if 475 nm (blue vertical bar) instead of the 480 nm (peak of green curve) excitation wavelength was used. Nevertheless, we were able to generate acceptable normalized values for our CALUX bioassay (Figures 14 and 15). The low relative standard deviations (RSD) obtained for all samples in two independent modified CALUX bioassay runs that used dsDNA concentration as

normalization factor indicated that our improved methodology could produce reproducible results.

In summary, we were able to identify a suitable plasma sample for analysis (1:100 plasma to media dilution), optimal exposure duration (24 hours), and validated a novel normalization method (Quant-iT™ Picogreen™ dsDNA concentration fluorescence assay) for the CALUX AhR bioassay. This optimization step was critical in generating reliable and reproducible results using multiple different human plasma samples collected and analyzed over an extended period, to identify the presence of bioactive petrogenic PAH that can serve as biomarker of human exposure to crude oil exposure.

C3-Naphthalene as a Potential Biomarker of Human Exposure to Petrogenic PAHs

The current biomarkers of exposure used in biomonitoring PAH exposure are pyrogenic PAHs (e.g. BaP) or metabolites of unsubstituted compounds (e.g. 1-hydroxypyrene)^{60,61,157}. Although valuable in approximating the potential toxicity of a PAH mixture generated through combustion processes, the presence of these biomarkers may not be an appropriate indicator of exposure to petrogenic PAHs found in crude oil which are composed predominantly of low molecular weight (2-4 rings) PAHs including their alkylated isomers^{17,20,26,158}. The US FDA and US EPA responses after the DWH oil spill were heavily criticized because they excluded alkylated isomers in the formulation of human PAH exposure threshold levels^{82,83}. Thus, aim 1 of this dissertation sought to address this deficiency in petrogenic PAH exposure-specific biomarkers by identifying highly bioactive petrogenic PAHs that can serve as potential biomarkers of human exposure to crude oil-contaminated seafood.

To address this aim, we first examined and compared the total PAH body burden and the CALUX AhR bioactivities associated with each cohort (Table 2 and Figure 16). Analyzing 30 randomly selected volunteer plasma sets (representing 3 waves) from each location revealed that there were no obvious associations between the total PAH

concentrations and the CALUX AhR bioactivities. The Houma, LA cohort consistently showed the highest total PAH body burden but produced the lowest CALUX AhR bioactivity measurements. On the other hand, Galveston, TX showed significantly lower total PAH body burdens than detected in the other cohorts yet still elicited readily detectable CALUX bioactivity responses compared to the samples obtained from the Houma, LA cohort. Taken together, these data indicate that the PAH mixtures present in the Houma, LA samples, although more abundant, lack congeners with robust AhR-mediated bioactivity in comparison to the other groups. The practicality of identifying a specific subset of AhR bioactive petrogenic PAHs (i.e. biomarkers of petrogenic PAH exposure) to screen populations that are exposed to potentially more toxic petrogenic PAHs is profoundly highlighted by these data.

Comparing the total PAH body burden and CALUX AhR bioactivities of the two Mississippi cohorts (Gulfport, MS and Biloxi, MS), we found no significant differences between their total PAH body burdens nor their CALUX AhR bioactivities. Our data showed that although coming from the same geographical region, the MVC Gulfport, MS cohort had a larger exposure as seen by the higher total PAH body burden and consequently generated higher CALUX AhR bioactivity. Jackson et. al. in 2019 reported that MVC Gulfport, MS Vietnamese fisherfolks consumed all edible parts of the fish, including the soft organs where PAHs may bioaccumulate. In contrast, the CEEJ Biloxi, MS group represented a largely African American population that generally only consumed the meaty part of fish less likely to be contaminated due to elevated PAH metabolism activity. Hence, seafood consumption preferences may thus serve to explain the differences in their PAH exposures and PAH bioavailabilities.

Next, we examined differences in total PAH body burdens and CALUX AhR bioactivities between waves or annual sample collection of the human plasma samples from the four cohorts (Table 3 and Figure 17). Our data showed that there was a general

increase in total PAH body burden in our cohorts in the years following the DWH oil spill, but the CALUX AhR bioactivity levels remained steady.

The increase in PAH bioavailability as years passed may potentially be due to an increase in consumption of contaminated seafood. Assessment of the data reported by Croisant et al. (2017) reporting on seafood consumption by the cohorts (Figure 11), revealed an improvement in perception of seafood quality and consequently an increase in consumption post-spill. However, the information presented in the paper only described responses collected in wave 1. As more GC-HARMS data are analyzed, it will be very interesting to see if the self-reported seafood consumption in waves 2 and 3 increased, as suggested by the trends seen in total PAH body burdens.

Galveston, TX cohort had a significant increase in total PAH body burden in wave 2 (plasma collected February 2015) which eventually decreased as another year passed. The timing of this increase in total PAH body burden may potentially be attributed to the March 2014 Galveston Bay (GB) heavy fuel oil spill. Although not as massive as the DWH spill, the bunker fuel that was spilled in the area contained 7 times more PAHs than the DWH light fuel oil¹⁵⁹. Interestingly, naphthalene and its alkylated isomers which were rarely detected in wave 1 plasma samples from Galveston, TX showed up in the wave 2 samples (data not shown). This transient increase in detectable petrogenic PAH plasma levels in the Galveston cohort coinciding with this regional spill increases confidence that these analytical approaches can distinguish exposures to petrogenic PAHs from other PAH sources.

MVC Gulfport, MS increases in total PAH concentrations was interesting because their seafood consumption levels post-spill did not change based on self-reporting in wave 1 (Figure 11B). If waves 2 and 3 consumption patterns in the MVC population remain unaltered, the increase in total PAH body burden may reflect consumption of seafood that is persistently exposed to petrogenic PAHs and/or has the ability to bioaccumulate the

PAHs, a change in the frequency and types of seafood consumed (e.g. filter-feeders such as oysters vs. mobile finfish), or other PAH exposure events.

MVC Gulfport, MS also had the lowest total PAH concentration in the wave 1 samples but elicited the highest CALUX bioactivity level for that period (Table 3). In fact, wave 1 data for all locations showed low PAH presence but comparable CALUX AhR bioactivities to waves 2 and 3. These results suggest that wave 1 exposures were likely to be due to more highly bioactive petrogenic PAHs than what were present in waves 2 and 3 and may potentially be enriched for potential biomarkers of petrogenic PAH exposure.

We subsequently analyzed samples at the individual level and obtained further evidence of the presence of more highly bioactive petrogenic PAHs (Figure 18). Wave 1 samples provided by volunteer 1 from MVC Gulfport, MS showed a very low total PAH body burden which generated a robust response in the CALUX assay. Wave 2 from the same volunteer contained substantially more PAHs but generated a lower signal in the CALUX assay. This relationship was seen in multiple plasma samples and differences in total PAH body burden and CALUX AhR bioactivities between each cohort became more obvious in Figure 18.

One strength of the GC-HARMS response to the DWH oil spill was the use of an expanded list of 36 parent PAHs and their alkylated isomers to characterize the extent of exposure. This analysis extended vastly on the EPA's 16 PAH priority list used by other agencies to determine crude oil exposure post DWH oil spill. Looking at the Pearson correlation results between the individual PAH concentrations and the corresponding CALUX AhR bioactivities of the plasma samples in Table 4, we were able to identify four PAHs that showed significant correlations – acenaphthylene, benz[a]anthracene, C3-naphthalene, and fluorene. It is important to note that not all of the 36 PAHs were detected in the human plasma samples which contributed to variations in the number of PAH-CALUX associations.

The association of the four PAHs with CALUX readouts varied greatly amongst the cohorts (Table 4). Benz[a]anthracene exhibited a significantly strong positive total PAH-CALUX association in Houma, LA samples but generated a significant inverse relationship in Galveston, TX cohorts. And although acenaphthylene and benz[a]anthracene showed moderate to strong PAH-CALUX associations in specific locations, they did not demonstrate a consistent association when data from all locations were combined. Since these exposures represent distinct mixtures of PAHs, it is formally possible that interactions between individual congeners in a given sample may lead to synergistic or antagonistic effects that serve to confound risk assessments based on the identification of individual PAH isomers in a mixture. These data also highlight that a biological readout that integrates these interactions, such as the CALUX assay, provides a more accurate assessment of exposures to harmful PAH mixtures. It should be noted that individual variables such as location, occupation, and other life-style characteristics in addition to seafood consumption patterns may influence exposures to bioactive PAHs from various sources. Personal data collected through the GC-HARMS survey instrument may shed valuable insights that will aid in our understanding of these PAH-CALUX associations.

Nevertheless, our data showed that the presence of C3-naphthalene generated positive associations in all the locations except for Biloxi, MS (Figure 19). Although Biloxi, MS demonstrated a very strong negative PAH-CALUX association, this data represented only 3 PAH-CALUX points and was not considered significant. The same can be said about the Houma, LA and Galveston, TX cohorts, where positive associations were based only on 4 and 3 PAH-CALUX pairs, respectively. However, when the data from all locations were analyzed together, detection of C3-naphthalene revealed a significant positive association ($r(26)=0.38$, $p=0.049$) particularly in the MVC Gulfport, MS cohort ($r(16)=0.48$, $p=0.045$). The abundance of C3-naphthalene in the MVC Gulfport, MS cohort

may potentially be due to differences in seafood consumption patterns that increased their exposure to petrogenic PAHs (i.e. eating seafood tissue prone to PAH bioaccumulation⁹).

The significant positive association of C3-naphthalene with AhR bioactivity makes it an attractive petrogenic PAH that should be included in future petrogenic PAH exposure biomonitoring. Of the four PAHs identified in this study, which notably are all petrogenic PAHs (i.e. 2-4 rings, with C3-naphthalene as an alkylated isomer), only C3-naphthalene is excluded from the EPA 16 priority PAH. It is not a surprise then that toxicity information regarding C3-naphthalene is severely limited. And if there are available data, these are mostly derived from studies done on the parent compound, naphthalene. Nevertheless, its presence in plasma samples is an important indicator of AhR activation and potential development of biological toxicity due to petrogenic PAH exposure. Thus, we consider C3-naphthalene as a potential biomarker of human exposure to bioactive petrogenic PAHs. This underscores the importance of updating the current priority PAHs used in human health risk assessments to include petrogenic alkylated PAHs, particularly in crude oil contamination exposures.

Although our study was able to identify C3-naphthalene as a potential biomarker of petrogenic PAH exposure, we could not discern the specific alkylation configuration that elicited this increased CALUX response. Nevertheless, alkyl arrangement is not an important consideration for screening exposed individuals as detecting this category of alkylated naphthalene in the plasma sample would suffice.

Identification of 36 parent and alkylated PAHs is an expanded but not exhaustive list of potentially highly bioactive petrogenic PAHs in the human plasma samples. Nitrated, oxygenated and higher alkylated PAH forms may also be present and can potentially be future considerations as biomarkers of exposure.

Controlled laboratory experiments using *in vitro* or animal models may provide some insights on a few PAH interactions but overall will not be generally applicable to the assessment of PAH toxicity in complex mixtures. They defy practical systematic analyses

in the laboratory setting. Interspecies variations may also confound the translatability of animal studies to human exposures. Thus, the results of our study using a CALUX AhR bioassay optimized for direct assessment of human plasma samples provides valuable data that can be directly applied to the monitoring of human exposures to crude oil.

CHAPTER SUMMARY

In this chapter, we outlined our strategy to identify highly bioactive PAHs in human plasma samples that can be used as biomarkers of exposure to potentially harmful petrogenic PAH mixtures such as those found in crude oil.

First, we optimized the CALUX bioassay to assure reliability of our data. We were able to identify the optimal plasma dilution (1:100 plasma to media) that showed robust luciferase signals without the induction of clotting, identified the optimal exposure duration for the H1L7.5c3 cell line for PAH detection (24 hours), and validated a novel normalization method that can be multiplexed with the CALUX luciferase bioassay (Quant-iT™ Picogreen™ dsDNA concentration fluorescence assay). All these modifications maintained the high throughput capacity of the bioassay to generate robust and reproducible results. This supported the use of the optimized CALUX bioassay to identify the human plasma samples with high CALUX AhR bioactivities.

In response to the DWH oil spill, the GC-HARMS consortium identified and measured the concentrations of 36 parent PAHs and their alkylated isomers from human plasma samples that were collected annually from 2013 to 2015 from four partner communities - MVC Gulfport, MS; CEEJ Biloxi, MS; UHN Houma, LA; with Galveston, TX as reference cohort using GC-MS. We then sought to identify the highly bioactive petrogenic PAHs in the plasma samples that can serve as biomarkers of exposures to petrogenic PAH contamination by correlating the individual PAH concentrations with the CALUX AhR bioactivities measured using our modified CALUX bioassay.

Analysis of the total PAH body burden and CALUX AhR bioactivities indicated the presence of more bioactive PAHs in some cohorts which needed to be identified. Analyzing the longitudinal sampling data also provided valuable insights that may yet benefit from further integration of the GC-HARMS survey data, health and clinical assessments that accompanied each collected plasma sample.

Four petrogenic PAHs showed significant PAH-CALUX bioactivity associations – acenaphthylene, benz[a]anthracene, C3-naphthalene, and fluorene. Of these four, the presence of C3-naphthalene showed a significant positive association with an increase in CALUX AhR bioactivity when data from all cohorts were combined. C3-naphthalene is a petrogenic alkylated PAH that is not included in the prioritized PAH list applied to general human health risk assessments, particularly in petrogenic PAH exposures. Thus, we recommend the inclusion of C3-naphthalene as a biomarker of exposure to petrogenic PAHs in biomonitoring efforts particularly after massive crude oil-exposure events such as oil spills. The results of this study highlight the importance of including alkylated petrogenic PAHs in biomonitoring panels used to screen affected populations after an oil spill and provides impetus to substantially update the EPA 16 priority PAH list that currently only includes parent and mostly pyrogenic PAHs.

CHAPTER 3: PLASMA MICRORNA AS BIOMARKER OF EFFECT

INTRODUCTION

Polycyclic aromatic hydrocarbons (PAHs) are a large group of persistent environmental pollutants that can activate the Aryl Hydrocarbon Receptor (AhR) pathway which mediates their metabolism and elimination^{17,30,64}. However, reactive intermediates can be generated during this process which can lead to genotoxic effects and consequently tumorigenesis and cancer development^{22,43,45}. Our current understanding of this diverse group of xenobiotics has largely been established using toxicological data on high molecular weight pyrogenic PAHs such as benzo[a]pyrene^{80,127,160}. As a consequence, biomonitoring of this diverse group has focused on detecting parent and high molecular weight PAHs and their metabolites^{60,86,157}. Petrogenic PAH mixtures found in crude oil predominantly contain low molecular weight PAHs and their alkylated conjugates^{15,24,26}. Thus, petrogenic PAH mixtures are not widely analyzed and current biomonitoring panels may not be applicable to petrogenic PAH or crude oil exposure biomonitoring.

As exposure to PAHs seldom involves a single compound, biomonitoring for toxicant mixtures is complicated. An emerging concept in the biomarker field is the use of microRNAs (miRNAs) as early biological indicators of exposure (i.e. biomarkers of effect) to complex mixtures such as petrogenic PAHs^{68,104}. MiRNAs are 18-22 nucleotide (nt) non-coding RNA sequences capable of post-transcriptional gene regulation. The 2-7nt seed sequence of these short biomolecules at their 5' end can form perfect RNA duplexes with the 3' untranslated region (3'UTR) of the target messenger RNA (mRNA) to promote transcript destabilization and degradation. Additionally, imperfect complementarity to the mRNA can lead to RNA duplexes that physically block the ribosome and repress protein translation. The miRNAs can affect a myriad of pathways including those involved in disease development^{104,107,109,161}.

Studies on circulating plasma miRNAs have burgeoned in recent years due to the therapeutic potential and ability of these biomolecules to act as early biomarkers of life-threatening and not easily detectable diseases such as cancer^{104,106,109,162}. Labelled as a “liquid biopsy”^{116–118}, miRNAs possess ideal biomarker characteristics – present in readily obtainable samples such as biofluids, levels are proportional to degree of xenobiotic exposure or disease, and are highly sensitive to dysregulation providing nuanced assessments using readily available facile techniques^{58,68}. MiRNAs are also resistant to degradation and are stable in plasma even after multiple freeze-thaw cycles^{163,164}.

The use of miRNAs as biomarkers of effect due to environmental toxicant exposure is supported by various studies. Choudhari et al. (2009) reported that miRNAs have promoter elements that are recognized and regulated by a variety of transcription factors. Xenobiotics that can activate these transcription factors in turn also regulate the expression of the miRNAs. In 2015, Hanieh showed that the miRNA 212/132 cluster in T47D human breast cancer cell line contains xenobiotic response elements (XREs) in its promoter that recruit the activated AhR. Gordon et al. (2015) reported seven *p53*-targetting miRNAs that were upregulated upon exposure to benzo[a]pyrene, which could have implications in multiple myeloma development. As a PAH mixture biomonitoring tool, Deng et al. (2014) explored the use of miRNAs to detect pyrogenic PAH exposure in coal coke oven workers in China. They reported on five miRNAs that correlated with micronuclei frequency in the blood cells of the exposed cohort. However, to our knowledge, there are currently no studies available on the changes in miRNA expression as attributed to petrogenic PAH exposure.

As humans continue to explore for petroleum sources, the probability of exposure to petrogenic PAH mixtures persists. Massive accidental oil spills such as the Deepwater Horizon oil rig explosion as well as innumerable smaller spills contaminate our waters and consequently the marine life that serve as a source of livelihood and sustenance^{1,5,10,165,166}. Unfortunately, currently available biomonitoring tools used on human exposures to crude

oil chemicals were developed without consideration for the mostly alkylated PAHs in petrogenic PAH mixtures^{82,83}. Thus, these tools are deemed inadequate in informing human health risk assessments that are needed in our preparation and response to crude oil exposure events.

The identification of differentially expressed miRNAs that are associated with exposures to crude oil will provide us with a novel minimally invasive biomonitoring panel more suitable for assessments of human exposure to complex petrogenic PAH mixtures. These studies will also provide critical improvements to human health risk assessments that were previously established using animal models and pyrogenic PAH exposure.

Hypothesis

We hypothesize that human plasma samples containing highly bioactive petrogenic PAHs that can activate the AhR pathway will have differentially expressed circulating plasma miRNAs that can serve as novel biomarkers of effect following exposure to petrogenic PAHs. Next Generation Sequencing (NGS) of paired high and low CALUX AhR bioactivity human plasma samples collected after the DWH oil spill will be used to identify these differentially expressed miRNAs, and quantitative Real-Time Polymerase Chain Reaction (qRT-PCR) will verify the utility of these miRNAs as petrogenic PAH exposure signature biomarkers of effect.

Significance

Completion of this aim will generate a novel petrogenic PAH biomonitoring tool developed using human plasma samples from populations that were exposed to crude oil contaminants. This will provide human petrogenic PAH exposure data to inform and update current health risk assessments focused on known and suspected exposures to crude

oil sources. To our knowledge, this study will be the first to demonstrate the utility of circulating plasma miRNAs as biomarkers of effect following human exposure to petrogenic PAHs.

Since miRNAs are involved in the regulation of numerous protein-coding genes, the identification of differentially expressed miRNAs associated to petrogenic PAH exposure may help uncover dysregulated biological pathways that can lead to diseases. Moreover, extensive population data that accompany each plasma sample (i.e. extensive survey data, thorough health assessments, and comprehensive clinical diagnostic information) can be used to further unravel associations between petrogenic PAH exposure and biological pathways leading to disease development.

RESEARCH STRATEGY

Small RNA Extraction

Paired plasma samples from a total of 35 volunteers (MVC Gulfport, MS - 13; CEEJ Biloxi, MS – 8; UHN Houma, LA - 5; and Galveston, TX – 9) which have the greatest differences between their highest and lowest CALUX AhR BaPTEQ ng/mL values in successive waves were selected for total small RNA extraction. The use of plasma paired samples from single individuals eliminated the need to match for several variables such as age and gender.

Total small RNA was extracted from the human plasma samples using manufacture recommended protocol for miRNeasy Serum/Plasma Advanced Kit (Qiagen, Cat # 217204) with several modifications: the plasma samples were thawed for 3 minutes in a 37°C water bath, incubation with 20 µL RNase-free water for elution was increased to 10 minutes, and the subsequent centrifugation step at full speed was set at 2 minutes. The protocol and reagents were also adjusted accordingly for extraction of total RNA from 200 µL of plasma sample. Synthetic *Caenorhabditis elegans* spike-in controls *cel-miR-39-3p*

(Qiagen, MSY0000010, 5'UCACCGGGUGUAAAUCAGCUUG3'), *cel-miR-54-3p* (Qiagen, MSY0000025, 5'UACCCGUAUAUCUUCAUAAUCCGAG3'), and *cel-miR-238-3p* (Qiagen, MSY0000293, 5'UUUGUACUCCGAUGCCAUUCAGA3') were added to the plasma samples after the Buffer RPL lysis step at 1.6×10^8 copies/ μ L working solution. Total RNA was eluted from the column with 20 μ L of the supplied RNase-free water and stored in 10 μ L aliquots at -80°C for Next Generation Sequencing (NGS) and quantitative Real Time – Polymerase Chain Reaction (qRT-PCR).

Plasma Sample Quality Control – Hemolysis QC

As hemolysis can contribute to the miRNA population in the plasma samples^{164,167–169}, a hemolysis quality control qRT-PCR run with *hsa-miR-451a* and *hsa-mir-23a* was performed. The total RNA samples were submitted to the Molecular Genomics Core Facility at the University of Texas Medical Branch (UTMB) in Galveston, TX for qRT-PCR. Applied Biosystems TaqMan MicroRNA Reverse Transcription Kit (Cat. #4366596) was used to generate the cDNA and qRT-PCR was performed using TaqMan small RNA assay *hsa-miR-451a* (Cat. #4427975, Assay ID 001141) and *hsa-miR-23a-3p* (Cat #4427975, Assay ID 000399) with the Fast Advanced Master Mix (Cat #4444558). The hemolysis ratio (ΔCq) described by Shah *et al.* (2016) was the difference between the quantitation cycle (Cq) values of the reference gene *mir-23a-3p* and red blood cell-enriched *mir-451a*.

$$\text{Hemolysis ratio } \Delta\text{Cq} = \text{Cq}_{\text{miR-23a-3p}} - \text{Cq}_{\text{miR-451}} \quad (\text{Eq. 3})$$

Hemolysis risk was considered low for $\Delta\text{Cq} < 5$, moderate for $5 \leq \Delta\text{Cq} \leq 7$, and high for $\Delta\text{Cq} > 7$ ¹⁶⁹. All human plasma samples used in the study passed the hemolysis test

(Appendix F), with a few samples at moderate hemolysis risk. The performance of these samples in subsequent tests was closely monitored.

Discovery Phase: Next Generation Sequencing (NGS)

We used Next Generation Sequencing (NGS) to identify differentially expressed microRNAs between the high AhR bioactivity and low AhR bioactivity samples. NGS is a high throughput sequencing technology that allows the profiling of miRNAs and detection of novel sequences¹⁰⁷. Total RNA from 20 volunteers (paired high and low CALUX bioactivity plasma samples from 5 volunteers in each of the four locations with the largest differences in CALUX levels, 40 total plasma samples) were submitted to the Next Generation Sequencing Core Facility at the University of Texas Medical Branch. The core facility used 8.0 uL total RNA with NEBNext® Small RNA Library Prep Set for Illumina to select for small miRNAs (with 5'phosphate and 3'hydroxyl ends). Briefly, miRNAs in the samples were ligated with 3' and 5' small RNA Illumina adapters, subjected to first strand synthesis to generate the adapter-extended cDNA, and then a series of PCR amplifications with P5 and P7 primers. This generates an Illumina-compatible double stranded DNA that contains P5 and P7 oligonucleotides that anchor the individual strands to complementary sequences on the flow cell, unique indices or barcode sequences at the P7 end that distinguish miRNAs from different sample libraries, and adapter sequences on which the sequencing primers bind. Size exclusion and library clean up assured the selection of small RNA¹⁷⁰.

The generated libraries were sequenced using the Illumina NextSeq 550 platform¹⁷¹ with single-end 75 nucleotide reads, for a total of 400 million reads for 40 samples. Templates were captured on a flow cell using immobilized oligonucleotides complementary to the P5 or P7 adapters. Bridge amplification then created clusters of the same template, and sequencing by synthesis generated the fluorescent signals that are

imaged to generate the sequences. This proprietary technology adds one fluorescent-labelled nucleotide per cycle, images the signal, and cleaves the probe to make way for the next nucleotide¹⁷². The generated sequencing files were analyzed with the help of Dr. Golovko and Dr. Khanipov (UTMB Sealy Center for Structural Biology) using the Small RNA Toolbox of the Qiagen CLC Genomics Workbench ver. 12.0.3 with miRbase ver. 22.1 as annotation database.

Verification Phase: Quantitative Real Time – Polymerase Chain Reaction (qRT-PCR)

To verify the candidate miRNA biomarkers identified by NGS in the Discovery Phase, quantitative Real Time – Polymerase Chain reaction (qRT-PCR) was performed. qRT-PCR provides a rapid and sensitive tool to quantitatively identify sets of miRNA¹⁰⁷. Applied Biosystems TaqMan Advanced miRNA assay kits for *hsa-miR-17-5p* (Cat.#A25576, Assay ID 478447), *hsa-miR-99b-5p* (Cat. #A25576, Assay ID 478343), TaqMan Non-coding RNA Assay for *mir-199a-1* (Cat. #4426961 Assay ID Hs07319370_s1), and a customized TaqMan Gene Expression assay (Cat. # 4331348) for *pre-let-7a-1* were used (Appendix G). TaqMan miRNA assays use a target-specific oligonucleotide with a fluorescent FAMTM on the 5' end, and a non-fluorescent quencher dye bound to a minor groove binder (MGB) on the 3' end. The TaqMan probe sits on the target miRNA in between sequences recognized by the forward and reverse primers. Once amplification occurs, the FAM dye is cleaved by the DNA polymerase and allowed to fluoresce as it separates from the quencher dye. This fluorescence can only occur if the TaqMan probe binds to its target sequence (i.e. mature miRNA), thereby preventing detection of non-targets¹⁷³.

Since the NGS data after CLC Genomics Workbench clean up merged the 5' and 3' data for each miRNA, we selected and ran the 5 prime (5p) arm assays that detect the

mature miRNA from the presumed predominant or more abundant 5' forward arm. However, it is important to note that recent studies have shown that the 3' mature miRNA can also be functional in post-transcriptional regulation^{174,175}.

Total RNA from the 20 volunteer sample sets (low and high CALUX AhR bioactivity) in the Discovery Phase were used together with 15 new volunteer sample sets (MVC Gulfport, MS - 8; CEEJ Biloxi, MS – 3; and Galveston, TX – 4). TaqMan Advanced miRNA cDNA Synthesis Kit (Applied Biosystems, Cat.# A28007) was used to generate ample cDNA for all eight assays from 2.0 uL of total RNA. From this small total RNA volume, this RT-PCR kit can generate cDNA suitable for 600 qRT-PCR reactions¹⁷³. To allow for normalization of all miRNAs (precursor and mature) to the *C. elegans* spike-in controls, a modified protocol by Buchholz (2017) was used. Briefly, the mature miRNA was subjected to 3'poly(A) tailing and 5'adapter ligation steps. Reverse transcription with a universal RT primer converted both mature and precursor miRNAs into cDNA, with random hexamers added to cover the precursor miRNA length. Using a modified reaction with pooled TaqMan assays, precursor miRNA-specific pre-amplification and mature miRNA amplification (miR-amp) were performed with 14 preamplification cycles. RT and cDNA pre-amplification steps were performed using the Bio-Rad T100 Thermal Cycler. The pre-amplified cDNA were stored at -20°C when not used for real-time PCR on the same day¹⁷⁶. Cycling settings for the cDNA preparation and real-time PCR are found in Appendix H.

Pre-amplified cDNA was diluted to 1:10 using molecular grade 1X TE buffer (10mM Tris, 1mM EDTA, pH 8.0; Invitrogen, AM9849) and were used as templates for the subsequent qRT-PCR runs. Since each individual served as his/her own control, cDNA of low and high CALUX bioactivity plasma from the same individual were ran in triplicate in the same plate for all miRNAs of interest (*miR-17-5p*, *miR-99b-5p*, *pre-mir-199a-1*, *pre-let-7a-1*, *cel-39-3p*, *cel-54-3p*, *cel-238-3p*). Real-time PCR was performed using the CFX96 Touch Real Time PCR Detection System (Bio-Rad). RNase-free water (No DEPC

treatment, Invitrogen, AM9938) was used as a no-template control (NTC) and was run in duplicate for each assay in all plates. To analyze the data, CFX Maestro Software for Mac 1.0 ver.4.02325.0418 (BioRad, 2017) was used¹⁷⁷.

Quality of the qRT-PCR data was assessed for every run. Amplification plots were visually checked particularly for elevated background noise. Baseline and threshold values were adjusted as needed. For the TaqMan assay run to be acceptable, the NTC should not yield amplification products, quantitation cycle (Cq) for all other samples should be under 37 cycles, and standard deviation between replicates should be below 0.5^{164,178}. Assays were repeated when QC standards were not met. Three volunteer sample sets were removed from analysis due to failed *cel-miR-39*, *cel-miR-238* and *pre-199-a1* assays.

The Cq values of the candidate microRNA biomarkers were normalized to the geometric means of the *cel-miR-39-3p*, *cel-miR-54-3p*, and *cel-miR-238-3p* spike-in controls. Calculation for relative quantification was performed based on the method described by Taylor et al. (2018)¹⁷⁹. Relative fold changes for each candidate miRNA are presented relative to the normalized expression of the candidate miRNA in low CALUX bioactivity plasma samples.

Verification of *miR-199a-1* and *let-7a-1* precursors

Although transcribed from different locations in the genome, *let-7a-1*, *let-7a-2*, and *let-7a-3* share the same mature miRNA sequence (5'UGAGGUAGUAGGUUGUAUAGUU3')¹⁸⁰. Likewise, for *miR-199a-1* and *miR-199a-2*. (5'CCCAGUGUUCAGACUACCUGUUC3'). Thus, the only way to verify *hsa-miR-199a-1* and *let-7a-1* as signature biomarkers of effect for petrogenic PAH exposure was to detect the presence of the precursor miRNAs *pre-miR-199a-1* and *pre-let-7a-1*. To do this, the TaqMan Non-coding RNA Assay for *mir-199a-1* (Applied Biosystems, Cat. # 4426961 Assay ID Hs07319370_s1), and a custom ordered TaqMan Gene Expression assay

(Applied Biosystems, Custom Cat. # 4331348) for *pre-let-7a-1* were used. The *pre-let-7a-1* primer and probe sequences were adapted from Schmittgen et al. (2005). Details of the custom-ordered *pre-let-7a-1* TaqMan assay are found in Appendix G. RT and qRT-PCR were performed as previously mentioned using the CFX96 Touch Real Time PCR Detection System. We previously attempted to identify endogenous controls that would be appropriate to use with precursor miRNAs using an 8-gene customized TaqMan Array plate (Cat. # 4413263). No reliable endogenous mRNAs were detected, which may potentially be due to the instability of mRNAs in long term storage compared to miRNAs.

Target Gene Prediction and Functional Enrichment Analysis

To identify pathways that may involve the candidate biomarker of effect identified in our discovery and verification strategies, we performed target prediction, gene ontology functional analysis, and disease prediction.

For the focused analyses, we used miRNet 2.0 (www.mirnet.ca)^{181–184} with miRTarbase v8.0 database¹⁸⁵ to predict the target genes of our candidate biomarker of effect. The terms used for the query were: Organism - *H. sapiens*; ID type - miRbase ID; Targets – Genes (miRTarBase v8.0), Diseases. Disease association using HMDD ver3.2¹⁸⁶, miR2Disease¹⁸⁷, and PhenomiR 2.0¹⁸⁸ databases and gene ontology (GO) enrichment analysis for biological processes were also analyzed using the same platform with significance set at $p < 0.001$. The GO ID terms were cross referenced using the EMBL-EBI Quick Go website (<https://www.ebi.ac.uk/QuickGO>)¹⁸⁹.

To investigate if there are more miRNAs that may be potentially dysregulated following exposure to highly bioactive petrogenic PAHs, a second analysis of the NGS data was conducted with the help of Drs. Golovko and Khanipov. Briefly, NGS files retrieved from the UTMB NGS Core Facility were analyzed using the CLC Genomics Workbench Small RNA toolbox¹⁹⁰ and full sequences were compared using NCBI-

BLAST¹⁹¹ to search for miRNA homologs. The combined miRNA reports were then analyzed using edgeR on the open sourced web-based platform Galaxy (www.usegalaxy.org)¹⁹². The resulting list of differentially expressed miRNAs ($p < 0.05$) were run on miRNA Enrichment Analysis and Annotation tool (miEAA)¹⁹³ to predict enriched biological functions. Ingenuity Path Analysis¹⁹⁴ was performed to predict top biological pathway hits (Appendix M) and create an interaction network (Appendix N).

Statistical Analyses

For the discovery phase, we performed a power analysis using an algorithm developed by Hart *et al.* (2013) for RNA sequencing studies and determined that a minimum of 14 samples per group were required to detect an at least 1.75 fold change with 80% statistical power, 0.05 level of significance (two-sided), and 0.5 coefficient of variation. Two-sided t-test with significance level set at $p < 0.05$ was performed on normalized microRNA expressions (scaling method) between low and high CALUX bioassay plasma samples using the Qiagen CLC Genomics Workbench ver. 12.0.3 with sequencing alignment set to miRbase ver. 22.1.

In the verification phase, two-tailed paired t-test was performed on the log2 transformed average normalized expressions of the candidate microRNAs. All statistical analyses in the study aside from the discovery phase t-test were performed using GraphPad Prism ver. 9.1 with statistical significance set at $p < 0.05$.

RESULTS

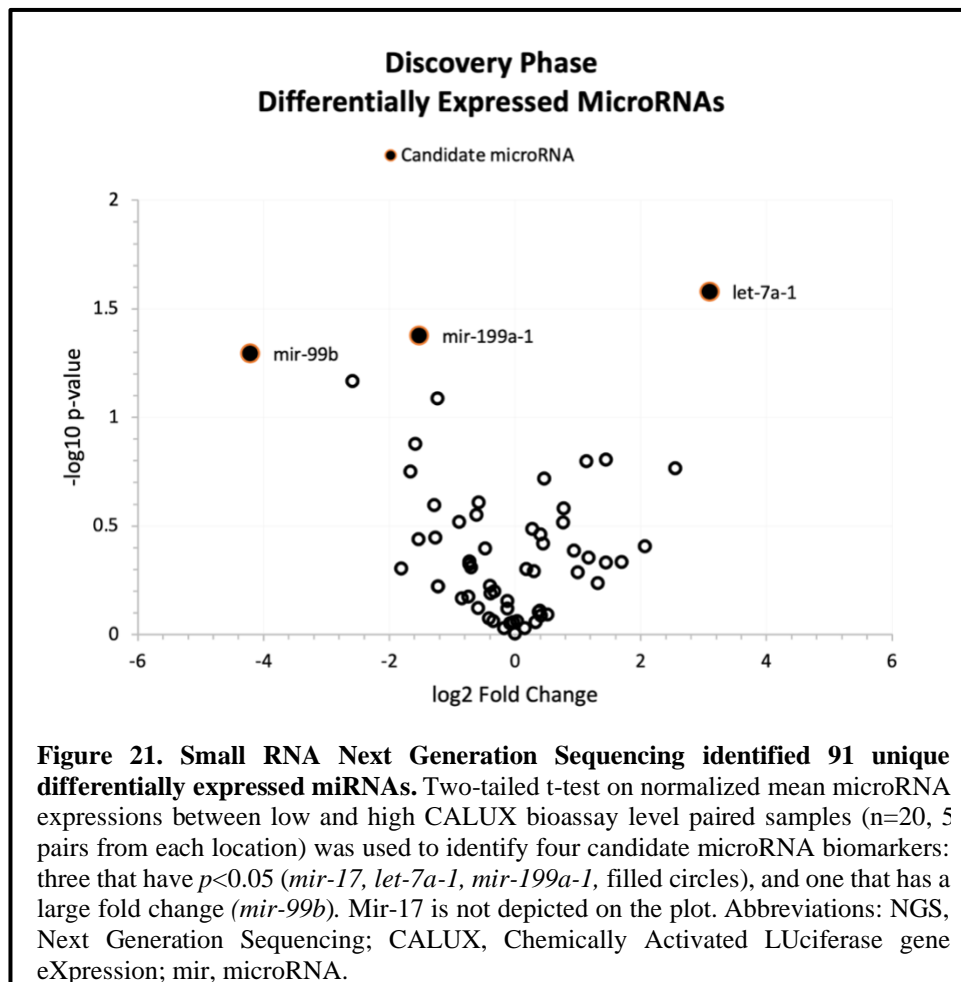
Phase I: Discovery with Small RNA Next Generation Sequencing

In the discovery phase, QIAGEN CLC Genomics Workbench 12.0.3 and miRbase v22.1 identified 196 microRNAs from the small RNA NGS data of paired low and high CALUX bioassay level plasma samples (n=20) that successfully mapped to known miRNAs in the database. Normalization using scaling average¹⁹⁵ revealed 91 differentially expressed miRNAs with 41 showing increased expression (Appendix I) and 50 showing decreased expression (Appendix J) between low and high CALUX bioactivity samples (Figure 21). From these, four candidate microRNA biomarkers of effect for PAH exposure were identified: *mir-17* ($p=0.040$), *mir-99b* ($p=0.051$), *mir-199a-1* ($p=0.042$) and *let-7a-1* ($p=0.026$). Table 5 shows a summary of the two-tailed t-test (significant at $p<0.05$) comparison of the microRNA expressions between low and high CALUX bioassay samples. Although the change in expression of *mir-99b* between low and high CALUX bioassay samples did not quite reach statistical significance, its large -18.5-fold decline made it an attractive candidate biomarker. On the other hand, *mir-17* yielded a significant fold change ($p=0.040$) but has an undefined value due to its undetectable (zero normalized) expression in high CALUX samples. Thus, *mir-17* is included on Table 5 but not depicted on Figure 21.

Table 5. Discovery Phase: Next Generation Sequencing (NGS).

miRNA ^a	Low CALUX ^b	High CALUX ^b	Fold Change ^c	p-value ^d
let-7a-1	0.045	0.385	8.556	0.026*
mir-199a-1	0.76	0.265	-2.868	0.042*
mir-17	0.185	0	-	0.040*
mir-99b	0.185	0.01	-18.5	0.051

^a n=20 paired low and high CALUX plasma samples ^b Normalized mean expression ^c Negative value indicates lower expression in high CALUX group, ^d Two-tailed t-test, significant at * $p<0.05$



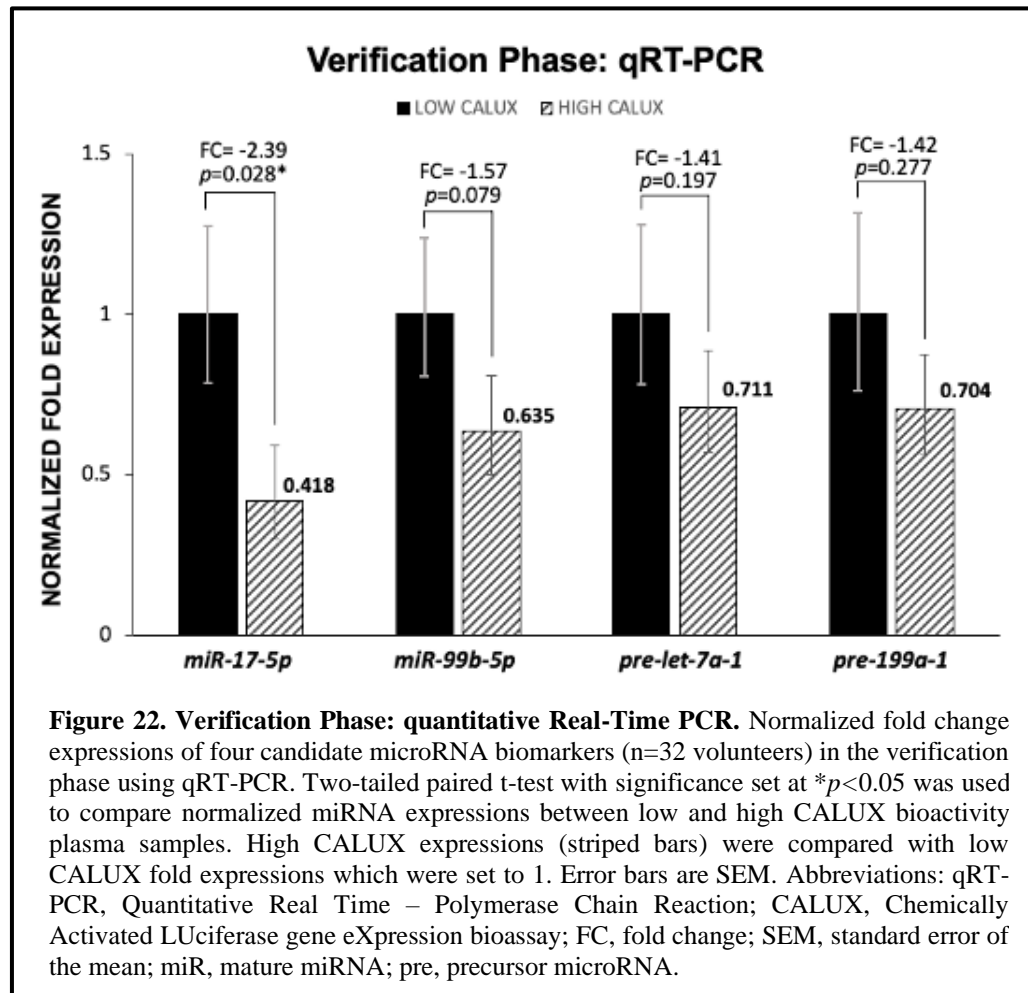
Phase II: Verification with qRT – PCR

Verification phase results of the four candidate microRNAs using qRT-PCR are shown in Figure 22. Verification Phase: quantitative Real-Time PCR.. Based on the NGS data, we should expect a decrease in expression for *hsa-miR-17-5p*, *hsa-miR-99b-5p*, and *hsa-miR-199a-1*, and an increase for *hsa-miR-let-7a-1* in high CALUX bioactivity samples relative to the levels detected in the low CALUX assay samples collected in a different wave isolated from single individuals.

Currently available TaqMan miRNA assays for *hsa-miR-199a-1* and *hsa-miR-let-7a-1* target mature miRNA sequences. Thus, other isoforms that share the same short

sequences (*hsa-miR-199b*; *let-7a-2*, *let-7a-3*, respectively) are also detected. One strategy to verify these two candidate biomarkers was to identify their respective precursor miRNAs (pre-miRNAs). Detecting precursor miRNAs may also be useful as biomarkers as these can be found in plasma. Custom ordered *pre-let-7a-1* TaqMan assays (Appendix G) and TaqMan Non-coding RNA Assay for *mir-199a-1* were used accordingly.

As shown in Figure 22, normalized fold expression between the paired low and high CALUX level plasma samples (n = 32 paired samples) verified *hsa-miR-17-5p* as significantly differentially expressed, with a fold change of - 2.39 ($p=0.028$). Fold changes in the other three candidate miRNA biomarkers did not reach significance threshold level. Moreover, qRT-PCR detected decreased *pre-let-7a-1* expression in the high CALUX AhR bioactivity group, which contradicts its increased expression observed in the NGS data.



Target Gene Prediction, Gene Ontology Functional Analyses

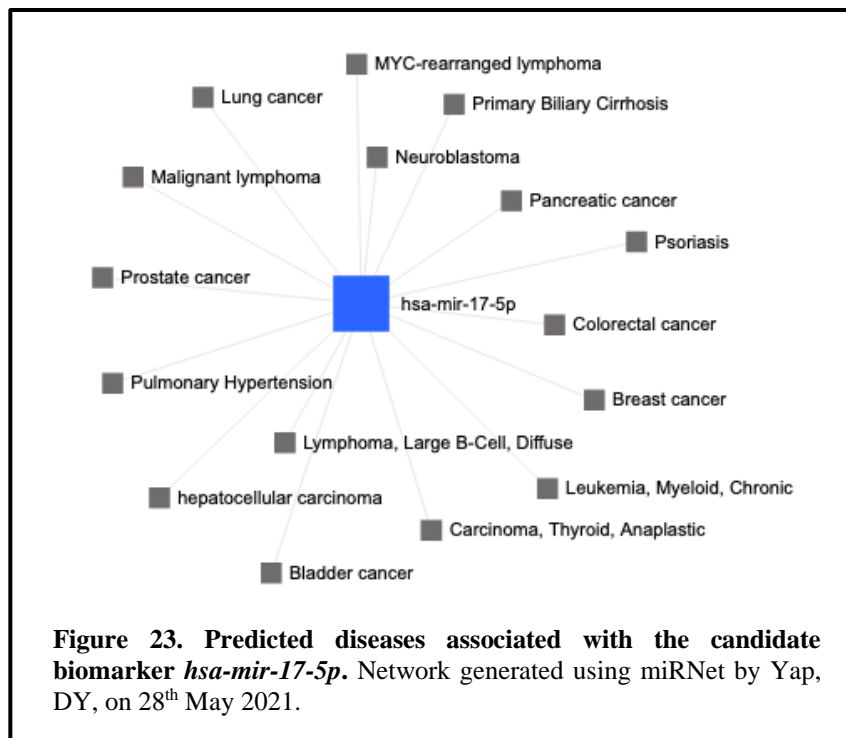
There were 1181 predicted target genes for the candidate human biomarker *hsa-miR-17-5p* using miRNET 2.0 (Appendix K). Gene ontology enrichment analysis shown in Table 6 revealed that these genes are involved in at least 30 biological processes ($p < 0.001$). Most of these biological functions were involved in cell cycle regulation, transcription regulation, signaling pathways involving growth factors, and epigenetic modifications.

Table 6. Biological pathways that involve *hsa-miR-17-5p* putative target genes.

GO ID	Biological Function	p-value
GO:0000080	G1 phase of mitotic cell cycle	9.0E-06
GO:0006417	regulation of translation	1.0E-05
GO:0051318	G1 phase	1.5E-05
GO:0045787	positive regulation of cell cycle	1.9E-05
GO:0016570	histone modification	6.6E-05
GO:0044265	cellular macromolecule catabolic process	8.5E-05
GO:0051726	regulation of cell cycle	8.6E-05
GO:0046777	protein autophosphorylation	9.0E-05
GO:0001666	response to hypoxia	1.1E-04
GO:0016569	covalent chromatin modification	1.2E-04
GO:0006606	protein import into nucleus	1.5E-04
GO:0051329	interphase of mitotic cell cycle	1.5E-04
GO:0060633	negative regulation of transcription from RNA polymerase II promoter	1.6E-04
GO:0051170	nuclear import	2.0E-04
GO:0051325	interphase	2.3E-04
GO:0009628	response to abiotic stimulus	2.5E-04
GO:0006352	DNA-dependent transcription, initiation	2.7E-04
GO:0017015	regulation of transforming growth factor beta receptor signaling pathway	2.7E-04
GO:0045727	positive regulation of translation	2.9E-04
GO:0016569	chromatin modification	4.0E-04
GO:0016567	protein ubiquitination	4.5E-04
GO:0044257	cellular protein catabolic process	5.3E-04
GO:0032268	regulation of cellular protein metabolic process	5.8E-04
GO:0040007	growth	6.2E-04
GO:0032269	negative regulation of cellular protein metabolic process	6.4E-04
GO:0097190	apoptotic signaling pathway	7.5E-04
GO:0009416	response to light stimulus	7.8E-04
GO:0006367	transcription initiation from RNA polymerase II promoter	8.2E-04
GO:0007173	epidermal growth factor receptor signaling pathway	8.4E-04
GO:0043280	positive regulation of cysteine-type endopeptidase activity involved in apoptotic process	9.1E-04

Using miRNet^{181–184} and the DisGeNET^{196,197} gene-disease association databases, an array of diseases such as psoriasis, pulmonary hypertension, hepatocellular carcinoma,

breast cancer, and leukemia ($p<0.05$) were identified in which the *hsa-miR-17-5p* putative target genes participate in (Figure 23). Note that disease nodes were placed arbitrarily around the central node and the distance does not indicate strength of association.



Supplemental Bioinformatics Analysis

A second analysis of the NGS data revealed 30 miRNAs which were significantly differentially expressed ($FDR<0.05$) out of 236 unique differentially expressed miRNAs (Appendix L). When these miRNAs were analyzed on the miEAA web-based application, two pathways were enriched – GO0017080 sodium channel regulator activity, and GO0005215 transporter activity ($p=0.048$). Ingenuity Path Analysis (IPA) findings presented in Appendix M show the top diseases associated to the miRNA set are involved in organismal injury and abnormalities, reproductive system disease, cancer, connective

tissue disorders, and gastrointestinal diseases. The miRNA sets were also found to be involved in organ toxicity such as in the liver, kidney, and heart.

The differential expression of *mir-17-5p* was re-analyzed using this same pipeline. It showed *mir-17-5p* decreased in expression with a fold change of 1.2 ($p=0.24$) from low to high CALUX AhR bioactivity plasma samples.

DISCUSSION

The use of plasma microRNAs (miRNAs) as indicators of biological changes due to environmental toxicant exposure (i.e. biomarkers of effect), particularly to complex PAH mixtures, is an emerging concept in biomarker development. Several studies have shown that miRNA expression can be regulated by the activation of transcription factors such as the Aryl Hydrocarbon Receptor (AhR), the primary biosensor for PAHs^{198–200}. Deng et. al. (2014) identified 5 differentially expressed plasma miRNAs that were associated with micronuclei frequency following pyrogenic PAH exposures in coal coke oven workers. To our knowledge, no studies have reported on the use of plasma microRNAs as biomarkers of effect for petrogenic PAH exposure. In this chapter, we described the discovery and verification of differentially expressed miRNAs found in human plasma samples that have highly bioactive petrogenic PAHs as candidate biomarkers of effect for petrogenic PAH exposure.

First, we used small RNA Next Generation Sequencing (NGS) to identify differentially expressed miRNAs in 20 volunteer sets of paired high and low AhR bioactivity human plasma samples as identified in chapter 2. NGS is a high throughput sequencing methodology that allows hypothesis-free profiling of miRNAs including novel sequences^{172,201}. Using the CLC Genomics Workbench version 12.0.3, with miRbase v22.1 as annotation database, the sequences were trimmed to 15-32 nucleotides, species set to *Homo sapiens*, and the default values of 2 additional or missing bases at both 5' or 3' ends

were used to allow alignment of mature miRNA variants called isomiRs. These variants were created due to alternative cleavage by the Dicer endoribonuclease during precursor miRNA processing to mature miRNAs or post-transcriptional addition of uridines or adenosines^{107,202}. Although originally thought as degradation products, isomiRs have been found to be stable and functional in gene expression regulation²⁰³.

The CLC Genomics Workbench (Qiagen) workflow generated 196 unique miRNAs that mapped to the precursor miRNAs in miRbase v22.1. Of these, 91 miRNAs were differentially expressed, with 50 miRNAs exhibiting decreased expression and 41 miRNAs being upregulated. We identified four candidate miRNAs: *mir-17* ($p=0.040$), *mir-99b* ($p=0.051$), *mir-199a-1* ($p=0.042$) and *let-7a-1* ($p=0.026$) (Table 5) based on their significance levels and meeting the cut-off threshold of 1.75 fold change (calculated during power analysis). *Mir-99b* was included despite its p -value of 0.051 as it produced a substantial fold change of -18.5. On the other hand, *mir-17* expression was significant but did not register a fold change due to its normalized expression value of zero in the high CALUX AhR bioactivity sample. Gene ontology to further narrow down the candidate biomarker pool was not employed since there were only 4 miRNAs that needed to be verified.

Verification of the NGS candidate biomarkers was performed using TaqMan-based qRT-PCR assays. We initially conducted the verification of the candidate biomarkers with TaqMan miRNA assays that used the stem-loop primer technology to identify and quantify mature miRNAs²⁰⁴. However, mature miRNAs that come from different chromosomes may share the same mature miRNA sequences, such as in the case of *mir-199a-1* and *mirR-199a2*, and *let-7a-1* with *let-7a-2* and *let-7a-3*^{180,205,206}. Thus, one way to verify the expression of *mir-199a-1* and *let-7a-1* is by targeting the precursor miRNAs (pre-miRNAs) using TaqMan gene expression assays. This posed a unique challenge, and our attempt to identify reliable short endogenous mRNA controls for the pre-miRNAs proved unsuccessful possibly due to degradation of mRNA in plasma samples. To facilitate the

use of the exogenous spike-in controls as normalization factors for both mature miRNA and pre-miRNAs, we used a modified protocol by Buccholz (2017) using the TaqMan Advanced miRNA assay for mature miRNA and TaqMan non-coding and gene expression assays to run all the assays on the same qRT-PCR plate. The TaqMan Advanced miRNA assays use 5' end adapter ligation, 3'poly(A) tailing, and universal primers to lengthen the short miRNAs, and TaqMan chemistry (previously described in the methods) to quantify the expression of these miRNAs.

Of the four candidate miRNAs, we were able to verify the decrease in expression seen in the NGS data for *hsa-miR-17-5p* when comparing high CALUX bioactivity plasma samples with the low CALUX bioactivity cohort. Figure 22 shows the expression of the four candidate microRNAs in high CALUX bioactivity plasma samples relative to expression in the low CALUX samples with paired t-test performed on the log2 transformed normalized expression. *Hsa-miR-99b-5p* also decreased in expression with a *p*-value of 0.079. A statistical significance in the relative expression levels for this miRNA may be revealed if the study was conducted on a larger sample set. *Pre-199a-1* expression detected by qRT-PCR decreased reflecting the observation in the NGS profile, but the change was not significant.

Interestingly, qRT-PCR detected a decrease in *pre-let-7a-1* expression which is contrary to its NGS profile. This discordance in expression pattern may be due to the presence of isomiRs or mature miRNA variants of *let-7a-1* that were pooled during analysis of the NGS expression profiles for the sample groups, which served to mask the decrease in *pre-let-7a-1* expression detected specifically by the qRT-PCR assay. A recent paper by Karlsen et al. (2019) that profiled canonical mature miRNA sequences and their isomiRs in human plasma exosomes from 46 individuals showed that there are 95 different isomiRs for *hsa-let-7a-5p*, 3 isomiRs for *hsa-miR-17-5p*, 4 isomiRs for *hsa-miR-99b-5p*, 1 isomiR for *hsa-199a-1-5p*, and 13 isomiRs for *hsa-miR-199a-1-3p*. In their study, *hsa-let-7a-5p* isomiR1 which lacks 2 bases at the 3' end (5'UGAGGUAGUAGGUUGUAUAG3')

compared to the canonical *hsa-let-7a-5p* (5'UGAGGUAGUAGGUUGUAUAGUUU3') was expressed in plasma by at least 52% of the canonical miRNA expression. IsomiR2 (5'UGAGGUAGUAGGUUGUAUAGU3') for *hsa-let-7a-5p* was also expressed at 40% of the canonical miRNA expression level²⁰⁷. Given the abundance of these various isomiRs, our candidate miRNA biomarker *hsa-let-7a-1*, which may also overlap with isomiRs for *hsa-let-7a-2* and *hsa-let-7a3*, renders use of the *hsa-let-7a* family as biomarkers of effect uncertain.

According to the same study, *hsa-miR-199a-5p* has 1 isomiR and has been found to be more prevalent than the canonical mature miRNA sequence (286% of canonical expression). Likewise, isomiR1 for *hsa-miR-199a-3p* is expressed at 148% of the canonical *hsa-miR-199a-3p* expression level. Due to their pervasiveness, Karlsen suggested that these isomiRs may have important gene regulation functions and were not just produced at random. Changes introduced to the first 2 nucleotides at the 5' end of the canonical miRNA may alter the seed sequence used in miRNA:mRNA binding to promote mRNA destabilization or translation repression²⁰³. This may lead to new mRNA targets that can potentially affect myriads of new pathways. Since isomiRs are essentially plasma-stable miRNA variants, these can also be potentially explored as biomarkers of effect following xenobiotic response. With the robustness of our NGS data, future efforts can be focused on revisiting these data and applying stringent parameters or using newly developed pipelines for isomiR identification that can reveal potential biomarkers of effect following petrogenic PAH exposure.

Karlsen's study showed that the 3 *hsa-miR-17-5p* isomers (with base differences found in the 3' end) can be present at a cumulative expression of 30% of total isomiR and canonical miRNA indicating that the canonical miRNA is still prevalent. Revisiting our NGS data, the high CALUX bioassay samples did not register expression of the *hsa-miR-17-5p* nor its variants, but the low CALUX bioactivity samples did. Although expression values may change due to the potential presence of the isomiRs, we are not sure how it

would affect the NGS test statistic. Nevertheless, there was certainly a decrease in the expression of *hsa-miR-17-5p* from low to high CALUX bioactivity samples which was verified using qRT-PCR. These 3 isomiRs also have the same 5' end which keep the original seed sequence of the *hsa-miR-17-5p*. Thus, these isomiRs and the canonical sequence would still affect the same mRNAs and consequently the same biological pathways.

To identify the potential pathways affected by the dysregulation of *hsa-miR-17-5p*, target gene prediction and gene ontology functional analysis was performed using miRNET 2.0 (Appendix K). Notable genes in the list are the Krüppel-like factors (*KLF*) 3, 6, and 10 that all come from different KLF groups with important functions in transcription activation for liver homeostasis and muscle development^{208,209}. *KLF6* is a major player in the non-canonical AhR pathway that is currently being investigated in the laboratory³⁵. Clock that controls the circadian rhythm cross-talks with AhR to promote metabolic syndrome²¹⁰.

Gene ontology biological function enrichment analysis in Table 6 shows that *hsa-miR-17-5p* target genes are involved in cell cycle regulation, metabolic processes, and epigenetic modifications. Diseases that were predicted to be associated with *hsa-miR-17-5p* putative target genes include numerous cancers (breast, pancreatic, lung, prostate, liver, blood cancers), psoriasis and pulmonary hypertension.

Downregulation of *hsa-miR-17-5p*, as seen in high AhR bioactivity plasma samples, has been implicated in disease progression in a couple of studies. Hossain et al. (2006) suggested that *hsa-miR-17-5p* acts as a tumor suppressor, inhibiting the expression of the oncogene *AIB1* (amplified in breast cancer 1) also known as *NCOA3* (nuclear receptor coactivator 3). Downregulation of *hsa-miR-17-5p* can lead to loss of *AIB1* suppression and increase in proliferation of the breast cancer cells through transcription activation of the Estrogen Receptor (ER) and E2F1 (E2F Transcription Factor 1)²¹¹. The study by Tsamou et al. (2020) suggested that downregulation of *miR-17/92* cluster

associated with PM_{2.5} exposure of pregnant women may lead to the production of reactive oxygen species that can generate DNA damage and profoundly affect the developing fetus²¹². Downregulation of *hsa-miR-17-5p* has also been associated with the development of drug resistance to erlotinib and paclitaxel in non-small cell lung cancer^{213,214}.

To look for more differentially expressed microRNAs, we reexamined the NGS files and used full sequences to annotate with NCBI-BLAST. We uncovered 236 unique differentially expressed miRNAs, 30 of which were significantly altered (FDR<0.05). A quick search of *hsa-miR-4637* and *hsa-miR-4697-5p* using miRbase revealed that these two miRNAs were only recently identified. Network analysis using IPA showed that the differentially expressed miRNAs are involved in the development of various organ toxicities. Future analysis that looks at other non-coding RNAs may also generate candidate biomarkers of effect to petrogenic PAHs.

Our study has several advantages over other miRNA biomarker discovery studies. First, the use of longitudinally collected plasma samples provided a unique opportunity to follow changes in exposure in single individuals, thus eliminating the need for matching cohorts based on age and gender. Second, compared to the miRNA biomarker for pyrogenic PAH study by Deng et. al. (2014) that sequenced miRNAs from pooled plasma samples, our study identified the miRNAs using individual plasma samples that received unique barcodes during library preparation with NEBNext® Small RNA Library Prep Set for Illumina. This allowed simultaneous sequencing of all the samples in the same flow cell, reducing costs while maintaining the ability to look at miRNA changes in each individual. Multiplexing retains information on biological variations within each volunteer which may have been lost if pooled plasma samples for each AhR bioactivity group was used. These indexed data will be useful in the reanalysis taking isomiR prevalence into consideration. Third, our study used sufficient sequencing depth of 10 million reads per sample which was enough to capture low expressed miRNAs²¹⁵.

We prioritized the verification of the 5p mature miRNA arm in this study. Future investigations may look at the differential expression of the 3p arm of the four identified candidate miRNAs, together with the identification of isomiRs that may also be valuable as biomarkers of effect.

Although we have identified and verified the downregulation of *hsa-miR-17-5p* as associated with the presence of highly bioactive and potentially harmful petrogenic PAHs in human plasma samples, validation using a larger cohort will help establish this differentially expressed miRNA as a signature biomarker of effect for petrogenic PAH exposures. The results of our study promise to form the basis of a rapid and sensitive high throughput biomonitoring tool that can be used in large-scale petrogenic PAH exposure events such as oil spills.

CHAPTER SUMMARY

We sought to identify biomarkers of effect in the form of differentially expressed microRNAs in high CALUX AHR bioactivity human plasma samples with highly bioactive petrogenic PAHs. To do this, we divided our strategy into two phases – discovery and verification.

As hemolysis can contribute to miRNA population in plasma samples, we first performed a hemolysis quality control test using the ΔCq of red blood cell-enriched *hsa-miR-451a* and the reference gene *hsa-miR-23a-3p*. All the plasma samples for discovery and verification phases passed the hemolysis test with a few at moderate hemolysis risk and were closely monitored in the next steps.

In the discovery phase, small RNA Next Generation Sequencing (NGS) and subsequent data analyses identified four differentially expressed miRNAs – *hsa-mir-17*, *hsa-mir-99b*, *hsa-mir-199a1*, and *hsa-let-7a-1* from 20 paired low and high CALUX

bioactivity human plasma samples. The use of paired samples from the same individual eliminated the need for matching factors like age and gender.

In the verification phase, we used quantitative Real Time – Polymerase Chain Reaction (qRT-PCR) to verify the expression of the candidate miRNAs using an additional 15 paired low and high CALUX bioactivity plasma samples. Of the four candidate miRNA biomarkers identified by NGS, *hsa-miR-17-5p* showed consistent significant downregulation concomitant with increased CALUX AHR bioactivity in plasma samples. Target gene prediction and functional enrichment analysis revealed the involvement of this miRNA in the development of cancer, psoriasis, and pulmonary hypertension. It will be interesting to ascertain in future studies whether these disease endpoints align with documented exposures to petrogenic PAHs.

The validation of *hsa-miR-17-5p* using a larger cohort may help establish it as a signature biomarker of effect following exposure to petrogenic PAH mixtures in crude oil and help focus future studies on the biological pathways in which this miRNA is involved. These will be valuable in the assessment and management of human health risks as these pertain to persistent petrogenic PAH exposures, particularly in the consumption of crude oil-contaminated seafood.

A secondary re-analysis of the NGS data using full sequences has identified 30 additional miRNAs that can be potentially used as biomarkers of effect pending verification. Although not statistically significant, this re-analysis verified the decrease in expression of *hsa-miR-17-5p*. Future re-analysis of the data should also consider the detection of functional mature miRNA variants called isomiRs which can also be prevalent in the plasma samples.

CHAPTER 4: DISSERTATION SUMMARY

The safety of seafood harvested from the Gulf of Mexico after the 2010 Deepwater Horizon (DWH) oil spill has been a primary concern particularly amongst the coastal populations that rely on the Gulf seafood for their economic livelihood and sustenance^{5,9,120}. Crude oil contains polycyclic aromatic hydrocarbons (PAHs), a large group of persistent environmental toxicants that can bioaccumulate in seafood and potentially cause adverse human health effects^{13,17,64,216}. A well-established concept in PAH toxicology is the formation of reactive metabolites by high molecular weight (HMW) pyrogenic PAHs through Aryl Hydrocarbon Receptor (AhR)-mediated pathways^{20,22,38}. Petrogenic PAH mixtures found in crude oil, which are primarily composed of 2-4 ring low molecular weight (LMW) PAHs and their alkylated analogues, have also been found to be potent activators of the AhR^{29,158,217}. Thus, it is important to monitor human exposures to these persistent xenobiotics that have the propensity to bioaccumulate in seafood.

Due to the differences in the composition of pyrogenic and petrogenic PAH mixtures (predominance of HMW parent PAHs vs. LMW alkylated PAHs), biomonitoring using pyrogenic PAH data may not be appropriate for petrogenic PAH exposures^{17,20,64}. Unfortunately, current biomonitoring protocols rely on the presence of the more well-studied pyrogenic PAHs (e.g. benzo[a]pyrene, BaP) and their metabolites (e.g. 3-hydroxyBaP) to indicate exposure and estimate human health risks^{61,86,218}. In their response to the DWH oil spill, the US Environmental Protection Agency (US EPA) and the US Food and Drug Administration (US FDA) were heavily criticized because they excluded alkylated petrogenic PAHs highly present in crude oil from their biomonitoring efforts^{82,83}. Hence, there is a need to develop new biomonitoring tools that better reflect human exposures to petrogenic PAHs.

The Gulf Coast Health Alliance: health Risks related to the Macondo Spill (GC-HARMS) was established in response to the DWH oil spill. One of their goals was to

understand the long-term health effects of the DWH oil spill, particularly as a result of consuming petrogenic PAH contaminated seafood. Aside from longitudinal seafood sample collection, the inter-institutional study annually collected human plasma samples spanning 2013 to 2015 from 100 volunteers at each of four partner communities: Mississippi Vietnamese Community (MVC) Gulfport, MS; Center for Environmental and Economic Justice (CEEJ) Biloxi, MS; United Houma Nation (UHN) Houma, LA; with Galveston, TX as comparison cohort not directly impacted by the DWH oil-spill. This plasma collection was accompanied by extensive survey data, comprehensive health assessments, and detailed clinical diagnostics of blood and urine samples^{5,9,120,129}. The group through Dr. Fernando and Dr. Ansari identified and measured the concentrations of 36 different PAHs including their alkylated analogues present in the human plasma samples using Gas Chromatography – Mass Spectrometry (GC-MS) (Fernando et al., unpublished results).

Together with the petrogenic PAHs identified and quantified by the GC-HARMS consortium, we sought to fill the gap in petrogenic PAH biomonitoring by addressing two aims: (1) identify highly bioactive petrogenic PAHs in human plasma samples longitudinally collected after the DWH oil spill that can serve as potential biomarkers of petrogenic PAH exposures, and (2) identify and verify plasma microRNAs associated with high CALUX bioactivity plasma samples that can serve as potential biomarkers of effect following exposures to petrogenic PAHs. These aims were presented in detail in Chapters 2 and 3 of this dissertation, respectively.

To identify highly bioactive petrogenic PAHs, we used the Chemically Activated Luciferase gene eXpression (CALUX) bioassay. This is a cell-based assay that uses a highly sensitive luciferase reporter gene construct regulated by 20 xenobiotic response elements (XREs) in its promoter to which the activated AhR binds and induces expression of the luciferase enzyme. This provides a quantifiable output for AhR activation by the bioactive PAHs in the plasma sample^{91,93,98,100,140}. Optimization of the bioassay resulted to

the determination of suitable human plasma sample volumes for analysis (1:100 dilution with culture media) that maintained robust and reliable luciferase signals without compromising the fidelity of the assay, and the ideal exposure duration (24 hours) of the H1L7.5c3 cell line that provided stable luciferase expression. We also introduced the use of the Quant-iT™ Picogreen™ dsDNA concentration fluorescence assay as a novel integrated normalization method that can be multiplexed with the CALUX bioassay^{142,143}. These modifications coupled with automation of the luciferase and fluorescent signal detection support high throughput processing of the bioassay whilst maintaining reliability and reproducibility.

The CALUX bioactivities of complete plasma sample sets (3 plasma samples per volunteer, 1 sample for each year/wave) from thirty volunteers in each partner location (total of 360 plasma samples) were measured. Analysis of the GC-MS spectra established total PAH body burdens and the CALUX AhR bioactivities using samples collected at each location demonstrated the presence of bioactive PAHs in some cohorts that warrant further characterization. A general increase in total PAH body burden was observed in all locations during the subsequent years post-spill, yet the CALUX bioactivities remained largely unchanged during this period. The reason for this discordance is unclear but suggests that assessment of PAH body burdens offer a poor approximation of the toxic potential, especially when PAH exposures may reflect both natural and anthropomorphic sources. Conceivably the increase in body burden may be associated to factors that will be revealed using the extensive survey data that accompanied each plasma sample, including other lifestyle considerations including diet and tobacco use etc.

Assessment of the individual PAH concentrations and plasma CALUX AhR bioactivities identified four petrogenic PAHs revealed significant associations in the different cohorts. The petrogenic PAHs are acenaphthylene, benz[a]anthracene, C3-naphthalene, and fluorene. Among these four, C3-naphthalene, which is the only alkylated PAH in the list, showed a significant positive association with increase in CALUX AhR

bioactivity when data from all cohorts were analyzed together. It is currently not included in the US EPA 16 priority PAHs that is routinely applied to human health risk assessments in PAH exposures⁸⁰. Our data strongly contends that an update of this list and the inclusion of C3-naphthalene as a biomarker of petrogenic PAH exposure are warranted.

Our second aim was to identify a novel human biomarker of effect to petrogenic PAH exposure in the form of differentially expressed plasma miRNAs that are associated with high CALUX AhR bioactivity. Next Generation Sequencing (NGS) of 20 paired high and low CALUX AhR bioactivity human plasma samples collected after the DWH oil spill identified four candidate microRNAs – *hsa-miR-17*, *hsa-miR-99b*, *hsa-miR-199a-1*, and *hsa-let-7a-1*. Verification using quantitative Real-Time Polymerase Chain Reaction (qRT-PCR) and the inclusion of an additional 15 paired low and high CALUX AhR bioactivity plasma samples established *hsa-miR-17-5p* as a candidate signature human biomarker of effect following exposure to highly bioactive petrogenic PAHs.

Target gene prediction and functional enrichment analysis revealed the association of *hsa-miR-17-5p* with important biological pathways that can ultimately lead the development of psoriasis, pulmonary hypertension, and different forms of cancer. Current literature also points at the downregulation of *hsa-miR-17-5p* as associated with the development of breast cancer²¹⁹, and chemotherapeutic drug resistance in non-small cell lung cancer^{213,214}. The downregulation of the *mir-17/92* cluster has also been associated with PM_{2.5} exposures of pregnant women^{212,220}. Large scale validation studies using a bigger cohort will help validate this miRNA as a reliable indicator of petrogenic PAH exposure.

This dissertation addressed several gaps in the biomonitoring of human exposures to petrogenic PAHs. First, the use of human plasma samples to address our two aims lends strength to the direct translational value of our results to support human health risk assessments. Second, we were able to identify biomarkers of exposure and effect that are associated specifically with human exposures to petrogenic PAHs. Using these biomarkers,

and the optimized methodologies developed in the process, we will be able to respond to future petrogenic PAH exposure events with more appropriate tools in hand. Third, the identification of an alkylated petrogenic PAH (C3-naphthalene) as a biomarker of petrogenic PAH exposure supports the update of the priority PAH list that is ill-suited for crude oil exposure biomonitoring. Fourth, we were able to optimize a US EPA certified bioassay by introducing a novel normalization method that maintains the reliability and reproducibility of the results. We were also able to directly use human plasma samples on the cell-based bioassay that did not require elaborate and expensive sample preparation. Lastly, we were able to identify and verify a biomarker of effect (*hsa-miR-17-5p*) associated with the presence of highly bioactive petrogenic PAHs in human plasma samples that can reflect overall petrogenic PAH mixture toxicity. Rapid advancements in molecular diagnostics have made testing for plasma miRNA faster, cheaper, and more sensitive even with small sample input. This provides several advantages particularly in large scale biomonitoring of human exposures to environmental toxicants such as the petrogenic PAHs in crude oil.

As humans continue to rely on petroleum for energy and plastics manufacturing, crude oil exposure events due to natural seepages and catastrophic oil spills will continue to happen. The identification of these biomarkers of exposure and effect associated with petrogenic PAH exposure will better inform our preparedness and response in the future. These will also provide the much-needed update on human health risk assessments which form the basis of policies and guidelines applied to crude oil exposure events. More importantly, results of this study and the optimized methodologies developed will be valuable in providing answers to the concerns of our partner communities in the Gulf that rely on the waters for their livelihood and sustenance.

APPENDICES

APPENDIX A: H1L7.5c3 LUCIFERASE REPORTER GENE CONSTRUCT

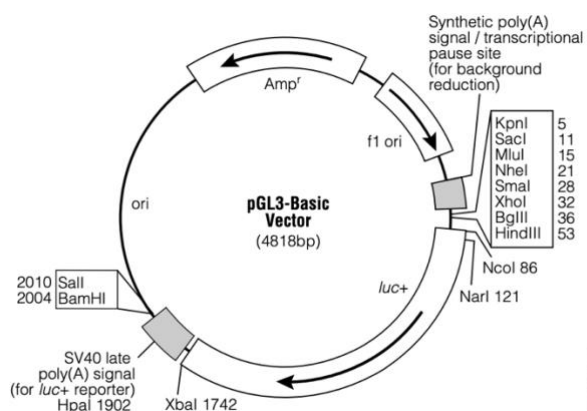


Figure 24. Luciferase reporter gene vector backbone used to construct the pGudLuc7.5. (Promega, 2015).

APPENDIX B: CALUX BIOASSAY PLATE LAYOUT

Table 7. CALUX Bioassay 96-well Plate Layout.

PLATE _____, LUC READ: _____, PICO: _____													
Treatment		1	2	3	4	5	6	7	8	9	10	11	12
STANDARDS	A	NO CELLS/ EMPTY			10E ⁻¹⁴ M BaP	10E ⁻¹³ M BaP	10E ⁻¹² M BaP	10E ⁻¹¹ M BaP	10E ⁻¹⁰ M BaP	10E ⁻⁹ M BaP	10E ⁻⁸ M BaP	10E ⁻⁷ M BaP	10E ⁻⁶ M BaP
	B	MEDIA											
	C	0.5% DMSO MEDIA											
PLASMA SAMPLES	D	Sample 1A	Sample 1B	Sample 1C	Sample 2A	Sample 2B	Sample 2C	Sample 3A	Sample 3B	Sample 3C	Sample 4A	Sample 4B	Sample 4C
	E												
	F												
DNA standards	G	BLANK (0 ng/mL)			50 ng/mL			250 ng/mL			750 ng/mL		
	H	25 ng/mL			100ng/mL			500 ng/mL			1000 ng/mL		

APPENDIX C: QUANT-iT™ PICOGREEN™ STANDARD CALCULATION

Table 8. Example calculation for Lambda DNA standards.

1 PLATE							EDITED 4/30/2018 DY	
Stock DNA (ug/mL)	Working DNA (ug/mL)	Total Volume (mL)	Vol stock DNA (uL)	Vol 1X TE (uL)		Total 1X TE (uL)	1016	
100	10	0.3	30	270				
Num of wells	Final standard (ng/mL)	Working DNA Standard Solution			Volume per well			
		Vol of working DNA (uL)	Vol 1X PLB (uL)	Vol 1X TE (uL)	Standard (uL/well)	Luc Substrate (uL/well)	Vol Diluted Picogreen (uL)	TOTAL/well
3	1000	80	80	40	50	50	100	200
3	750	60	80	60	50	50	100	200
3	500	40	80	80	50	50	100	200
3	250	20	80	100	50	50	100	200
3	100	8	80	112	50	50	100	200
3	50	4	80	116	50	50	100	200
3	25	2	80	118	50	50	100	200
3	0	0	80	120	50	50	100	200
TOTAL		214	640	746				
*Extra 1 well in calculations			1600					

APPENDIX D: TOTAL PAH vs CALUX AhR BIOACTIVITY OF ALL 30 VOLUNTEERS IN EACH COHORT

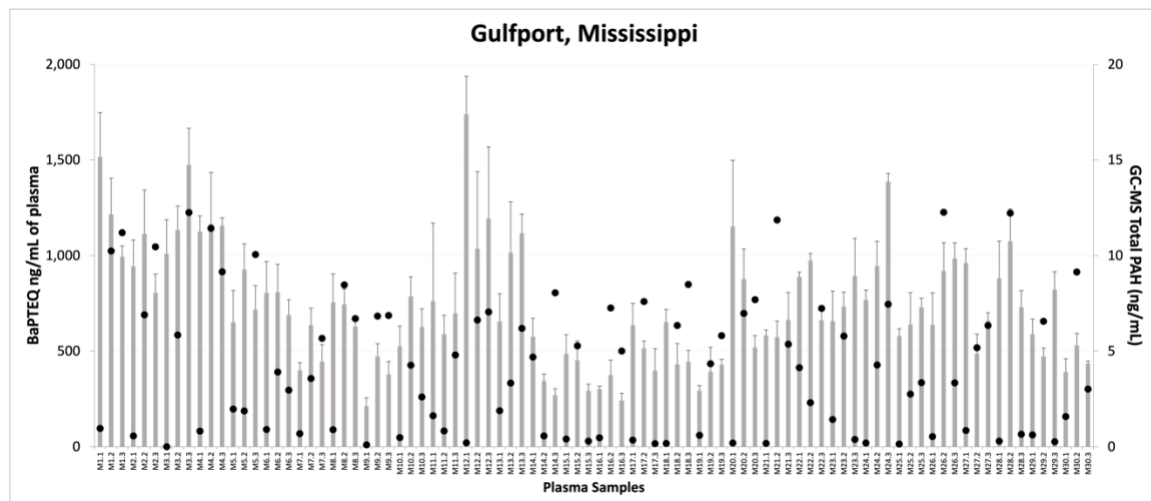


Figure 25. MVC Gulfport, MS CALUX AhR bioactivities and GC-MS Total PAH body burdens.

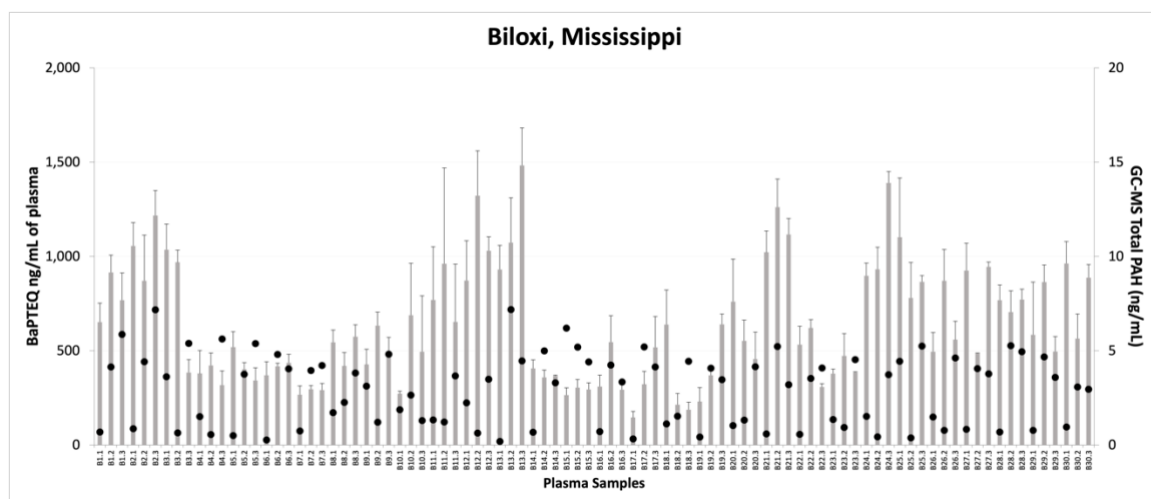


Figure 26. CEEJ Biloxi, MS CALUX AhR bioactivities and GC-MS Total PAH body burdens.

APPENDIX D (CONTINUED)

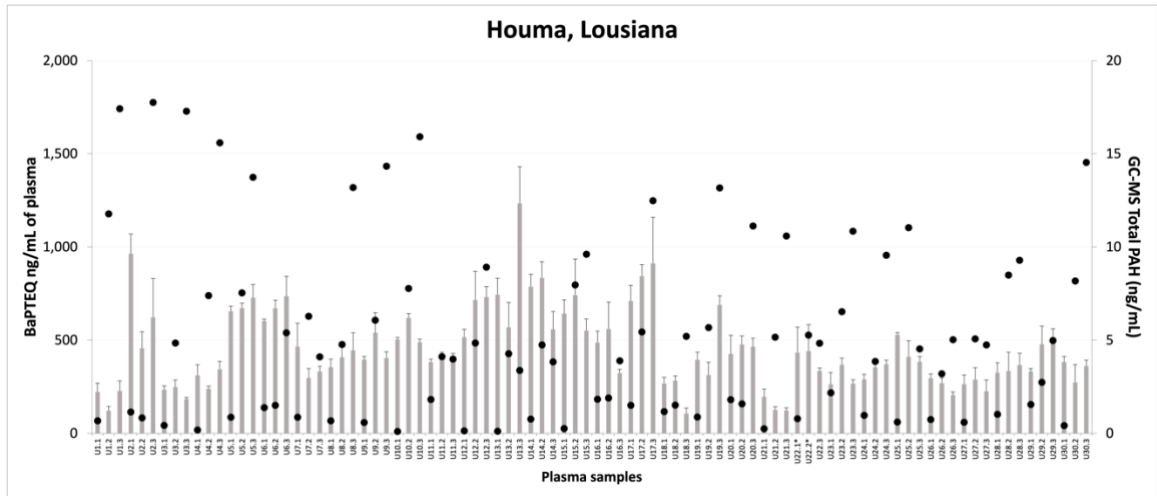


Figure 27. UHN Houma, LA CALUX AhR bioactivities and GC-MS Total PAH body burdens.

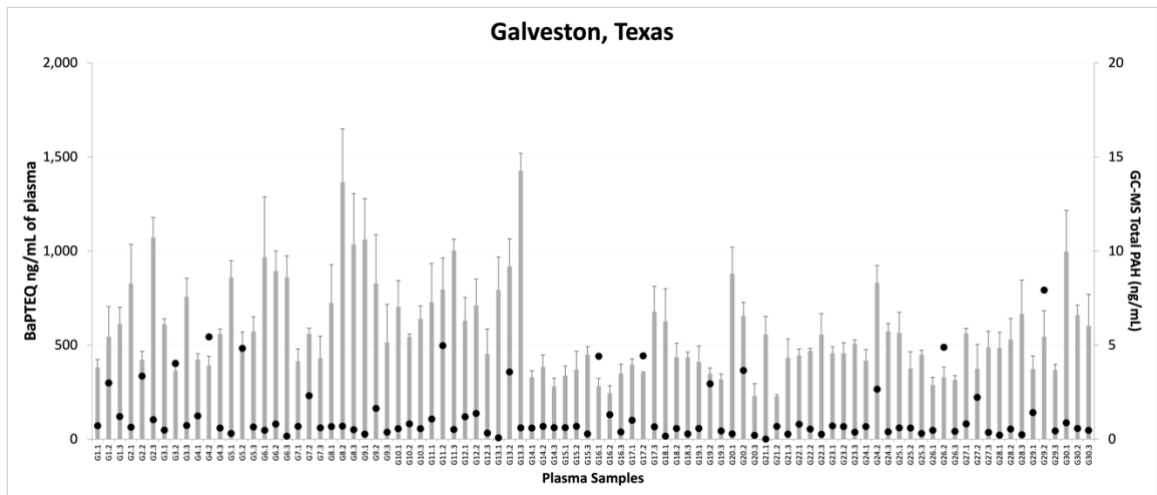


Figure 28. Galveston, TX CALUX AhR bioactivities and GC-MS Total PAH body burdens.

APPENDIX E: PEARSON CORRELATIONS OF ALL DETECTED PAHS

Table 9. Pearson correlation results generated from analyzing ALL locations.

Pearson correlation (ALL LOCATIONS)	r	95% confidence interval	R squared	P (two-tailed)	P value summary	Significant? (alpha = 0.05)	Number of XY Pairs
CALUXvs.Naphthalene (10.02)	0.008929	-0.1020 to 0.1197	7.97E-05	0.875	ns	No	313
CALUXvs.C-1 Naph	-0.003974	-0.1249 to 0.1170	1.58E-05	0.9489	ns	No	263
CALUXvs.C-2 Naph	-0.05056	-0.1755 to 0.07603	0.002556	0.4336	ns	No	242
CALUXvs.C-3 Naph	0.3759	0.003243 to 0.6568	0.1413	0.0487	*	Yes	28
CALUXvs.C-4 Naph							
CALUXvs.Acenaphthylene (15.79)	0.03303	-0.1486 to 0.2125	0.001091	0.7225	ns	No	118
CALUXvs.Acenaphthene (16.46)	0.09485	-0.8593 to 0.9017	0.008997	0.8794	ns	No	5
CALUXvs.Flourene (18.48)	-0.7594	-0.9463 to -0.1923	0.5767	0.0176	*	Yes	9
CALUXvs.C1Flourene							
CALUXvs.C2Flourene (22.62)	0.6394	-0.5572 to 0.9729	0.4088	0.2454	ns	No	5
CALUXvs.C3Flourene							
CALUXvs.Phenanthrene (22.16)	-0.07773	-0.2377 to 0.08637	0.006042	0.3527	ns	No	145
CALUXvs.C1Phenanthrene							
CALUXvs.C2Phenanthrene							
CALUXvs.C3Phenanthrene							
CALUXvs.C4Phenanthrene							
CALUXvs.Anthracene (22.36)							1
CALUXvs.Flouanthene (26.78)	0.8598	-0.5832 to 0.9970	0.7393	0.1402	ns	No	4
CALUXvs.Pyrene (27.63)	-0.005271	-0.1257 to 0.1153	2.78E-05	0.9319	ns	No	265
CALUXvs.C1Pyrene							
CALUXvs.C2Pyrene							
CALUXvs.C3Pyrene							
CALUXvs.C4Pyrene							
CALUXvs.Benzo(a)anthracene (32.32)	-0.07558	-0.3137 to 0.1715	0.005713	0.5496	ns	No	65
CALUXvs.Chrysene (32.43)	0.5048	-0.6806 to 0.9597	0.2548	0.3857	ns	No	5
CALUXvs.C1-Chrys							
CALUXvs.C2-Chrys							
CALUXvs.C3-Chrys							
CALUXvs.C4-Chrys							
CALUXvs.Perylene-d12 IS#3 (37.45)							
CALUXvs.Benzo(b) flouranthene (36.2)							
CALUXvs.Benzo(k) flouranthene (36.29)							
CALUXvs.Benzo(a)pyrene (37.26)							1
CALUXvs.Indeno(1,2,3-cd)pyrene (40.71)							
CALUXvs.Dibenzo(a,h)anthracene (40.81)							
CALUXvs.Benzo(g,h,i)perylene (41.53)							

APPENDIX E (CONTINUED)

Table 10. Pearson correlation results for MVC Gulfport, MS.

Pearson Correlation (GULFPORT, MS)	r	95% confidence interval	R squared	P (two-tailed)	P value summary	Significant? (alpha = 0.05)	Number of XY Pairs
CALUXvs.Naphthalene (10.02)	0.1897	-0.03427 to 0.3955	0.03599	0.0962	ns	No	78
CALUXvs.C-1 Naph	0.1916	-0.05489 to 0.4160	0.0367	0.1263	ns	No	65
CALUXvs.C-2 Naph	0.06864	-0.1908 to 0.3191	0.004711	0.6055	ns	No	59
CALUXvs.C-3 Naph	0.4776	0.01376 to 0.7723	0.2281	0.045	*	Yes	18
CALUXvs.C-4 Naph							
CALUXvs.Acenaphthylene (15.79)	-0.0764	-0.4370 to 0.3054	0.005837	0.6992	ns	No	28
CALUXvs.Acenaphthene (16.46)							1
CALUXvs.Flourene (18.48)	-0.6849	-0.9618 to 0.2851	0.4692	0.1333	ns	No	6
CALUXvs.C1Flourene							
CALUXvs.C2Flourene (22.62)							
CALUXvs.C3Flourene							
CALUXvs.Phenanthrene (22.16)	0.0521	-0.2486 to 0.3437	0.002715	0.737	ns	No	44
CALUXvs.C1Phenanthrene							
CALUXvs.C2Phenanthrene							
CALUXvs.C3Phenanthrene							
CALUXvs.C4Phenanthrene							
CALUXvs.Anthracene (22.36)							
CALUXvs.Flouanthene (26.78)							1
CALUXvs.Pyrene (27.63)	0.1533	-0.08121 to 0.3718	0.02351	0.1984	ns	No	72
CALUXvs.C1Pyrene							
CALUXvs.C2Pyrene							
CALUXvs.C3Pyrene							
CALUXvs.C4Pyrene							
CALUXvs.Benzo(a)anthracene (32.32)	-0.2579	-0.6287 to 0.2084	0.06651	0.2723	ns	No	20
CALUXvs.Chrysene (32.43)							
CALUXvs.C1-Chrys							
CALUXvs.C2-Chrys							
CALUXvs.C3-Chrys							
CALUXvs.C4-Chrys							
CALUXvs.Perylene-d12 IS#3 (37.45)							
CALUXvs.Benzo(b) flouranthene (36.2)							
CALUXvs.Benzo(k) flouranthene (36.29)							
CALUXvs.Benzo(a)pyrene (37.26)							
CALUXvs.Indeno(1,2,3-cd)pyrene (40.71)							
CALUXvs.Dibenzo(a,h)anthracene (40.81)							
CALUXvs.Benzo(g,h,i)perylene (41.53)							

APPENDIX E (CONTINUED)

Table 11. Pearson correlation results of CEEJ Biloxi, MS.

Pearson correlation (BILOXI,MS)	r	95% confidence interval	R squared	P (two-tailed)	P value summary	Significant? (alpha = 0.05)	Number of XY Pairs
CALUXvs.Naphthalene (10.02)	0.201	-0.01813 to 0.4017	0.04041	0.072	ns	No	81
CALUXvs.C-1 Naph	-0.08567	-0.3096 to 0.1473	0.007338	0.4711	ns	No	73
CALUXvs.C-2 Naph	-0.1525	-0.3913 to 0.1055	0.02325	0.2448	ns	No	60
CALUXvs.C-3 Naph	-0.994		0.988	0.0699	ns	No	3
CALUXvs.C-4 Naph							
CALUXvs.Acenaphthylene (15.79)	0.0117	-0.3009 to 0.3220	0.0001368	0.9429	ns	No	40
CALUXvs.Acenaphthene (16.46)	0.02451	-0.9592 to 0.9629	0.000601	0.9755	ns	No	4
CALUXvs.Flourene (18.48)							1
CALUXvs.C1Flourene							
CALUXvs.C2Flourene (22.62)							
CALUXvs.C3Flourene							
CALUXvs.Phenanthrene (22.16)	-0.2022	-0.5146 to 0.1576	0.04089	0.267	ns	No	32
CALUXvs.C1Phenanthrene							
CALUXvs.C2Phenanthrene							
CALUXvs.C3Phenanthrene							
CALUXvs.C4Phenanthrene							
CALUXvs.Anthracene (22.36)							1
CALUXvs.Flouanthrene (26.78)							
CALUXvs.Pyrene (27.63)	-0.0968	-0.3311 to 0.1487	0.00937	0.4394	ns	No	66
CALUXvs.C1Pyrene							
CALUXvs.C2Pyrene							
CALUXvs.C3Pyrene							
CALUXvs.C4Pyrene							
CALUXvs.Benzo(a)anthracene (32.32)	-0.002124	-0.7057 to 0.7036	4.51E-06	0.996	ns	No	8
CALUXvs.Chrysene (32.43)	0.872	-0.5502 to 0.9973	0.7604	0.128	ns	No	4
CALUXvs.C1-Chrys							
CALUXvs.C2-Chrys							
CALUXvs.C3-Chrys							
CALUXvs.C4-Chrys							
CALUXvs.Perylene-d12 IS#3 (37.45)							
CALUXvs.Benzo(b) flouranthene (36.2)							
CALUXvs.Benzo(k) flouranthene (36.29)							
CALUXvs.Benzo(a)pyrene (37.26)							
CALUXvs.Indeno(1,2,3-cd)pyrene (40.71)							
CALUXvs.Dibenzo(a,h)anthracene (40.81)							
CALUXvs.Benzo(g,h,i)perylene (41.53)							

APPENDIX E (CONTINUED)

Table 12. Pearson correlation results for UHN Houma, LA.

Pearson Correlation (HOUMA, LA)	r	95% confidence interval	R squared	P (two-tailed)	P value summary	Significant? (alpha = 0.05)	Number of XY Pairs
CALUXvs.Naphthalene (10.02)	-0.04069	-0.2581 to 0.1806	0.001656	0.72	ns	No	80
CALUXvs.C-1 Naph	-0.02042	-0.2559 to 0.2173	0.0004169	0.8677	ns	No	69
CALUXvs.C-2 Naph	-0.1439	-0.3764 to 0.1056	0.02072	0.2565	ns	No	64
CALUXvs.C-3 Naph	0.9049	-0.4313 to 0.9980	0.8188	0.0951	ns	No	4
CALUXvs.C-4 Naph							
CALUXvs.Acenaphthylene (15.79)	0.4404	0.01075 to 0.7328	0.1939	0.0457	*	Yes	21
CALUXvs.Acenaphthene (16.46)							
CALUXvs.Fluorene (18.48)							1
CALUXvs.C1Flourene							
CALUXvs.C2Flourene (22.62)							2
CALUXvs.C3Flourene							
CALUXvs.Phenanthrene (22.16)	-0.09526	-0.4349 to 0.2681	0.009074	0.6102	ns	No	31
CALUXvs.C1Phenanthrene							
CALUXvs.C2Phenanthrene							
CALUXvs.C3Phenanthrene							
CALUXvs.C4Phenanthrene							
CALUXvs.Anthracene (22.36)							
CALUXvs.Fluoranthene (26.78)							1
CALUXvs.Pyrene (27.63)	-0.03429	-0.2607 to 0.1957	0.001176	0.7718	ns	No	74
CALUXvs.C1Pyrene							
CALUXvs.C2Pyrene							
CALUXvs.C3Pyrene							
CALUXvs.C4Pyrene							
CALUXvs.Benzo(a)anthracene (32.32)	0.7177	0.3772 to 0.8873	0.5151	0.0008	***	Yes	18
CALUXvs.Chrysene (32.43)							1
CALUXvs.C1-Chrys							
CALUXvs.C2-Chrys							
CALUXvs.C3-Chrys							
CALUXvs.C4-Chrys							
CALUXvs.Perylene-d12 IS#3 (37.45)							
CALUXvs.Benzo(b) flouranthene (36.2)							
CALUXvs.Benzo(k) flouranthene (36.29)							
CALUXvs.Benzo(a)pyrene (37.26)							1
CALUXvs.Indeno(1,2,3-cd)pyrene (40.71)							
CALUXvs.Dibenzo(a,h)anthracene (40.81)							
CALUXvs.Benzo(g,h,i)perylene (41.53)							

APPENDIX E (CONTINUED)

Table 13. Pearson correlation results for Galveston, TX.

Pearson correlation (GALVESTON, TX)	r	95% confidence interval	R squared	P (two-tailed)	P value summary	Significant? (alpha = 0.05)	Number of XY Pairs
CALUXvs.Naphthalene (10.02)	-0.1138	-0.3336 to 0.1178	0.01294	0.3345	ns	No	74
CALUXvs.C-1 Naph	-0.1743	-0.4181 to 0.09281	0.0304	0.1987	ns	No	56
CALUXvs.C-2 Naph	-0.05975	-0.3111 to 0.1994	0.00357	0.6531	ns	No	59
CALUXvs.C-3 Naph	0.4196		0.1761	0.7243	ns	No	3
CALUXvs.C-4 Naph							
CALUXvs.Acenaphthylene (15.79)	-0.1915	-0.5214 to 0.1882	0.03667	0.3197	ns	No	29
CALUXvs.Acenaphthene (16.46)							
CALUXvs.Flourene (18.48)							1
CALUXvs.C1Flourene							
CALUXvs.C2Flourene (22.62)	0.826		0.6822	0.3812	ns	No	3
CALUXvs.C3Flourene							
CALUXvs.Phenanthrene (22.16)	-0.1476	-0.4462 to 0.1806	0.02178	0.3766	ns	No	38
CALUXvs.C1Phenanthrene							
CALUXvs.C2Phenanthrene							
CALUXvs.C3Phenanthrene							
CALUXvs.C4Phenanthrene							
CALUXvs.Anthracene (22.36)							
CALUXvs.Fluoranthene (26.78)							2
CALUXvs.Pyrene (27.63)	0.01719	-0.2543 to 0.2862	0.0002955	0.9028	ns	No	53
CALUXvs.C1Pyrene							
CALUXvs.C2Pyrene							
CALUXvs.C3Pyrene							
CALUXvs.C4Pyrene							
CALUXvs.Benzo(a)anthracene (32.32)	-0.5196	-0.7879 to -0.08558	0.27	0.0226	*	Yes	19
CALUXvs.Chrysene (32.43)							
CALUXvs.C1-Chrys							
CALUXvs.C2-Chrys							
CALUXvs.C3-Chrys							
CALUXvs.C4-Chrys							
CALUXvs.Perylene-d12 I5#3 (37.45)							
CALUXvs.Benzo(b) flouranthene (36.2)							
CALUXvs.Benzo(k) flouranthene (36.29)							
CALUXvs.Benzo(a)pyrene (37.26)							
CALUXvs.Indeno(1,2,3-cd)pyrene (40.71)							
CALUXvs.Dibenzo(a,h)anthracene (40.81)							
CALUXvs.Benzo(g,h,i)perylene (41.53)							

APPENDIX F: HEMOLYSIS RATIOS OF PLASMA SAMPLES FOR NGS AND QRT-PCR

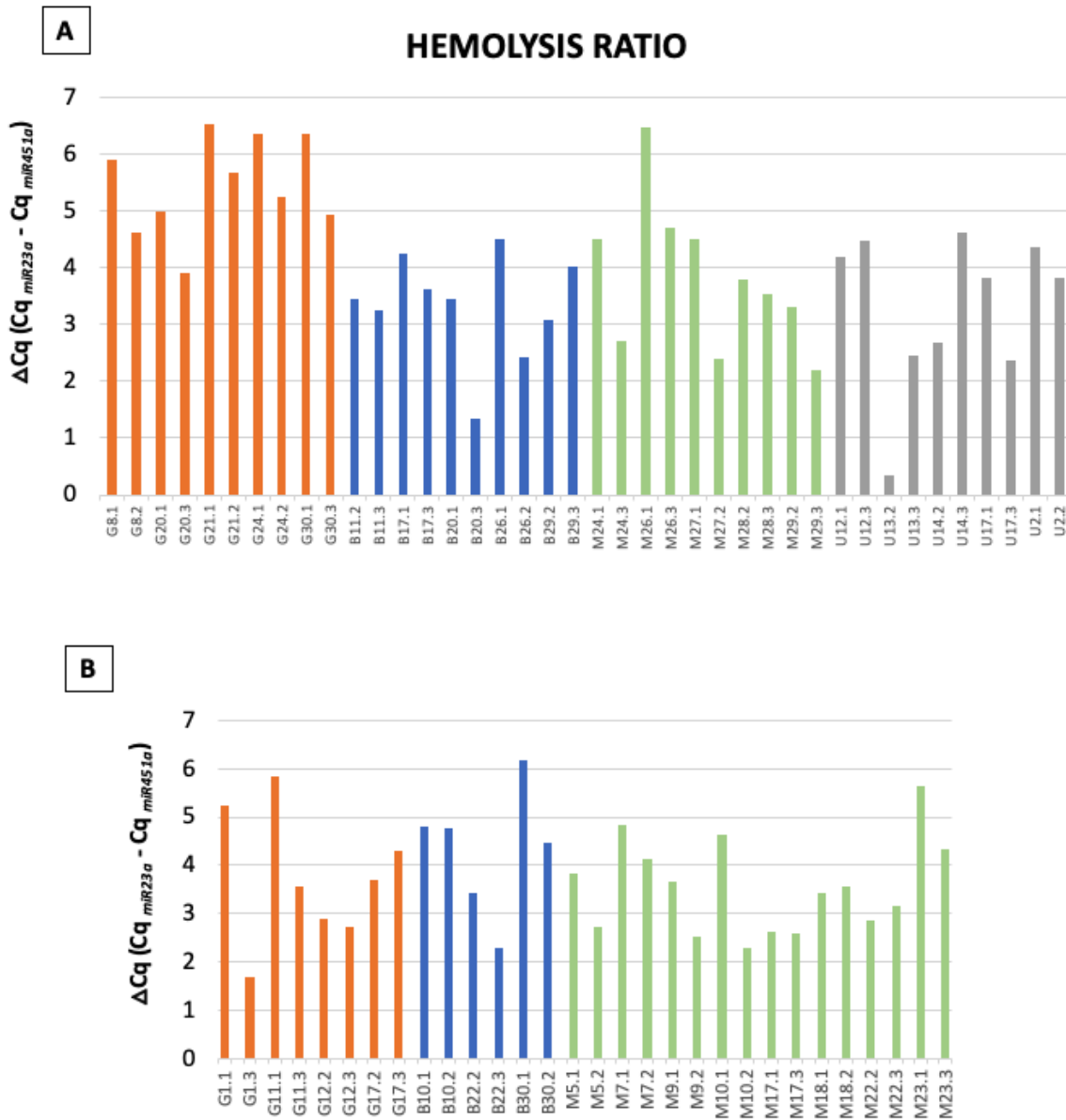
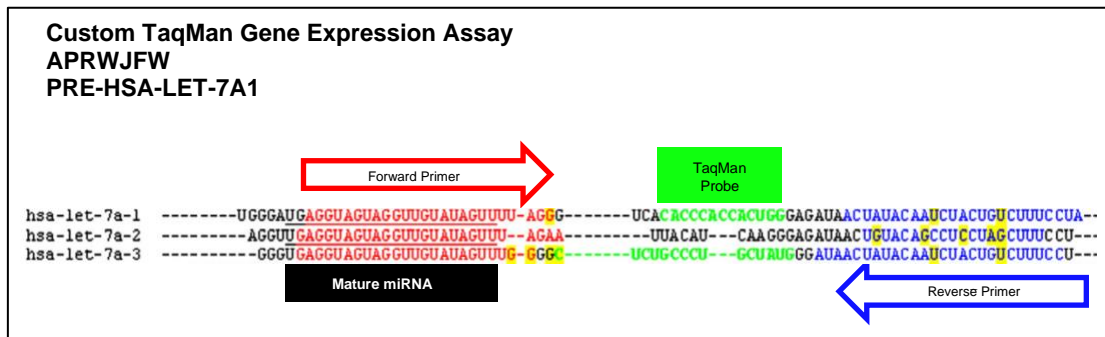


Figure 29. GC-HARMS human plasma samples passed hemolysis test. Hemolysis ratio of (A) paired low and high CALUX plasma samples for discovery phase (n=20 volunteers), and (B) additional paired samples (n=15 volunteers) for the verification phase. A ΔCq of <5 is considered low hemolysis risk, $5 \leq \Delta Cq \leq 7$ moderate risk, and $\Delta Cq > 7$ as high hemolysis risk. Samples are labeled as G - Galveston, TX, B - Biloxi, MS; M - MVC Gulfport, MS; and U - UHN Houma, LA. Digit before decimal place corresponds to sample number and digit after decimal place represents wave or annual sample collection. Abbreviations: Cq, quantitation cycle; miR, microRNA.

APPENDIX G: TAQMAN miRNA ASSAY TARGET SEQUENCES

Assay Kit	Target miRNA	Cat#/Assay ID	Target Sequence
TaqMan small RNA assay	<i>hsa-miR-451a</i>	Cat. #4427975, Assay ID 001141	AAACCGUUACCAUUACUGAGUU
	<i>hsa-miR-23a-3p</i>	Cat #4427975, Assay ID 000399	AUCACAUUGCCAGGGAUUUUCC
	<i>hsa-miR-17-5p</i>	Cat. # A25576, Assay ID 478447	CAAAGUGCUUACAGUGCAGGUAG
	<i>hsa-miR-99b-5p</i>	Cat. # A25576, Assay ID 478343	CACCCGUAGAACCGACCUUGCG
TaqMan Non-coding RNA Assay	<i>mir-199a-1</i>	Cat. # 4426961 Assay ID Hs07319370_s1	NCBI Reference Sequence: NR_029586.1
Customized TaqMan Gene Expression assay	<i>pre-let-7a-1</i>	Cat. # 4331348 APRWJFW PRE-HSA-LET-7A1	Forward Primer: 5'AGGTAGTAGGTTGTATAGTTTTAGG Reverse Primer: 5'TAGGAAAGACAGTAGATTGTATAGT Probe Sequence: 5'-FAM-CACCCACCACTGG-MGB-3' Mature miRNA Sequence: 5'UGAGGUAGUAGGUUGUAUAGUU



Modified image, reproduced with permission, Jiang et al, 2005

APPENDIX H: PCR AND QRT-PCR PROTOCOLS

Poly(A) Tailing		
STEP	Temp (°C)	Time (minutes)
Polyadenylation	37°C	45 minutes
Stop Reaction	65°C	10 minutes
Hold	4°C	Hold

Adapter Ligation		
STEP	Temp (°C)	Time (minutes)
Ligation	16°C	60 minutes
Hold	4°C	Hold

Reverse Transcription		
STEP	Temp (°C)	Time (minutes)
Reverse Transcription	42°C	15 minutes
Stop Reaction	85°C	5 minutes
Hold	4°C	Hold

miR-Amplification and Pre-amplification			
STEP	Temp (°C)	Time	Cycles
Enzyme Activation	95°C	10 minutes	1
Denature	95°C	15 seconds	} 14
Anneal and Extend	60°C	4 minutes	
Stop Reaction	99°C	10 minutes	1
Hold	4°C	Hold	1

Real-Time PCR Protocol (BioRad CFX96)			
STEP	Temp (°C)	Time	Cycles
UNG Activation (Optional)	50°C	2 minutes	1
Enzyme activation	95°C	20 seconds	1
Denature	95°C	3 seconds	} 40
Anneal/Extend	60°C	30 seconds	

APPENDIX I: UPREGULATED miRNAs IN HIGH CALUX BIOACTIVITY SAMPLES

miRNA	LOW CALUX Normalized	HIGH CALUX Normalized	Fold Change	P-value
*let-7a-1	0.045	0.385	8.555555556	0.026438698
<i>mir-22</i>	0.035	0.205	5.857142857	0.171549319
<i>mir-186</i>	0.09	0.375	4.166666667	0.392829462
<i>mir-148a</i>	0.04	0.13	3.25	0.464993031
<i>mir-20a</i>	0.14	0.38	2.714285714	0.467089191
<i>mir-122</i>	0.19	0.515	2.710526316	0.1570823
<i>mir-11400</i>	0.02	0.05	2.5	0.580735349
<i>mir-3613</i>	0.06	0.135	2.25	0.441521055
<i>let-7b</i>	0.295	0.65	2.203389831	0.159496934
<i>mir-125a</i>	0.085	0.17	2	0.51732676
<i>mir-10a</i>	0.115	0.22	1.913043478	0.411572992
<i>mir-142</i>	0.38	0.65	1.710526316	0.263024946
<i>mir-340</i>	0.28	0.475	1.696428571	0.304735789
<i>mir-106b</i>	0.035	0.05	1.428571429	0.807183238
<i>mir-423</i>	2.72	3.755	1.380514706	0.191293514
<i>let-7a-3</i>	0.955	1.3	1.361256545	0.381075756
<i>mir-584</i>	0.06	0.08	1.333333333	0.818286968
<i>let-7g</i>	1.27	1.68	1.322834646	0.346477424
<i>mir-744</i>	0.3	0.395	1.316666667	0.779633148
<i>mir-16-2</i>	0.085	0.11	1.294117647	0.786227457
<i>mir-32</i>	0.04	0.05	1.25	0.876722804
<i>mir-432</i>	0.04	0.05	1.25	0.876722804
<i>let-7f-2</i>	1.09	1.345	1.233944954	0.509508881
<i>mir-30d</i>	2.425	2.94	1.212371134	0.325477249
<i>mir-146a</i>	3.345	3.785	1.131539611	0.49930285
<i>mir-139</i>	0.045	0.05	1.111111111	0.932825299
<i>mir-26a-1</i>	3.72	3.81	1.024193548	0.865715103
<i>mir-26a-2</i>	4.215	4.22	1.00118624	0.993629161
<i>mir-182</i>	0	0.025	#DIV/0!	0.163264694
<i>mir-532</i>	0	0.085	#DIV/0!	0.237434954
<i>mir-181c</i>	0	0.01	#DIV/0!	0.323636084
<i>mir-431</i>	0	0.01	#DIV/0!	0.323636084
<i>mir-192</i>	0	0.035	#DIV/0!	0.323636084
<i>mir-361</i>	0	0.05	#DIV/0!	0.323636084
<i>mir-6891</i>	0	0.055	#DIV/0!	0.323636084
<i>mir-181d</i>	0	0.055	#DIV/0!	0.323636084
<i>mir-378d-2</i>	0	0.065	#DIV/0!	0.323636084
<i>mir-224</i>	0	0.07	#DIV/0!	0.323636084
<i>mir-181b-2</i>	0	0.075	#DIV/0!	0.323636084
<i>mir-556</i>	0	0.095	#DIV/0!	0.323636084
<i>mir-505</i>	0	0.115	#DIV/0!	0.323636084

* $p < 0.05$

APPENDIX J: DOWNREGULATED miRNAs IN HIGH CALUX BIOACTIVITY SAMPLES

miRNA	LOW CALUX Normalized	HIGH CALUX Normalized	Fold Change	P-value
<i>mir-21</i>	6.585	6.43	8.555555556	0.877540585
<i>mir-30e</i>	0.955	0.905	5.857142857	0.887384624
<i>let-7i</i>	3.125	2.88	4.166666667	0.761365909
<i>mir-451a</i>	5.65	5.2	3.25	0.702837024
<i>mir-629</i>	0.045	0.04	2.714285714	0.934251306
<i>let-7a-2</i>	0.44	0.35	2.710526316	0.632689097
<i>mir-3135b</i>	0.07	0.055	2.5	0.867086138
<i>mir-126</i>	0.3	0.23	2.25	0.648552827
<i>let-7f-1</i>	0.495	0.375	2.203389831	0.595471249
<i>mir-140</i>	0.02	0.015	2	0.842547069
<i>mir-486-2</i>	0.76	0.545	1.913043478	0.403937251
<i>mir-191</i>	0.805	0.54	1.710526316	0.246931613
<i>mir-194-1</i>	0.06	0.04	1.696428571	0.757833202
<i>mir-26b</i>	0.875	0.575	1.428571429	0.280506853
<i>mir-181a-1</i>	0.195	0.12	1.380514706	0.491101562
<i>mir-16-1</i>	0.38	0.23	1.361256545	0.46214021
<i>let-7d</i>	0.215	0.13	1.333333333	0.473069512
<i>mir-181a-2</i>	0.1	0.06	1.322834646	0.668664887
<i>mir-30c-1</i>	0.045	0.025	1.316666667	0.681229557
<i>mir-486-1</i>	0.58	0.315	1.294117647	0.302242458
<i>mir-134</i>	0.035	0.015	1.25	0.602477635
<i>mir-335</i>	0.035	0.015	1.25	0.602477635
<i>mir-199a-2</i>	0.47	0.2	1.233944954	0.082037688
<i>mir-93</i>	0.145	0.06	1.212371134	0.358477056
<i>mir-99a</i>	0.315	0.13	1.131539611	0.253340746
*mir-199a-1	0.76	0.265	1.111111111	0.042145156
<i>mir-98</i>	0.13	0.045	1.024193548	0.363943496
<i>mir-185</i>	0.345	0.115	1.00118624	0.132705968
<i>mir-146b</i>	0.19	0.06	#DIV/0!	0.177759546
<i>mir-30c-2</i>	0.035	0.01	#DIV/0!	0.496377109
<i>mir-374a</i>	0.15	0.025	#DIV/0!	0.068384653
*mir-99b	0.185	0.01	#DIV/0!	0.050914884
*mir-17	0.185	0	#DIV/0!	0.03967066
<i>mir-425</i>	0.175	0	#DIV/0!	0.104122317
<i>mir-223</i>	0.12	0	#DIV/0!	0.14663067
<i>mir-30a</i>	0.125	0	#DIV/0!	0.163264694
<i>mir-125b-1</i>	0.085	0	#DIV/0!	0.190522882
<i>mir-345</i>	0.07	0	#DIV/0!	0.1933728
<i>mir-221</i>	0.055	0	#DIV/0!	0.235028425
<i>mir-1843</i>	0.055	0	#DIV/0!	0.235028425
<i>mir-4433b</i>	0.02	0	#DIV/0!	0.323636084
<i>mir-3913-1</i>	0.045	0	0	0.323636084
<i>mir-11401</i>	0.07	0	0	0.323636084
<i>mir-382</i>	0.06	0	0	0.323636084
<i>let-7e</i>	0.05	0	0	0.323636084
<i>mir-7855</i>	0.05	0	0	0.323636084
<i>mir-130b</i>	0.04	0	0	0.323636084
<i>mir-450a-2</i>	0.02	0	0	0.323636084
<i>mir-151a</i>	0.01	0	0	0.323636084
<i>mir-378i</i>	0.01	0	0	0.323636084

* $p < 0.05$

APPENDIX K: *hsa-miR-17-5p* PUTATIVE TARGET GENES

Target Gene	TargetID	Target Gene	TargetID	Target Gene	TargetID	Target Gene	TargetID
<i>A1CF</i>	29974	<i>ANKRD12</i>	23253	<i>ATG2A</i>	23130	<i>BTF3L4</i>	91408
<i>AAK1</i>	22848	<i>ANKRD13C</i>	81573	<i>ATG2B</i>	55102	<i>BTG2</i>	7832
<i>ABCA1</i>	19	<i>ANKRD27</i>	84079	<i>ATL3</i>	25923	<i>BTG3</i>	10950
<i>ABCG8</i>	64241	<i>ANKRD33B</i>	651746	<i>ATP1B3</i>	483	<i>BTN3A1</i>	11119
<i>ABHD15</i>	116236	<i>ANKRD50</i>	57182	<i>ATP2B1</i>	490	<i>BTN3A2</i>	11118
<i>ABHD18</i>	80167	<i>ANKRD52</i>	283373	<i>ATP5B</i>	506	<i>BTN3A3</i>	10384
<i>ABHD2</i>	11057	<i>ANKS4B</i>	257629	<i>ATP6</i>	4508	<i>BZW1</i>	9689
<i>ABI2</i>	10152	<i>AP1G1</i>	164	<i>ATP6V0E1</i>	8992	<i>C11orf54</i>	28970
<i>ACADSB</i>	36	<i>APEX1</i>	328	<i>ATRX</i>	546	<i>C12orf65</i>	91574
<i>ACAP2</i>	23527	<i>APOH</i>	350	<i>ATXN1</i>	6310	<i>C14orf119</i>	55017
<i>ACBD5</i>	91452	<i>APP</i>	351	<i>ATXN7</i>	6314	<i>C14orf28</i>	122525
<i>ACER2</i>	340485	<i>ARAP2</i>	116984	<i>ATXN7L3B</i>	552889	<i>C15orf40</i>	123207
<i>ACOT2</i>	10965	<i>ARCN1</i>	372	<i>AZIN1</i>	51582	<i>C15orf41</i>	84529
<i>ACOT9</i>	23597	<i>ARHGAP1</i>	392	<i>B2M</i>	567	<i>C16orf52</i>	730094
<i>ACOX1</i>	51	<i>ARHGAP10</i>	79658	<i>BACE1</i>	23621	<i>C16orf70</i>	80262
<i>ACSL4</i>	2182	<i>ARHGAP12</i>	94134	<i>BAGE5</i>	85316	<i>C17orf75</i>	64149
<i>ACTR2</i>	10097	<i>ARHGAP35</i>	2909	<i>BAZ2A</i>	11176	<i>C18orf32</i>	497661
<i>ACVR1B</i>	91	<i>ARHGAP5</i>	394	<i>BBX</i>	56987	<i>C1orf50</i>	79078
<i>ADAR</i>	103	<i>ARHGEF18</i>	23370	<i>BCAS4</i>	55653	<i>C2orf69</i>	205327
<i>ADARB1</i>	104	<i>ARHGEF7</i>	8874	<i>BCL2</i>	596	<i>C3orf38</i>	285237
<i>ADAT2</i>	134637	<i>ARID4B</i>	51742	<i>BCL2L11</i>	10018	<i>C6orf120</i>	387263
<i>ADD1</i>	118	<i>ARIH1</i>	25820	<i>BCL2L2</i>	599	<i>C7orf43</i>	55262
<i>AFF1</i>	4299	<i>ARL1</i>	400	<i>BHMT2</i>	23743	<i>C9orf40</i>	55071
<i>AGFG2</i>	3268	<i>ARL9</i>	132946	<i>BICD2</i>	23299	<i>CABLES1</i>	91768
<i>AGMAT</i>	79814	<i>ARMT1</i>	79624	<i>BLOC1S3</i>	388552	<i>CADM2</i>	253559
<i>AGO1</i>	26523	<i>ARPC2</i>	10109	<i>BLVRA</i>	644	<i>CAMK2N2</i>	94032
<i>AGO3</i>	192669	<i>ARSJ</i>	79642	<i>BMP2</i>	650	<i>CAMTA1</i>	23261
<i>AK1</i>	203	<i>ASB1</i>	51665	<i>BMP8B</i>	656	<i>CANX</i>	821
<i>AKAP11</i>	11215	<i>ASB16</i>	92591	<i>BMPR2</i>	659	<i>CAP1</i>	10487
<i>AKR7A2</i>	8574	<i>ASH1L</i>	55870	<i>BMT2</i>	154743	<i>CAPN15</i>	6650
<i>AKTIP</i>	64400	<i>ASNS</i>	440	<i>BNIP2</i>	663	<i>CAPRIN2</i>	65981
<i>ALDH9A1</i>	223	<i>ATAD2</i>	29028	<i>BRCA2</i>	675	<i>CAPZA2</i>	830
<i>AMD1</i>	262	<i>ATAT1</i>	79969	<i>BRI3BP</i>	140707	<i>CASP2</i>	835
<i>ANKFY1</i>	51479	<i>ATF3</i>	467	<i>BRMS1L</i>	84312	<i>CAV1</i>	857
<i>ANKH</i>	56172	<i>ATG14</i>	22863	<i>BSCL2</i>	26580	<i>CAVIN1</i>	284119
<i>ANKIB1</i>	54467	<i>ATG16L1</i>	55054	<i>BTBD7</i>	55727	<i>CBL</i>	867

Target Gene	TargetID	Target Gene	TargetID	Target Gene	TargetID	Target Gene	TargetID
CBX1	10951	CHSY1	22856	CRY2	1408	DNM1L	10059
CBX5	23468	CHTF8	54921	CSDE1	7812	DNMBP	23268
CBX8	57332	CHURC1	91612	CSNK1A1	1452	DNMT1	1786
CCDC125	202243	CIT	11113	CTSA	5476	DNTTIP2	30836
CCDC137	339230	CKAP2	26586	CTSS	1520	DPP9	91039
CCDC198	55195	CLEC12B	387837	CXorf38	159013	DPYSL2	1808
CCDC30	728621	CLIC4	25932	CYB5A	1528	DRAXIN	374946
CCDC47	57003	CLIP4	79745	CYBRD1	79901	DSPP	1834
CCDC6	8030	CLOCK	9575	CYCS	54205	DSTYK	25778
CCDC71L	168455	CLPTM1	1209	CYLD	1540	DUSP18	150290
CCL1	6346	CLU	1191	CYP7B1	9420	DUSP2	1844
CCL5	6352	CMPK1	51727	DAPK3	1613	DYNC1LI2	1783
CCND1	595	CMTR2	55783	DCAF1	9730	DYRK2	8445
CCND2	894	CNEP1R1	255919	DCAF8	50717	E2F1	1869
CCP110	9738	CNKSR3	154043	DCBLD2	131566	E2F2	1870
CCSER2	54462	CNOT4	4850	DCTN5	84516	E2F3	1871
CCT6A	908	CNOT6L	246175	DCTN6	10671	E2F5	1875
CD28	940	CNOT7	29883	DCTPP1	79077	EARS2	124454
CD47	961	COA1	55744	DCUN1D4	23142	EEA1	8411
CDIPT	10423	COIL	8161	DDHD1	80821	EEF1A1	1915
CDKN1A	1026	COPS3	8533	DDI2	84301	EFCAB11	90141
CDKN2AIPNL	91368	COX19	90639	DDX5	1655	EFCAB14	9813
CENPQ	55166	COX2	4513	DEGS1	8560	EFHC1	114327
CEP104	9731	COX6B1	1340	DENND5B	160518	EGLN3	112399
CEP170	9859	COX7B	1349	DEPDC1	55635	EGR2	1959
CEP57	9702	CPE	1363	DHODH	1723	EHMT2	10919
CEP72	55722	CPOX	1371	DHX33	56919	EIF2B2	8892
CEP97	79598	CPS1	1373	DIS3L	115752	EIF2S1	1965
CERCAM	51148	CPSF1	29894	DNAJB13	374407	EIF4A2	1974
CETN2	1069	CPT1A	1374	DNAJB4	11080	EIF4G2	1982
CFL2	1073	CRCP	27297	DNAJB6	10049	EIF4G3	8672
CHAF1A	10036	CREB1	1385	DNAJB9	4189	EIF4H	7458
CHD4	1108	CRISPLD2	83716	DNAJC10	54431	EIF5A2	56648
CHD9	80205	CRK	1398	DNAJC27	51277	ELAVL2	1993
CHIC1	53344	CROT	54677	DNAJC28	54943	ELK4	2005
CHST14	113189	CRTC3	64784	DNAL1	83544	ELMO2	63916

Target Gene	TargetID	Target Gene	TargetID	Target Gene	TargetID	Target Gene	TargetID
<i>ELMSAN1</i>	91748	<i>FAM213A</i>	84293	<i>FOXK1</i>	221937	<i>GOLGA2</i>	2801
<i>ELOC</i>	6921	<i>FAM241A</i>	132720	<i>FOXK2</i>	3607	<i>GPAM</i>	57678
<i>ELP2</i>	55250	<i>FAM46C</i>	54855	<i>FOXQ1</i>	94234	<i>GPI</i>	2821
<i>EMC1</i>	23065	<i>FAM57A</i>	79850	<i>FOXRED2</i>	80020	<i>GPM6A</i>	2823
<i>EMSY</i>	56946	<i>FAM83D</i>	81610	<i>FRMD6</i>	122786	<i>GPR137B</i>	7107
<i>ENPP5</i>	59084	<i>FAM89A</i>	375061	<i>FRS2</i>	10818	<i>GPR155</i>	151556
<i>ENTPD4</i>	9583	<i>FAM8A1</i>	51439	<i>FTH1</i>	2495	<i>GPR157</i>	80045
<i>ENTPD7</i>	57089	<i>FANCA</i>	2175	<i>FUT10</i>	84750	<i>GPR183</i>	1880
<i>EPAS1</i>	2034	<i>FAS</i>	355	<i>FXR2</i>	9513	<i>GPRIN3</i>	285513
<i>EPB41L2</i>	2037	<i>FASN</i>	2194	<i>FXYD5</i>	53827	<i>GRAMD1A</i>	57655
<i>EPB41L5</i>	57669	<i>FAXC</i>	84553	<i>FYCO1</i>	79443	<i>GRK3</i>	157
<i>EPHA4</i>	2043	<i>FBXL5</i>	26234	<i>FZD9</i>	8326	<i>GRK7</i>	131890
<i>EPS15L1</i>	58513	<i>FBXL7</i>	23194	<i>GAB1</i>	2549	<i>GRPEL2</i>	134266
<i>ERAP1</i>	51752	<i>FBXO10</i>	26267	<i>GABBR1</i>	2550	<i>GTF2H2C</i>	728340
<i>ERCC2</i>	2068	<i>FBXO21</i>	23014	<i>GABPB1</i>	2553	<i>GTF2H3</i>	2967
<i>EREG</i>	2069	<i>FBXO28</i>	23219	<i>GAK</i>	2580	<i>GTF2IRD2</i>	84163
<i>ERGIC2</i>	51290	<i>FBXO31</i>	79791	<i>GANAB</i>	23193	<i>GTF2IRD2B</i>	389524
<i>ERLIN1</i>	10613	<i>FBXO48</i>	554251	<i>GAPDH</i>	2597	<i>HAS2</i>	3037
<i>ESR2</i>	2100	<i>FCHO2</i>	115548	<i>GATA6</i>	2627	<i>HAUS8</i>	93323
<i>ETF1</i>	2107	<i>FEM1A</i>	55527	<i>GATAD1</i>	57798	<i>HBP1</i>	26959
<i>ETV1</i>	2115	<i>FEM1B</i>	10116	<i>GBF1</i>	8729	<i>HCP5</i>	10866
<i>EXO5</i>	64789	<i>FEM1C</i>	56929	<i>GBP3</i>	2635	<i>HDAC10</i>	83933
<i>EZH1</i>	2145	<i>FER</i>	2241	<i>GDAP1</i>	54332	<i>HECA</i>	51696
<i>F2R</i>	2149	<i>FEZ2</i>	9637	<i>GDF11</i>	10220	<i>HEXIM1</i>	10614
<i>F2RL1</i>	2150	<i>FGFR1OP</i>	11116	<i>GDF5OS</i>	554250	<i>HIF1A</i>	3091
<i>F2RL3</i>	9002	<i>FHDC1</i>	85462	<i>GEMIN8</i>	54960	<i>HIF1AN</i>	55662
<i>F3</i>	2152	<i>FICD</i>	11153	<i>GID4</i>	79018	<i>HIP1</i>	3092
<i>FAAP24</i>	91442	<i>FJX1</i>	24147	<i>GIGYF1</i>	64599	<i>HIST1H2AM</i>	8336
<i>FAF2</i>	23197	<i>FKBP14</i>	55033	<i>GIN54</i>	84296	<i>HIST1H2BD</i>	3017
<i>FAHD1</i>	81889	<i>FMNL2</i>	114793	<i>GLO1</i>	2739	<i>HIST1H2BG</i>	8339
<i>FAM102A</i>	399665	<i>FMNL3</i>	91010	<i>GNAS</i>	2778	<i>HIST1H2BJ</i>	8970
<i>FAM117B</i>	150864	<i>FNBP1L</i>	54874	<i>GNB1</i>	2782	<i>HIST1H4C</i>	8364
<i>FAM126B</i>	285172	<i>FOPNL</i>	123811	<i>GNB5</i>	10681	<i>HIST2H2AA3</i>	8337
<i>FAM129A</i>	116496	<i>FOXC1</i>	2296	<i>GNPTAB</i>	79158	<i>HIST2H3A</i>	333932
<i>FAM160B1</i>	57700	<i>FOXJ2</i>	55810	<i>GNS</i>	2799	<i>HIST2H4B</i>	554313
<i>FAM210A</i>	125228	<i>FOXJ3</i>	22887	<i>GOLGA1</i>	2800	<i>HIST3H2A</i>	92815

Target Gene	TargetID	Target Gene	TargetID	Target Gene	TargetID	Target Gene	TargetID
<i>HMBOX1</i>	79618	<i>ITGB1</i>	3688	<i>KLRD1</i>	3824	<i>LZIC</i>	84328
<i>HMGB1</i>	3146	<i>ITGB8</i>	3696	<i>KMT2A</i>	4297	<i>M6PR</i>	4074
<i>HMGB2</i>	3148	<i>ITPKB</i>	3707	<i>KMT2B</i>	9757	<i>MAK16</i>	84549
<i>HMGB3</i>	3149	<i>JAK1</i>	3716	<i>KMT5B</i>	51111	<i>MAN2B2</i>	23324
<i>HNRNPR</i>	10236	<i>JPT1</i>	51155	<i>KPNA2</i>	3838	<i>MANEAL</i>	149175
<i>HNRNPU</i>	3192	<i>KANSL1</i>	284058	<i>KPNA6</i>	23633	<i>MAP3K12</i>	7786
<i>HOOK3</i>	84376	<i>KAT2A</i>	2648	<i>KRT10</i>	3858	<i>MAP3K14</i>	9020
<i>HOXD11</i>	3237	<i>KAT2B</i>	8850	<i>LAMC1</i>	3915	<i>MAP3K2</i>	10746
<i>HS3ST1</i>	9957	<i>KATNAL1</i>	84056	<i>LAMTOR1</i>	55004	<i>MAP3K3</i>	4215
<i>HSP90AA1</i>	3320	<i>KCNA7</i>	3743	<i>LAPTM4A</i>	9741	<i>MAP3K8</i>	1326
<i>HSPA4L</i>	22824	<i>KCNB1</i>	3745	<i>LARP1</i>	23367	<i>MAP3K9</i>	4293
<i>HSPA8</i>	3312	<i>KCND3</i>	3752	<i>LAS1L</i>	81887	<i>MAP7</i>	9053
<i>HSPB2</i>	3316	<i>KCNJ8</i>	3764	<i>LASP1</i>	3927	<i>MAPK1</i>	5594
<i>HTT</i>	3064	<i>KCNK6</i>	9424	<i>LCOR</i>	84458	<i>MAPK14</i>	1432
<i>HUWE1</i>	10075	<i>KCNMA1</i>	3778	<i>LDHD</i>	197257	<i>MAPK9</i>	5601
<i>HYPK</i>	25764	<i>KCNMB1</i>	3779	<i>LDLR</i>	3949	<i>MAPKAPK5</i>	8550
<i>ICA1L</i>	130026	<i>KCTD7</i>	154881	<i>LEPROT</i>	54741	<i>MAPRE3</i>	22924
<i>ICAM1</i>	3383	<i>KDM4A</i>	9682	<i>LGSN</i>	51557	<i>MARCH4</i>	57574
<i>ICMT</i>	23463	<i>KDM6B</i>	23135	<i>LIAS</i>	11019	<i>MARCH6</i>	10299
<i>IER3</i>	8870	<i>KIAA0232</i>	9778	<i>LIMA1</i>	51474	<i>MASTL</i>	84930
<i>IFNAR1</i>	3454	<i>KIAA0513</i>	9764	<i>LIMK1</i>	3984	<i>MAVS</i>	57506
<i>IFNAR2</i>	3455	<i>KIAA1147</i>	57189	<i>LINC00598</i>	646982	<i>MBNL1</i>	4154
<i>IGFBP3</i>	3486	<i>KIAA1191</i>	57179	<i>LLPH</i>	84298	<i>MCC</i>	4163
<i>IGFBP5</i>	3488	<i>KIAA1551</i>	55196	<i>LPAR2</i>	9170	<i>MCL1</i>	4170
<i>ILF3</i>	3609	<i>KIAA1841</i>	84542	<i>LPGAT1</i>	9926	<i>MDK</i>	4192
<i>IMMT</i>	10989	<i>KIF1A</i>	547	<i>LPIN1</i>	23175	<i>MDM2</i>	4193
<i>INPP5F</i>	22876	<i>KIF23</i>	9493	<i>LRIF1</i>	55791	<i>MECP2</i>	4204
<i>IPP</i>	3652	<i>KIF5C</i>	3800	<i>LRP12</i>	29967	<i>MED12</i>	9968
<i>IQSEC1</i>	9922	<i>KIF6</i>	221458	<i>LRPAP1</i>	4043	<i>MED13</i>	9969
<i>IRAK1</i>	3654	<i>KLF10</i>	7071	<i>LRRC58</i>	116064	<i>MED16</i>	10025
<i>IRAK4</i>	51135	<i>KLF3</i>	51274	<i>LRRD1</i>	401387	<i>MED17</i>	9440
<i>ISCA2</i>	122961	<i>KLF6</i>	1316	<i>LSM14A</i>	26065	<i>MED18</i>	54797
<i>ISOC1</i>	51015	<i>KLHL15</i>	80311	<i>LSM3</i>	27258	<i>MEF2D</i>	4209
<i>ISY1</i>	57461	<i>KLHL20</i>	27252	<i>LUZP2</i>	338645	<i>MELK</i>	9833
<i>ITCH</i>	83737	<i>KLHL28</i>	54813	<i>LY6G5B</i>	58496	<i>MEN1</i>	4221
<i>ITGA2</i>	3673	<i>KLHL36</i>	79786	<i>LYSMD3</i>	116068	<i>METTL8</i>	79828

Target Gene	TargetID	Target Gene	TargetID	Target Gene	TargetID	Target Gene	TargetID
MFHAS1	9258	MXI1	4601	NONO	4841	PARD6B	84612
MFN1	55669	MYC	4609	NOTCH2	4853	PBXIP1	57326
MFN2	9927	MYCBP2	23077	NPAS2	4862	PCBP2	5094
MFSD2A	84879	MYH9	4627	NPAS3	64067	PCGF5	84333
MFSD4B	91749	MYLIP	29116	NPAT	4863	PCLAF	9768
MFSD8	256471	MYLK3	91807	NPNT	255743	PCMTD1	115294
MGEA5	10724	MYO1D	4642	NR2C2	7182	PCNX1	22990
MICB	4277	MYO1F	4542	NR2F6	2063	PCNX2	80003
MIDN	90007	MYPN	84665	NR3C1	2908	PDE4C	5143
MINK1	50488	N4BP1	9683	NRBP1	29959	PDGFB	5155
MINOS1-NBL1	100532736	N4BP2L2	10443	NRIP3	56675	PDHB	5162
MIXL1	83881	NAA50	80218	NSD2	7468	PDLIM5	10611
MKI67	4288	NABP1	64859	NTN4	59277	PDLIM7	9260
MKKNK2	2872	NACC2	138151	NUCKS1	64710	PDPK1	5170
MKRN1	23608	NAGK	55577	NUDT3	11165	PDRG1	81572
MLLT1	4298	NAP1L1	4673	NUFIP2	57532	PDZD11	51248
MLXIP	22877	NAPEPLD	222236	NUGGC	389643	PEA15	8682
MMP2	4313	NARS	4677	NUP35	129401	PEAK1	79834
MNT	4335	NAT8L	339983	NUP98	4928	PELI1	57162
MOB1B	92597	NBL1	4681	OCIAD1	54940	PER1	5187
MORC1	27136	NBR1	4077	OCRL	4952	PFKFB2	5208
MORF4L1	10933	NCAPD2	9918	OFD1	8481	PFKP	5214
MORF4L2	9643	NCOA3	8202	OLAH	55301	PGAM1	5223
MPPE1	65258	ND2	4536	OPTN	10133	PGM2L1	283209
MRPL40	64976	ND4	4538	ORAI1	84876	PHF6	84295
MRPS10	55173	NEK8	284086	ORAI2	80228	PHLPP1	23239
MRPS6	64968	NETO2	81831	ORMDL3	94103	PHLPP2	23035
MSH3	4437	NFAT5	10725	OSTM1	28962	PHTF2	57157
MSMO1	6307	NFATC2IP	84901	OTUD4	54726	PIGO	84720
MTF1	4520	NFIB	4781	OXR1	55074	PIGS	94005
MTMR3	8897	NHLRC3	387921	PAFAH1B1	5048	PIK3CA	5290
MTMR9	66036	NIN	51199	PAIP1	10605	PIP4K2A	5305
MTPAP	55149	NIPA1	123606	PAK6	56924	PIP4K2C	79837
MTRF1L	54516	NKIRAS1	28512	PANK3	79646	PITPNA	5306
MUC17	140453	NME6	10201	PAPD5	64282	PIWIL2	55124
MUC21	394263	NOC2L	26155	PARD3	56288	PKD2	5311

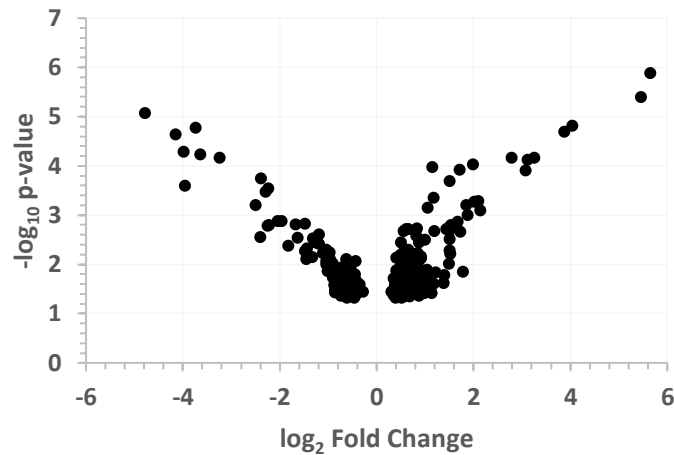
Target Gene	TargetID	Target Gene	TargetID	Target Gene	TargetID	Target Gene	TargetID
<i>PKMYT1</i>	9088	<i>PRKCB</i>	5579	<i>RAB5B</i>	5869	<i>RNF146</i>	81847
<i>PKNOX1</i>	5316	<i>PRNP</i>	5621	<i>RABEP1</i>	9135	<i>RNF19B</i>	127544
<i>PLAG1</i>	5324	<i>PRPF4</i>	9128	<i>RABGAP1L</i>	9910	<i>RNF216</i>	54476
<i>PLAGL2</i>	5326	<i>PRPF8</i>	10594	<i>RACGAP1</i>	29127	<i>RNF34</i>	80196
<i>PLEKHM1</i>	9842	<i>PRR14L</i>	253143	<i>RAD21</i>	5885	<i>RORA</i>	6095
<i>PLEKHO2</i>	80301	<i>PRRG1</i>	5638	<i>RALGAPA1</i>	253959	<i>RPA2</i>	6118
<i>PLRG1</i>	5356	<i>PRRG4</i>	79056	<i>RAN</i>	5901	<i>RPF2</i>	84154
<i>PLS1</i>	5357	<i>PSD3</i>	23362	<i>RANGAP1</i>	5905	<i>RPL14</i>	9045
<i>PLXNA1</i>	5361	<i>PTBP1</i>	5725	<i>RAP2C</i>	57826	<i>RPL17</i>	6139
<i>PMAIP1</i>	5366	<i>PTEN</i>	5728	<i>RAPGEF4</i>	11069	<i>RPL17-C18orf32</i>	100526842
<i>PNPLA4</i>	8228	<i>PTENP1</i>	11191	<i>RB1</i>	5925	<i>RPL21</i>	6144
<i>PNRC1</i>	10957	<i>PTGER4</i>	5734	<i>RBBP7</i>	5931	<i>RPL37</i>	6167
<i>POFUT1</i>	23509	<i>PTGES3</i>	10728	<i>RBL1</i>	5933	<i>RPL7</i>	6129
<i>POGK</i>	57645	<i>PTGFRN</i>	5738	<i>RBL2</i>	5934	<i>RPRD2</i>	23248
<i>POGZ</i>	23126	<i>PTGIS</i>	5740	<i>RBM12B</i>	389677	<i>RPS15A</i>	6210
<i>POLM</i>	27434	<i>PTP4A1</i>	7803	<i>RBM20</i>	282996	<i>RPS27A</i>	6233
<i>POLQ</i>	10721	<i>PTPDC1</i>	138639	<i>RBM41</i>	55285	<i>RPS6KA5</i>	9252
<i>POLR1B</i>	84172	<i>PTPN4</i>	5775	<i>RBM5</i>	10181	<i>RPSA</i>	3921
<i>POLR3A</i>	11128	<i>PTPRO</i>	5800	<i>RCCD1</i>	91433	<i>RRAGD</i>	58528
<i>POLR3F</i>	10621	<i>PTTG1</i>	9232	<i>REEP3</i>	221035	<i>RRAS2</i>	22800
<i>POLR3G</i>	10622	<i>PUDP</i>	8226	<i>REEP5</i>	7905	<i>RRM2</i>	6241
<i>PPP1CA</i>	5499	<i>PURB</i>	5814	<i>REST</i>	5978	<i>RRN3</i>	54700
<i>PPP1R12B</i>	4660	<i>PVR</i>	5817	<i>REV1</i>	51455	<i>RSRP1</i>	57035
<i>PPP1R15A</i>	23645	<i>PXK</i>	54899	<i>RFK</i>	55312	<i>RTCA</i>	8634
<i>PPP1R15B</i>	84919	<i>QARS</i>	5859	<i>RFXANK</i>	8625	<i>RTN2</i>	6253
<i>PPP1R3B</i>	79660	<i>QKI</i>	9444	<i>RFXAP</i>	5994	<i>RTN3</i>	10313
<i>PPP2R1A</i>	5518	<i>QRFPR</i>	84109	<i>RGMB</i>	285704	<i>RTTN</i>	25914
<i>PPP2R2A</i>	5520	<i>QSOX1</i>	5768	<i>RHOC</i>	389	<i>RUFY2</i>	55680
<i>PPP3R1</i>	5534	<i>RAB10</i>	10890	<i>RLF</i>	6018	<i>RUNDC1</i>	146923
<i>PPP6C</i>	5537	<i>RAB11FIP1</i>	80223	<i>RLIM</i>	51132	<i>RUNX1</i>	861
<i>PPP6R3</i>	55291	<i>RAB12</i>	201475	<i>RMDN1</i>	51115	<i>RUNX3</i>	864
<i>PRICKLE1</i>	144165	<i>RAB22A</i>	57403	<i>RMND1</i>	55005	<i>RYR2</i>	6262
<i>PRICKLE4</i>	29964	<i>RAB23</i>	51715	<i>RNASEH1</i>	246243	<i>SACS</i>	26278
<i>PRIM1</i>	5557	<i>RAB30</i>	27314	<i>RND3</i>	390	<i>SALL3</i>	27164
<i>PRKACB</i>	5567	<i>RAB3IP</i>	117177	<i>RNF115</i>	27246	<i>SAMD12</i>	401474
<i>PRKAR1A</i>	5573	<i>RAB42</i>	115273	<i>RNF145</i>	153830	<i>SAMD8</i>	142891

Target Gene	TargetID	Target Gene	TargetID	Target Gene	TargetID	Target Gene	TargetID
<i>SAMD9L</i>	219285	<i>SLC16A2</i>	6567	<i>SON</i>	6651	<i>TBC1D15</i>	64786
<i>SCAMP2</i>	10066	<i>SLC16A9</i>	220963	<i>SORCS2</i>	57537	<i>TBC1D17</i>	79735
<i>SCAMP5</i>	192683	<i>SLC1A5</i>	6510	<i>SOX4</i>	6659	<i>TBC1D2</i>	55357
<i>SCD</i>	6319	<i>SLC22A23</i>	63027	<i>SP2</i>	6668	<i>TBL1XR1</i>	79718
<i>SDHA</i>	6389	<i>SLC25A28</i>	81894	<i>SP4</i>	6671	<i>TCEA2</i>	6919
<i>SEC16A</i>	9919	<i>SLC25A3</i>	5250	<i>SPCS1</i>	28972	<i>TCEAL1</i>	9338
<i>SEC23A</i>	10484	<i>SLC25A33</i>	84275	<i>SPIB</i>	6689	<i>TCF3</i>	6929
<i>SELE</i>	6401	<i>SLC25A37</i>	51312	<i>SPOPL</i>	339745	<i>TCF4</i>	6925
<i>SEMA4B</i>	10509	<i>SLC25A44</i>	9673	<i>SPRED1</i>	161742	<i>TCF7L2</i>	6934
<i>SEMA7A</i>	8482	<i>SLC25A46</i>	91137	<i>SPTLC2</i>	9517	<i>TEFM</i>	79736
<i>SENP1</i>	29843	<i>SLC28A1</i>	9154	<i>SQSTM1</i>	8878	<i>TET3</i>	200424
<i>SEPT11</i>	55752	<i>SLC30A1</i>	7779	<i>SRCAP</i>	10847	<i>TFAM</i>	7019
<i>SEPT2</i>	4735	<i>SLC30A7</i>	148867	<i>SREK1IP1</i>	285672	<i>TGFB1</i>	7040
<i>SERF1A</i>	8293	<i>SLC35E2B</i>	728661	<i>SRSF2</i>	6427	<i>TGFBR2</i>	7048
<i>SERF1B</i>	728492	<i>SLC35F5</i>	80255	<i>SSH2</i>	85464	<i>TGOLN2</i>	10618
<i>SERF2</i>	10169	<i>SLC35F6</i>	54978	<i>SSX2IP</i>	117178	<i>THBS1</i>	7057
<i>SERINC1</i>	57515	<i>SLC4A7</i>	9497	<i>STAC2</i>	342667	<i>THEM4</i>	117145
<i>SESN1</i>	27244	<i>SLC5A3</i>	6526	<i>STAT3</i>	6774	<i>TIMM17A</i>	10440
<i>SESN2</i>	83667	<i>SLC6A4</i>	6532	<i>STIL</i>	6491	<i>TIMP3</i>	7078
<i>SESN3</i>	143686	<i>SLC7A11</i>	23657	<i>STK11</i>	6794	<i>TJP1</i>	7082
<i>SF3B3</i>	23450	<i>SLCO5A1</i>	81796	<i>STK11IP</i>	114790	<i>TLR7</i>	51284
<i>SGMS1</i>	259230	<i>SLK</i>	9748	<i>STK17B</i>	9262	<i>TM4SF5</i>	9032
<i>SGPL1</i>	8879	<i>SMAD3</i>	4088	<i>STX4</i>	6810	<i>TMBIM6</i>	7009
<i>SGTB</i>	54557	<i>SMAD4</i>	4089	<i>STX6</i>	10228	<i>TMED10</i>	10972
<i>SH3BP5</i>	9467	<i>SMAD5</i>	4090	<i>SUCO</i>	51430	<i>TMEM100</i>	55273
<i>SH3GLB1</i>	51100	<i>SMAD6</i>	4091	<i>SUGP1</i>	57794	<i>TMEM123</i>	114908
<i>SH3GLB2</i>	56904	<i>SMIM13</i>	221710	<i>SUOX</i>	6821	<i>TMEM127</i>	55654
<i>SHOC2</i>	8036	<i>SMIM15</i>	643155	<i>SURF4</i>	6836	<i>TMEM131L</i>	23240
<i>SIK1</i>	150094	<i>SMOC1</i>	64093	<i>SUSD6</i>	9766	<i>TMEM133</i>	83935
<i>SIKE1</i>	80143	<i>SMURF1</i>	57154	<i>SYNDIG1</i>	79953	<i>TMEM134</i>	80194
<i>SIPA1L3</i>	23094	<i>SNAP47</i>	116841	<i>SYNJ2BP</i>	55333	<i>TMEM138</i>	51524
<i>SIRPA</i>	140885	<i>SNTB2</i>	6645	<i>SYNPO2L</i>	79933	<i>TMEM165</i>	55858
<i>SKI</i>	6497	<i>SOCS5</i>	9655	<i>TADA2B</i>	93624	<i>TMEM167A</i>	153339
<i>SKIL</i>	6498	<i>SOCS6</i>	9306	<i>TAF9B</i>	51616	<i>TMEM19</i>	55266
<i>SLAIN2</i>	57606	<i>SOCS7</i>	30837	<i>TANC1</i>	85461	<i>TMEM196</i>	256130
<i>SLC12A6</i>	9990	<i>SOD2</i>	6648	<i>TAX1BP1</i>	8887	<i>TMEM200C</i>	645369

Target Gene	TargetID	Target Gene	TargetID	Target Gene	TargetID	Target Gene	TargetID	Target Gene	Target ID
TMEM242	729515	TRIM11	81559	UGCG	7357	YES1	7525	ZNF385A	25946
TMEM245	23731	TRIM32	22954	ULK1	8408	YIPF4	84272	ZNF417	147687
TMEM267	64417	TRIM37	4591	UNK	85451	YOD1	55432	ZNF426	79088
TMEM38A	79041	TRIM44	54765	UQCRCF51	7386	YTHDC1	91746	ZNF446	55663
TMEM64	169200	TRIM65	201292	USP16	10600	YWHAZ	7534	ZNF454	285676
TMEM67	91147	TRIM71	131405	USP28	57646	ZBED1	9189	ZNF507	22847
TMEM9	252839	TRIM8	81603	USP3	9960	ZBED3	84327	ZNF514	84874
TMEM9B	56674	TRIOBP	11078	USP32	84669	ZBTB18	10472	ZNF532	55205
TMOD3	29766	TRIP10	9322	USP38	84640	ZBTB25	7597	ZNF578	147660
TMSB10	9168	TRRAP	8295	USP48	84196	ZBTB33	10009	ZNF597	146434
TMTC4	84899	TRUB1	142940	UXS1	80146	ZBTB37	84614	ZNF598	90850
TMX3	54495	TSC2	7249	VCPKMT	79609	ZBTB4	57659	ZNF652	22834
TNF	7124	TSG101	7251	VDAC1	7416	ZBTB5	9925	ZNF665	79788
TNFAIP1	7126	TSKU	25987	VEGFA	7422	ZBTB6	10773	ZNF681	148213
TNFAIP8L1	126282	TSPAN6	7105	VLDLR	7436	ZBTB7A	51341	ZNF682	91120
TNFRSF10B	8795	TSR1	55720	VPS13C	54832	ZBTB9	221504	ZNF689	115509
TNFRSF21	27242	TTC9	23508	VPS26A	9559	ZC3H12C	85463	ZNF7	7553
TNFSF12	8742	TTPAL	79183	VPS50	55610	ZC3H18	124245	ZNF70	7621
TNIP3	79931	TUBB4B	10383	VPS53	55275	ZDHHC20	253832	ZNF770	54989
TNKS2	80351	TUT1	64852	VTI1A	143187	ZFYVE21	79038	ZNF780A	284323
TNPO3	23534	TWF1	5756	WAC	51322	ZFYVE26	23503	ZNF785	146540
TNRC6A	27327	TXK	7294	WASL	8976	ZFYVE9	9372	ZNF786	136051
TNRC6B	23112	TXLNA	200081	WDR1	9948	ZIK1	284307	ZNF800	168850
TOB1	10140	TXNIP	10628	WDR37	22884	ZMAT3	64393	ZNF805	390980
TOPORS	10210	U2SURP	23350	WDR53	348793	ZMYM1	79830	ZNF93	81931
TP53	7157	UBC	7316	WDR73	84942	ZNF107	51427	ZNFX1	57169
TP53COR1	102800311	UBE2C	11065	WDR82	80335	ZNF12	7559	ZRANB1	54764
TP53INP1	94241	UBE2Q2	92912	WDR89	112840	ZNF174	7727	ZSWIM3	140831
TPK1	27010	UBE2S	27338	WDR92	116143	ZNF180	7733	ZYG11A	440590
TPM4	7171	UBE2V2	7336	WEE1	7465	ZNF202	7753		
TPRG1L	127262	UBE3C	9690	WIPF2	147179	ZNF264	9422		
TRA2B	6434	UBE4A	9354	WNK3	65267	ZNF280B	140883		
TRAF3IP2	10758	UBFD1	56061	WSB1	26118	ZNF280C	55609		
TRAP1	10131	UBOX5	22888	WWC1	23286	ZNF347	84671		
TRAPPC10	7109	UBR5	51366	XIAP	331	ZNF35	7584		
TRAPPC2	6399	UBXN2A	165324	XIRP2	129446	ZNF354B	117608		


APPENDIX L: SECOND NGS ANALYSIS – TOP 30 DIFFERENTIALLY EXPRESSED MIRNAS

Low vs High CALUX




miRNA	log2FC	-log10 p value	FDR
<i>hsa-miR-8071</i>	-5.65692542	5.857355864	0.00181518
<i>hsa-miR-4637</i>	-10.02289807	5.50156152	0.00181518
<i>hsa-miR-8055</i>	7.679561245	5.391860781	0.002650848
<i>hsa-miR-10527-5p</i>	-5.454490864	5.27787745	0.002650848
<i>hsa-miR-3689a-3p</i>	-9.293049279	5.056527507	0.002757136
<i>hsa-miR-5002-3p</i>	4.778647312	5.04660425	0.003354311
<i>hsa-miR-4697-5p</i>	10.72124883	4.797846602	0.003354311
<i>hsa-miR-7705</i>	-4.040673497	4.765343712	0.004985622
<i>hsa-miR-302a-3p</i>	3.74285097	4.686493048	0.004985622
<i>hsa-miR-3065-3p</i>	-3.878678011	4.624945104	0.005380375
<i>hsa-miR-5591-3p</i>	4.148984952	4.279899811	0.005635959
<i>hsa-miR-6847-3p</i>	3.98507015	4.255553479	0.010816583
<i>hsa-miR-3924</i>	6.424461136	4.219126597	0.010816583
<i>hsa-miR-3662</i>	3.644060798	4.158955853	0.010816583
<i>hsa-miR-135b-5p</i>	-2.783397708	4.157865962	0.010816583
<i>hsa-miR-8066</i>	3.236201798	4.152766562	0.010816583
<i>hsa-miR-3085-5p</i>	-3.262219004	4.117759002	0.010816583
<i>hsa-miR-5587-5p</i>	-3.111786809	4.014989721	0.011073224
<i>hsa-miR-7113-5p</i>	-1.99504252	3.963343921	0.013291141
<i>hsa-miR-660-3p</i>	-1.153241816	3.915514453	0.01422105
<i>hsa-miR-3126-3p</i>	-1.715895813	3.896385134	0.015083349
<i>hsa-miR-2278</i>	-3.072931447	3.733797878	0.015083349
<i>hsa-miR-590-3p</i>	2.386704688	3.682809267	0.020978763
<i>hsa-miR-3663-5p</i>	-1.507374081	3.582245802	0.022609194
<i>hsa-miR-455-3p</i>	3.951818465	3.531236507	0.027360232
<i>hsa-miR-4457</i>	2.242566278	3.462164629	0.029586646
<i>hsa-miR-410-5p</i>	2.285162653	3.344036385	0.033402363
<i>hsa-miR-4689</i>	-1.175011205	3.272629738	0.042277682
<i>hsa-miR-3921</i>	-2.096383261	3.256586897	0.048114707
<i>hsa-miR-4327</i>	-2.020480056	3.187525723	0.048261124

APPENDIX M: INGENUITY PATHWAY ANALYSES (IPA)



INGENUITY
PATHWAY ANALYSIS



Analysis Name: edgeR_L-H - 2021-05-13 09:59 AM
 Analysis Creation Date: 2021-05-13
 Build version: exported
 Content version: 62089861 (Release Date: 2021-02-17)

Experiment Metadata

Name	Value

Analysis Settings

Reference set: Ingenuity Knowledge Base (Genes Only)
 Relationship to include: Direct and Indirect
 Includes Endogenous Chemicals
 Optional Analyses: My Pathways My List

Filter Summary:
 Consider only molecules and/or relationships where
 (species = Mouse OR Rat OR Human OR Uncategorized) AND
 (confidence = Experimentally Observed) AND
 (tissues/cell lines = Smooth Muscle OR Testis OR CNS Cell Lines not otherwise specified OR Gray Matter OR Macrophages not otherwise specified OR MDA-MB-468 OR Dorsal Root Ganglion OR Swiss 3T3 cells OR Pro-B lymphocytes OR Other Fibroblast cell lines OR 293 cells OR Cell Line not otherwise specified OR Kidney OR Mast cells OR Fibroblast cell lines not otherwise specified OR Other Neuroblastoma Cell

(c) 2000-2021 QIAGEN. All rights reserved.
1
INGENUITY
PATHWAY ANALYSIS

Summary of Analysis - edgeR_L-H - 2021-05-13 09:59 AM

Lines OR Small Intestine OR MEF cells OR Thymus OR SK-MEL-5 OR HOP-92 OR Cortical neurons OR MDA-MB-435 OR Other Hepatoma Cell Lines OR 786-O OR NCI-H522 OR Epithelial cells not otherwise specified OR Other NK cells OR Blood platelets OR Microglia OR PC-3 OR B lymphocytes not otherwise specified OR Other Kidney cell lines OR Astrocytes OR Peripheral blood leukocytes not otherwise specified OR NCI-ADR-RES OR Pancreatic Cancer Cell Lines not otherwise specified OR Cerebellum OR Trachea OR RBL-2H3 OR HEL OR Other Monocytes OR Other Immune cells OR Th1 cells OR A375 OR Dermis OR PBMCs OR BDCA-1+ dendritic cells OR Other Osteosarcoma Cell Lines OR Skin OR Parietal Lobe OR Substantia Nigra OR Teratocarcinoma Cell Lines not otherwise specified OR LOX IMVI OR Cartilage Tissue OR CD4+ T-lymphocytes OR Pyramidal neurons OR Putamen OR Activated Vd2 Gamma-delta T cells OR Crypt OR BT-474 OR Naive B cells OR Other Macrophages OR Caco2 cells OR Other Leukemia Cell Lines OR Other Teratocarcinoma Cell Lines OR Other Cervical cancer cell line OR Osteoblasts OR Immune cells not otherwise specified OR RXF-393 OR EK VX OR Myeloid dendritic cells OR IGROV1 OR MOLT-4 OR SNB-75 OR K-562 OR Thyroid Gland OR Ovarian Cancer Cell Lines not otherwise specified OR Brain OR Heart OR MG-63 OR RPMI-8266 OR PC-12 cells OR LNCaP cells OR U87MG OR Bone marrow-derived macrophages OR Other Memory T lymphocytes OR Bone marrow cells not otherwise specified OR MDA-MB-231 OR Vascular smooth muscle cells OR Cerebral Ventricles OR Caudate Nucleus OR Cos-7 cells OR Hepatoma Cell Lines not otherwise specified OR Langerhans cells OR Megakaryocytes OR HepG2 OR Epidermis OR Other Melanoma Cell Lines OR Granulocytes not otherwise specified OR MDA-N OR NCI-H23 OR Cerebral Cortex OR Other Stem cells OR Vd1 Gamma-delta T cells OR KM-12 OR Other Kidney Cancer Cell Lines OR Other Endothelial cells OR Oocytes OR Olfactory Bulb OR Other Mononuclear leukocytes OR Lymphocytes not otherwise specified OR HeLa OR Pituitary Gland OR T47-D OR Other Lymphocytes OR Microvascular endothelial cells OR Cervical cancer cell line not otherwise specified OR Placenta OR Large Intestine OR Lymphoma Cell Lines not otherwise specified OR A549-ATCC OR Stomach OR Thalamus OR Natural T-regulatory cells OR Corpus Callosum OR Neuroblastoma Cell Lines not otherwise specified OR Memory B cells OR Th17 cells OR Mesenchymal stem cells OR Immune cell lines not otherwise specified OR Other Ovarian Cancer Cell Lines OR Adipose OR HOP-62 OR 3T3-L1 cells OR NB4 OR MCF7 OR Calvaria OR Dendritic cells not otherwise specified OR Colon Cancer Cell Lines not otherwise specified OR Monocyte-derived macrophage OR BA/F3 OR Other Granulocytes OR Other Neurons OR UACC-62 OR Other Dendritic cells OR Plasma cells OR NCI-H332M OR U2OS OR Adrenal Gland OR Purkinje cells OR Effector memory RA+ cytotoxic T cells OR Forestomach OR NT2/D1 OR Activated CD56bright NK cells OR CCRF-CEM OR SK-MEL-28 OR SW-480 OR Other Cells OR HuH7 OR J774 OR Other Lung Cancer Cell Lines OR Prostate Gland OR Striatum OR MDA-MB-361 OR CD56dim NK cells OR Cells not otherwise specified OR Spleen OR Central memory cytotoxic T cells OR Other Smooth muscle cells OR Breast Cancer Cell Lines not otherwise specified OR Other Myeloma Cell Lines OR PANC-1 OR Chondrocytes OR SK-OV-3 OR Neutrophils OR Central memory helper T cells OR Memory T lymphocytes not otherwise specified OR Stromal cells OR Activated helper T cells OR Osteosarcoma Cell Lines not otherwise specified OR Effector memory helper T cells OR HCT-116 OR Myeloma Cell Lines not otherwise specified OR Other Bone marrow cells OR Ovary OR Peripheral blood lymphocytes OR Melanoma Cell Lines not otherwise specified OR SR OR Peritoneal macrophages OR Keratinocytes OR MALME-3M OR

(c) 2000-2021 QIAGEN. All rights reserved.
2
INGENUITY
PATHWAY ANALYSIS

Fibroblasts OR RKO OR Other Organ Systems OR Effector memory cytotoxic T cells OR Monocytes not otherwise specified OR UACC-257 OR Pancreas OR Salivary Gland OR Kidney cell lines not otherwise specified OR Other CNS Cell Lines OR Esophagus OR Macrophage Cancer Cell Lines not otherwise specified OR Cornea OR HS 578T OR OVCAR-3 OR Trigeminal Ganglion OR CD34+ cells OR Min6 OR U937 OR INS-1 OR Pheochromocytoma cell lines not otherwise specified OR Effector T cells OR COLO205 OR OVCAR-5 OR Cytotoxic T cells OR H460 OR Kidney Cancer Cell Lines not otherwise specified OR Other Epithelial cells OR Plasmacytoid dendritic cells OR Murine NKT cells OR Nucleus Accumbens OR Th2 cells OR SF-539 OR Activated Vd1 Gamma-delta T cells OR TK-10 OR Mature monocyte-derived dendritic cells OR UO-31 OR Other Nervous System OR Cardiomyocytes OR A2780 OR Subventricular Zone OR Lung Cancer Cell Lines not otherwise specified OR SW-620 OR Peripheral blood monocytes OR P19 OR SK-N-SH OR Retina OR HCT-15 OR Other Prostate Cancer Cell Lines OR Other Pancreatic Cancer Cell Lines OR Endothelial cells not otherwise specified OR Stem cells not otherwise specified OR HT29 OR Granulosa cells OR Melanocytes OR HMC-1 OR Monocyte-derived dendritic cells not otherwise specified OR NIH/3T3 cells OR SF-268 OR T lymphocytes not otherwise specified OR Splenocytes OR Other Pheochromocytoma cell lines OR Eosinophils OR Lens OR Other Lymphoma Cell Lines OR Beta islet cells OR NK cells not otherwise specified OR J-774A.1 OR CD56bright NK cells OR BDCA-3+ dendritic cells OR Tissues and Primary Cells not otherwise specified OR Medulla Oblongata OR HCC-2998 OR SN12C OR Choroid Plexus OR Granule Cell Layer OR HUVEC cells OR Hypothalamus OR Uterus OR Brainstem OR Other B lymphocytes OR Sertoli cells OR Other Immune cell lines OR Embryonic stem cells OR Hep3B OR Prostate Cancer Cell Lines not otherwise specified OR RAW 264.7 OR Pre-B lymphocytes OR Other Cell Line OR Other Macrophage Cancer Cell Lines OR Naive helper T cells OR Thymocytes OR U266 OR Sciatic Nerve OR Other Peripheral blood leukocytes OR White Matter OR Smooth muscle cells not otherwise specified OR Hepatocytes OR Lymph node OR Organ Systems not otherwise specified OR M14 OR ACHN OR BT-549 OR WEHI-231 OR Neurons not otherwise specified OR Other Colon Cancer Cell Lines OR Bone marrow-derived dendritic cells OR Lung OR Skeletal Muscle OR SF-295 OR SK-MEL-2 OR A498 OR Intraepithelial T lymphocytes OR Other Tissues and Primary Cells OR Adipocytes OR Jurkat OR HL-60 OR Hematopoietic progenitor cells OR NCI-H226 OR U251 OR Other Monocyte-derived dendritic cells OR Vd2 Gamma-delta T cells OR Ventricular Zone OR Hippocampus OR Immature monocyte-derived dendritic cells OR OVCAR-4 OR Other T lymphocytes OR OVCAR-8 OR Activated CD56dim NK cells OR Nervous System not otherwise specified OR Spinal Cord OR CAKI-1 OR Amygdala OR Mononuclear leukocytes not otherwise specified OR Leukemia Cell Lines not otherwise specified OR Bladder OR Mammary Gland OR Liver OR DU-145 OR Granule cells OR Other Breast Cancer Cell Lines OR THP-1) AND (mol. types = biologic drug OR canonical pathway OR chemical - endogenous mammalian OR chemical - endogenous non-mammalian OR chemical - kinase inhibitor OR chemical - other OR chemical - protease inhibitor OR chemical drug OR chemical reagent OR chemical toxicant OR complex OR cytokine OR disease OR enzyme OR function OR fusion gene/product OR G-protein coupled receptor OR group OR growth factor OR ion channel OR kinase OR ligand-dependent nuclear receptor OR mature microRNA OR microRNA OR other OR peptidase OR phosphatase OR transcription regulator OR translation regulator OR transmembrane receptor OR transporter) AND

(data sources = An Open Access Database of Genome-wide Association Results OR BIND OR BioGRID OR Catalogue Of Somatic Mutations In Cancer (COSMIC) OR Chemical Carcinogenesis Research Information System (CCRIS) OR ClinicalTrials.gov OR ClinVar OR Cognia OR DIP OR DrugBank OR Gene Ontology (GO) OR GVK Biosciences OR Hazardous Substances Data Bank (HSDB) OR HumanCyc OR Ingenuity Expert Findings OR Ingenuity ExpertAssist Findings OR IntAct OR Interactome studies OR MIPS OR miRBase OR miRecords OR Mouse Genome Database (MGD) OR Obesity Gene Map Database OR Online Mendelian Inheritance in Man (OMIM) OR TarBase OR TargetScan Human)

Top Canonical Pathways

Top Upstream Regulators

Top Diseases and Bio Functions

Diseases and Disorders

Name	p-value range	# Molecules
Organismal Injury and Abnormalities	4.87E-02 - 3.32E-14	21
Reproductive System Disease	4.87E-02 - 3.32E-14	14
Cancer	4.19E-02 - 3.10E-06	10
Connective Tissue Disorders	3.44E-02 - 3.10E-06	5
Gastrointestinal Disease	4.19E-02 - 7.31E-05	5

Molecular and Cellular Functions

Name	p-value range	# Molecules
Cellular Movement	2.85E-03 - 1.18E-05	5
Cellular Development	4.19E-02 - 4.89E-04	6
Cellular Growth and Proliferation	4.19E-02 - 4.89E-04	6
Cell Death and Survival	3.36E-02 - 5.32E-04	2
Cell Cycle	3.50E-02 - 6.75E-04	4

Physiological System Development and Function

Name	p-value range	# Molecules
Digestive System Development and Function	7.31E-05 - 7.31E-05	3
Hepatic System Development and Function	7.31E-05 - 7.31E-05	3
Organ Development	3.91E-02 - 7.31E-05	4
Connective Tissue Development and Function	8.92E-04 - 8.92E-04	2
Skeletal and Muscular System Development and Function	2.53E-02 - 8.92E-04	3

Top Tox Functions**Cardiotoxicity**

Name	p-value range	# Molecules
Cardiac Regeneration	1.98E-02 - 1.98E-02	1
Cardiac Proliferation	2.53E-02 - 2.53E-02	1

(c) 2000-2021 QIAGEN. All rights reserved.

5

INGENUITY
PATHWAY ANALYSIS

Cardiac Dilation	1.35E-01 - 1.35E-01	2
Cardiac Enlargement	1.35E-01 - 1.35E-01	2

Hepatotoxicity

Name	p-value range	# Molecules
Liver Inflammation/Hepatitis	7.31E-05 - 7.31E-05	3
Liver Cirrhosis	5.82E-03 - 1.61E-03	4
Liver Fibrosis	5.82E-03 - 5.82E-03	3
Hepatocellular carcinoma	5.71E-01 - 5.71E-01	3
Liver Hyperplasia/Hyperproliferation	1.00E00 - 5.71E-01	4

Nephrotoxicity

Name	p-value range	# Molecules
Glomerular Injury	1.23E-02 - 1.23E-02	2
Renal Inflammation	1.23E-02 - 1.23E-02	2
Renal Nephritis	1.23E-02 - 1.23E-02	2

Top Regulator Effect Networks**Top Networks**

(c) 2000-2021 QIAGEN. All rights reserved.

6

INGENUITY
PATHWAY ANALYSIS

ID	Associated Network Functions	Score
1	Organismal Injury and Abnormalities, Reproductive System Disease, Gene Expression	21
2	Organismal Injury and Abnormalities, Reproductive System Disease, Gene Expression	19
3	Gene Expression, Organismal Injury and Abnormalities, Reproductive System Disease	15
4	Cellular Movement, Cancer, Organismal Injury and Abnormalities	11
5		4

Top Tox Lists

Top My Lists

Top My Pathways

Top Analysis-Ready Molecules

(c) 2000-2021 QIAGEN. All rights reserved.

7

INGENUITY
PATHWAY ANALYSIS

Expr Log Ratio

Molecules	Expr. Value	Chart
miR-1237-5p (and other miRNAs w/seed GGGGCG)*	↑ 10.721	
miR-8055 (miRNAs w/seed UUUGAGC)	↑ 7.680	
miR-3924 (miRNAs w/seed UAUGUUAU)	↑ 6.424	
miR-5002-3p (miRNAs w/seed GACUGCC)	↑ 4.779	
miR-5591-3p (miRNAs w/seed UACCCAU)	↑ 4.149	
miR-6847-3p (miRNAs w/seed GCUCAUG)	↑ 3.985	
miR-455-3p (miRNAs w/seed CAGUCCA)	↑ 3.952	
miR-291a-3p (and other miRNAs w/seed AAGUGCU)*	↑ 3.743	
miR-3662 (miRNAs w/seed AAAAUGA)	↑ 3.644	
miR-8066 (miRNAs w/seed AAUGUGA)	↑ 3.236	

Expr Log Ratio

Molecules	Expr. Value	Chart
miR-4637 (miRNAs w/seed ACUAAACU)	↓ -10.023	
miR-1273h-5p (and other miRNAs w/seed UGGGAGG)*	↓ -9.293	
miR-8071 (and other miRNAs w/seed GGUGGAC)	↓ -5.657	
miR-10527-5p (miRNAs w/seed AAGCAAA)	↓ -5.454	
miR-7705 (miRNAs w/seed AUAGCUC)	↓ -4.041	
miR-3065-3p (miRNAs w/seed CAGCACC)	↓ -3.879	
miR-3085-5p (miRNAs w/seed GGUGCCA)	↓ -3.262	
miR-3073a-5p (and other miRNAs w/seed UGGUCAC)	↓ -3.112	
miR-2278 (miRNAs w/seed AGAGCAG)	↓ -3.073	

(c) 2000-2021 QIAGEN. All rights reserved.

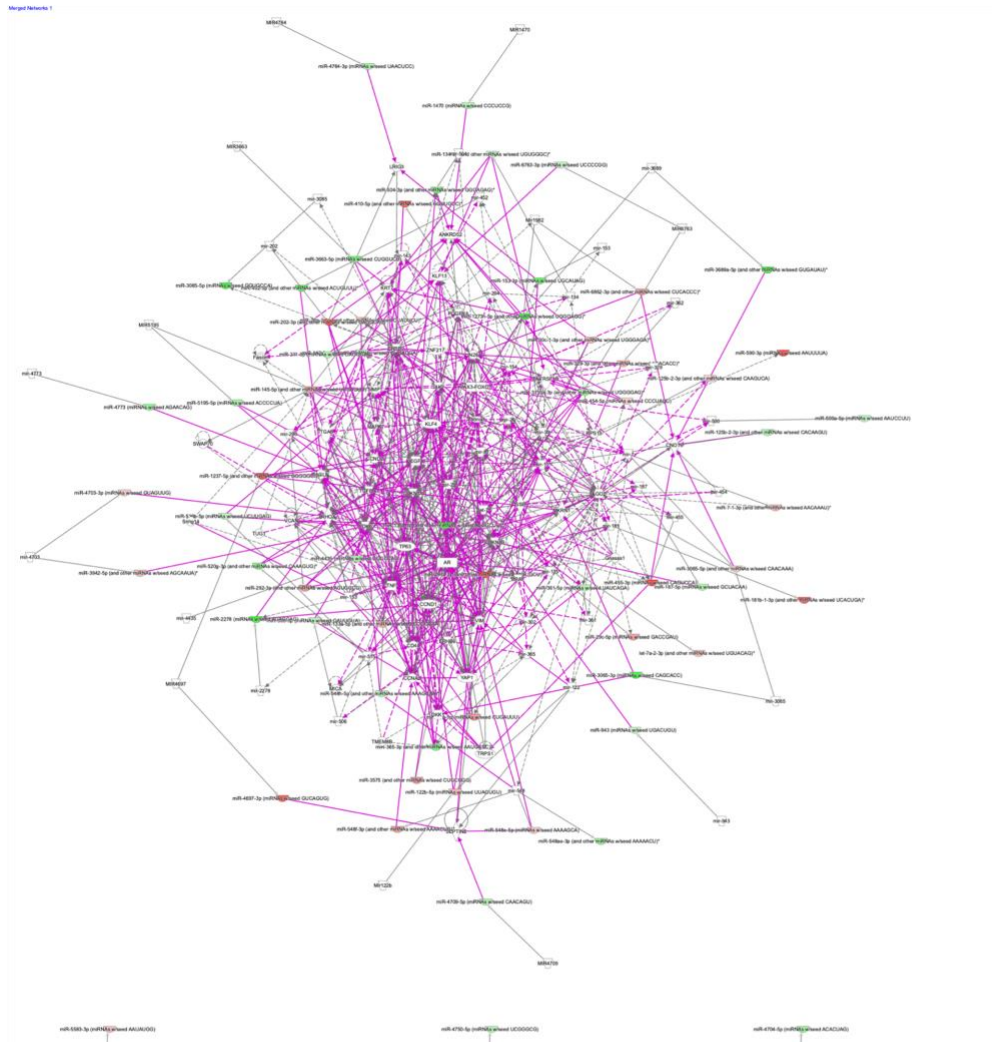
8

INGENUITY
PATHWAY ANALYSIS

miR-135a-5p (and other miRNAs w/seed AUGGCUU)*

↓ -2.783

APPENDIX N: INGENUITY PATHWAY ANALYSIS NETWORK



APPENDIX O: INSTITUTIONAL REVIEW BOARD LETTER




Institutional Review Board
301 University Blvd.
Galveston, TX 77555-0158

19-Sep-2017

MEMORANDUM

TO: Cornelis Elferink
Pharmacology 146400

FROM: 
Dwight Wolf, MD
Chairman, IRB #2

RE: Final Approval of Continuing Review

IRB #: IRB # 11-194

TITLE: Gulf Coast Health Alliance: Risks Associated with the Macondo Spill
(GC-HARMS)

DOCUMENTS: 1. Research Protocol

The UTMB Institutional Review Board (IRB) reviewed the above-referenced research protocol via an expedited review procedure on **02-Aug-2017** in accordance with 45 CFR 46.110(a)-(b)(1). Having met all applicable requirements, the research protocol; is approved for continuation for a period of 12 months. The approval period for this research protocol begins on **19-Sep-2017** and lasts until **02-Aug-2018**.

The research protocol cannot continue beyond the approval period without continuing review and approval by the IRB. In order to avoid a lapse in IRB approval, the Principal Investigator must apply for continuing review of the protocol and related documents before the expiration date. A reminder will be sent to you approximately 90 days prior to the expiration date.

If you have any questions, please do not hesitate to contact the IRB office via email at IRB@utmb.edu.

Comments: It is the understanding of the IRB that enrollment has ended on this project; therefore, written informed consent is no longer needed. If however, you intend to begin enrolling subjects

again or need to re-consent subjects already enrolled, the latest version of the consent form will have to be resubmitted for IRB review and approval.

1. Removing Ghulam Ansari, Joseph Shearer, Missy Huyen Bui, and Wendy Nguyen. (All other personnel were removed previously)
2. The addition of Bret Howrey and Dianne Yep.

General Instructions

To maintain IRB approval in good standing, please observe the following requirements:

1. Obtain prior IRB approval for any modifications including addition of new recruiting materials, changes in research personnel or site location, sponsor amendments or other changes to the protocol or associated documents. Only those changes that are necessary to avoid an immediate apparent hazard to a subject may be implemented without prior IRB approval.
2. Report all adverse events, protocol violations, DSMB reports, external reports and study closures promptly to the IRB.
3. Make study records available for inspection. All research-related records and documentation may be inspected by the IRB for the purpose of ensuring compliance with UTMB policies and procedures and federal regulations governing the protection of human subjects. The IRB has authority to suspend or terminate its approval if applicable requirements are not strictly adhered to by all research study personnel.

REFERENCES

1. Summerhayes, C. *Deep Water – The Gulf Oil Disaster and the Future of Offshore Drilling. Underwater Technology* **30**, (2011).
2. Kolian, S. R. *et al.* Oil in the Gulf of Mexico after the capping of the BP/Deepwater Horizon Mississippi Canyon (MC-252) well. *Environ. Sci. Pollut. Res.* **22**, 12073–12082 (2015).
3. McNutt, M. K. *et al.* Applications of science and engineering to quantify and control the Deepwater Horizon oil spill. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 20222–20228 (2012).
4. Environmental Response Management Application. *National Oceanic and Atmospheric Administration* (2014). Available at: <http://response.restoration.noaa.gov/erma/>. (Accessed: 15th July 2021)
5. Croisant, S. A. *et al.* The gulf coast health Alliance: health risks related to the macondo spill (GC-HARMS) study: Self-reported health effects. *Int. J. Environ. Res. Public Health* **14**, 1–18 (2017).
6. Adcroft, A. *et al.* Simulations of underwater plumes of dissolved oil in the Gulf of Mexico. *Geophys. Res. Lett.* **37**, (2010).
7. Graham, L. *et al.* The Deepwater Horizon Oil Spill’s Impact on Bottlenose Dolphins. *Gulf Mex. Res. Inst.* 1–8 (2010).
8. Wickliffe, J. *et al.* Evaluation of polycyclic aromatic hydrocarbons using analytical methods, toxicology, and risk assessment research: Seafood safety after a petroleum spill as an example. *Environ. Health Perspect.* **122**, 6–9 (2014).
9. Jackson, D. *et al.* Using Precision Environmental Health Principles in Risk Evaluation and Communication of the Deepwater Horizon Oil Spill. *New Solut.* **28**, 599–616 (2019).
10. Sumaila, U. R. *et al.* Impact of the deepwater horizon well blowout on the economics of US gulf fisheries. *Can. J. Fish. Aquat. Sci.* **69**, 499–510 (2012).
11. Horizon, D. Oil budget calculator deepwater Horizon. *Fate Oil from Deep. Horiz. Spill* 21–186 (2011).
12. Rico-Martínez, R., Snell, T. W. & Shearer, T. L. Synergistic toxicity of Macondo crude oil and dispersant Corexit 9500A ® to the *Brachionus plicatilis* species complex (Rotifera). *Environ. Pollut.* **173**, 5–10 (2013).
13. Upton, H. F. The deepwater horizon oil spill and the gulf of Mexico fishing industry. *Impacts Gulf Oil Spill Fish. Wildl.* 1–21 (2011).
14. Ylitalo, G. M. *et al.* Federal seafood safety response to the Deepwater Horizon oil spill. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 20274–20279 (2012).

15. Wang, Z., Stout, S. A. & Fingas, M. Forensic fingerprinting of biomarkers for oil spill characterization and source identification. *Environ. Forensics* **7**, 105–146 (2006).
16. Patel, A. B., Shaikh, S., Jain, K. R., Desai, C. & Madamwar, D. Polycyclic Aromatic Hydrocarbons: Sources, Toxicity, and Remediation Approaches. *Front. Microbiol.* **11**, 1–23 (2020).
17. Abdel-shafy, H. I. & Mansour, M. S. M. REVIEW A review on polycyclic aromatic hydrocarbons : Source , environmental impact , effect on human health and remediation. *Egypt. J. Pet.* **25**, 107–123 (2016).
18. Srogi, K. Monitoring of environmental exposure to polycyclic aromatic hydrocarbons: a review. *Environ. Chem. Lett.* **5**, 169–195 (2007).
19. Mojiri, A., Zhou, J. L., Ohashi, A., Ozaki, N. & Kindaichi, T. Comprehensive review of polycyclic aromatic hydrocarbons in water sources, their effects and treatments. *Sci. Total Environ.* **696**, 1–16 (2019).
20. Murray, J. R. & Penning, T. M. Carcinogenic Polycyclic Aromatic Hydrocarbons. *Compr. Toxicol.* 87–153 (2018). doi:10.1016/B978-0-12-801238-3.95691-5
21. Quoc, A., Suzuki, G., Michinaka, C., Minh, N. & Huu, L. Characterization of unsubstituted and methylated polycyclic aromatic hydrocarbons in settled dust : Combination of instrumental analysis and in vitro reporter gene assays and implications for cancer risk assessment. *Sci. Total Environ.* **788**, 1–10 (2021).
22. Munoz, B. & Albores, A. DNA Damage Caused by Polycyclic Aromatic Hydrocarbons: Mechanisms and Markers. *Sel. Top. DNA Repair* (2011). doi:10.5772/22527
23. Wolska, L., Mechlińska, A., Rogowska, J. & Namieśnik, J. Sources and fate of PAHs and PCBs in the marine environment. *Crit. Rev. Environ. Sci. Technol.* **42**, 1172–1189 (2012).
24. Saha, M., Takada, H. & Bhattacharya, B. Establishing Criteria of Relative Abundance of Alkyl Polycyclic Aromatic Hydrocarbons (PAHs) for Differentiation of Pyrogenic and Petrogenic PAHs: An Application to Indian Sediment. *Environ. Forensics* **13**, 312–331 (2012).
25. Boehm, P. D., Pietari, J., Cook, L. L. & Saba, T. Improving rigor in polycyclic aromatic hydrocarbon source fingerprinting. *Environ. Forensics* **19**, 172–184 (2018).
26. Boehm, P. & Saba, T. Environmental forensics. *Environ. Forensics* **4**, 1–5 (2008).
27. Wang, Z. *et al.* Quantitative characterization of PAHs in burn residue and soot samples and differentiation of pyrogenic PAH1 from petrogenic PAHs - The 1994 mobile burn study. *Environ. Sci. Technol.* **33**, 3100–3109 (1999).
28. Ramesh, A. *et al.* Bioavailability and risk assessment of orally ingested polycyclic aromatic hydrocarbons. *International Journal of Toxicology* **23**, (2004).
29. Marvanová, S. *et al.* Toxic effects of methylated benz[a]anthracenes in liver cells.

- Chem. Res. Toxicol.* **21**, 503–512 (2008).
30. Hýzdalová, M. *et al.* Aryl Hydrocarbon Receptor-Dependent Metabolism Plays a Significant Role in Estrogen-Like Effects of Polycyclic Aromatic Hydrocarbons on Cell Proliferation. *Toxicol. Sci.* **165**, 447–461 (2018).
 31. Gastelum, G. *et al.* Polycyclic aromatic hydrocarbon-induced pulmonary carcinogenesis in cytochrome P450 (CYP)1A1- And 1A2-null mice: Roles of CYP1A1 and CYP1A2. *Toxicol. Sci.* **177**, 347–361 (2020).
 32. Nguyen, N. T. *et al.* Aryl hydrocarbon receptor and kynurenine: Recent advances in autoimmune disease research. *Front. Immunol.* **5**, 551 (2014).
 33. Kaiser, H., Parker, E. & Hamrick, M. W. Kynurenine signaling through the aryl hydrocarbon receptor: Implications for aging and healthspan. *Exp. Gerontol.* **130**, 1–5 (2020).
 34. Mulero-Navarro, S. & Fernandez-Salguero, P. M. New trends in Aryl hydrocarbon receptor biology. *Front. Cell Dev. Biol.* **4**, 45 (2016).
 35. Wright, E. J., Pereira De Castro, K., Joshi, A. D. & Elferink, C. J. Canonical and non-canonical aryl hydrocarbon receptor signaling pathways. *Curr. Opin. Toxicol.* **2**, 87–92 (2017).
 36. Jackson, D. P., Joshi, A. D. & Elferink, C. J. Ah receptor pathway intricacies; signaling through diverse protein partners and DNA-motifs. *Toxicol. Res. (Camb)*. **4**, 1143–1158 (2015).
 37. Beischlag, T. V., Morales, J. L., Hollingshead, B. D. & Perdew, G. H. The aryl hydrocarbon receptor complex and the control of gene expression. *Crit. Rev. Eukaryot. Gene Expr.* **18**, 207–250 (2008).
 38. Zhang, L., Jin, Y., Huang, M. & Penning, T. M. The role of human aldo-keto reductases in the metabolic activation and detoxication of polycyclic aromatic hydrocarbons: Interconversion of PAH catechols and PAH o-quinones. *Front. Pharmacol.* **3 NOV**, 1–12 (2012).
 39. Pott, P. Chirurgical Observations Relative to... Cancer of the Scrotum.. *CA: A Cancer Journal for Clinicians* **24**, 110–116 (1974).
 40. Benmoussa, N., Rebibo, J.-D., Conan, P. & Charlier, P. Chimney-sweeps' cancer—early proof of environmentally driven tumourigenicity. *Lancet Oncol.* **20**, 338 (2019).
 41. Yamagiwa, K. & Ichikawa, K. Experimental study of the pathogenesis of carcinoma. *J. Cancer Res.* **3**, 155–161 (1917).
 42. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Chemical agents and related occupations. *IARC Monogr. Eval. Carcinog. Risks Hum.* **100**, 9–562 (2012).
 43. Arlt, V. M. *et al.* Metabolic activation of benzo[a]pyrene in vitro by hepatic cytochrome P450 contrasts with detoxification in vivo: Experiments with hepatic cytochrome P450 reductase null mice. *Carcinogenesis* **29**, 656–665 (2008).

44. Ewa, B. & Danuta, M. Š. Polycyclic aromatic hydrocarbons and PAH-related DNA adducts. *J. Appl. Genet.* **58**, 321–330 (2017).
45. Banks, L. D. *et al.* Polycyclic aromatic hydrocarbons. *Biomarkers Toxicol.* 451–458 (2014). doi:10.1016/B978-0-12-404630-6.00026-9
46. Kim, K.-H., Jahan, S. A., Kabir, E. & Brown, R. J. C. A review of airborne polycyclic aromatic hydrocarbons (PAHs) and their human health effects. *Environ. Int.* **60**, 71–80 (2013).
47. Rengarajan, T. *et al.* Exposure to polycyclic aromatic hydrocarbons with special focus on cancer. *Asian Pacific Journal of Tropical Biomedicine* **5**, (2015).
48. Zhang, Y., Dong, S., Wang, H., Tao, S. & Kiyama, R. Biological impact of environmental polycyclic aromatic hydrocarbons (ePAHs) as endocrine disruptors. *Environ. Pollut.* **213**, 809–824 (2016).
49. Pfeifer, G. P. & Hainaut, P. On the origin of G → T transversions in lung cancer. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **526**, 39–43 (2003).
50. Lopes, H. & Proença, S. Insights into PCDD/Fs and PAHs in biomass boilers envisaging risks of ash use as fertilizers. *Appl. Sci.* **10**, 1–32 (2020).
51. Bláha, L., Kapplová, P., Vondráček, J., Upham, B. & Machala, M. Inhibition of gap-junctional intercellular communication by environmentally occurring polycyclic aromatic hydrocarbons. *Toxicol. Sci.* **65**, 43–51 (2002).
52. Rummel, A. M., Trosko, J. E., Wilson, M. R. & Upham, B. L. Polycyclic aromatic hydrocarbons with bay-like regions inhibited gap junctional intercellular communication and stimulated MAPK activity. *Toxicol. Sci.* **49**, 232–240 (1999).
53. Vondráček, J., Kozubik, A. & Machala, M. Modulation of estrogen receptor-dependent reporter construct activation and G0/G1-S-phase transition by polycyclic aromatic hydrocarbons in human breast carcinoma MCF-7 cells. *Toxicol. Sci.* **70**, 193–201 (2002).
54. Santodonato, J. REVIEW OF THE ESTROGENIC AND ANTIESTROGENIC ACTIVITY OF POLYCYCLIC AROMATIC HYDROCARBONS: RELATIONSHIP TO CARCINOGENICITY. *Chemosphere* **34**, 835–848 (1997).
55. Bolden, A. L., Rochester, J. R., Schultz, K. & Kwiatkowski, C. F. Polycyclic aromatic hydrocarbons and female reproductive health: A scoping review. *Reprod. Toxicol.* **73**, 61–74 (2017).
56. Kamelia, L., de Haan, L., Ketelslegers, H. B., Rietjens, I. M. C. M. & Boogaard, P. J. In vitro prenatal developmental toxicity induced by some petroleum substances is mediated by their 3- to 7-ring PAH constituent with a potential role for the aryl hydrocarbon receptor (AhR). *Toxicol. Lett.* **315**, 64–76 (2019).
57. Mastrangelo, G., Fadda, E. & Marzia, V. Polycyclic aromatic hydrocarbons and cancer in man. *Environmental Health Perspectives* **104**, 1166–1170 (1996).
58. Gupta, R. C. Introduction. *Biomarkers Toxicol.* 3–5 (2014). doi:10.1016/B978-0-12-404630-6.00001-4

59. Sogorb, M. A., Estévez, J. & Vilanova, E. Biomarkers in biomonitoring of xenobiotics. *Biomarkers Toxicol.* 965–973 (2014). doi:10.1016/B978-0-12-404630-6.00057-9
60. Smith, C. J. *et al.* Quantification of monohydroxy-PAH metabolites in urine by solid-phase extraction with isotope dilution-GC-MS. *Fresenius. J. Anal. Chem.* **372**, 216–220 (2002).
61. Kuang, D. *et al.* Dose-response relationships of polycyclic aromatic hydrocarbons exposure and oxidative damage to DNA and lipid in coke oven workers. *Environ. Sci. Technol.* **47**, 7446–7456 (2013).
62. Schneider, K., Roller, M., Kalberlah, F. & Schuhmacher-Wolz, U. Cancer risk assessment for oral exposure to PAH mixtures. *J. Appl. Toxicol.* **22**, 73–83 (2002).
63. Eichbaum, K. *et al.* In vitro bioassays for detecting dioxin-like activity - Application potentials and limits of detection, a review. *Sci. Total Environ.* **487**, 37–48 (2014).
64. Lawal, A. T. Polycyclic aromatic hydrocarbons. A review. *Cogent Environ. Sci.* **3**, 1–89 (2017).
65. Wise, S. A., Sander, L. C. & Schantz, M. M. Analytical Methods for Determination of Polycyclic Aromatic Hydrocarbons (PAHs) — A Historical Perspective on the 16 U.S. EPA Priority Pollutant PAHs. *Polycycl. Aromat. Compd.* **35**, 187–247 (2015).
66. My, T. T. A. *et al.* Evaluation of the dioxin-like toxicity in soil samples from Thua Thien Hue province using the AhR-CALUX bioassay – An update of Agent Orange contamination in Vietnam. *Ecotoxicol. Environ. Saf.* **212**, 1–8 (2021).
67. Deng, Q. *et al.* Polycyclic aromatic hydrocarbons-associated micrnas and their interactions with the environment: Influences on oxidative dna damage and lipid peroxidation in coke oven workers. *Environ. Sci. Technol.* **48**, 4120–4128 (2014).
68. Penman, A. D., Kaufman, G. E. & Daniels, K. K. *MicroRNA expression as an indicator of tissue toxicity. Biomarkers in Toxicology* (Elsevier Inc., 2014). doi:10.1016/B978-0-12-404630-6.00060-9
69. Deng, Q. *et al.* Plasma microRNA expression and micronuclei frequency in workers exposed to polycyclic aromatic hydrocarbons. *Environ. Health Perspect.* **122**, 719–725 (2014).
70. Beedanagari, S. R., Vulimiri, S. V., Bhatia, S. P. & Mahadevan, B. *Genotoxicity biomarkers: Molecular basis of genetic variability and susceptibility. Biomarkers in Toxicology* (Elsevier Inc., 2014). doi:10.1016/B978-0-12-404630-6.00043-9
71. Moldovan, L. *et al.* Methodological challenges in utilizing miRNAs as circulating biomarkers. *J. Cell. Mol. Med.* **18**, 371–390 (2014).
72. Wang, J., Chen, J. & Sen, S. MicroRNA as Biomarkers and Diagnostics. *Journal of Cellular Physiology* (2016). doi:10.1002/jcp.25056
73. Balasubramanian, S., Gunasekaran, K., Sasidharan, S., Jeyamanickavel Mathan, V.

- & Perumal, E. MicroRNAs and Xenobiotic Toxicity: An Overview. *Toxicol. Reports* **7**, 583–595 (2020).
74. Banerjee, A., Waters, D., Camacho, O. M. & Minet, E. Quantification of plasma microRNAs in a group of healthy smokers, ex-smokers and non-smokers and correlation to biomarkers of tobacco exposure. *Biomarkers* **20**, 123–131 (2015).
 75. Costa, L. G. & Pellacani, C. *Central nervous system toxicity biomarkers. Biomarkers in Toxicology* (Elsevier Inc., 2014). doi:10.1016/B978-0-12-404630-6.00009-9
 76. Aguilera, F., Méndez, J., Pásaroa, E. & Laffona, B. Review on the effects of exposure to spilled oils on human health. *J. Appl. Toxicol.* **30**, 291–301 (2010).
 77. Torres, J. & Bobst, S. *Toxicological Risk Assessment for Beginners. Toxicological Risk Assessment for Beginners* (2015). doi:10.1007/978-3-319-12751-4
 78. Saengtienchai, A. *et al.* Identification of interspecific differences in phase II reactions: Determination of metabolites in the urine of 16 mammalian species exposed to environmental pyrene. *Environ. Toxicol. Chem.* **33**, 2062–2069 (2014).
 79. EPA. Toxicological Review of Benzo[a]pyrene (Final Report). *US EPA* (2017).
 80. Keith, L. H. The Source of U.S. EPA's Sixteen PAH Priority Pollutants. *Polycycl. Aromat. Compd.* **35**, 147–160 (2015).
 81. Interim Report: Suspect Carcinogens in Water Supplies. *US EPA Office of Research and Development* (1975). Available at: <https://nepis.epa.gov/Exe/ZyNET.exe/20013N1N.txt?ZyActionD=ZyDocument&Client=EPA&Index=Prior to 1976&Docs=&Query=&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&UseQField=&IntQFieldOp=0&ExtQFieldOp=0>. (Accessed: 17th July 2021)
 82. Rotkin-Ellman, M., Wong, K. K. & Solomon, G. M. Seafood contamination after the BP gulf oil spill and risks to vulnerable populations: A critique of the FDA risk assessment. *Environ. Health Perspect.* **120**, 157–161 (2012).
 83. Rotkin-Ellman, M. & Solomon, G. FDA Risk Assessment of Seafood Contamination after the BP Oil Spill: Rotkin-Ellman and Solomon Respond. *Environ. Health Perspect.* **120**, (2012).
 84. Dickey, R. W. FDA risk assessment of seafood contamination after the BP oil spill. *Environ. Health Perspect.* **120**, (2012).
 85. Balcıoğlu, E. B. Potential effects of polycyclic aromatic hydrocarbons (PAHs) in marine foods on human health: a critical review. <http://dx.doi.org/10.1080/15569543.2016.1201513> **35**, 98–105 (2016).
 86. Nethery, E. *et al.* Urinary polycyclic aromatic hydrocarbons as a biomarker of exposure to PAHs in air: A pilot study among pregnant women. *J. Expo. Sci. Environ. Epidemiol.* **22**, 70–81 (2012).
 87. Radmacher, P. G., Looney, S. W. & Myers, S. R. Polycyclic aromatic

- hydrocarbons in maternal and cord blood plasma. *Polycycl. Aromat. Compd.* **30**, 113–128 (2010).
88. Jira, W., Ziegenhals, K. & Speer, K. Gas chromatography-mass spectrometry (GC-MS) method for the determination of 16 european priority polycyclic aromatic hydrocarbons in smoked meat products and edible oils. *Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess.* **25**, 704–713 (2008).
 89. Poster, D. L., Schantz, M. M., Sander, L. C. & Wise, S. A. Analysis of polycyclic aromatic hydrocarbons (PAHs) in environmental samples: A critical review of gas chromatographic (GC) methods. *Anal. Bioanal. Chem.* **386**, 859–881 (2006).
 90. Wang, S. W. *et al.* Determination of polycyclic aromatic hydrocarbons (PAHs) in cosmetic products by gas chromatography-tandem mass spectrometry. *J. Food Drug Anal.* **27**, (2019).
 91. Sakthivel, S., Balasubramanian, P., Nakamura, M., Ko, S. & Chakraborty, P. CALUX bioassay: A cost-effective rapid screening technique for screening dioxins like compounds. *Rev. Environ. Health* **31**, 149–152 (2016).
 92. Pieterse, B., Felzel, E., Winter, R., Van Der Burg, B. & Brouwer, A. PAH-CALUX, an optimized bioassay for AhR-mediated hazard identification of polycyclic aromatic hydrocarbons (PAHs) as individual compounds and in complex mixtures. *Environ. Sci. Technol.* **47**, 11651–11659 (2013).
 93. Murk, A. J. *et al.* Chemical-activated luciferase gene expression (CALUX): A novel in vitro bioassay for Ah receptor active compounds in sediments and pore water. *Fundam. Appl. Toxicol.* **33**, 149–160 (1996).
 94. US EPA. *Screening of Dioxin-like Chemical Activity in soils and sediments using the CALUX bioassay and TEQ determinations.* **SW-846**, (2014).
 95. Windal, I. *et al.* Validation and Discussion of CALUX Analysis for Marine Samples. *Organohalogen Compd.* **60**, 215–218 (2003).
 96. He, G. *et al.* Cell-based assays for identification of aryl hydrocarbon receptor (AhR) activators. *Methods Pharmacol. Toxicol.* 221–235 (2014). doi:10.1007/978-1-62703-742-6_13
 97. Han, D., Nagy, S. R. & Denison, M. S. Comparison of recombinant cell bioassays for the detection of Ah receptor agonists. *BioFactors* **20**, 11–22 (2004).
 98. Ziccardi, M. H., Gardner, I. A. & Denison, M. S. Development and modification of a recombinant cell bioassay to directly detect halogenated and polycyclic aromatic hydrocarbons in serum. *Toxicol. Sci.* **54**, 183–193 (2000).
 99. TF, B. *et al.* Validation and use of the CALUX-bioassay for the determination of dioxins and PCBs in bovine milk. *Food Addit. Contam.* **15**, 863–875 (1998).
 100. He, G. *et al.* Third-generation ah receptor-responsive luciferase reporter plasmids: Amplification of dioxin-responsive elements dramatically increases CALUX bioassay sensitivity and responsiveness. *Toxicol. Sci.* **123**, 511–522 (2011).
 101. Nisbet, I. C. T. & LaGoy, P. K. Toxic equivalency factors (TEFs) for polycyclic

- aromatic hydrocarbons (PAHs). *Regul. Toxicol. Pharmacol.* **16**, 290–300 (1992).
102. Andersson, J. T. & Achten, C. Time to Say Goodbye to the 16 EPA PAHs? Toward an Up-to-Date Use of PACs for Environmental Purposes. *Polycycl. Aromat. Compd.* **35**, 330–354 (2015).
 103. Stout, S. A., Uhler, A. D. & Boehm, P. D. Recognition of and Allocation Among Multiple Sources of PAH in Urban Sediments. *Environ. Claims J.* **13**, 141–158 (2001).
 104. Choudhuri, S. Small noncoding RNAs: Biogenesis, function, and emerging significance in toxicology. *J. Biochem. Mol. Toxicol.* **24**, 195–216 (2010).
 105. Xu, Y. *et al.* Association between serum concentrations of perfluoroalkyl substances (PFAS) and expression of serum microRNAs in a cohort highly exposed to PFAS from drinking water. *Environ. Int.* **136**, 105446 (2020).
 106. Liu, B., Li, J. & Cairns, M. J. Identifying miRNAs, targets and functions. *Brief. Bioinform.* **15**, 1–19 (2014).
 107. Hammond, S. M. An overview of microRNAs. *Advanced Drug Delivery Reviews* **87**, 3–14 (2015).
 108. Catalanotto, C., Cogoni, C. & Zardo, G. MicroRNA in control of gene expression: An overview of nuclear functions. *Int. J. Mol. Sci.* **17**, (2016).
 109. Schulte, C., Westermann, D., Blankenberg, S. & Zeller, T. Diagnostic and prognostic value of circulating microRNAs in heart failure with preserved and reduced ejection fraction. *World J. Cardiol.* **7**, 843 (2015).
 110. Bartel, D. P. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **116**, 281–297 (2004).
 111. Witwer, K. W. Circulating MicroRNA biomarker studies: Pitfalls and potential solutions. *Clin. Chem.* **61**, 56–63 (2015).
 112. Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).
 113. Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**, 855–862 (1993).
 114. Friedman, R. C., Farh, K. K. H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105 (2009).
 115. Gruszka, R. & Zakrzewska, M. The Oncogenic Relevance of miR-17-92 Cluster and Its Paralogous miR-106b-25 and miR-106a-363 Clusters in Brain Tumors. *Int. J. Mol. Sci.* **19**, (2018).
 116. Komatsu, S., Kiuchi, J., Imamura, T., Ichikawa, D. & Otsuji, E. Circulating microRNAs as a liquid biopsy: a next-generation clinical biomarker for diagnosis of gastric cancer. *J. Cancer Metastasis Treat.* **4**, 36 (2018).
 117. Urabe, F. *et al.* Large-scale circulating microRNA profiling for the liquid biopsy

- of prostate cancer. *Clin. Cancer Res.* **25**, 3016–3025 (2019).
118. Shigeyasu, K., Toden, S., Zumwalt, T. J., Okugawa, Y. & Goel, A. Emerging role of microRNAs as liquid biopsy biomarkers in gastrointestinal cancers. *Clin. Cancer Res.* **23**, 2391–2399 (2017).
 119. Hanieh, H. Aryl hydrocarbon receptor-microRNA-212 / 132 axis in human breast cancer suppresses metastasis by targeting SOX4. *Mol. Cancer* 1–13 (2015). doi:10.1186/s12943-015-0443-9
 120. Fernando, H. *et al.* Distribution of petrogenic polycyclic aromatic hydrocarbons (PAHs) in seafood following Deepwater Horizon oil spill. *Mar. Pollut. Bull.* **145**, 200–207 (2019).
 121. Ji, Z. G., Li, Z., Johnson, W. R. & Auad, G. Progress of the oil spill risk analysis (OSRA) model and its applications. *J. Mar. Sci. Eng.* **9**, 1–22 (2021).
 122. Rowe, G. T. *et al.* Polycyclic aromatic hydrocarbons (PAHs) cycling and fates in Galveston Bay, Texas, USA. *PLoS One* **15**, 1–21 (2020).
 123. Safety Consumption Calculator. Available at: <https://www.utmb.edu/scg/>. (Accessed: 31st July 2021)
 124. Munoz, B. & Albores, A. DNA Damage Caused by Polycyclic Aromatic Hydrocarbons: Mechanisms and Markers. *Sel. Top. DNA Repair* 125–144 (2011). doi:10.5772/22527
 125. Sun, K. *et al.* A review of human and animals exposure to polycyclic aromatic hydrocarbons: Health risk and adverse effects, photo-induced toxicity and regulating effect of microplastics. *Sci. Total Environ.* **773**, 145403 (2021).
 126. Samburova, V., Zielinska, B. & Khlystov, A. Do 16 polycyclic aromatic hydrocarbons represent PAH air toxicity? *Toxics* **5**, (2017).
 127. Coppock, R. W. & Dziwenka, M. M. Biomarkers of petroleum products toxicity. *Biomarkers Toxicol.* 647–654 (2014). doi:10.1016/B978-0-12-404630-6.00037-3
 128. Bak, S. M. *et al.* In vitro and in silico AHR assays for assessing the risk of heavy oil-derived polycyclic aromatic hydrocarbons in fish. *Ecotoxicol. Environ. Saf.* **181**, 214–223 (2019).
 129. GC-HARMS | The Institute for Translational Sciences | UTMB Home. Available at: <https://www.utmb.edu/its/communities/our-communities/gc-harms>. (Accessed: 31st July 2021)
 130. Promega. Luciferase Assay System: Tech Bulletin. (2015).
 131. Takeichi, M. & Okada, T. S. Roles of magnesium and calcium ions in cell-to-substrate adhesion. *Exp. Cell Res.* **74**, 51–60 (1972).
 132. Tenland, E. & Hillman, M. Adding calcium to EDTA plasma samples prior to analysis could solve the compatibility issue in commercially available ELISAs that are standardized for serum. *Clin. Chem. Lab. Med.* **51**, 145–147 (2013).
 133. Cornell, R. Cell-substrate adhesion during cell culture. An ultrastructural study. *Exp. Cell Res.* **58**, 289–295 (1969).

134. Curtis, A. S. G. Cell adhesion. *Prog. Biophys. Mol. Biol.* **27**, (1973).
135. Zhao, B., Baston, D. S., Khan, E., Sorrentino, C. & Denison, M. S. Enhancing the response of CALUX and CAFLUX cell bioassays for quantitative detection of dioxin-like compounds. *Sci. China Chem.* **53**, 1010–1016 (2010).
136. Amakura, Y. *et al.* Detection of Aryl Hydrocarbon Receptor Activation by Some Chemicals in Food Using a Reporter Gene Assay. *Foods* **5**, 15 (2016).
137. Feeney, K. A., Putker, M., Brancaccio, M. & O'Neill, J. S. In-depth Characterization of Firefly Luciferase as a Reporter of Circadian Gene Expression in Mammalian Cells. *J. Biol. Rhythms* **31**, 540–550 (2016).
138. Promega. TECHNICAL MANUAL pGL3 Luciferase Reporter Vectors. 30 (2015).
139. Allard, S. T. M. & Kopish, K. Luciferase Reporter Assays : Powerful , Adaptable Tools for Cell Biology Research. *Cell Notes* 23–26 (2008).
140. He, G. *et al.* Cell-based assays for identification of aryl hydrocarbon receptor (AhR) activators. in *Methods in Pharmacology and Toxicology* (eds. Caldwell, G. W. & Yan, Z.) 221–235 (Humana Press, 2014). doi:10.1007/978-1-62703-742-6_13
141. Murk, A. J. *et al.* *Chemical-Activated Luciferase Gene Expression (CALUX): A Novel in Vitro Bioassay for Ah Receptor Active Compounds in Sediments and Pore Water* *Chemical-Activated Luciferase Expression (CALUX): A Novel in Vitro Bioassay for Ah Receptor Active Compounds in Sediments and Pore Water. FUNDAMENTAL AND APPLIED TOXICOLOGY* **33**, (1996).
142. Invitrogen. Quant-iT™ Assays for high-throughput quantitation of DNA, RNA, and oligos. *Brochure* (2006).
143. Quant-it, M. *et al.* Quant-iT™ PicoGreen® dsDNA Reagent and Kits. 1–7 (2005). Available at: <https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FSLSG%2Fmanuals%2Fmp07581.pdf&title=UXVhbnQtaVQgUGljB0dyZWVuIGRzRE5BIFJIYWdlbnQgYW5kIEtpdHM=>. (Accessed: 20th July 2021)
144. 455 OECD/OCDE ANNEX 4 Stably Transfected Human Estrogen Receptor- α Transactivation Assay for Detection of Estrogenic Agonist and Antagonist Activity of Chemicals using the ER α CALUX cell line.
145. Croes, K. *et al.* Quantification of PCDD/Fs and dioxin-like PCBs in small amounts of human serum using the sensitive H1L7.5c1 mouse hepatoma cell line: Optimization and analysis of human serum samples from adolescents of the Flemish Environment and Health Study (FLEHS II). *Talanta* **85**, 2484–2491 (2011).
146. Diagnostics, Q. Specimen-handling, Serum, Plasma, Whole blood : Serum, Plasma, Whole blood. Available at: <http://www.questdiagnostics.com/home/physicians/testing-services/specialists/hospitals-lab-staff/specimen-handling/serum-plasma-whole-blood.html>. (Accessed: 20th July 2021)

147. Green, D., McMahon, B., Foiles, N. & Tian, L. Measurement of hemostatic factors in EDTA plasma. *Am. J. Clin. Pathol.* **130**, 811–815 (2008).
148. TOCANTINS, L. M., CARROLL, R. T. & HOLBURN, R. H. The clot accelerating effect of dilution on blood and plasma. Relation to the mechanism of coagulation of normal and hemophilic blood. *Blood* **6**, 720–739 (1951).
149. Rogers, J. M. & Denison, M. S. 2000 Rogers and Denison Recombinant Cell estrogenic chemicals.pdf. *Vitr. Mol. Toxicol.* **13**, 67–82 (2000).
150. Promega. *Bioluminescent Reporter Gene Assays*.
151. Thompson, J. F., Hayes, L. S. & Lloyd, D. B. Modulation of firefly luciferase stability and impact on studies of gene regulation. *Gene* **103**, 171–177 (1991).
152. Eggers, C., Hook, B., Lewis, S., Strayer, C. & Landreman, A. *Designing a Bioluminescent Reporter Assay: Normalization. Promega Application Notes* (2016).
153. Leitão, J. M. M. & Esteves da Silva, J. C. G. Firefly luciferase inhibition. *J. Photochem. Photobiol. B Biol.* **101**, 1–8 (2010).
154. Marques, S. M. & Esteves Da Silva, J. C. G. Firefly bioluminescence: A mechanistic approach of luciferase catalyzed reactions. *IUBMB Life* **61**, 6–17 (2009).
155. ThermoFisher. Quant-i(TM) PicoGreen (R) dsDNA Reagent and Kits. 1–7 (2008).
156. Fluorescence SpectraViewer. *Thermo Fisher Scientific* 2011 (2011). Available at: <https://www.thermofisher.com/order/fluorescence-spectraviewer#!/>. (Accessed: 26th July 2021)
157. Fan, A. M. *Biomarkers in toxicology, risk assessment, and environmental chemical regulations. Biomarkers in Toxicology* (Elsevier Inc., 2014). doi:10.1016/B978-0-12-404630-6.00064-6
158. Baird, S. J. S., Bailey, E. A. & Vorhees, D. J. Evaluating human risk from exposure to alkylated PAHs in an aquatic system. *Hum. Ecol. Risk Assess.* **13**, 322–338 (2007).
159. Yin, F., Hayworth, J. S. & Clement, T. P. A tale of two recent spills-comparison of 2014 Galveston Bay and 2010 Deepwater Horizon oil spill residues. *PLoS One* **10**, e0118098 (2015).
160. Agency, F. & Draft, I. Toxicological Review of Benzo[a]pyrene. *US EPA Off. Res. Dev.* **76**, 57033–57034 (2016).
161. Lema, C. & Cunningham, M. J. Mini review MicroRNAs and their implications in toxicological research. *Toxicol. Lett.* **198**, 100–105 (2010).
162. Lam, J. K. W., Chow, M. Y. T., Zhang, Y. & Leung, S. W. S. siRNA versus miRNA as therapeutics for gene silencing. *Mol. Ther. - Nucleic Acids* **4**, e252 (2015).
163. Glinge, C. *et al.* Stability of circulating blood-based microRNAs-Pre-Analytic methodological considerations. *PLoS One* **12**, 1–16 (2017).

164. Blondal, T. *et al.* Assessing sample and miRNA profile quality in serum and plasma or other biofluids. *Methods* **59**, 164–169 (2013).
165. Xia, K. *et al.* Polycyclic aromatic hydrocarbons (PAHs) in Mississippi seafood from areas affected by the deepwater horizon oil spill. *Environ. Sci. Technol.* **46**, 5310–5318 (2012).
166. Pallardy, R. Deepwater Horizon oil spill | Summary, Effects, Cause, Clean Up, & Facts | Britannica. *Britannica* (2021). Available at: <https://www.britannica.com/event/Deepwater-Horizon-oil-spill>. (Accessed: 21st July 2021)
167. McDonald, J. S., Milosevic, D., Reddi, H. V., Grebe, S. K. & Algeciras-Schimmich, A. Analysis of circulating microRNA: Preanalytical and analytical challenges. *Clin. Chem.* **57**, 833–840 (2011).
168. Meerson, A. & Ploug, T. Assessment of six commercial plasma small RNA isolation kits using qRT-PCR and electrophoretic separation: Higher recovery of microRNA following ultracentrifugation. *Biol. Methods Protoc.* **1**, 1–9 (2016).
169. Shah, J. S., Soon, P. S. & Marsh, D. J. Comparison of methodologies to detect low levels of hemolysis in serum for accurate assessment of serum microRNAs. *PLoS One* **11**, 1–12 (2016).
170. Preparation, L. Instruction Manual LIBRARY PREPARATION NEBNext® Multiplex Small RNA Library Prep Set for Illumina® Set 1, Set 2, Index Primers 1-48 and Multiplex Compatible. (2018).
171. NextSeq 550 System | For everyday genomics. Available at: <https://www.illumina.com/systems/sequencing-platforms/nextseq.html>. (Accessed: 31st July 2021)
172. Illumina. Next-Generation Sequencing (NGS) | Explore the technology. *Innovative technologies* (2021). Available at: <https://www.illumina.com/science/technology/next-generation-sequencing.html>. (Accessed: 31st July 2021)
173. Applied Biosystems. TaqMan® Advanced miRNA Assays User Guide. *Thermo Fish. Sci.*
174. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68–D73 (2014).
175. Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36**, D154–D158 (2008).
176. Buchholz, M. Simultaneous detection of miRNA and mRNA on TaqMan Array Cards using the TaqMan Advanced miRNA workflow. *Thermo Fish. Sci.* (2017).
177. Laboratories, B.-R. CFX Maestro™ Software User Guide. (2017).
178. Biosystems, A. Data Analysis on the ABI PRISM 7700: Setting Baselines and Thresholds. *Life Technol.* 1–12 (2002).
179. Taylor, S. C. *et al.* The Ultimate qPCR Experiment: Producing Publication

- Quality, Reproducible Data the First Time. *Trends Biotechnol.* **37**, 761–774 (2019).
180. Jiang, J., Lee, E. J., Gusev, Y. & Schmittgen, T. D. Real-time expression profiling of microRNA precursors in human cancer cell lines. *Nucleic Acids Res.* **33**, 5394 (2005).
 181. Fan, Y. *et al.* miRNet - dissecting miRNA-target interactions and functional associations through network-based visual analysis. *Nucleic Acids Res.* **44**, W135–W141 (2016).
 182. Fan, Y. & Xia, J. miRNet—Functional analysis and visual exploration of miRNA–target interactions in a network context. *Methods Mol. Biol.* **1819**, 215–233 (2018).
 183. Fan, Y., Habib, M. & Xia, J. Xeno-mirNet: A comprehensive database and analytics platform to explore xeno-miRNAs and their potential targets. *PeerJ* **2018**, e5650 (2018).
 184. Chang, L., Zhou, G., Soufan, O. & Xia, J. miRNet 2.0: Network-based visual analytics for miRNA functional analysis and systems biology. *Nucleic Acids Res.* **48**, W244–W251 (2020).
 185. Huang, H. Y. *et al.* MiRTarBase 2020: Updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res.* **48**, D148–D154 (2020).
 186. Huang, Z. *et al.* HMDD v3.0: A database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.* **47**, D1013–D1017 (2019).
 187. Jiang, Q. *et al.* miR2Disease: A manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* **37**, (2009).
 188. Ruepp, A. *et al.* PhenomiR: A knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol.* **11**, (2010).
 189. QuickGO. Available at: <https://www.ebi.ac.uk/QuickGO/>. (Accessed: 27th July 2021)
 190. QIAGEN Digital Insights. QIAGEN CLC Genomics Workbench. (2020). Available at: <https://digitalinsights.qiagen.com/>. (Accessed: 29th July 2021)
 191. Rédei, G. P. Blast (basic local alignment search tool). *Encyclopedia of Genetics, Genomics, Proteomics and Informatics* 221–221 (2008). doi:10.1007/978-1-4020-6754-9_1879
 192. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
 193. Backes, C., Khaleeq, Q. T., Meese, E. & Keller, A. MiEAA: MicroRNA enrichment analysis and annotation. *Nucleic Acids Res.* **44**, W110–W116 (2016).
 194. QIAGEN. Ingenuity Pathway Analysis | QIAGEN Digital Insights. Available at: <https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/>. (Accessed: 29th July 2021)
 195. Bolstad, B. M., Irizarry, R. A., Åstrand, M. & Speed, T. P. A comparison of

- normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
196. Queralt-Rosinach, N., Piñero, J., Bravo, À., Sanz, F. & Furlong, L. I. DisGeNET-RDF: Harnessing the innovative power of the Semantic Web to explore the genetic basis of diseases. *Bioinformatics* **32**, 2236–2238 (2016).
 197. Piñero, J. *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* **48**, D845–D855 (2020).
 198. Li, Y., Wei, Y., Guo, J., Cheng, Y. & He, W. Interactional role of microRNAs and bHLH-PAS proteins in cancer (Review). *Int. J. Oncol.* **47**, 25–34 (2015).
 199. Gordon, M. W. *et al.* Regulation of p53-targeting microRNAs by polycyclic aromatic hydrocarbons: Implications in the etiology of multiple myeloma. *Mol. Carcinog.* **54**, 1060–1069 (2015).
 200. Tumolo, M. R. *et al.* The expression of microRNAs and exposure to environmental contaminants related to human health: a review. *Int. J. Environ. Health Res.* **00**, 1–23 (2020).
 201. Behjati, S. & Tarpey, P. S. What is next generation sequencing? *Arch. Dis. Child. Educ. Pract. Ed.* **98**, 236–238 (2013).
 202. Creemers, E. E., Tijssen, A. J. & Pinto, Y. M. Circulating MicroRNAs. *Circ. Res.* **110**, 483–495 (2012).
 203. Tan, G. C. *et al.* 5' isomiR variation is of functional and evolutionary importance. *Nucleic Acids Res.* **42**, 9424–9435 (2014).
 204. Biosystems, A. TaqMan MicroRNA Assays Protocol. *Design* 1–36 (2006).
 205. Wang, Q. *et al.* Overview of microRNA-199a regulation in cancer. *Cancer Manag. Res.* **11**, 10327–10335 (2019).
 206. Schmittgen, T. D. *et al.* Real-time PCR quantification of precursor and mature microRNA. *Methods* **44**, 31–38 (2008).
 207. Karlsen, T. A., Aae, T. F. & Brinchmann, J. E. Robust profiling of microRNAs and isomiRs in human plasma exosomes across 46 individuals. *Sci. Reports* **2019** **9**, 1–9 (2019).
 208. Oishi, Y. & Manabe, I. Krüppel-Like Factors in Metabolic Homeostasis and Cardiometabolic Disease. *Front. Cardiovasc. Med.* **5**, 69 (2018).
 209. Pollak, N. M., Hoffman, M., Goldberg, I. J. & Drosatos, K. Krüppel-Like Factors: Crippling and Uncrippling Metabolic Pathways. *JACC Basic to Transl. Sci.* **3**, 132–156 (2018).
 210. Jaeger, C. & Tischkau, S. A. Role of Aryl Hydrocarbon Receptor in Circadian Clock Disruption and Metabolic Dysfunction. *Environ. Health Insights* **10**, 133 (2016).
 211. Hossain, A., Kuo, M. T. & Saunders, G. F. Mir-17-5p Regulates Breast Cancer Cell Proliferation by Inhibiting Translation of AIB1 mRNA. *Mol. Cell. Biol.* **26**, 8191–8201 (2006).

212. Tsamou, M. *et al.* Prenatal particulate air pollution exposure and expression of the miR-17/92 cluster in cord blood: Findings from the ENVIRONAGE birth cohort. *Environ. Int.* **142**, (2020).
213. Zhang, W., Lin, J., Wang, P. & Sun, J. miR-17-5p down-regulation contributes to erlotinib resistance in non-small cell lung cancer cells. *J. Drug Target.* **25**, 125–131 (2017).
214. Chatterjee, A., Chattopadhyay, D. & Chakrabarti, G. miR-17-5p downregulation contributes to paclitaxel resistance of lung cancer cells through altering Beclin1 expression. *PLoS One* **9**, e95716 (2014).
215. Sun, Z. *et al.* CAP-miRSeq: A comprehensive analysis pipeline for microRNA sequencing data. *BMC Genomics* **15**, 1–10 (2014).
216. Neff, J. M., Stout, S. A. & Gunster, D. G. Ecological risk assessment of polycyclic aromatic hydrocarbons in sediments: identifying sources and ecological hazard. *Integr. Environ. Assess. Manag.* **1**, 22–33 (2005).
217. Fallahtafti, S., Rantanen, T., Brown, R. S., Snieckus, V. & Hodson, P. V. Toxicity of hydroxylated alkyl-phenanthrenes to the early life stages of Japanese medaka (*Oryzias latipes*). *Aquat. Toxicol.* **106–107**, 56–64 (2012).
218. Bouchard, M. & Viau, C. Urinary excretion kinetics of pyrene and benzo(a)pyrene metabolites following intravenous administration of the parent compounds or the metabolites. *Toxicol. Appl. Pharmacol.* **139**, 301–309 (1996).
219. Hossain, A., Kuo, M. T. & Saunders, G. F. Mir-17-5p Regulates Breast Cancer Cell Proliferation by Inhibiting Translation of AIB1 mRNA . *Mol. Cell. Biol.* **26**, 8191–8201 (2006).
220. Dejmek, J., Solanský, I., Beneš, I., Leníček, J. & Šrám, R. J. The impact of polycyclic aromatic hydrocarbons and fine particles on pregnancy outcome. *Environ. Health Perspect.* **108**, 1159–1164 (2000).

VITA

Dianne Aster Yunque – Yap was born on June 25th 1986, to Nestor G. Yunque and Lorena T. Yunque in Iloilo City, Philippines. She is the older sister of Dr. Vanessa Aura T. Yunque. Dianne is married to Jan Paul Vincent S. Yap with whom she has a daughter, Astrid Catalina Y. Yap. She graduated *cum laude* with a Bachelor of Science in Molecular Biology and Biotechnology from the University of the Philippines in Diliman, Quezon City, Philippines in 2007. After receiving her college degree, Dianne joined the Algal Production Laboratory at the Southeast Asian Fisheries Development Center – Aquaculture Department (SEAFDEC-AQD) in Tigbauan, Iloilo, Philippines as a research assistant where she published 1 primary author paper and co-authored 2 papers. She mentored students who worked on their thesis projects in collaboration with the scientists in the department. After two years, she moved to Kota Kinabalu, Sabah, Malaysia to spearhead the development of a seaweed tissue culture laboratory as research manager. Dianne moved to the McAllen, Texas, USA in 2011 and gained employment as a Medical Technologist at the Histocompatibility Laboratory at The University of Texas MD Anderson Cancer Center in 2012. She mentored several clinical laboratory scientists and molecular genetics students completing their internships. Her work resulted to a co-author status on a journal article. She received her certification as a Molecular Biology Technologist from the ASCP in 2016. On the same year, she joined the Pharmacology and Toxicology Graduate Program at the University of Texas Medical Branch (UTMB). In 2018, Dianne joined Dr. Cornelis (Kees) Elferink's laboratory where she pursued a dissertation project on the identification of human biomarkers of exposure and effect to petrogenic polycyclic aromatic hydrocarbons.

Permanent address: 3211 Carriage Cove Ct., Dickinson, TX 77539

This dissertation was typed by Dianne Aster Yunque - Yap.