Regulation of Long Non-Coding RNA Expression by Aryl Hydrocarbon Receptor Activation

Aeshah Alluli

Faculty of Medicine
Department of Pathology,
Meakins-Christie Laboratories
McGill University
December 2021

A Thesis Submitted to McGill University in Partial Fulfillment of the Requirements of the Degree of Master of Science
# TABLE OF CONTENTS

LIST OF FIGURES .......................................................................................................................... 6

List of Abbreviations ..................................................................................................................... 7

ABSTRACT ..................................................................................................................................... 12

RESUMÉ .......................................................................................................................................... 14

Acknowledgement .......................................................................................................................... 16

CHAPTER 1: INTRODUCTION ....................................................................................................... 17

1.1 The Aryl Hydrocarbon Receptor ............................................................................................. 17

1.1.1 History .................................................................................................................................. 17

1.1.2 Structure of the AhR ............................................................................................................ 18

1.1.3 AhR signaling ....................................................................................................................... 20

1.1.3.1 AhR ligands ..................................................................................................................... 20

1.1.3.1.1 Xenobiotic AhR ligands ............................................................................................ 21

1.1.3.1.2 Other Exogenous AhR ligands .................................................................................. 24

1.1.3.1.3 Endogenous AhR ligands .......................................................................................... 25

1.1.3.1.3.1 Heme metabolites .................................................................................................. 25

1.1.3.1.3.2 Arachidonic acid metabolites (AA) ....................................................................... 26

1.1.3.1.3.3 Tryptophan metabolites ....................................................................................... 27

1.1.3.1.4 AhR antagonists ........................................................................................................ 27

1.1.3.2 The Canonical AhR Pathway .......................................................................................... 30

1.1.3.3 The non-canonical AhR pathway ..................................................................................... 32

1.1.3.3.1 Protein interaction with AhR ................................................................................... 33
1.1.3.3.2 Ligand-independent AhR signaling ................................................................. 36
1.1.4 AhR Expression ................................................................................................. 37
1.1.5 Regulation of AhR Expression ......................................................................... 38
  1.1.5.1 Transcriptional regulation of AhR ................................................................. 38
  1.1.5.2 Suppressors of the AhR .............................................................................. 39
  1.1.5.3 Epigenetic regulation .................................................................................. 41
    1.1.5.3.1 Histone modifications ............................................................................ 41
    1.1.5.3.2 DNA methylation ................................................................................ 42
    1.1.5.3.3 Non-coding RNAs (ncRNAs) ................................................................. 43
  1.1.6 General and cellular physiological functions of AhR ....................................... 44
    1.1.6.1 Xenobiotic metabolism ............................................................................ 44
    1.1.6.2 Organ development .................................................................................. 45
    1.1.6.3 Immunity .................................................................................................. 46
    1.1.6.4 AhR regulation of ncRNAs expression ....................................................... 47
1.2 Long non-coding RNAs (lncRNAs) .................................................................. 49
  1.2.1 Subcellular location of lncRNAs .................................................................... 50
  1.2.2 Classification of lncRNAs .............................................................................. 50
    1.2.2.1 Classification of lncRNAs based on biogenesis ....................................... 50
    1.2.2.2 Classification of lncRNA based on whether they act in cis or in trans .... 51
    1.2.2.3 Classification of lncRNAs based on their mode of action ....................... 52
  1.2.3 Function of lncRNAs ...................................................................................... 53
1.2.3.1 Role of lncRNAs in chromatin modifications and transcription .................................. 53
1.2.3.2 Role lncRNA in post-transcriptional regulation ..................................................... 55
1.2.3.2.1 LncRNA and mRNA stability ................................................................. 55
1.2.3.2.2 LncRNAs and alternative splicing ........................................................... 57
1.2.3.2.3 LncRNA and RNA degradation ................................................................. 58
1.2.3.2.4 LncRNA act as miRNA Sponges ............................................................... 59
1.2.3.3 Role of lncRNA in translational regulation ......................................................... 60
1.2.3.4 Role of lncRNA in post-translational modifications (PTMs) ................................. 60
1.2.3.4.1 LncRNA and protein ubiquitination ......................................................... 61
1.2.3.4.2 LncRNA and protein phosphorylation .................................................... 61
1.2.4 Regulation of lncRNA expression ............................................................................ 64
1.3 Project Rationale ........................................................................................................... 65

CHAPTER 2: HYPOTHESIS AND AIMS ............................................................................ 66
2.1 Hypothesis: .................................................................................................................. 66
2.2 Aims: ............................................................................................................................ 66

CHAPTER 3: MATERIALS AND METHODS .................................................................... 67
3.1 Reagents ....................................................................................................................... 67
3.2 Cell culture .................................................................................................................. 67
3.3. Generation of A549-AhRKO cells ........................................................................... 67
3.4 Cell treatment ............................................................................................................. 68
3.5 MTT assay .................................................................................................................. 68
LIST OF FIGURES

Chapter 1 INTRODUCTION

Figure 1.1 Structure of the AhR protein................................................................. 20
Figure 1.2 The canonical AhR pathway ................................................................. 32
Figure 1.3 Non-genomic AhR pathway ................................................................. 36
Figure 1.4 LncRNA execute their functions through different mechanisms............... 63

Chapter 5 RESULTS

Figure 5.1B[a]P dose-dependently decreases A549 cell viability.................................. 75
Figure 5.2 RT-qPCR validation of CYP1A1 and CYP1B1 expression in A549 cells treated with B[a]P................................................................. 76
Figure 5.3 AhR expression in A549WT and A549-AhRKO cells.................................. 78
Figure 5.4 Knockout of AhR does not change the expression of selected lncRNA........ 80
Figure 5.5 Principal component analysis of untreated and B[a]P treated cells............... 83
Figure 5.6 MA plots of differentially up- and downregulated genes in response to B[a]P . 85
Figure 5.7 Hierarchical heat maps of the DEG in response to B[a]P.............................. 86
Figure 5.8 GO analysis for 6h treatment with B[a]P................................................ 94
Figure 5.9 GO analysis for 24h exposure to B[a]P.................................................. 95
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-MC</td>
<td>3-methylcholanthrene</td>
</tr>
<tr>
<td>3′,4′-DMF</td>
<td>3′,4′-dimethoxyflavone</td>
</tr>
<tr>
<td>3′-UTR</td>
<td>3′-untranslated region</td>
</tr>
<tr>
<td>6,2′,4′-TMF</td>
<td>6,2′,4′-trimethoxyflavone</td>
</tr>
<tr>
<td>A549</td>
<td>adenocarcinoma human alveolar basal epithelial cells</td>
</tr>
<tr>
<td>AA</td>
<td>Arachidonic acid metabolites</td>
</tr>
<tr>
<td>Ah</td>
<td>aryl hydrocarbon</td>
</tr>
<tr>
<td>AhH</td>
<td>aryl hydrocarbon hydroxylase</td>
</tr>
<tr>
<td>AhR</td>
<td>aryl hydrocarbon receptor</td>
</tr>
<tr>
<td>AhRE</td>
<td>aryl hydrocarbon response element</td>
</tr>
<tr>
<td>AhRR</td>
<td>aryl hydrocarbon receptor repressor</td>
</tr>
<tr>
<td>AIP</td>
<td>AhR-interacting protein</td>
</tr>
<tr>
<td>Akt</td>
<td>AKT serine/threonine kinase</td>
</tr>
<tr>
<td>ALDH3A1</td>
<td>aldehyde dehydrogenase</td>
</tr>
<tr>
<td>AML</td>
<td>acute myeloid leukemia</td>
</tr>
<tr>
<td>AMPK</td>
<td>adenosine monophosphate-activated protein kinase</td>
</tr>
<tr>
<td>ANF</td>
<td>α-naphtoflavone</td>
</tr>
<tr>
<td>ANOVA</td>
<td>one-way analysis of variance</td>
</tr>
<tr>
<td>AQP2</td>
<td>aquaporin 2</td>
</tr>
<tr>
<td>AQP3</td>
<td>aquaporin 3</td>
</tr>
<tr>
<td>ARNT</td>
<td>aryl hydrocarbon receptor nuclear translocator</td>
</tr>
<tr>
<td>AS</td>
<td>alternative splicing</td>
</tr>
<tr>
<td>B[a]P</td>
<td>benzo[a]pyrene</td>
</tr>
<tr>
<td>BAFF</td>
<td>B cell-activating factor</td>
</tr>
<tr>
<td>BCA</td>
<td>bicinchinonic acid</td>
</tr>
<tr>
<td>BTG2</td>
<td>BTG anti-proliferation factor 2</td>
</tr>
<tr>
<td>C/EBP</td>
<td>CCAAT-enhancer-binding proteins</td>
</tr>
<tr>
<td>CAVIN2</td>
<td>caveolae associated protein 2</td>
</tr>
<tr>
<td>CC</td>
<td>cellular component</td>
</tr>
<tr>
<td>CD</td>
<td>Crohn’s disease</td>
</tr>
<tr>
<td>ceRNAs</td>
<td>competitive endogenous RNAs</td>
</tr>
<tr>
<td>CH-223191</td>
<td>2-methyl-2H-pyrazole-3-carboxylic acid-amide</td>
</tr>
<tr>
<td>CH3</td>
<td>methyl group</td>
</tr>
<tr>
<td>CNS</td>
<td>central nervous system</td>
</tr>
<tr>
<td>COPD</td>
<td>chronic obstructive pulmonary disease</td>
</tr>
<tr>
<td>CRISPR/Cas9</td>
<td>clustered regularly interspaced short palindromic repeats/Cas9</td>
</tr>
<tr>
<td>CROCC2</td>
<td>ciliary rootlet coiled-coil</td>
</tr>
<tr>
<td>CS</td>
<td>cigarette smoke</td>
</tr>
<tr>
<td>CSE</td>
<td>cigarette smoke extract</td>
</tr>
<tr>
<td>CYP</td>
<td>Cytochrome P450</td>
</tr>
<tr>
<td>DC</td>
<td>dendritic cells</td>
</tr>
<tr>
<td>DCX</td>
<td>doublecortin</td>
</tr>
<tr>
<td>DE</td>
<td>differential expression</td>
</tr>
<tr>
<td>DEG</td>
<td>differential expression gene</td>
</tr>
<tr>
<td>DHR3</td>
<td>dehydrogenase/reductase 3</td>
</tr>
<tr>
<td>DLCs</td>
<td>dioxin-like chemicals</td>
</tr>
<tr>
<td>DLL4</td>
<td>Delta like canonical Notch ligand 4</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s Modified Eagle Medium</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNMT</td>
<td>DNA methyltransferase</td>
</tr>
<tr>
<td>DRE</td>
<td>dioxin response element</td>
</tr>
</tbody>
</table>
bHLH: basic helix loop helix
BP: biological process
DV: ductus venosus
E2: estradiol
ECL: enhanced chemiluminescence
EGFLAM: EGF like, fibronectin type III and laminin G domains
eIF4E: eukaryotic initiation factor-4E
ER: estrogen receptor
ERE: estrogen-responsive element
eRNA: enhancer RNA
EZH2: enhancer of zeste homolog 2
FANTOM5: functional annotation of mammalian genomes 5
FICZ: 6-formylindolo[3,2-b] carbazole
GAS5: growth arrest-specific 5
GC: gastric cancer
GNF351: N-(2-(1H-indol-3-yl) ethyl)-9-isopropyl-2-(5-methylpyridin-3-yl)-9H-purin-6-amine
GO: Gene ontology
GPER1: G protein-coupled estrogen receptor
GR: glucocorticoid receptor
GSTA2: glutathione-S-transferase A2
HAH: Halogenated aromatic hydrocarbons
HAT: histone acetyltransferases
HCC: hepatocellular carcinoma
HDACs: histone deacetylases
HMGCR: 3-hydroxy-3-methylglutaryl-CoA reductase
HMT: histone methyltransferases
hnRNPs: heterogeneous nuclear ribonucleoproteins
HOTAIR: HOX antisense intergenic RNA
HOTTIP: HOXA transcript
HSP: heat-shock protein
HULC: hepatocellular carcinoma up-regulated long non-coding RNA
HuR: human antigen R
I3C: indole-3-carbinol
IAA: Indole acetic acid
IDO: indoleamine 2,3-dioxygenase
IFFO2: intermediate filament family orphan 2
IL: interleukin
ITE: 2-(1-H-indole-3-carbonyl)-thiazole-4-carboxylic acid methyl ester
IκBs: inhibitors of NF-κB
KEGG: Kyoto Encyclopedia of Genes and Genomes
KO: knockout
LCP2: lymphocyte cytosolic protein 2
LD50: median lethal dose
LINC01533: long intergenic non-protein coding RNA
LINC02532: long intergenic non-protein coding RNA
lincRNAs: long intergenic RNAs
LIPG: lipase G
LncRNAs: long non-coding RNAs
LncRNPs: lncRNA-protein complexes
LPS: lipopolysaccharides
LSCC: laryngeal squamous cell carcinoma
hnRNPLL: heterogeneous nuclear RNA-binding protein LL
LTβ: lymphotoxin beta
LXA4: Lipoxin A4
MACC-AS: Metastasis-associated in colon cancer antisense
MACC1: Metastasis-associated in colon cancer 1
MALAT1: metastasis associated lung adenocarcinoma transcript 1
MEG3: maternally expressed gene 3
MEGF6: multiple EGF like domains 6
MF: molecular function
MIR4290HG: microRNA 4290 host gene
miRNA: microRNA
MLF: mouse lung fibroblasts
MMP: Matrix metalloproteinase
MNF: 3’-methoxy-4’-nitroflavone mouse embryonic fibroblasts (MEFs)
MREs: miRNA response elements
mRNA: messenger RNA
MTT: (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide)
MYH4: myosin heavy chain 4
NADPH: nicotinamide adenine dinucleotide phosphate
NATs: natural antisense transcripts
Ncoa1,2,3: nuclear coactivators
ncRNA non-coding RNA
NEAT1: nuclear paraspeckle assembly transcript 1
NES: nuclear export signal
NF-κB: nuclear factor-κB
NK: Natural killer
NLS: N-terminal nuclear localization signal
NORAD: non-coding RNA activated by DNA damage
NQO1: NAD(P)H quinone oxidoreductase 1
NR4A2: nuclear receptor subfamily 4 group A member2
NRF2: nuclear factor-erythroid 2-related factor 2
NSCLC: Non-small-cell lung carcinoma
Oct4: octamer-binding transcription factor 4
ORF: open reading frame
P23: prostaglandin E synthase 3
PAH: polycyclic aromatic hydrocarbons
PAIP2: polyadenylate-binding protein-interacting protein-2
PancRNAs: promoter-associated non-coding RNAs
PAS: PER-ARNT-SIM
PCA: Principal component analysis
PCB: polychlorinated biphenyls
PCDD: polychlorinated dibenzo-p-dioxin
PCDF: polychlorinated dibenzofurans
PCR: Polymerase chain reaction
PER: Period
PG: prostaglandins
PML: promyelocytic leukemia
Pol II: RNA polymerase II
PPARα: peroxisome proliferator-activated receptor-α
NEAT2: nuclear paraspeckle assembly transcript2
PRC2: Polycomb repressive complexes 2
pre-mRNA: precursor mRNA
PTMs: post-translational modifications
PTP: protein tyrosine phosphatase
PUF: Pumilio-Fem3-binding factor
PVDF: Polyvinylidene difluoride
RASAL2-AS1: RAS protein activator like 2 antisense RNA 1
RB: retinoblastoma protein
RBPs: RNA-binding proteins
reaction
RepA: Repeat A
RIN: RNA integrity number
RIP: receptor-interacting protein
RIPA: Radio-immunoprecipitation assay
RNA: Ribonucleic acid
RNases: ribonucleases
RNP: ribonucleoprotein complexes
RPMI: Roswell Park Memorial Institute
rRNA: ribosomal RNA
RT-qPCR: Real time quantitative
SAMD12-AS1: sterile alpha motif domain containing 12 antisense RNA 1
SATB1-AS1: special AT-rich sequence binding protein 1 antisense RNA 1
SBS: STAU1-binding site
SDS-PAGE: sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SFs: splicing factors

SIM: single-minded
siRNA: small-interfering RNA
SLCO4C1: solute carrier organic anion transporter family member 4C1
SMD: STAU1-mediated mRNA decay
snRNPs: small nuclear ribonucleoproteins
SOCS-2: suppressor of cytokine signaling 2
SP: specificity protein
SRC1: steroid receptor coactivator 1
SRSF1: Serine/arginine-rich splicing factor 1
STAT3: signal transducer and activator of transcription 3
STAU1: short interspersed nuclear element half-Staufen1
STC1: stanniocalcin 1
Suz12: suppressor of zeste 12
TA: tryptamine
TAD: transactivation domain
TCDD: 2,3,7,8-tetrachlorodibenzo-p-dioxin
TCP: trichlorophenol
TF: Transcription factor
TFBS: transcription factor binding sites
TGF-β: transforming growth factor-beta
Th: Helper T lymphocyte
TiPARP: TCDD-inducible poly ADP-ribose polymerase
TMEM132D: transmembrane protein 132D
TNF-α: tumor necrosis factor-α
Treg: Regulatory T lymphocyte
tRNA: transfer
Tyr705: tyrosine-705
SHP-1: Src homology region 2 domain-containing phosphatase-1
TYMSOS: thymidylate synthetase opposite strand
UGT: UDP-glucuronosyltransferases
UV: ultraviolet
VCP: valosin-containing protein
VIPR1-AS1: vasoactive intestinal peptide receptor 1 antisense RNA 1
VST: variance stabilized transformed
WT: wild type
XAP2: immunophilin-like X-associated protein 2
XCI: X chromosome inactivation
xeno-lncs: xenobiotic-responsive IncRNAs
XKR6: XK related 6
XME: xenobiotic-metabolizing enzyme
XRE: Xenobiotic-response element
ABSTRACT

The aryl hydrocarbon receptor (AhR) is a cytosolic transcription factor that can be activated by endogenous (e.g., 6-formylindolo(3,2-b) carbazole (FICZ)) or xenobiotic (e.g., dioxin and benzo[a]pyrene (B[a]P)) ligands. Once the AhR is activated, it translocates into the nucleus, dimerizes with the AhR nuclear translator (ARNT), and binds to specific DNA sequences called xenobiotic response elements (XRE) to promote the target gene transcription, including cytochrome P450 (CYP1A1). The AhR not only mediates the toxic effects of dioxin, but also has physiological functions such as suppressing inflammation, oxidative stress, and apoptosis. We have previously shown that AhR-dependent regulation of endogenous miR-196a expression suppressed cigarette smoke (CS)-induced apoptosis. However, the exact mechanism of how AhR regulates apoptosis through miR-196a is still poorly understood. We think that part of the mechanism of action for the AhR is through regulation of long non-coding RNA (lncRNAs) expression. LncRNAs may play important roles in the response to xenobiotics. LncRNAs are defined as transcript more than 200 nucleotides in length that do not encode a protein but are implicated in many physiological processes such as cell differentiation, cell proliferation, and apoptosis. A few studies revealed that the AhR affects the expression of lncRNA. Therefore, we hypothesize that AhR activation regulates the expression of lncRNA in response to B[a]P in the human alveolar epithelial cell line (A549). The aims of this study are: (1) investigate the effect of B[a]P on lncRNAs expression and (2) determine whether B[a]P regulates lncRNAs expression is in an AhR-dependent manner. First, to confirm whether B[a]P regulates the well-characterized lncRNAs (NEAT1, HOTTIP, SOX2OT, MALAT1, H19, and Linc00673) is in AhR-dependent manner, A549WT and AhR knock-out (A549-AhRKO) cells were treated with B[a]P (8μM) for 6 and 24h. The expression of well-characterized mRNA and lncRNAs was measured by RT-qPCR.
Despite AhR activation, as indicated by induction of \textit{CYP1A1} and \textit{CYP1B1} mRNA, we found that neither treatment nor AhR expression affected the expression of NEAT1, HOTTIP, SOX2OT, MALAT1, H19, and Linc00673. Next, to evaluate transcriptome profiles of B[a]P in A549 cells, RNA sequencing was performed. We found that 704 and 811 lncRNAs (e.g., SATB1-AS1, MIR4290HG, AC008969.1, LINC01533, VIPR1-AS1) and protein-coding RNAs (e.g., CYP1A1, BX005266.2, AQP3, BTG2, DCX, and AhRR), respectively, were differentially-expressed in A549 cells treated with B[a]P (8µM) after 6 and 24h. GO analyses indicated that DE protein-coding RNAs are associated with distinct molecular functions, whereas KEGG analyses showed the hsa01100 pathway was associated with differently-expressed lncRNAs. In summary, our research revealed that B[a]P affects lncRNAs expression; however, further investigations are required to reveal the role of these lncRNA in modulating AhR function. This study profiles the lncRNAs expression in A549 cells exposed to B[a]P, which could assist in establishing the basis for the physiological functions of AhR.
RESUMÉ

Le récepteur d'hydrocarbure aryle (AhR) est un facteur de transcription cytosolique qui peut être activé par des ligands endogènes (ex: 6-formylindolo (3,2-b) carbazole (FICZ)) ou exogènes (ex: dioxine et benzo[a]pyrène (B[a]P)). En présence du ligant, AhR est alors transloquée dans le noyau, où il se dimérise avec son partenaire ARNT (pour *Aryl Hydrocarbon Receptor Nuclear Translocator*). Ce dimère ensuite se lie à des séquences d'ADN spécifiques appelées éléments de réponse aux xénobiotiques (XRE). Ceci favorise la transcription des gènes cibles tels que le cytochrome P450 1A1 (CYP1A1). L'AhR médie non seulement les effets toxiques des dioxines, mais il a également des fonctions physiologiques. Par exemple, AhR est impliqué dans la régulation de l'inflammation, le stress oxydatif et l'apoptose. Nous avons récemment montré que l’AhR peut réguler l’expression d’un micro-ARN nommait miR-196a dans les poumons des souris exposées à la fumée de cigarette. Ce dernier participe à la réduction de l’apoptose cellulaire. Cependant, le mécanisme exact par lequel AhR régule l'apoptose par miR-196a n’est compris. Nous pensons qu'une partie de ce mécanisme se fait via les ARN long non-codants ou ARNlnc. Ces derniers pourraient jouer un rôle important dans la réponse aux xénobiotiques. Les ARNlnc sont définis comme des transcrits de plus de 200 nucléotides de long qui ne codent pas pour une protéine. Ils sont impliqués dans de nombreux processus physiologiques tels que la différenciation cellulaire, la prolifération cellulaire et l'apoptose. Peu des études ont montré que l'AhR affecte l'expression des ARNlnc. Par conséquent, nous avons proposés que l'activation d'AhR par ses ligands active l'expression des ARNlnc dans les cellules épithéliales afin de réguler des processus physiologiques. Les objectifs de cette étude sont: (1) Étudier l'effet de B[a]P sur l'expression des ARNlnc et (2) Déterminer si B[a]P régule l'expression des ARNlnc de manière dépendante d’AhR.

Premièrement, afin de vérifier si B[a]P régule l’expression des ARNlnc bien caractérisés (NEAT1,
Ensuite, pour évaluer les profils de transcriptome de B[a]P dans les cellules épithéliales alvéolaires A549, un séquençage d'ARN a été utilisé. Nous avons trouvé que 704 et 811 ARNlnc (ex : SATB1-AS1, MIR4290HG, AC008969.1, LINC01533, VIPR1-AS1) et ARN codants pour protéines (ex : CYP1A1, BX005266.2, AQP3, BTG2, DCX, and AhRR) étaient exprimés de manière significativement différentielle dans les cellules traitées avec B[a]P pour 6 et 24 heures. Les analyses GO ont indiqué que ces ARN codants pour des protéines sont associés à des fonctions moléculaires distinctes. Tandis que, les analyses KEGG ont montré que la voie hsa01100 est associée à des ARNlnc qui sont exprimée différemment entre les cellules contrôles et les cellules traitées avec le B[a]P.

En résumé, notre recherche a révélé que le B[a]P affecte l'expression de certains ARNlnc. Cependant, d'autres investigations sont nécessaires pour révéler le rôle ces transcrit dans la modulation de la fonction d'AhR. Cette étude caractérise l'expression des ARNlnc dans les cellules A549 exposées aux B[a]P, ce qui pourrait aider à établir une meilleure compréhension aux rôles physiologiques d'AhR.
Acknowledgement

Foremost, I would like to express my sincere gratitude to Allah for giving me the strength and patience to finish this unique journey. I would like to thank Dr. Carolyn Baglole for her supervision, generous support, and accepting me in her lab. I would like to express my deepest appreciation to my committee advisors: Dr. Carlos Telleria, Dr. Sabah Hussain, and Dr. John White, for their valuable criticism, comments, and guidance that helped me throughout my project. Moreover, I would like to express my special gratitude and thanks to Dr. Edith Zorychta for her support and help with my thesis editing.

I would also like to acknowledge Dr. Gregory Fonseca for providing indispensable information and advice on the RNA sequence analysis for my project. I would like to thank faculty members of the Meakins-Christie Laboratories, particularly Dr. David Eidelman, for providing generous advice and endless kind help during my project. My sincere gratitude to all colleagues, particularly Hussein and Noof, for providing support, helpful advice, and insightful comments on my project. In addition, my sincere thanks go to the Saudi Ministry of education and health for the financial support and for providing all the help I needed. I would also like to thank the Department of Pathology for having me as a grad student.

I owe my deepest gratitude to my special friend, Najla, for her generous help and support while in Montreal and Saudi Arabia. I am also grateful to my friends in Montreal: Aeshah, Dareen, Halema, and Wedad, as well as my friends back at home: Ryhanh and Jassem.

This thesis would not have been possible without the help, support, and love of my parents, sisters, and brothers. Finally, my deepest gratitude to my loving, supportive, and caring son, Rakan. He was encouraging, proud, with extraordinary patience, and sacrificing most of his weekends for me to work and finish up this project.
CHAPTER 1: INTRODUCTION

1.1 The Aryl Hydrocarbon Receptor

1.1.1 History

The aryl hydrocarbon receptor (AhR) is a cytosolic transcription factor that is highly conserved among mammalian species [1]. In 1956, Allan Conney was the first to identify the AhR in the liver of rats treated with polycyclic aromatic hydrocarbons (PAHs) [2]. At that time, the AhR was known as benzpyrene hydroxylase and later became aryl hydrocarbon hydroxylase (AHH) [2, 3]. AHH is a microsomal enzyme consisting of a reductase fraction, a lipid fraction, and a hemoprotein fraction (cytochrome P450) [4]. This enzyme relies on a cofactor (nicotinamide adenine dinucleotide phosphate (NADPH)) and oxygen to metabolize specific types of polycyclic hydrocarbon compounds [5, 6].

Early observations revealed that the basal level and inducibility of AHH activity varied between inbred mouse strains [7]. In 1969, Nebert and Gelboin observed the inducibility of AHH activity in different tissues of six mouse strains (e.g., the Swiss, C-57GK, C3H/HEN, A/HEN, AKR/N, and DBA strains) [7]. AHH activity was highly responsive to PAHs in the C57 strain, which was described as a responsive mouse strain [7]. However, AHH was not inducible in DBA and AKR strains, and thus were described as nonresponsive mouse strains [7]. Subsequently, genetic studies showed that a single autosomal dominant gene, which has multiple alleles, controlled AHH inducibility [8]. Halogenated aromatic hydrocarbons (HAHs), in particular 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), were found to be more potent inducers of AHH than PAHs [9]. In 1975, Poland et al. speculated that TCDD and 3-methylcholanthrene (3-MC) might share the same mechanism of induction of AHH activity and act on the same receptor since both
chemicals induce the same maximal responses in mice [5, 9, 10]. After examining the dose-response curve of TCCD and 3-MC in both responsive and nonresponsive mouse strains, a theory was proposed [5, 9, 10]. This theory suggested the presence of a receptor for these chemicals [5, 9, 10]. Nonresponsive strains do not respond to 3-MC and require higher doses of TCCD to induce AHH enzyme activity similar to that seen in responsive strains [5, 9, 10]. These observations clarified that the nonresponsive strains were, in fact, less responsive and less sensitive to TCDD [5, 9, 10]. These results support Poland’s hypothesis that both TCDD and 3-MC act on the same receptor.

In 1976, Poland confirmed existence of the AhR using radiolabeled dioxin congeners in vitro and in vivo to characterize a large class of chemical inducers based on their potency to induce AHH activity and binding affinity [11]. Indeed, the binding affinity of TCCD and dioxin congeners for AhR parallels their toxic capacity in vivo [11]. In the early 1980s, studies to characterize the AhR protein using rat liver cytosol revealed that this was a cytosolic protein bound to a chaperone complex [12, 13]. Upon TCCD binding, this cytosolic protein translocates to the nucleus [13-15]. Then, the aryl hydrocarbon (Ah) locus was described as a regulatory gene for the cytosolic receptor [11]. Polymorphism of the Ah locus between different inbred strain mice determines the sensitivity of the strains to the effects produced by compounds such as TCDD and PAHs [9, 11]. By the end of the 1980s, this protein was named the AhR [16].

1.1.2 Structure of the AhR

The basic helix loop helix (bHLH) and PER-ARNT-SIM (PAS) family of transcription factors are defined by the Period (PER), AhR nuclear translocator (Arnt), and single-minded (SIM) proteins
The bHLH-PAS family consist of six major sets of proteins (A-F), with the AhR being a member of the bHLH group C that has two PAS domains, PAS-A and PAS-B [17]. The AhR protein possesses two structural motifs. The basic helix loop helix (bHLH) domain is located in the N-terminal and a PER-ARNT-SIM (PAS) domain that is located at the C-terminal [18, 19]. The bHLH domain further subdivides into two domains: a basic domain that allows AhR binding to DNA and an HLH domain that is essential for protein-protein interaction [19]. The bHLH domain also serves as a harbor for both an N-terminal nuclear localization signal (NLS) and a nuclear export signal (NES) [20]. The PAS domain is a docking site for other PAS proteins and the molecular chaperone heat-shock protein 90 (HSP90) [18, 21]. Furthermore, the PAS-B domain is characterized by a ligand-binding site, which partly interferes with the binding site of HSP90 [21]. The AhR protein also has a glutamine-rich transactivation domain (TAD) that is located in the C-terminal [22, 23]. The TAD is a crucial region for the activation of target genes when it binds with transcriptional coactivators (Figure 1.1) [22, 23].
Figure 1.1 Structure of the AhR protein. The AhR is comprised of an amino-terminus (known as N-terminal (NH2)) and carboxy-terminus (known as C-terminal (COOH)). A bHLH/PAS domain is found in N-terminal, while C-terminal consists of a TAD domain. Adapted from [24].

1.1.3 AhR signaling

1.1.3.1 AhR ligands

A wide variety of compounds are recognized as ligands for the AhR. Identifying a range of chemical inducers that act as ligands of the AhR is crucial towards understanding both the physiological and pathologic roles of this protein. Indeed, it is now understood that AhR ligands produce differential effects based on tissue-specific differences of AhR in developmental, physiological, and disease states [25-30].
AhR ligands are hydrophobic molecules that enter the plasma membrane through simple diffusion [31]. In general, AhR ligands are planar and have an estimated length between 1.2 and 1.4 nm [32, 33]. AhR ligands differ greatly in their binding affinity to the AhR, which ranges from from $10^{-12}$ and $10^{-3}$ M [29]. Although AhR ligands with high affinity are believed to be toxic, there are other high-affinity ligands that do not produce toxic outcomes [34]. The AhR protein had a binding site for a wide range of compounds capable of modulating AhR activity, acting as an agonist or antagonist. Generally, AhR ligands can be classified into xenobiotic, exogenous, and endogenous ligands (Table 1.1).

### 1.1.3.1.1 Xenobiotic AhR ligands

Xenobiotics are external chemicals that are not naturally found within an organism but are typically produced by human activities and are thus considered as environmental pollutants [35]. In general, examples of xenobiotics include food additives, drugs, environmental pollutants, carcinogens, and insecticides [36, 37]. The best characterized xenobiotic AhR ligands are PAHs and HAHs, which are also known as exogenous ligands [25, 29, 38]. PAHs are found in coal, crude oil, and gasoline [29, 39, 40]. PAHs are the product of incomplete combustion of organic compounds and are generated during the burning of wood or fuel, for example [29]. As such, PAHs are present in vehicle exhaust, charbroiled foods, and CS [29, 39, 40]. PAHs are found as complex mixtures and includes benz[a]anthracene, benzo[a]pyrene (B[a]P) and dibenz [a, h] anthracene (Table 1.1) [29, 41, 42]. The binding affinities of PAHs are within the nM to μM range, which is a lower affinity compared to HAHs [29].
HAHs are produced from a wide range of industrial processes including the incineration of municipal solid waste, medical waste, sewage sludge, hazardous waste, and burning of fuel [29]. HAHs also can be produced through manufacturing (smelting, chlorine bleaching of paper pulp, and manufacturing pesticides and herbicides) [29]. Natural processes (e.g., volcanic eruption and forest fires) and CS are other potential sources of HAHs [29]. Chlorinated HAHs are called dioxin-like chemicals (DLCs) [29]. DLCs are classified into three main groups: polychlorinated dibenzo-p-dioxin (PCDDs), polychlorinated dibenzofurans (PCDFs), and polychlorinated biphenyls (PCB) [29, 25]. There 419 congeners that make up DLCs and are generally characterized as having high affinity for lipids, being insoluble in water, having similar chemical structure, and execute their effects through the AhR pathway [43]. Of the 419 dioxin congeners, 29 exhibit toxicity in experimental animals [43].

TCDD (dioxin) is the best-known AhR ligand, which belongs to PCDDs group of DLCs [29, 34, 43]. TCDD has two benzene rings and a central oxygenated ring, and has high affinity for AhR, estimated to be between 1 pM - 10 nM (Table 1.1) [29, 34, 43]. TCDD is colorless, solid, and unscented [34, 43]. TCDD is considered one of the most toxic human-made chemicals due to its eight chlorine atoms that can be laterally placed on carbons 2, 3, 7, and 8 [34, 43]; this configuration results in impaired metabolism and clearance [43]. Exposure to TCDD affects various species (e.g., guinea pigs, mice, horses, hamster, mink, rats, turkeys, trout, non-human primates, salmon, and chickens) of both sexes and in different organs, tissues, and cells [44-46]. Examples of toxic effects of TCDD exposure to animals are chick edema disease, lymphoid and gonadal atrophy, metaplasia, endocrine disruption, teratogenesis, liver and immune toxicity, tumor promotion, and wasting syndrome followed by death [44-46]. TCDD sensitivity is highly variable
between species which is based on the median lethal dose (LD$_{50}$) [45, 47]. For example, in the guinea pig, the LD$_{50}$ for TCDD is 1 µg/kg; however, the same dose in a hamster has little effect, with the LD$_{50}$ dose being > 1000 µg/kg [45-47]. It is established that the toxic effects of TCDD exposure are mediated by AhR [44-46, 48] as confirmed using AhR deficient mice, which are resistant to TCDD toxicity [46, 47, 49]. In these studies, mice which express AhR and exposed to TCDD have thymic atrophy, hepatic lesions (e.g., fatty infiltration, hepatomegaly, hepatocellular and bile duct hyperplasia) while AhR deficient mice are resistant to these effects [50].

Knowledge of the effects of TCDD in human comes from exposures to both individuals as well as populations. For example, one of the largest incidence of human exposure to TCDD was the use of the herbicide Agent Orange during the US-Vietnam war [51-54]. During the Second Indochina War (1961–1975), the US military used herbicides contaminated with TCDD [52, 54]. From 1961 to 1975, the amount of herbicides that were sprayed by US military was approximately 20 million gallons [52, 54]. Herbicides were sprayed in areas around military bases and the forests (e.g., inland hardwood forests and coastal mangrove forests) by helicopters and other aircraft [52, 54]. This inadvertently resulted in significant human and environmental exposures to TCDD [52-54]. The long-term effects of TCDD due to Agent Orange use include increased risk of diabetes, congenital abnormalities in children, and cancers such as Hodgkin’s disease, prostate cancer, and leukemia [52-54].

In industrial accidents, such as that which occurred in Seveso (Italy) in 1976, also resulted in individuals being exposed to TCDD [55]. An increase in the temperature and pressure during trichlorophenol (TCP) manufacturing, an antiseptic against gram-positive bacteria, resulted in the
accidental release of TCDD, a by-product of the manufacturing process [56]. One of the manifestations was the development of chloracne, particularly in children [55]. The long-term effects of TCDD exposure in Seveso includes increased incidence of diabetes, adverse cardiovascular effects, and altered endocrine function [55, 57]. Twenty-five years of follow-up with this exposed population showed an increase in lymphatic and hematopoietic neoplasms, observations that are consistent with outcomes found in the mouse model [57-60].

Finally, the assassination attempt on former Ukrainian president Viktor Yushchenko in 2004 is another example of human exposure to a high level of TCDD [48, 51, 57, 61]. It should be noted that the maximum accepted daily dose of TCDD in humans is 4 pg/kg [48, 51, 57, 61]. Yushchenko was poisoned by a single oral dose of 20 µg/kg TCDD [48, 61]. Two weeks after exposure, facial inflammatory edema and skin inflammation appeared [46, 48, 61]; he subsequently developed severe chloracne [48, 61] and exhibited digestive tract symptoms (e.g., hepatitis, pancreatitis, gastritis, and colitis with multiple ulcers) [46, 61]; these symptoms improved within six weeks [61]. Collectively, these studies show that TCDD exhibits toxic effects across various species including humans.

1.1.3.1.2 Other Exogenous AhR ligands

Besides HAH and PAH, exogenous ligands such as dietary and bacterial products have been identified. In general, dietary compounds which activate the AhR have a lower affinity compared to TCDD, with affinities of natural exogenous ligands ranging between µM to mM concentrations (Table 1.1) [32, 62]. Dietary AhR ligands include natural flavonoids ubiquitously found in vegetables and fruits [26, 34, 63, 64]. Flavonoids are one of the richest types of plant polyphenols.
and have many biological properties such as antioxidant, antiestrogenic, anti-inflammatory, and antiproliferative effects that are linked to health benefits [34, 65, 66]. Flavonoids can be classified into six main groups, including flavones, flavanols, isoflavones, flavanones, catechins, and anthocyanidins [26, 65, 67]. Several well-characterized examples of flavonoids that have been reported as activators of AhR are the flavonol quercetin and the flavonoids apigenin, galanin, chrysin, daidzein, baicalein, and genistein [26, 34, 65, 67]. Another example of a dietary AhR ligand is indole-3-carbinol (I3C), abundant in leafy vegetables such as broccoli and Brussels sprouts, cabbage, and cauliflower [26, 34, 68]. A previous report showed that I3C has anti-cancer properties either by inducing apoptosis of cancer cells or inhibiting tumor growth [68].

Besides nutritional products, bacteria pigment virulence factors such as the phenazines from *Pseudomonas aeruginosa* and the naphthoquinone phthiocol from *Mycobacterium tuberculosis* are examples of other natural exogenous AhR ligands [46, 69]. Biologically, the AhR may participate in host defense against bacterial infection through sensing of these virulence factors [46, 69]. Together, these are a few examples of naturally-occurring AhR ligands.

### 1.1.3.1.3 Endogenous AhR ligands

#### 1.1.3.1.3.1 Heme metabolites

In addition to exogenous and xenobiotic ligands, there exist a plethora of endogenous ligands, including heme metabolites. Heme metabolites such as bilirubin, biliverdin, and hemin are examples of endogenous AhR ligands, although these have weaker affinity for AhR compared to TCDD (*Table 1.1*) [26, 34, 46, 70]. In normal physiological conditions, heme metabolites such as plasma bilirubin- with an average level ranging from 5–20 µM- do not induce AhR-dependent gene expression [29, 70], likely because most plasma bilirubin is already bound to serum albumin.
However, in pathological conditions, there is an increase in plasma bilirubin where levels can reach up to 300-800 µM, including Crigler–Najjar syndrome and Gilbert’s syndrome [70]. In these two syndromes, the elevation of bilirubin results in induction of AhR target genes including CYP1A1 and CYP1A2 [70]. In addition, heme metabolites are capable of binding and activating AhR to stimulate cyp1a1 transcription in mouse hepatoma cells [70, 71]. Furthermore, after liver transplantation, plasma bilirubin level increase, which is associated with an elevation in CYP1A1 expression [70].

1.1.3.1.3.2 Arachidonic acid metabolites (AA)

Lipoxin A4 (LXA4), an arachidonic acid metabolite produced by lipoxygenase, has anti-inflammatory properties [72]. LXA4 has high affinity for the AhR, resulting in the upregulation of AhR target gene expression [29, 72]. LXA4 can bind and activate the AhR in a murine hepatoma cell line, resulting in increased expression of CYP1A1 [72]. In addition, LXA4 mediates its anti-inflammatory properties by modulating innate and acquired immune responses, responses that require the AhR [72, 73]. Here, in mouse dendritic cells (DCs), LXA4 attenuates IL-12 production via the AhR, which increases the expression of suppressor of cytokine signaling 2 (SOCS-2) [73, 74]. Other arachidonic acid metabolites such as leukotriene and prostaglandins (PG) (prostaglandins B2, D2, F3α, G2, H1, and H2) are also endogenous AhR ligands [46, 75, 76]. These can bind AhR to promote AhR target gene induction [46, 75-77]. Many prostaglandins can transform AhR into its active DNA-binding form at concentrations of 10 µM or greater [77]. Thus, prostaglandins are considered to be weak ligands, as they need µM concentration to trigger AhR activation, unlike LXA4 which can activate AhR with nM concentrations (Table 1.1) [29, 72].
1.1.3.1.3 Tryptophan metabolites

Tryptophan metabolites are also AhR agonists [28]. Indole acetic acid (IAA) and tryptamine (TA), two tryptophan metabolites, are now recognized as endogenous AhR ligands (Table 1.1) [28]. IAA and TA are relatively weak AhR ligands compared to TCDD, but can bind to and cause AhR translocation to the nucleus, subsequent DNA binding and stimulation of target gene induction [28]. In addition to these two tryptophan metabolites, two other high affinity AhR ligands produced from tryptophan metabolism are kynurenine and 6-formylindolo[3,2-b] carbazole (FICZ) [34, 46]. Kynurenine is generated from the first step of tryptophan degradation catalyzed by indoleamine 2,3-dioxygenase (IDO) [78, 79] whereas FICZ is formed from tryptophan photoproducts upon ultraviolet B (UVB) irradiation [26, 79, 80]. FICZ can also be generated from tryptophan upon exposure of visible light [81]. Both kynurenine and FICZ are high-affinity ligands since they require pM to nM concentration to activate the AhR [29], causing AhR translocation to the nucleus and upregulation of CYP1A1 transcription (Table 1.1) [80-82].

Another potential endogenous AhR ligand is 2-(1-H-indole-3-carbonyl)-thiazole-4-carboxylic acid methyl ester (ITE), which was isolated from porcine lung, and is generated by the condensation of cysteine and tryptophan [30, 83]. ITE is a high-affinity ligand that binds AhR and induces CYP1A1 expression with concentration ranges from pM to nM, thereby exhibiting similar potency for AhR binding as TCDD (Table 1.1) [30, 84].

1.1.3.1.4 AhR antagonists

In addition to agonist which activate AhR, there are a number of chemicals that exhibit partial or complete AhR antagonism. AhR antagonists vary in their binding affinity to the AhR and lead to
inhibition of AhR activation [85]. Some AhR antagonists have low affinity for AhR binding that ranges between μM to mM (Table 1.1) [62]. AhR antagonists that are high-affinity ligands have also been recognized, including N-(2-(1H-indol-3-yl) ethyl)-9-isopropyl-2-(5-methylpyridin-3-yl)-9H-purin-6-amine (GNF351), which constrains AhR activity in the nM range (Table 1.1) [85].

An example of a synthetic AhR antagonist is α-napthoflavone (ANF), which inhibits TCDD-induced AhR activity (Table 1.1) [86]. However, high ANF concentration results in ANF acting as a partial agonist of AhR [86]. 2-methyl-2H-pyrazole-3-carboxylic acid-amide (CH-223191) is a well-studied synthetic AhR antagonist (Table 1.1) [87]. CH-223191 does not act as a partial agonist of AhR [87]. Additionally, CH-223191 blocks AhR activity by preventing xenobiotic ligands (e.g., TCCD, B[a]P, and CS) from binding AhR [88, 89]. However, CH-223191 does not prevent other PAHs from binding AhR (e.g., benzo[a]anthracene and benzo[k]fluoranthene [90], suggesting that CH-223191 is a ligand-specific antagonist [87, 90].

Some AhR antagonists prevent formation of the AhR nuclear complex and subsequent CYP1A1 induction by TCDD. These antagonists include 3′-methoxy-4′-nitroflavone (MNF), 3′,4′-dimethoxyflavone (3′,4′-DMF) and 6,2′,4′-trimethoxyflavone (6,2′,4′-TMF) [34]. In addition to synthetic antagonists, naturally occurring AhR antagonists have been discovered, such as flavonoids in flowers, vegetables, and fruits (e.g., kaempferol, alpha-naphthoflavone and resveratrol) (Table 1.1) [91-93].
### Table 1.1 AhR ligands

<table>
<thead>
<tr>
<th>Ligand type</th>
<th>Affinity</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1.1.3.1.1 Xenobiotic AhR ligands</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAHs (e.g., (B[a]P), benz[a]anthracene,</td>
<td>between nM to µM concentration.</td>
<td>[29, 41, 42].</td>
</tr>
<tr>
<td>and dibenz [a, h] anthracene)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLCs (e.g., TCDD)</td>
<td>between pM to nM concentration.</td>
<td>[29, 34, 43]</td>
</tr>
<tr>
<td><strong>1.1.3.1.2 Other Exogenous AhR ligands</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacterial products and dietary compounds</td>
<td>between µM to mM concentration.</td>
<td>[46, 69]</td>
</tr>
<tr>
<td>(e.g., flavonoids and I3C)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1.1.3.1.3 Endogenous AhR ligands</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1.1.3.1.3.1 Heme metabolites</strong></td>
<td>within µM concentration.</td>
<td>[26, 34, 46, 70]</td>
</tr>
<tr>
<td>(e.g., bilirubin, biliverdin, and hemin)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1.1.3.1.3.2 Arachidonic acid metabolites</strong></td>
<td>LXA4- within nM concentrations. (PGs) need µM concentration.</td>
<td>[29, 72]</td>
</tr>
<tr>
<td>(e.g., LXA4, leukotriene and PGs)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1.1.3.1.3.3 Tryptophan metabolites</strong></td>
<td>IAA and TA- within µM concentration.</td>
<td>[28-30]</td>
</tr>
<tr>
<td>(e.g., IAA, TA, FICZ, kynurenine, and ITE)</td>
<td>-FICZ, kynurenine, and ITE- within pM to nM concentration.</td>
<td></td>
</tr>
<tr>
<td><strong>1.1.3.1.4 AhR antagonists</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e.g., GNF351, CH-223191, MNF, 3’,4’-</td>
<td>GNF351 and CH-223191- within nM</td>
<td>[34, 62, 85-87,</td>
</tr>
<tr>
<td>DMF, 6,2’,4’-TMF, ANF, and flavonoids)</td>
<td>concentrations. MNF, 3’,4’-DMF, 6,2’,4’-TMF, ANF, and flavonoids- between µM to mM.</td>
<td>91-93]</td>
</tr>
</tbody>
</table>
1.1.3.2 The Canonical AhR Pathway

Typically, the AhR is an inactive cytosolic receptor associated with a multiprotein chaperone complex, the function of which is to stabilize AhR in the cytoplasm [94]. This chaperone complex consists of two HSP90, prostaglandin E synthase 3 (p23) and immunophilin-like X-associated protein 2 (XAP2) [94]. The homodimer of HSP90 is essential to give the newly-synthesized AhR protein appropriate folding [94]. p23 has a significant role in protecting inactive AhR from degradation [94, 95] via the ubiquitin-proteasome system [94, 95]. Finally, XAP2 (also called ARA9 or AhR-interacting protein (AIP)) stabilizes unliganded cytosolic AhR, and stimulates the transcriptional activity of the AhR [94].

When a ligand crosses the plasma membrane, it binds to the AhR, resulting in AhR translocating from the cytoplasm to the nucleus and disassociating from its chaperone complex [96]. Upon dissociation, the AhR interacts with its partner the aryl hydrocarbon receptor nuclear translocator (ARNT) [96-98]. The active heterodimer of AhR-ARNT binds to specific DNA sequences, either ACGTG or GCGTG, located in the promoter regions of target genes [99]. These specific DNA recognition sequences are known as xenobiotic responsive elements (XRE) but are also called dioxin response element (DRE) or aryl hydrocarbon response element (AhRE) [99]. After the AhR-ARNT heterodimer binds to the core sequence of the XRE (5′-GCGTG-3′), transcription of target genes is induced [100-102]. The initiation of transcription is affected by coactivators (e.g., steroid receptor coactivator 1 (SRC1), nuclear coactivators (Ncoa1,2,3), CBP/p300 (cAMP response element-binding protein-binding protein), coactivator receptor-interacting protein (RIP), specificity protein (Sp1), retinoblastoma protein (RB), and promyelocytic leukemia (PML)) that facilitate the recruitment of RNA poly II to the AhR-
ARNT dimer [100-102]. This DNA binding by AhR: ARNT induces the transcription of target genes such as CYP1A1 and CYP1B1 [103]. Six members of the AhR gene battery were initially described to code for xenobiotic-metabolizing enzymes (XMEs) including CYP1A1, CYP1A2, NAD(P)H quinone oxidoreductase 1 (NQO1), aldehyde dehydrogenase (ALDH3A1), glutathione-S-transferase A2 (GSTA2), and UDP-glucuronosyltransferases (UGT) [104]. Nevertheless, the number of AhR genes that are induced upon AhR activation are approximately 600 [105-108]. Activation of the AhR also induces the transcription of suppressors of AhR expression and activity, including the AhRR and TCDD-inducible poly ADP-ribose polymerase (TiPARP) [109]. Finally, once transcription of the target gene is induced, AhR activity is terminated by dissociation from ARNT and is exported out the nucleus and back to the cytoplasm through the NES [20]. Once in the cytoplasm, the AhR is degraded by 26S ubiquitin-proteasome system [110] (Figure 1.2).
Figure 1.2 The canonical AhR pathway. Cytosolic AhR is found inactive in a complex with chaperones. Characteristics of the canonical AhR pathway are: (1) ligand binding, (2) AhR changes conformation, (3) nuclear translocation, (4) AhR-ARNT dimerization, (5) AhR-ARNT binding to XRE sequence in promoter target genes to activate transcription. The figure is based on [111].

1.1.3.3 The non-canonical AhR pathway

In addition to canonical AhR signaling, non-canonical pathways for the AhR have been described and include (1) protein interactions and (2) ligand-independent AhR signaling. These non-canonical pathways may be crucial to understanding the endogenous regulatory roles for the AhR in normal physiology and development.
1.1.3.3.1 Protein interaction with AhR

The AhR can dimerize with a wide range of proteins that results in crosstalk between the AhR and other intracellular signaling pathways [112]. One of these pathways involves nuclear factor-κB (NF-κB). The NF-κB family of proteins plays prominent roles in coordination of the inflammatory response, innate and adaptive immunity, cell survival, proliferation, and differentiation [113]. Five family members of the NF-κB pathway have been identified in mammals; these include RelA (p65), RelB, cRel, NF-κB1 (p50), and NF-κB2 (p52) [113]. These proteins form homo- or heterodimers that bind DNA and are divided into two main signaling pathways: the canonical and the non-conical pathways [113, 114]. In the canonical NF-κB pathway, the heterodimer of p65/p50 is inactive in the cytoplasm due to sequestration by inhibitory proteins known as inhibitors of NF-κB (IκBs) [113, 114]. Upon stimulation, these inhibitors are degraded by the proteasome [113, 114]. Due to IκB degradation, these NF-κB dimers are released and translocate to the nucleus, resulting in gene transcription [113, 114]. This canonical pathway of NF-κB is stimulated by a wide variety of proinflammatory agonists including interleukin 1(IL-1), lipopolysaccharides (LPS), and tumor necrosis factor-α (TNF-α) [113, 114]. Conversely, the non-canonical NF-κB pathway is activated by a selected number of agonists including CD40 ligand, B cell-activating factor (BAFF) and lymphotoxin beta (LTβ) [113-115]. This pathway is comprised of p52/RelB heterodimers [113-115]. Under normal conditions, RelB is sequestered by the inhibitory protein p100 in the cytosol [113, 115, 116]. However, activation induces p100 processing into p52, allowing p52:RelB dimers to be released [114-116]. p52/RelB then translocates to the nucleus to bind to DNA, also resulting in gene transcription [114, 115].
The AhR interacts with RelA [p65], where heterodimerization with the AhR to RelA results in binding to NF-kB elements and transactivation of the c-Myc gene [117, 118]. c-Myc is an oncogene that belongs to the bHLH transcription family [17]. c-Myc regulates the expression of a variety of genes involved in cellular proliferation and thus is implicated in cancer formation [119, 120]. Crosstalk between the AhR and RelA may promote breast cancer proliferation and tumorigenesis [117]. A similar association between AhR-RelA has been shown in lung cancer, in which the AhR-RelA complex induces IL-6, a pro-inflammatory cytokine that has a role in tumor progression [121] (Figure 1.3).

Another example of protein interaction with the AhR involves the estrogen receptor (ER) [122, 123]. Estrogens are classically perceived as ovarian sex hormones responsible for cell proliferation and tissue growth of the reproductive system [124]. The ER is a member of a nuclear family of intercellular receptors that plays a vital role in biological, physiological, and pathological processes [124, 125]. Three types of estrogen-bound receptors have been discovered, including estrogen receptor alpha (ERα), estrogen receptor beta (ERβ), and the G protein-coupled estrogen receptor (GPER1) [124]. ERα and ERβ are nuclear estrogen receptors, while GPER1 is a membrane estrogen receptor [124]. The classical genomic (canonical) estrogen pathway is mainly mediated by ERα and ERβ acting as ligand-activated transcription factors [124]. Upon ligand-binding (e.g., 17β-estradiol (E2)) to the ERα and ERβ in the cytoplasm, a conformational alteration results in complex dimerization [124, 126]. This dimer translocates to the nucleus, where it binds to the estrogen-responsive element (ERE) and induces target gene transcription [124, 126]. ERs can also bind DNA through protein-protein dimerization with other classes of transcription factors at their respective response elements and
activate or suppress target gene transcription [124, 126]. In contrast, the non-genomic (noncanonical) estrogen pathway is mostly mediated by the membrane-bound receptor and certain variants of nuclear receptors [124, 126]. The non-genomic signaling is associated with cytoplasmic cascades, including the production of intracellular second messengers, activation of protein-kinases, and phosphorylation of transcription factors [124, 126]. These events result in indirect alterations in gene transcription [124].

Environmental toxicants can negatively influence ER function through receptor crosstalk, including those that bind AhR [122]. TCDD, for example, induces physical interaction between AhR and ERα human breast cancer cells [122]; TCDD inhibits ER activity by activating nuclear AhR that recruits ERα and the proteasome complex [122]. These data showed that the activated AhR inhibits the activity of ER through degradation as E3 ligase ubiquitination [122, 127, 128] (Figure 1.3).
Figure 1.3 Non-genomic AhR pathway. Alternative or nongenomic pathways of AhR occur due to the AhR forming complexes with other proteins such as the ER and NF-κB protein RelA. AhR crosstalk with other proteins can result in the induction of gene transcription such as IL-6 and c-Myc. The figure is based on [62].

1.1.3.3.2 Ligand-independent AhR signaling.

A component of the non-canonical AhR pathway may occur through ligand-independent signaling. For example, the AhR regulates cell proliferation in mouse embryonic fibroblasts (MEFs) where MEFs expressing AhR have a higher cell proliferation rate than the MEFs that lack AhR, even in in the absence of exogenous ligands. This was due to AhR prolonging the cell cycle, particularly the G1 and S phases [129]. As such, this ability of the AhR to control
fibroblast proliferation was proposed to be ligand-independent. Further evidence came from studies showing that differences in proliferation rates are not influenced either by exposure to AhR ligands or depletion of the ligand-binding domain of AhR [130]. In fact, immortalized hepatocytes lacking the capacity for AhR ligand binding showed that the AhR can still interact with ARNT and induce the transcription of target genes [131, 132]. Hence, these findings support that ligand-independent activation of the AhR is important in regulating cellular proliferation, further implicating a role for the AhR beyond its ability to mediate toxic responses to xenobiotics [131, 132].

1.1.4 AhR Expression

After its discovery, a question about the distribution of the AhR among mammalian species and within tissues emerged. Expression of the AhR has now been characterized in numerous species including humans, mice, hamsters, rats, rabbits, and monkeys [7, 133-136] as well as various nonmammalian species [134, 137-139] including fish, birds, and invertebrates [140, 141]. Most animals have one AhR isoform, except birds and fish which express two AhR isoforms: AhR1 and AhR2 [140, 141].

The AhR has variable tissue and developmental expression. AhR expression has been characterized in human fetal and adult tissues. In the fetus, the AhR is expressed ubiquitously in the kidneys, testicles, liver, retinas, esophagus, thymus, and lungs [32, 142, 143]. However, AhR is expressed at a low level in some fetal tissues including the brain, sclera, heart, aorta, and choroids [32, 142, 143]. The expression level of AhR in the adult is high in the liver, placenta, pancreas, lung, and spleen, but lower in skeletal muscles, brain, and heart [32, 142, 143].
Developmental expression of AhR is available for several animal species. It is known that at birth, AhR is expressed at a high level in prostate, but rapidly decreases before weaning [144]. Carlstedt-Duke et al. found during postnatal development around the time of weaning, AhR expression in the liver is highest but decreases with age [136, 145]. Gasiewicz et al. further supported this observation; after weaning, the level of AhR decreases in the liver and lung, while the AhR level stays high in the thymus [146]. Thus, the AhR is highly expressed in younger animals and gradually decreases with age in some organs [62, 147].

Furthermore, the level of AhR is differentially expressed in immune cells, with expression being relatively high in T helper (Th) 17 cells, mast cells, DCs, macrophages, and plasma cells [62, 148]. Natural killer (NK) cells and B lymphocytes express AhR at moderate levels [62, 149]. In non-hematopoietic cells, AhR expression is constitutive, including in various cells such as fibroblasts, keratinocytes, endothelial and epithelial cells [150-152].

1.1.5 Regulation of AhR Expression

Although the AhR is expressed in a wide range of various cells and tissues, the mechanisms controlling AhR expression are not fully understood. It is important to understand what dictates AhR expression, given that is appears to be indispensable for numerous physiological functions.

1.1.5.1 Transcriptional regulation of AhR

Transcription factors (TFs) are one mechanism that can impact AhR expression, although few studies have shown TFs regulate AhR expression. One group of TFs that can affect AhR expression is Sp1. Sp1 a zinc-finger transcription factor with high affinity for binding guanine
and cytosine (GC)-rich sequences in the promoter of numerous genes, results in activation of transcription [153, 154]. Sp1 interacts with any four Sp1 consensus motifs within the mouse Ahr gene 5' promoter region [154], supporting that Sp1 regulates the basal expression of Ahr mRNA [153, 154].

Another transcription factor that may control AhR expression is peroxisome proliferator-activated receptor-α (PPARα), a transcription factor that is necessary for cellular differentiation and regulation of lipid metabolism in the liver [155]. PPARα induces AhR expression and its target gene CYP1A1 in human Caco-2 cells treated with the potent PPARα agonist WY-14643 [156]. This finding was confirmed using PPARα-deficient mice, where there was no change in AhR expression upon WY-14643 exposure [156]. This supports that PPARα influences expression of the AhR [156].

Another transcription factor that can control AhR expression is RelB, a member of the NF-κB family. RelB interacts with the AhR promoter [157, 158] to increase AhR expression and activation [158]. A functional consequence of the RelB-dependent increase in AhR results in the ability of RelB to attenuate lung cell apoptosis by CS [158]. Thus, control over AhR expression by TF binding likely contributes not only to its expression, but the ability of the AhR to control various cellular processes.

1.1.5.2 Suppressors of the AhR

Two AhR target genes, AhRR and TiPARP, are negative regulators of AhR activity and expression [159]. There is similarity in the structure of AhR protein and AhRR; however, AhRR
lacks the PAS-B domain in the N-terminal site, meaning that the AhRR cannot bind to a ligand [160-162]. Instead, the AhRR has a transrepression domain in the C-terminal region, unlike the AhR, which has a transactivation domain [160]. One way that the AhRR influences activity of the AhR is by competing with its ability to interact with ARNT, thereby decreasing transcription of AhR target genes [162, 163]. As expected, in AhRR overexpressing mice, the induction of CYP1A1 upon TCCD exposure is diminished in adipose tissue, kidneys, and spleen [164].

However, the AhRR may also negatively affect AhR activity in an ARNT-independent fashion. Findings support that ARNT overexpression does not attenuate the ability of the AhRR to repress AhR activity [165]. Overexpression of AhRR could also disrupt AhR activity by interfering with AhR-protein interactions. Vogel et al. illustrated that TCDD-induced CCAAT-enhancer-binding proteins (C/EBP) and NF-κB DNA binding were reduced in AhRR overexpressing mice compared to TCDD-treated WT mice [164]. This suggested that AhRR competes with AhR to bind with NF-κB and C/EBP [166, 167]. This inhibitory mechanism is called transrepression, which is caused by interfering of AhRR with AhR-protein dimerization [164]. Indeed, this negative feedback could be through the non-canonical pathway of AhR.

Another negative suppressor is TiPARP, which is an AhR target gene induced by TCDD [168]. TiPARP negatively affects AhR expression and activity by promoting AhR protein degradation [168]. MEFs deficient in TiPARP showed an increase in AhR protein expression [168]. Conversely, overexpression of TiPARP facilitated AhR protein degradation and decreased its expression [168]. These findings suggest that the AhR expression can be negatively regulated via the AhR target genes AhRR and TiPARP.
1.1.5.3 Epigenetic regulation

Epigenetic modifications also affect the AhR expression. Epigenetic modifications are defined as heritable changes in gene expression without changes in the DNA sequence [169]. There are three central mechanisms of epigenetic modifications: histone modification, DNA methylation, and non-coding RNAs (ncRNAs). These modifications control gene expression and chromatin structure throughout normal development and in a variety of diseases [169].

It is well-known the DNA is packed into chromatin in eukaryotes [170-172]. Chromatin is a structural component of the cell nucleus and a complex that consists of DNA and proteins [170-172]. The primary unit of chromatin is the nucleosome that has 147 base pairs of DNA wrapped around eight histone proteins involving H2A, H2B, H3, and H4 [170-172]. These four histone proteins are known as the histone octamer, which helps to organize DNA into a structural nucleosome linked by the H1 protein [170-172]. In addition, chromatin has two types: heterochromatin and euchromatin [170-172]. Heterochromatin is very condensed and cannot be transcribed whereas euchromatin is less condensed and can be involved in transcription [170-172]. Chromatin undergoes structural changes and opens for polymerases and TFs in order for gene transcription to occur [173, 174]. Changes in chromatin occurs because of the negative charge of RNA, which is able to neutralize histone tails that have a positive charge, resulting in chromatin de-compaction [173, 174].

1.1.5.3.1 Histone modifications

Some modifications can impact histone proteins and result in the activation or silencing of gene expression [175]. Examples of such histone modifications are methylation, acetylation,
ubiquitylation, and phosphorylation [175]. In normal conditions, activated genes are associated with methylation at histone H3K4 and H3K36, while silenced genes are associated with methylation at histone H3K9, H3K27, and H4K20 [176]. The human genome encodes a diversity of histone-modifying enzymes that modify histones and regulate genes; these include histone methyltransferases (HMT), histone acetyltransferases (HATs), and histone deacetylases (HDACs) [176, 177]. As such, histone modifications help to recruit these regulators to chromatin, which controls transcription of genes [175, 178].

Histone modification can impact AhR expression. Previous study has demonstrated that AhR expression is increased in long-term culture of MCF-7 cells with estrogen (approximately 4–20 months of continuous cell culture). This increase in AhR expression was due to a reduction in histone modifications, specifically trimethylation of histone three at lysine 27 (H3K27me3) at the AHR promoter [179].

1.1.5.3.2 DNA methylation

DNA methylation is an epigenetic modification that requires covalent addition of a methyl group (CH3) to the carbon at position five of the cytosine ring of DNA [169, 180, 181]. In mammalian cells, DNA methylation usually occurs at a cytosine followed by guanine (CpG dinucleotide) that are usually found in clusters called CpG islands [169, 180, 181]. DNA methylation is established and maintained by DNA methyltransferases (DNMTs). DNMTs catalyze DNA methylation by transferring methyl groups from S-adenosyl-methionine to the cytosine ring [169, 180-182]. CpG islands are found unmethylated in the promoter regions of active genes, whereas methylation of CpG islands causes silencing of gene transcription [169, 180, 181].
Changes in DNA methylation of the AhR protomer disrupts its expression. For example, a decrease in AhR expression was seen in human acute lymphoblastic leukemia cells due to hypermethylation in the AhR promoter region, which prevents binding of Sp1 that is necessary to maintain constitutive AhR levels [183]. Additionally, alterations in DNA methylation in the AhR promoter region negatively affects its expression in non-mammalian species, as detected in PBC-exposed Atlantic killifish [184]. Two AHR isoforms are expressed in killifish: AhR1, which is expressed in a tissue-specific manner, and AhR2, which is ubiquitously expressed. The CpG island of the AhR1 isoform was hypermethylated within the promoter and was associated with lower AhR1 expression [184]. In contrast, the CpG island of AhR2 isoform was hypomethylated in the AhR2 promoter and associated with abundant hepatic AhR1 expression [184]. Collectively, these findings support that AhR promoter methylation can influence AhR expression in multiple species.

1.1.5.3.3 Non-coding RNAs (ncRNAs)

ncRNA are types of RNA do not translate into proteins. In 1960, the first regulatory ncRNAs were discovered in bacteria [185], with the first eukaryote IncRNAs being found in late 1980s [185]. ncRNAs are now classified into housekeeping and regulatory non-coding RNAs [186]. Examples of housekeeping ncRNAs are snoRNAs, small nuclear (snRNA), transfer (tRNA), and ribosomal (rRNA) [186]. Regulatory ncRNAs are grouped into short ncRNAs (<200 nucleotides), which includes miRNAs, piwi-associated RNAs (piRNAs), small interfering RNAs (siRNAs), and long ncRNAs (lncRNAs) which are > than 200 nucleotides [186]. microRNA (miRNA) is ncRNA that guide messenger RNA (mRNA) to translational repression and induce mRNA cleavage. Numerous studies now show that miRNAs control expression of the AhR [187-189]. For example, miR-124 inhibits AhR protein levels by directly targeting the AHR 3’-untranslated region (3’-
Additional evidence that miRNAs control AhR expression comes from studies in A549 and HepG2 cells treated with TCDD or B[a]P, which show that miR-203 suppresses AhR expression at both the mRNA and protein levels through interaction with the 3'UTR of AhR promotor [189]. In this study, overexpression of miR-203 not only reduced AhR expression, but also reduced downstream target gene expression (e.g., CYP1A1, CYP1A2, NQO1) [189]. These results implicate miRNA in the regulation of AhR expression. In addition to miRNA, recent evidence supports that IncRNAs also control AhR expression. Changes in expression of the AhR occur with knockdown of the IncRNA AK170409 [190]. However, the mechanism through which IncRNA control AhR expression is still not well understood.

1.1.6 General and cellular physiological functions of AhR
1.1.6.1 Xenobiotic metabolism

AhR activation leads to the induction of AhR target genes, which includes many XMEs [42, 29]. XMEs are grouped into Phase I and Phase II proteins [35, 36]. Phase I XMEs include CYP1A1, CYP1A2, and CYP1B1, and Phase II enzymes include NQO1, ALDH3A1, GSTA2, and several UGT such as UGT1A1 and UGT1A6 [42, 191, 192]. Phase I XMEs oxidize PAHs into polar compounds [191] whereas phase II XMEs conjugate these polar compounds to form metabolites that are more water soluble for excretion from the body [191, 193]. Typically, XMEs have a crucial role in the metabolism and detoxification of PAHs [42, 191]. In contrast, the metabolism of several HAHs is not affected by XMEs, especially DLCs [29]. XMEs also metabolize food products and endogenous compounds such as arachidonic acid, tryptophan, and heme metabolites [29, 71, 194].
1.1.6.2 Organ development

The AhR has a prominent role in the development of various organ systems, including the heart, liver, and nervous system. The embryonic heart of Ahr-deficient mice have a reduction in weight and changes in myocardial development, including a decrease in ventricle wall thickness, compared to wild-type (WT) mice [195]. Interestingly, WT mice exposed to TCDD showed similar alteration in the heart as that seen in Ahr-deficient mice [195]. The AhR is also required for vascular remodeling and developmental closure of the ductus venosus (DV) [196]. In the fetus, the DV closes typically after birth and acts as a shunt to direct oxygenated blood from the placenta away from the fetal liver to the inferior vena cava [196]. Failure of DV to occlude after birth leads to the development of a patent DV, which results in liver disease in adulthood [197]. In support of its importance in liver development, mice with a mutation which prevents AhR nuclear localization or DRE binding demonstrate abnormalities in the development of hepatocytes, such as a reduction in liver weight and the presence of the DV [198, 199]. Ahr-deficient mice also exhibit a reduction in liver weight as well as hepatic portal fibrosis [200, 201]. These findings highlight that the development of normal liver requires the canonical AhR pathway [199]. Taken together, these outcomes suggest that the AhR is required for heart and liver development as well as function.

An important role for the AhR in development of the central nervous system (CNS) has also emerged. In the early developmental stages of neuronal cells, AhR is expressed in neural progenitor cells in the hippocampus, and its expression increases throughout development, suggesting that the AhR may participate in neuronal cell differentiation [202]. Knockout (KO) studies demonstrate that disruption of AhR expression impairs hippocampal neuronal differentiation, leading to memory defects [203]. Additionally, AhR WT mice exposed to TCDD
have similar effects in memory to KO mice [203]. Other findings indicate that ablation of the AhR in cerebellar granule neuron precursor cell suppresses the growth and development of nervous tissue by inhibiting proliferation and increasing differentiation [204]. These results demonstrate that AhR deletion as well as AhR activation by TCDD may disrupt the physiological functions of AhR in the CNS [203].

1.1.6.3 Immunity

The AhR has important roles in immune regulation [205]. AhR is expressed in most types of immune cells, including macrophages, DCs, granulocytes (neutrophils), NK cells as well as T and B lymphocytes [206]. In microbial defense, the AhR modulates host defense against bacterial infection, including by *Streptococcus pneumonia*, *P. aeruginosa*, and *M. tuberculosis* [69, 207, 208]. In support of its role in controlling infection, mice lacking AhR are more susceptible to these microbial pathogens than mice that express the AhR [69]. This may be due to the ability of the AhR to bind to bacterial virulence factors that leads to the AhR activation [69]. AhR activation results in the degradation of bacterial virulence factors and production of pro-inflammatory cytokines and chemokines [69]. AhR activation also initiates the recruitment of neutrophils to an infected site in order to control bacteria replication [69]. Thus, the AhR plays a role in regulating the innate immune system by detecting the presence of pathogens [69].

The AhR also regulates the adaptive immune system. Here, the AhR promotes B cell proliferation, a central part of humoral immunity [209]. AhR deficient mouse B cells have a slow proliferation rate compared to the cells which express AhR [209, 210]. Additionally, the
AhR governs B cell fate decisions, including differentiation into either plasma B cell or memory B cell [210]. This is important, as memory B cells provide faster antibody production in response to invading pathogens upon antigen re-exposure [211]. In addition, the AhR regulates T-cell differentiation in a ligand-dependent manner [212]. For example, activation of the AhR by TCDD promotes T regulatory (Treg) cell differentiation, which limits the immune response in autoimmune, allergic and infectious diseases (e.g., influenza infection, autoimmune uveoretinitis) [206, 213-216]. However, activation of the AhR by FICZ promotes Th17 differentiation, which produces IL-17 and IL-22 to stimulate the immune response [217]. Together, these results support that the AhR has important roles in homeostatic maintenance of various physiological function.

1.1.6.4 AhR regulation of ncRNAs expression

Physiological functions of AhR are still not well-understood but may involve ncRNAs. One study using AhR deficient cells showed both downregulation and upregulation of numerous miRNAs compared to WT cells [88, 218]. Functionally, this control over miRNA may explain some function of AhR. For example, AhR reduces CS-induced apoptosis through regulation of miRNA-196a expression [88]. Besides miRNA, lncRNA may be regulated by AhR. Previous study showed that AhR binds to promoter region of TiPARP and regulates its expression in MCF-7 cells [219]. Reporter gene assay showed that AhR not only drives TiPARP expression, but also its IncRNA TIPARP-AS1 by binding to the AHRE cluster II [219]. These studies showed the effect of AhR on the expression of these ncRNAs, however further studies are needed to understand the role of the AhR in regulating cellular functions through ncRNAs.
Even though there remains limited information on the role of the AhR in regulating IncRNA, both the AhR and IncRNA control shared biological functions including inflammation, proliferation, differentiation, apoptosis, migration, and invasion [33, 62, 209, 220-226]. Furthermore, there is dichotomy in the AhR response to ligands, suggesting that changes in gene transcription alone cannot explain ligand-specific effects. For example, the consequence of AhR activation on osteoclast differentiation occurs both in response to different AhR ligands as well as in the concentration of the ligand [227-232]. In response to low concentrations of B[a]P (0.5 μM), the AhR promotes osteoclast differentiation, whereas differentiation is inhibited with high concentration of B[a]P (10^{-5} M) [232, 233]. Another example occurs in response to TCDD, where a concentration of 10 nM TCDD does not impact osteoclast differentiation for a short exposure (48h) [234, 235]. However, AhR inhibits osteoclast differentiation upon long exposure (after 7 and 10 days) of TCDD in a dose-dependent manner[234, 235]. How the AhR exerts differential effects based on ligand exposure remains unclear. It is noteworthy that IncRNA can also affect differentiation [236-242]. For example, HOTTIP has a vital role in developing bone and differentiation of bone marrow mesenchymal stem cells (BMSCs) [242]. HOTTIP expression correlates with the increased deposition of calcium and level of osteogenic markers such as runt-related transcription factor 2 (Runx2), a transcription factor, and the primary regulator of osteoblast differentiation [242, 243]. HOTTIP may also participate in bone formation through activation Wnt/β-catenin pathway that contributes to development processes and regulation of cell fate during embryogenesis [242, 244]. Thus, it is possible that there is a relationship between IncRNA and AhR, and that its ligands might contribute to numerous physiological processes and pathologies via changes in IncRNA expression and/or function. Indeed, the role of the AhR in the regulation of toxic versus physiological outcomes are still elusive and may not be the result of the
classical AhR signaling pathway. This is further supported by discoveries that several complex interactions exist between the AhR and other cellular components, including ncRNA, the Wnt/β-catenin pathway and NF-κB pathway [117, 118, 121-123, 245]. Thus, ligand-dependent regulation of IncRNA through the AhR requires explorations due to the fact that a wide range of AhR ligands produce diverse functional outcomes [25-34]. Such investigations may lead to a better understanding of roles of AhR beyond its ability to regulate xenobiotic metabolism.

1.2 Long non-coding RNAs (IncRNAs)

The transcriptional genome of all organisms is composed of a collection of protein coding and ncRNA [246, 247]. It is estimated that more than 85% of the human genome is transcribed into RNA that do not code for protein [248]. In fact, ncRNA represent the largest group of RNA produced from the genome [185]. IncRNA are a class of ncRNAs that contain more than 200 nucleotides in length and do not translate into protein [185, 247, 249]. According to NONCODEV5 database, there are 131,697 mouse IncRNA and 172,216 humans [250]. In addition, 27,919 human IncRNA have been identified in 1,829 samples from a variety of human primary cells by using functional annotation of mammalian genomes 5 (FANTOM5) [251]. Generally, a large fraction of IncRNAs shares common characteristics with mRNA [185, 246, 249] such as the presence of promoter regions, transcription by RNA pol II, and subsequent post-transcriptional modifications (splicing, capping on the 5′-end with methyl-guanosine, and 3′-polyadenylation) [185, 246, 249]. However, IncRNAs differ from mRNAs in that IncRNA do not have polyadenylation tails or open reading frame (ORF). Also, IncRNA have fewer and longer exons compared to mRNA [185, 246, 249] and most IncRNAs are in low abundance, with better cell-, tissue-, and disease-specificity in comparison to mRNAs [252] [253].

49
1.2.1 Subcellular location of lncRNAs

lncRNAs can be located in the nucleus, cytoplasm, or both [220]. However, most intronic lncRNA and lncRNA that lack poly-A tails are located in the nucleus [220]. lncRNA have varying patterns of cellular localization [254], with five major groups of lncRNA being identified based on their subcellular localization; these include lncRNA with: (i) large nuclear foci; (ii) large nuclear foci with single fragments distributed over the nucleus; (iii) predominantly nuclear deprived of foci; (iv) predominantly cytoplasmic; and (v) nuclear and cytoplasmic [254]. Nuclear lncRNAs occur within particular subnuclear structures, such as nucleoli, paraspeckles, and chromatin speckles [220] whereas cytoplasmic lncRNAs are found in the mitochondria, extracellular membranes, exosomes, and ribosomes [220].

1.2.2 Classification of lncRNAs
1.2.2.1 Classification of lncRNAs based on biogenesis

lncRNAs can be classified based on their genomic proximity to protein-coding genes [220, 249]. lncRNA are sub-grouped into five main classes, including 1) antisense lncRNAs or natural antisense transcripts (NATs); 2) sense lncRNAs; 3) intronic lncRNAs; 4) long intergenic RNAs (lincRNAs), and 5) bidirectional lncRNAs [220, 249]. The most common classes of lncRNAs in humans are antisense and intergenic lncRNAs [220].

**Antisense lncRNA** are transcribed from the opposite strand of protein-coding genes, and have sequence elements which can base pair to other RNA [220]. lncRNA can also be transcribed from the 3’-end of a protein-coding transcript or in the opposite direction of protein-coding transcripts [220]. In the mouse genome, 87% of coding genes have antisense lncRNAs, while in the human
genome, approximately 32% of lncRNAs are antisense to coding transcripts [255, 256]. Sense lncRNA are transcribed from the sense strand of protein-coding transcripts, containing exons from protein-coding genes [220]. These might cover the entire protein-coding sequences through an intron, or they may overlap with part of protein-coding transcripts [220]. Intronic lncRNAs are transcribed from introns of protein-coding transcripts without overlapping with exons of protein-coding genes [220]. Intronic lncRNAs may be transcribed from both the sense and antisense strands of intergenic regions [257]. Intergenic lncRNA are transcribed from intergenic regions between two protein-coding transcripts [220]. A large portion of intergenic lncRNAs is considered enhancer RNA (eRNA) because they are located in the enhancer regions of protein-coding genes [220]. Bidirectional lncRNA are transcribed from the opposite strand of protein-coding genes and located 1 kb away from the promoter of protein-coding genes in the opposite direction [220]. Bidirectional lncRNA are also called promoter-associated non-coding RNAs (pancRNAs) [258].

1.2.2.2 Classification of lncRNA based on whether they act in cis or in trans

One group of nuclear lncRNA has the ability to regulate target gene expression [252]. This group can be divided based on their action into two subgroups: cis- meaning they are located at the same chromosome, or trans, meaning they are located at the same or different chromosomes [249, 259]. LncRNA that work in cis and nearby their transcription sites could modulate one gene or many genes, which are locally present on the same chromosome [249, 259]. However, trans-regulatory lncRNA can work at a distance from their transcription site to regulate the expression of numerous genes, which can be globally present in the same or different chromosomes [249, 259].
1.2.2.3 Classification of lncRNAs based on their mode of action

Based on the molecular mechanisms of action, lncRNAs can also be categorized into decoy, scaffold, and guide [249, 260]. Decoy lncRNA bind and capture different molecules, including proteins, transcription factors, and other regulatory RNA, which result in inhibition of functions [249, 260]. Decoy lncRNA can positively or negatively affect transcription [249, 260]. Decoy lncRNA can titrate and prevent transcription factors or repressors from binding their target gene promoters [249, 260] (Figure 1.4). Scaffold lncRNA have binding sites that interact with distinct effector molecules and thus can serve as a platform for connecting various protein complexes [249, 260]. The action of scaffold lncRNA can be through interacting and recruiting different molecules such as protein complexes and chromatin-modifying complexes to their target gene promoters [249, 260]. Consequently, this action can carry out activation or suppression of the transcription [249, 260] (Figure 1.4).

Guide lncRNA have a capacity to induce or silence gene expression through binding with transcription factors and guiding them to a direct location, such as the promoter of target genes [249, 260]. Additionally, this type of lncRNA can also interact with regulatory molecules such as chromatin-modifying enzymes and ribonucleoprotein complexes (RNP) which recruits them to target gene promoters or loci to promote or suppress gene transcription [249, 260]. Guide lncRNA can therefore regulate gene expression by acting either in cis or trans [249, 260] (Figure 1.4).
1.2.3 *Function of lncRNAs*

lncRNA have various functions based on their subcellular localization [261, 262]. Some lncRNA require exportation to the cytoplasm in order to execute their function [262]. The following section describes the mechanisms of lncRNA function.

1.2.3.1 *Role of lncRNAs in chromatin modifications and transcription*

lncRNA can guide chromatin to specific genomic regions to regulate gene expression by acting either in cis or trans [222, 223, 263] (**Figure 1.4**). For example, the best-characterized lncRNA to mediate chromatin regulation and gene suppression is Xist [264]. Xist is transcribed from the X chromosome and is involved in X chromosome inactivation (XCI), a process that silences one X chromosome in female somatic cells [265, 266]. Xist is also expressed from the X inactive chromosome of female cells, leading to dysregulation of dose compensation, which is a regulatory process needed reach an equal expression of sex chromosome genes [265, 266]. The reason for this is that female mammalian cells have two X chromosomes (XX), while males have one chromosome (XY) [265, 266]. In order for female cells to express an equal amount of X-linked genes as male cells, one X-chromosome in each female cell becomes inactivated [265, 266]. Xist causes an alteration in the inactive X chromosome chromatin through its physical interaction with the polycomb repressive complexes 2 (PRC2) [264]. PRC2 proteins control chromatin structure to regulate gene expression [267]. The primary function of PRC2 is transcriptional suppression by introducing chromatin marks, in particular H3K27me3 [267]. The mammalian PRC2 group is composed of five core components, including extra sex combs (Eed); enhancer of zeste (Ezh1/2); a nucleosome remodeling factor (Rbbp7/4); Jumonji, ARID domain-containing protein 2 (Jarid2); and suppressor of zeste 12 (Suz12) [267]. Physical interaction between Xist and PRC2 is through
a 1.6-kb ncRNA, Repeat A (RepA), which has specific sequences found in the 5' region of the Xist transcript and bound to Ezh2 [264]. As a result, RepA recruits PCR2 mediated histone mark H3K27me3 to Xist and leading to methylation of the promoter of Xist and activation of the inactive X chromosome located in cis [264]. Other lncRNA that have similar mechanisms include HOXA transcript (HOTTIP), Air, and COLDAIR [263]. Other examples of lncRNA that mediate chromatin remodeling [222, 223, 268] are HOTAIR, Jpx, and lincRNA-p21 [269].

The best-studied lncRNA in trans is HOX antisense intergenic RNA (HOTAIR), located on chromosome 12 [270]. HOTAIR is transcribed from the HOXC locus [270]. HOX genes are grouped into four different categories called HOXA, HOXB, HOXC, and HOXD [271]. HOX genes are highly conserved and include 39 genes which encode for HOX proteins, homeodomain transcription factors that activate or repress gene expression [271]. The main function of HOX proteins is to regulate morphogenesis [271, 272]. HOTAIR acts as a transcriptional suppresser of the HOXD locus [271] by interacting with the subunit of PCR2 called SUZ12, resulting in localization H3K27me3 and subsequent inhibition of HOXD gene transcription [273].

lncRNA can also increase DNA methylation by binding with DNMTs [274]. Three active DNMTs have been recognized in mammals: DNMT3a, DNMT3b, and DNMT1 [275]. LncRNA can affect the regulation of gene expression through the recruitment of DNMTs to their target gene [274]. ATB, located on human chromosome 14, is a lncRNA activated by transforming growth factor-beta (TGF-β) [274]. ATB promotes the proliferation and inhibits apoptosis in part by inhibiting the tumor suppressor p53 at the mRNA and protein level [274, 276, 277]. Mechanistically, ATB
binds to DNMT1 to increase its expression, resulting in downregulation of p53 via binding to the promoter region [274].

lncRNA can also modulate gene expression by directly interacting with either RNA Pol II or TFs [222, 223, 268]. lncRNA interfere with binding site of RNA pol II to the promoter region or with coactivators [278]. For example, growth arrest-specific 5 (GAS5) is a lncRNA encoded by the GAS5 gene on human chromosome 1 [279, 280]. GAS5 acts as a decoy for specific DNA sequences located on response elements of the activated glucocorticoid receptor (GR) [279, 280]. GR is a ligand-binding receptor and a transcription factor that regulates genes controlling development, immune responses, metabolism, and cell survival [280]. GAS5 has a hairpin sequence motif that resembles the DNA-binding site of the GR response elements [279, 280]. GAS5 represses the GR by competing for the DNA binding site of GR response element, leading to inhibition of target gene expression [279, 280] (Figure 1.4).

1.2.3.2 Role lncRNA in post-transcriptional regulation

In addition to transcriptional regulation, lncRNA regulate post-transcriptional events [222, 223, 268] to control mRNA function by changing its stability, splicing patterns, and translation [268].

1.2.3.2.1 LncRNA and mRNA stability

The stability of mRNA is an important mechanism that controls gene expression by balancing production with degradation of the RNA [281, 282]. mRNA stability is controlled in part by RNA-binding proteins (RBPs) [281, 282], a large group of more than 2,000 proteins including human antigen R (HuR), heterogeneous nuclear RNA-binding protein LL (hnRNPLL), and polyadenylate-binding protein-interacting protein-2 (PAIP2) [283-285]. The ability of lncRNA to
regulate mRNA stability at the post-transcriptional level have recently been identified [286] (Figure 1.4). lncRNA can alter mRNA stability by interacting with a specific sequence motif of an RBP, resulting in the formation of lncRNA-protein complexes (lncRNPs) [287]. For example, non-coding RNA activated by DNA damage (NORAD) is a lincRNA located on human chromosome 20. NORAD is conserved and highly-expressed in various species and cells [287]. NORAD is a 5.3kb transcript with multiple conserved binding sites for RBPs [287], which function to maintain genome stability through regulation of Pumilio-Fem3-binding factor (PUF) activity [287]. PUF belongs to a family of RBPs which negatively regulates gene expression [288, 289]. PUF proteins possess PUMILIO homology domains, which can interact through specific sequences in the 3' UTRs of target mRNA [288, 289]. NORAD acts as a molecular decoy for PUF proteins by binding them through conserved elements found in the NORAD repeats, thereby preventing them from binding to mRNAs [287].

Another example of a lncRNA that regulates mRNA stability is metastasis-associated in colon cancer antisense (MACC-AS). Metastasis-associated in colon cancer 1 (MACC1) is an oncogene initially discovered in colon cancer [290]. The expression of MACC1 is increased in different types of cancer such as osteosarcoma, gallbladder cancer, and gastric cancer (GC) [290-293]. MACC1-AS1 has a binding site for MACC1 mRNA, which positively affects the expression and stability of MACC1 mRNA through interacting with MACC1-AS1 binding site [286]. Further analysis has shown that MACC1 mRNA has a binding site for Lin28, a RBP involved in post-transcriptional regulation [286, 294]. MACC1-AS1 phosphorylates 5' adenosine monophosphate-activated protein kinase (AMPK), thereby increasing Lin28 translocation from the nucleus to the
cytoplasm [286]. Overall, the interaction of MACC1-AS1 and MACC1 mRNA forms RNA-RNA complexes that regulates MACC1 mRNA stability [286].

1.2.3.2 LncRNAs and alternative splicing

RNA splicing is a critical post-transcriptional mechanism for RNA processing [295]. In eukaryotic cells, immature mRNA, also known as precursor mRNA (pre-mRNA), is a primary transcript from DNA that is converted into a mature mRNA [296]. During the maturation of mRNA, pre-mRNA undergoes splicing in the nucleus [296]. Two distinctive types of splicing have been described: constitutive splicing and alternative splicing [297]. Constitutive splicing is a process where introns (non-coding regions) are removed, while exons (coding regions) are fused to produce mature mRNA [297]. In contrast, alternative splicing (AS) is a process that can produce a diverse array of mRNA isoforms from an individual gene by either excluding or including exons in different patterns [298, 299]. Thus, AS expands the range of the transcriptome [298, 299]. In human cells, more than 95% of genes undergo AS [300-302]. Different categories of RBPs and splicing factors (SFs) participate in the regulation of AS (e.g., the serine/arginine-rich (SR), SR-related proteins, small nuclear ribonucleoproteins (snRNPs), and the heterogeneous nuclear ribonucleoproteins (hnRNPs)) [303]. RNA splicing is controlled by the spliceosome, a complex of five small nuclear ribonucleoproteins (snRNPs) which includes U1, U2, U4, U5, and U6 that function to remove introns [296].

A class of lncRNA can also regulate AS by directly interacting with splicing factors or proteins [303] (Figure 1.4). The highly conserved lncRNA metastasis associated lung adenocarcinoma transcript 1 (MALAT1), also called nuclear paraspeckle assembly transcript2 (NEAT2) [304], is primarily localization in subnuclear domains enriched with pre-mRNA splicing factors that are
called nuclear speckles [304]; nuclear speckles are also known as splicing speckles [305]. MALAT1 interacts with different pre-mRNA splicing factors thereby affecting their distribution within splicing speckles [304]. MALAT1 has an RNA binding site for Serine/arginine-rich splicing factor 1 (SRSF1) in both humans and mice [304]. SRSF1 is an SR protein encoded by the SRSF1 gene in humans, and it is also named alternative splicing factor 1, pre-mRNA-splicing factor SF2, or ASF1/SF2 [306]. SRSF1 has two functional domains, including the RS domain (arginine-serine rich region) and RNA recognition motif region, which binds with RNA and other splicing factors [306]. SRSF1 interacts with the 5’region of MALAT1, and this interaction affects the cellular phosphorylation of SR protein levels [304]. MALAT1 also plays a fundamental role in recruiting SR proteins to active transcription sites which impacts the AS of pre-mature mRNA that undergoes alternate exon exclusion in normal cells [304]. However, the level of the dephosphorylated form of the SR protein increases and alters AS of pre-mature mRNA by the inclusion of an alternative exon containing weak splice sites in MALAT1 depleted cells [304].

1.2.3.2.3 LncRNA and RNA degradation

RNA decay controls RNA fate in the cytoplasm [307]. The rates of mRNA decay play a vital role in the post-transcriptional regulation of gene expression [307]. Three main groups of enzymes that play a role in RNA degradation are endonucleases, 3’ exonucleases, and 5’ exonucleases [308]. These enzymes are called ribonucleases (RNases), and RNA can be degraded at both ends of a transcript [308]. RNA degradation can also be mediated by LncRNA through RBPs (Figure 1.4) such as Staufen1 (STAU1) [309, 310]. In mammals, STAU1 interacts with the 3’ untranslated region of mRNA called the Alu element, leading to RNA degradation [309, 310]. The Alu element is a 300 base pairs long segment of DNA repetitive elements found in the human genome,
transcribed by RNA pol II, which does not encode for proteins [311]. The Alu element of the 3’ end of STAU1-mediated mRNA decay (SMD) can target and interact with another Alu element of poly-A end of lncRNAs [309]. This group of lncRNAs is known as short interspersed nuclear element half-Staufen1 (STAU1)-binding site RNAs (1/2-sbsRNAs) [309]. When 1/2-sbsRNAs interacts with the Alu element of a target gene in the SMD, the STAU1-binding site (SBS) will form and activate the SMD pathway [309]. As a result, the recruitment of Upf1 protein to STAU1 leads to mRNA degradation [309].

1.2.3.4 LncRNA act as miRNA Sponges

Cytoplasmic lncRNA can also regulate mRNA expression by sequestering miRNAs [268]. This group of lncRNA is known as competitive endogenous RNAs (ceRNAs) because they compete with miRNA to regulate gene expression [312, 313]. Indeed, ceRNAs have miRNA response elements (MREs) where miRNAs bind [312, 313]. Thus, these lncRNA act as miRNA sponges [312, 313] (Figure 1.4). An example of a ceRNA is nuclear paraspeckle assembly transcript 1 (NEAT1), a lncRNA that was discovered in 2007 [314]. NEAT1 is an un-spliced, non-coding transcript located on human chromosome 11 that is transcribed by RNA pol II from the familial tumor syndrome multiple endocrine neoplasia type 1 loci [314]. NEAT1 has two isoforms, including a 3.7 kb length isoform referred to as NEAT1-1 and a 23 kb in length isoform called NEAT1-2 [314]. NEAT1 can serve as an oncogene or tumor suppresser gene [315]. In fact, NEAT1 is over-expressed in several types of cancer, such as lung cancer, breast cancer, ovarian cancer and hepatocellular carcinoma (HCC) [316-319]. One mechanism implicated in the upregulation of NEAT1 in cancer is its ability to act as a ceRNA by targeting miRNA [315]. For example, overexpression of NEAT1 induces breast cancer growth through downregulation of miR-101

59
Indeed, miR-101 regulates enhancer of zeste homolog 2 (EZH2) which plays a vital role in histone methylation and transcriptional regulation [320-322]. Knockdown of NEAT1 also suppresses cancer cell growth through the upregulation of miR-101, resulting in diminished levels of EZH2 [320, 321]. In HCC, NEAT1 negatively affects miR-129-5p by increasing valosin-containing protein (VCP) and decreasing IkB, thereby causing tumor progression through the NF-kB pathway [319].

1.2.3.3 Role of lncRNA in translational regulation

LncRNA play a critical role in regulating gene expression at the translational level [323]. For example, GAS5 is a lncRNA that controls mRNA translation by interacting with the translation initiation machinery [323]. GAS5 has two RNA-binding motifs which interact with eukaryotic initiation factor-4E (eIF4E), part of the translation initiation complex [323]. eIF4E interacts with structural elements of c-Myc mRNA and regulates its translation [323]. GAS5 suppresses c-Myc translation through interaction with its mRNA, a mechanism known as lncRNA-mRNA interaction [323]. However, GAS5 does not affect c-Myc mRNA levels or its protein stability [323]. Therefore, GAS5 regulates c-Myc mRNA by binding and collaborating with the eIF4E complex [323].

1.2.3.4 Role of lncRNA in post-translational modifications (PTMs)

Post-translation modification of proteins contributes to the diversity of protein function and coordination of their signaling networks [324, 325]. More than 20 types of modifications have been recognized in eukaryotes, including phosphorylation, ubiquitination, acetylation, and
glycosylation [324, 325]. Regulatory roles of lncRNA in PTMs include protein ubiquitination and phosphorylation [326].

1.2.3.4.1 LncRNA and protein ubiquitination

Protein ubiquitination is the most common type of PTMs. Ubiquitination labels proteins for degradation via the proteasome system and can influence the activity of proteins, change their cellular location, and promote or inhibit protein interaction [327, 328]. In protein ubiquitination, three enzymes are required, including ubiquitin-activating enzyme (E1), ubiquitin-binding enzyme (E2), and ubiquitin-protein ligase (E3) [327, 328]. lncRNA can regulate the ubiquitination of proteins by sequestering SFs, guiding RNP complexes, and titrate miRNA and RBPs [326, 329]. An example of a lncRNA that modulate protein ubiquitination is HOTAIR. HOTAIR positively influences protein ubiquitination by interacting with two subunits of E3 ubiquitin ligases that possess RNA-binding domains (Dzip3 and Mex3b), forming a complex of two E3 ubiquitin ligases [326]. Once the complex forms, HOTAIR binds to Ataxin-1 and Snurportin-1, the ubiquitination substrates of Dzip3 and Mex3b [326]. Then, HOTAIR stimulates the degradation of Ataxin-1 and Snurportin-1 by acting as a scaffold molecule carrying their ubiquitination (Dzip3 and Mex3b) and substrates, leading to an increase in Dzip3 and Mex3b ubiquitination and subsequent degradation [326].

1.2.3.4.2 LncRNA and protein phosphorylation

Protein phosphorylation is a reversible biochemical process that adds a phosphate group to an amino acid via a covalent bond, thereby regulating signaling pathways and numerous cellular
processes [330, 331]. LncRNA are now known to contribute to protein phosphorylation [332, 333]. Lnc-DC is a lncRNA that is expressed only in human DCs during their differentiation from monocytes; Lnc-DC controls DC differentiation and function via phosphorylation of signal transducer and activator of transcription 3 (STAT3) [332]. STAT3 is a transcription factor that controls gene expression in response to cytokines and growth factors, acting as a regulator for many physiological processes including development and apoptosis [334]. In the cytoplasm, Lnc-DC interacts with the C-terminus of STAT3. Indeed, Lnc-DC binding to STAT3 stimulates phosphorylation of STAT3 on tyrosine-705 (Tyr705) by preventing STAT3 from interacting with Src homology region 2 domain-containing phosphatase-1 (SHP-1) [332]. SHP-1 can enhance dephosphorylation of tyrosine, which is a protein tyrosine phosphatase (PTP) and expresses in all lineages of hematopoietic cells [335, 336]. In the absence of Lnc-DC, SHP-1 binds to STAT3 and promotes STAT3 dephosphorylation [332]. Taken together, Lnc-DC can impact DC differentiation through interacting with STAT3 pathway.
Figure 1.4 LncRNA execute their functions through different mechanisms. (A) LncRNA regulate gene expression by recruiting chromatin remodeling complex and chromatin marks. (B) LncRNA sequester miRNA by acting as an endogenous sponge and reduce miRNA effects. (C) LncRNA suppress specific genes by titrating transcription factors away from their promoters. (D) LncRNA initiate transcription by acting as guides for TFs to their promoters. (E) LncRNA provide docking sites for proteins, which work together in the same biological pathway by playing scaffolding roles. (F) LncRNA can base-pair with mRNAs and regulate their functions by inhibiting their translation, changing their splicing pattern, or guiding them to degradation. The figure is based on [337, 338].
1.2.4 Regulation of lncRNA expression

There are two described mechanisms that regulate IncRNA expression: epigenetic regulation and action of TFs [339-341]. Epigenetic modifications such as methylation and miRNA might affect the expression of IncRNA maternally expressed gene 3 (MEG3) in HCC cell lines [341]. The downregulation of MEG3 is through hypermethylation of the MEG3 promoter [341]. MEG3 expression is increased with absence of DNMT-1 and DNMT-3B [341]. Control of DNMT-1 and DNMT-3B can occur via the miR-29 family [342, 343] suggesting that the mechanism of MEG3 dysregulation in HCC could be related to the downregulated miR-29a [341]. Thus, IncRNA expression can be regulated through distinct epigenetic mechanisms.

Another mechanism regulating IncRNA expression is through the action of core TFs. Numerous TFs, including SP1, p53, cMyc, ERα, NF-kB, Nanog, Sox2, and Oct4, regulate IncRNA expression through transcription factor binding sites (TFBS) located within the promoter or ncRNA genes [316, 339, 340, 344-348]. For example, octamer-binding transcription factor 4 (Oct4), a homeodomain transcription factor of the POU family that serves as a marker for undifferentiated stem cells, binds to the promoter and enhancer of NEAT1 and MALAT1, respectively, thereby inducing their expression [316]. This increase in NEAT1 and MALAT1 then promotes lung cancer proliferation, migration, and invasion [316]. These data suggest that the expression of IncRNAs is regulated by TFs. However, the regulation of IncRNA expression is still poorly understood.
1.3 Project Rationale

The AhR is a primary mediator of toxic outcomes in response to xenobiotics like TCDD. However, the AhR may also control numerous other physiological processes including proliferation and differentiation via non-canonical pathways. It remains poorly understood how the AhR controls toxic versus physiological outcomes. Moreover, emerging studies show that various ligands of AhR such as TCDD, CS, and B[a]P influence IncRNA expression [349-352]. However, studies have yet to confirm the direct relationship between AhR and IncRNA expression, and exceedingly few studies reported in the literature show a direct correlation between AhR and IncRNA [219, 221, 353]. One of these is a recent study demonstrating that there is a decrease in Linc00673 expression upon knockdown of AhR in treated cells with B[a]P [221]. Another such study showed that TCDD treatment increases expression of MALAT1 [353]. In short, there are limited studies investigating the regulation of IncRNA expression via AhR. Since the exact mechanism(s) of AhR regulation of physiological functions remain unclear, understanding AhR-dependent mechanism(s) will provide insight into its physiologically mechanisms, mechanisms that may be different from its known toxicological ones, and provide a new level of AhR-dependent control.
CHAPTER 2: HYPOTHESIS AND AIMS

2.1 Hypothesis:
AhR activation by xenobiotic ligands regulates the expression of lncRNA.

2.2 Aims:

Aim 1:
To transcriptionally profile gene expression using the classic AhR ligand benzo[a]pyrene (B[a]P) in A549 cells by RNA-sequencing (RNA-seq).

Aim 2:
To determine the extent to which the transcription regulation of IncRNA is dependent on AhR expression by using AhR A549 cells (A549-AhRKO).
CHAPTER 3: MATERIALS AND METHODS

3.1 Reagents

All chemicals were purchased from Sigma unless otherwise indicated. B[a]P (≥ 96%, Sigma-Aldrich, USA) was dissolved in dimethyl sulfoxide (DMSO) and stored at -20 °C.

3.2 Cell culture

The human NSCLC adenocarcinoma cell line (A549) was purchased from American Type Culture Collection (ATCC, USA). A549 cells were cultured in n Dulbecco’s modified Eagle’s medium (DMEM) medium (Gibco, USA) supplemented with 10% fetal bovine serum (FBS; Gibco, USA) and antibiotics/antimycotics (penicillin G, streptomycin, and amphotericin; Invitrogen). All cells were cultured in DMEM media and maintained at 37 °C in a humidified atmosphere containing 5% CO₂.

3.3. Generation of A549-AhRKO cells

A549 cells were utilized to generate A549-AhRKO cells using clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 technology targeting the AhR. In brief, A549 cells were transfected with a CRISPR/Cas plasmid at passage 9. Cells were then exposed to puromycin for 5 days, allowed to recover and then plated at 3 cells per well of a 96-well plate; cells were then expanded. A549WT and A549 AhRKO cells were a kind gift from Dr. Jason Matthews (Department of Pharmacology and Toxicology, University of Toronto).
3.4 Cell treatment

A549 cells were cultured with serum free DMEM overnight (18 hours (h)) before treating with B[a]P for 6 and 24h. For B[a]P exposure, cells were treated with increasing concentrations of B[a]P (2, 4, 8, 16, and 32 μM) for indicated time (6 and 24h). DMSO was used a vehicle control.

3.5 MTT assay

The 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide (MTT) assay was performed to evaluate the effects of B[a]P on A549 cell viability as we have described [354, 355]. In brief, equivalent numbers of A549 cells were cultured in triplicate in flat-bottom 96-well plates in 200 μl of complete medium. Cells were rinsed once in serum-free DMEM and then treated with 140 μl/well of varying concentrations of B[a]P for 24 h. Following this, 10 μl of 5 mg/ml of MTT reagent was added to each well for 4 h at 37 °C. Next, the plate was centrifuged at 700-800 RPM for 5 minutes. Finally, the culture medium was removed, and the insoluble precipitate was dissolved by adding 200 μl of DMSO to each well and read with a Bio-Rad microplate reader at 510 nm.

3.6 Quantitative RT-PCR (RT-qPCR)

A549WT and A549-AhRKO cells were cultured with serum free DMEM for overnight (18h) before treating with B[a]P for 6 and 24h. Subsequently, total RNA was isolated in accordance with the manufacturer’s instructions for the AurumTM total RNA mini kit (Bio-Rad, CA). RNA quantity and quality were measured by using a Nanodrop 2000 spectrophotometer (Thermo Scientific). Reverse transcription of RNA was performed using iScript™ Reverse Transcription Supermix.
The reaction was carried out at 25 °C for 5 min followed by 42 °C for 30 min. Then, the mRNA levels were analyzed using this cDNA template and gene-specific primers (Table 1.3). Quantitative PCR (qPCR) was performed by addition of 1μl cDNA and 0.5μM primers with SsoFast™ Eva Green ® (Bio-Rad Laboratories, Mississauga, ON), and PCR amplification was performed using a CFX96 Real-Time PCR Detection System (Bio-Rad, CA). Thermal cycling was initiated at 95°C for 3 29 minutes and followed by 39 cycles denaturation at 95°C for 10 seconds and annealing for 5 seconds. Gene expression was analyzed using the ΔΔCt method, and results are presented as fold-change normalized to housekeeping gene (GAPDH).

Table 3.1 Primer sequences used for qRT-PCR analysis

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer sequence</th>
<th>Reverse primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP1A1</td>
<td>TTGGTCTCCCTCTCTACACTCTTGAATA</td>
<td>GCAAGCTCAATGCAGGCTAGAATAG</td>
</tr>
<tr>
<td>CYP1B1</td>
<td>TATAGACACATACACCAAAACACTTACACC</td>
<td>CCTATCAACCTCAATCACACACTTTACAC</td>
</tr>
<tr>
<td>NEAT1</td>
<td>GATGGGGGTATGCAGCTTGGA</td>
<td>TGGCTCACCCACGCACCTAAA</td>
</tr>
<tr>
<td>HOTTIP</td>
<td>GCTTTCGGGCGGCACTTCCCT</td>
<td>CGCTGCTTGGGTTCTTCAGCA</td>
</tr>
<tr>
<td>SOX2OT</td>
<td>GAAAGGTGTGCAGGCAGAATG</td>
<td>GAAGGCCAGCTCTCTTGCT</td>
</tr>
<tr>
<td>MALAT1</td>
<td>AGCTGTCCCTATAGGCTGCCCATT</td>
<td>TGTGACACTGCTTGGGTTCTG</td>
</tr>
<tr>
<td>H19</td>
<td>CACACTCAGCGACTCGTA</td>
<td>CACCACCCTCCTTCTCTTT</td>
</tr>
<tr>
<td>Linc00673</td>
<td>GGAGUCCAGCCAGAUCAUTT</td>
<td>AUGAUCUGGCAUGGACCUU</td>
</tr>
<tr>
<td>GAPDH</td>
<td>GTCTCCTCTGAATCTTCAACAGC</td>
<td>ACCACCCGTTTGCTGAGCCA</td>
</tr>
</tbody>
</table>
3.7 Western Blot

A549\textsuperscript{WT} and A549 AhR-A549\textsuperscript{KO} cells were grown to approximately 75-85\% confluence and the cultured with serum free DMEM for 18h. Total cellular protein was extracted using radio-immunoprecipitation assay (RIPA) lysis buffer (Thermo Fisher Scientific, USA) in conjunction with Protease Inhibitor Cocktail (Roche, USA). Following extraction, protein concentration was measured by the bicinchoninic acid (BCA) protein kit (Thermo Fisher Scientific, USA). Protein lysate at a concentration of 20 or 60 μg was electrophoresed on 10\% SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) gels and transferred onto Immuno-blot PVDF membranes (Bio-Rad Laboratories, USA). After transfer, the membrane was blocked with a blocking solution of 5\% Oxoid\textsuperscript{TM} skim milk powder (Thermo Fisher Scientific) in 1x PBS/0.1\% Tween-20 for one hour at room temperature. Primary antibodies were applied to membranes for 1 hour at room temperature or overnight at 4 °C. The following is a list of the antibodies used: anti-AhR (1:2000; Enzo Life Sciences) and anti-Tubulin (1:10000; Sigma-aldrich, USA). After the application of the primary antibody, the secondary antibodies used were antirabbit IgG, horseradish peroxidase (HRP) linked (1:10000; Cell Signaling Technologies, USA) and HRP conjugated anti-mouse IgG (1:10000; Cell Signaling Technologies, USA). Membrane visualization was performed by using Clarity western enhanced chemiluminescence (ECL) substrate (Bio-Rad Laboratories, Canada) or Super Signal West Femto Maximum Sensitivity Substrate (Thermo Scientific, USA). Detection of protein bands was done by the ChemiDoc MP Imaging System (Bio-Rad Laboratories, Canada). Densitometric analysis was analyzed through Image Lab Software Version 5 (Bio-Rad Laboratories, Canada) and protein expression was normalized to tubulin and presented as fold change compared to the A549 parent cells.
3.8 RNA library construction and RNA sequencing

Total RNA was extracted from A549 cells treated with 8uM B[a]P or with DMSO for 6 or 24h in accordance with the manufacturer’s instructions using the Aurum™ total RNA mini kit (Bio-Rad, CA). RNA quantity and quality were measured by using a Nanodrop 2000 spectrophotometer (Thermo Scientific). Library preparation and RNA-sequencing (RNA-seq) were performed at the McGill University and Genome Quebec Innovation Centre. RNA integrity was evaluated using a Bioanalyzer (Agilent) and the RNA integrity number (RIN⁰) of each RNA sample was between 9-10. Ribosomal RNAs were depleted from 250 ng of total RNA using QIAseq FastSelect kit (Qiagen). cDNA synthesis was achieved with the NEBNext RNA First Strand Synthesis and NEBNext Ultra Directional RNA Second Strand Synthesis Modules (New England BioLabs, USA). The remaining steps of library preparation were done using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs, USA). Adapters and PCR primers were purchased from New England BioLabs. Libraries were quantified using the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (Kapa Biosystems) and average size fragment was determined using a LabChip GX (PerkinElmer) instrument. The libraries were normalized and pooled and then denatured in 0.05N NaOH and neutralized using HT1 buffer. The pool was loaded at 225pM on an Illumina NovaSeq S4 lane using Xp protocol as per the manufacturer’s recommendations. The run was performed for 2x100 cycles (paired-end mode). A phiX library was used as a control and mixed with libraries at 1% level. Base calling was performed with RTA v3. Program bcl2fastq2 v2.20 was then used to demultiplex samples and generate fastq reads. The paired-end flow cells were sequenced on Illumina NovaSeq 6000 S4 PE100 bp and PE150 bp.
3.9 RNA Seq Workflow Analysis

The reads were trimmed using fastp and then aligned using the STAR aligner. From the aligned reads, HTSeq was used to get the raw reads counts. Batch effect was accounted for using the sva R package. Then, DESeq2 R package was used to normalize the counts and run a differential expression (DE) analysis between the different conditions. From the results of DE analysis, PCA plot and heatmaps were generated. The Gage and Pathview R packages were used on the most significant differentially expressed genes based on a log2 fold change cut-off >0.6 and a p-value<0.05 to get pathways and associated gene sets.

3.10 Statistical Analysis

Statistical differences among the means of more than two groups were determined using one-way analysis of variance (ANOVA) followed by a Turkey’s multiple comparisons test. Groups of two were analyzed by two-tailed unpaired t-test using GraphPad Prism 6 (v.602; GraphPad Software Inc, USA). Results are presented as mean ± standard error of the mean (SEM) of the fold changes compared to A549WT treated with only DMSO. Statistical significance was considered in all cases which had a p value < 0.05.
CHAPTER 4: CONTRIBUTION OF CO-AUTHORS

This thesis is based on my original work and is presented as a manuscript currently in preparation for peer-review and submission. Below are those individuals who also contributed to this body of work:

1- Dr. Jason Matthews provided the A549-AhR\(^{KO}\) cells that were used in this project.

2- Dr. David H. Eidelman provided assistance with experimental design and manuscript editing.

3 - Dr. Carolyn Baglole provided support for planning and designing experiments in addition to editing my thesis.

Bioinformatics and RNA sequence analysis was provided by Mathieu Simard and Daniel Jimenez.
CHAPTER 5: RESULTS

5.1 Effect of B[a]P on A549 cell viability and AhR activation

B[a]P is an established AhR ligand and a toxic component of CS and other combusted organic material [158]. One of the main target organs of B[a]P is the lungs, which also expresses high levels of AhR. Activation of the AhR by B[a]P may have global effects on transcription, including alterations in the expression of lncRNA. However, there is little information on AhR and regulation of lncRNA by B[a]P, which activates the canonical AhR pathway. To understand AhR regulation of lncRNA by B[a]P, we first needed to establish the optimum dose of B[a]P that does not affect A549 cell viability but can still activate the AhR. For this, A549 cells were first treated with increasing concentrations of B[a]P for 24h and viability assessed by MTT. There was a significant decrease in viability in response to 32 μM of B[a]P compared to DMSO (Figure 5.1). There was a similar reduction in viability at 16 μM of B[a]P, but this did not reach statistical significance (p = 0.08563). There was no significant cytotoxic effect of 8 μM of B[a]P (Figure 5.1).
Figure 5.1. B[a]P dose-dependently decreases A549 cell viability. The effect of B[a]P on cytotoxicity of human lung epithelial cells (A549) was determined by MTT assay. After A549 cells were seeded for 24h, they were treated with a set of concentrations of B[a]P (2, 4, 8, 16, and 32 μM) for 24h. Data are expressed as the percentage compared to the cells treated with vehicle control (0.1% DMSO). There was a significant reduction in viability in response to 32 μM B[a]P (*p<.05) but not 16 μM (p = 0.0863) Data represented as means ± SEM, n = 4 independent experiments.
Because 8 μM B[a]P did not cause cytotoxicity, we next confirmed that this concentration of B[a]P caused the transcription of the AhR target genes CYP1A1 and CYP1B1 in A549 cells. These data show that there is a significant increase in CYP1A1 mRNA induction in response to 8 μM B[a]P at 24h of exposure but not 6h (Figure 5.2A). Similarly, there was a significant increase in CYP1B1 mRNA expression in response to 8 μM B[a]P at 24h (Figure 5.2B). Therefore, we chose 8 μM of B[a]P to conduct the remainder of experiments in which we investigated the regulation of lncRNA by AhR activation.

**Figure 5.2** RT-qPCR validation of CYP1A1 and CYP1B1 expression in A549 cells treated with B[a]P. A549 cells were treated for 6 and 24 h with vehicle control (DMSO) or with B[a]P (8μM) and cell lysates collected for RNA analysis of CYP1A1 and CYP1B1. B[a]P significantly increased CYP1A1 (A) and CYP1B1 (B) in A549 cells (**** p < 0.0001) compared to untreated cells. Results are expressed as the mean ± SEM of 10 independent experiments; all values were normalized to GAPDH.
5.2 AhR-dependent control of gene expression by B[a]P

To confirm the importance of AhR expression on target gene induction, we utilized A549 cells whereby AhR protein was deleted using CRISPR/Cas9 (Figure 5.3A). Next, we assessed the mRNA level of CYP1A1 and CYP1B1 in A549WT and A549-AhRKO cells treated with B[a]P (8uM) for 6 and 24h. These results show that there is a significant induction in CYP1A1 and CYP1B1 mRNA expression in A549WT cells in response to B[a]P at 24h, but not in the A549-AhRKO cells (Figure 5.3B and 5.3C). These results indicate that B[a]P induces the expression of CYP1A1 and CYP1B1 in an AhR-dependent manner.
Figure 5.3 AhR expression in A549WT and A549-AhRKO cells. (A) AhR protein is absent in A549-AhRKO cells but is present in WT cells. Representative western blot is shown. The mRNA level of CYP1A1 (B) and CYP1B1 (C) is induced in A549WT cells treated with 8μM of B[a]P for 6 and 24h compared to control; there was negligible CYP1A1 or CYP1B1 mRNA in A549-AhRKO cells. Data are represented as the mean ± SEM; n = 3-7) independent experiment (**** p < 0.0001).
5.3 AhR-independent regulation of select lncRNA

Previous literature has shown that AhR ligands, including B[a]P, affect the expression of lncRNA [221, 351-353, 356]. Many of these lncRNA are known to regulate functions similar to AhR, suggesting their involvement in endogenous functions controlled by AhR [225, 357]. Therefore, we first asked whether the expression of some of these lncRNA changed in response to B[a]P in an AhR-dependent manner. We assessed the expression of NEAT1, HOTTIP, SOX2OT, MALAT1, H19 and Linc00673 in A549WT and A549-AhRKO cells. Although there were marginal increases in the expression of several lncRNA in response to B[a]P, none reached statistical significance (Figure 5.4). Moreover, there was little difference in the expression of these lncRNA based on presence of absence of AhR. Taken together, these data suggest that AhR activation by B[a]P did not influence the expression in our selected lncRNAs in A549 cells.
Figure 5.4 Knockout of AhR does not change the expression of selected lncRNA. A549WT and A549-AhRKO cells were treated with DMSO or with B[a]P (8μM) for 6 and 24h; cell lysates were collected for RNA analysis of NEAT1 (A), HOTTIP (B), SOXT2OT (C), MALAT1 (D), H19 (E), and Linc00673 (F). Results are represented as the mean ± SEM of 6-7 independent experiments,
and all values were normalized to GAPDH. Means are expressed as fold change from the control (6 or 24h).
5.4 Differential regulation of gene expression by B[a]P

We then used next-generation RNA sequencing to obtain a more comprehensive understanding of the transcriptomic changes in response to B[a]P in A549 cells expressing AhR. RNA-seq libraries were constructed using total RNA (with rRNA depletion) and sequenced by the paired-end method, after which we conducted a PCA to evaluate variability between B[a]P and untreated cells (Figure 5.5A and 5.5B). The first and second principal components explained 47 and 30 percent of the variance, respectively (Figure 5.5A). We also observed some variation between replicates from the same treatment group. The first and second principal components of A549 cells treated with B[a]P for 24h (versus untreated cells) explained 69 and 15 percent of the variance, respectively (Figure 5.5B). We also used Pearson correlation to test the clustering of our samples based on their Euclidean distance using the regularized log-transformed count data (Figure 5.5C and 5.5D). Taken together, these data show that although there is some variability between replicates of the same experimental group, B[a]P-treated cells form a distinct cluster of genes after 6 or 24h of treatment compared to cells without B[a]P. These RNA-seq expression data were thus used to identify genes differentially regulated between B[a]P and untreated cells at these two time points.
Figure 5.5 Principal component analysis of untreated and B[a]P treated cells. The PCA plot shows the first two principal components (PCA1 and PCA2). (A) PCA1 and PCA2 were 47 and 30 percent of the variance (6h B[a]P and control groups; n=3). (B) PCA1 and PCA2 were 69 and 15 percent of the variance, respectively (24h B[a]P and control groups; n=3). (C, D) Pearson correlation showing a summary of a positive correlation between biological replicates.
5.5 B[a]P changes the expression of both protein-coding and non-coding RNA in A549 cells

Our objective was to identify lncRNA that are differentially expressed upon AhR activation, using B[a]P as a representative agonist of the classic AhR pathway. We therefore conducted DE analysis between B[a]P and untreated cells for 6 and 24h of exposure; these two timepoints would allow us to capture early and late genes that may be controlled by B[a]P. Overall, we detected 1981 genes that were expressed in A549 cells exposed to B[a]P compared to control (DMSO) (see Supplemental Tables S1 and S2 for complete gene lists). Furthermore, the DE genes (DEG) with the lowest p-value were shown in the MA plot (where M is the difference between the log intensity values and A is the average of the log intensity values) (Figure 5.6A and 5.6B). The blue dots that are above zero represent upregulated genes, and blue dots that are below zero represent down-regulated gene between control (DMSO) and treated (B[a]P) cells at the 6- and 24h-timepoints (Figure 5.6A and 5.6B). The grey dots represent the genes whose expressions did not differ between B[a]P and untreated cells (Figure 5.6A and 5.6B). DEGs were determined using a log2 fold-change cut-off of \( \geq 0.6 \) and a p-value<0.05 [358]. This analysis provided an overview of the distribution of gene expression in response to B[a]P.
Figure 5.6 MA plots of differentially up- and downregulated genes in response to B[a]P. The MA plots demonstrate DEGs after 6 (A) and 24h (B) treated with B[a]P. The y-axis corresponds to the mean of normalized counts (p-value), and the x-axis displays the log\(_2\) fold change cut-off 0.6 value. There are 1035 and 946 genes that are significant after 6 and 24h of B[a]P exposure, respectively.
Next, we performed hierarchical heat map analysis for the DEG between B[a]P and untreated cells after 6 and 24h (Figure 5.7A and 5.7B) to visualize and identify the DEG regulated after B[a]P treatment. These data indicate that 1981 genes are differentially expressed after 6 and 24h of B[a]P exposure (see Supplemental Tables S1 and S2).

**Figure 5.7 Hierarchical heat maps of the DEG in response to B[a]P.** Heat maps of DEGs after 6 (A) and 24h (B) of B[a]P treatment. Heatmaps were generated from variance stabilized transformed (VST) scaled RNA-seq normalized count of 1982 DE genes among the B[a]P treated groups compared DMSO. Data is presented where each row represents one gene, and each column represents an individual B[a]P and DMSO group. The color and intensity of the heat maps represent changes in gene expression. The color code indicates that red are upregulation genes, blue are downregulated genes, and yellow are unchanged. Case: A549 cells treated with B[a]P (8μM) for 6 or 24h and control refers to DMSO.
These data also show that there were 704 DE protein-coding RNA, of which 339 were upregulated and 21 were downregulated by B[a]P after 6h. In addition, there were 39 upregulated and 305 downregulated protein-coding RNA by B[a]P after 24h. Note that among the top genes induced by B[a]P were CYP1A1 and CYP1B1 (Table 5.1), which is consistent with published literature [158] as well as with our RT-qPCR data (see Figures 5.2 and 5.3). The log2 fold change of the top significant differentially expressed protein coding RNAs is summarized in Tables 5.1 and 5.2. Other genes that were also changed in response to B[a]P included VSIR, STRA6, ZNF667, LRRC3B, LAIR1, MVD, PCSK9, NR4A1, PDE10A, and SMO (see Supplemental Tables S1 and S2).
Table 5.1 Top DE protein coding RNA that were upregulated in response B[a]P in A549 cells after 6 and 24h of exposure.

<table>
<thead>
<tr>
<th>Transcript ID (Ensembl_Gene_ID)</th>
<th>Log₂ Fold Change</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Upregulated protein-coding RNAs (6h)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSG00000140465-CYP1A1</td>
<td>0.91</td>
<td>1.05E-06</td>
</tr>
<tr>
<td>ENSG00000258991-DUX4L19</td>
<td>4.97</td>
<td>0.0001337</td>
</tr>
<tr>
<td>ENSG00000236138-DUX4L26</td>
<td>2.21</td>
<td>0.00017695</td>
</tr>
<tr>
<td>ENSG00000197408-CYP2B6</td>
<td>5.69</td>
<td>0.00029738</td>
</tr>
<tr>
<td>ENSG00000211459-MT-RNR1</td>
<td>0.68</td>
<td>0.00046158</td>
</tr>
<tr>
<td>ENSG00000043462-LCP2</td>
<td>5.32</td>
<td>0.00064446</td>
</tr>
<tr>
<td>ENSG00000226321-CROCC2</td>
<td>3.61</td>
<td>0.00074603</td>
</tr>
<tr>
<td>ENSG00000171044-XKR6</td>
<td>4.18</td>
<td>0.00099646</td>
</tr>
<tr>
<td>ENSG00000171804-WDR87</td>
<td>2.72</td>
<td>0.00109269</td>
</tr>
<tr>
<td>ENSG00000164318-EGFLAM</td>
<td>5.20</td>
<td>0.00131023</td>
</tr>
<tr>
<td><strong>Upregulated protein-coding RNAs (24h)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSG00000162496-DHRS3</td>
<td>1.84</td>
<td>5.63E-71</td>
</tr>
<tr>
<td>ENSG00000163659-TIPARP</td>
<td>1.21</td>
<td>3.34E-36</td>
</tr>
<tr>
<td>ENSG00000140465-CYP1A1</td>
<td>3.15</td>
<td>1.48E-20</td>
</tr>
<tr>
<td>ENSG00000168497-CAVIN2</td>
<td>0.88</td>
<td>1.76E-16</td>
</tr>
<tr>
<td>ENSG00000159388-BTG2</td>
<td>1.06</td>
<td>1.83E-14</td>
</tr>
<tr>
<td>ENSG00000165272-AQP3</td>
<td>0.67</td>
<td>1.03E-08</td>
</tr>
<tr>
<td>ENSG00000162591-MEGF6</td>
<td>0.65</td>
<td>3.18E-08</td>
</tr>
<tr>
<td>ENSG00000169991-IFFO2</td>
<td>0.67</td>
<td>2.75E-07</td>
</tr>
<tr>
<td>ENSG00000286169-AHRR</td>
<td>0.84</td>
<td>1.47E-06</td>
</tr>
<tr>
<td>ENSG00000248487-ABHD14A</td>
<td>0.99</td>
<td>6.66E-06</td>
</tr>
</tbody>
</table>
**Table 5.2** Top DE protein coding RNA that were downregulated in response B[a]P in A549 cells after 6 and 24h of exposure.

<table>
<thead>
<tr>
<th>Transcript ID (Ensembl_Gene_ID)</th>
<th>Log₂ Fold Change</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Downregulated protein-coding RNAs (6h)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSG00000101255-TRIB3</td>
<td>-0.71</td>
<td>2.90E-38</td>
</tr>
<tr>
<td>ENSG00000112972-HMGCS1</td>
<td>-0.69</td>
<td>1.08E-36</td>
</tr>
<tr>
<td>ENSG00000113161-HMGCR</td>
<td>-0.63</td>
<td>9.73E-34</td>
</tr>
<tr>
<td>ENSG00000159167-STC1</td>
<td>-0.65</td>
<td>9.96E-12</td>
</tr>
<tr>
<td>ENSG00000173930-SLCO4C1</td>
<td>-0.84</td>
<td>1.55E-08</td>
</tr>
<tr>
<td>ENSG00000153234-NR4A2</td>
<td>-0.65</td>
<td>5.12E-08</td>
</tr>
<tr>
<td>ENSG00000128917-DLL4</td>
<td>-0.67</td>
<td>0.00099855</td>
</tr>
<tr>
<td>ENSG00000101670-LIPG</td>
<td>-0.69</td>
<td>0.00251786</td>
</tr>
<tr>
<td>ENSG00000119508-NR4A3</td>
<td>-0.81</td>
<td>0.00292378</td>
</tr>
<tr>
<td>ENSG00000108551-RASD1</td>
<td>-0.62</td>
<td>0.00410468</td>
</tr>
<tr>
<td><strong>Downregulated protein-coding RNAs (24h)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSG00000153162-BMP6</td>
<td>-1.09</td>
<td>1.63E-10</td>
</tr>
<tr>
<td>ENSG00000119508-NR4A3</td>
<td>-1.39</td>
<td>4.01E-08</td>
</tr>
<tr>
<td>ENSG00000120211-INSL4</td>
<td>-0.66</td>
<td>1.54E-07</td>
</tr>
<tr>
<td>ENSG00000153234-NR4A2</td>
<td>-0.80</td>
<td>6.16E-07</td>
</tr>
<tr>
<td>ENSG00000144057-ST6GAL2</td>
<td>-0.67</td>
<td>5.07E-06</td>
</tr>
<tr>
<td>ENSG00000157542-KCNJ6</td>
<td>-0.90</td>
<td>1.03E-05</td>
</tr>
<tr>
<td>ENSG00000250731-TPM3P6</td>
<td>-2.59</td>
<td>5.15E-05</td>
</tr>
<tr>
<td>ENSG00000213763-ACTBP2</td>
<td>-5.66</td>
<td>7.29E-05</td>
</tr>
<tr>
<td>ENSG00000077279-DCX</td>
<td>-0.60</td>
<td>0.0001795</td>
</tr>
<tr>
<td>ENSG00000159167-STC1</td>
<td>-0.94</td>
<td>0.0002021</td>
</tr>
</tbody>
</table>
In addition, we also detected DE lncRNA by B[a]P exposure. A total of 811 lncRNA were differentially expressed, of which 395 were upregulated in response to B[a]P after 6h, whereas 9 lncRNA were upregulated after 24h. Eight lncRNA were downregulated in cells treated with B[a]P for 6h while 399 lncRNA were downregulated by B[a]P after 24h. The log₂ fold change of the top differentially expressed lncRNA are summarized in Tables 5.3 and 5.4. In summary, B[a]P changed the expression of both protein-coding and non-coding RNA in A549 cells.
Table 5.3. Top DE IncRNA that were upregulated in response B[a]P in A549 cells after 6 and 24h of exposure.

<table>
<thead>
<tr>
<th>Transcript ID (Ensembl  Gene ID)</th>
<th>Log₂ Fold Change</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Upregulated IncRNAs (6h)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSG00000228956-SATB1-AS1</td>
<td>5.92</td>
<td>6.17E-05</td>
</tr>
<tr>
<td>ENSG00000227555-MIR4290HG</td>
<td>5.84</td>
<td>0.00025144</td>
</tr>
<tr>
<td>ENSG00000176593-AC008969.1</td>
<td>0.61</td>
<td>0.00044477</td>
</tr>
<tr>
<td>ENSG00000227706-AL713998.1</td>
<td>5.91</td>
<td>0.0005048</td>
</tr>
<tr>
<td>ENSG00000257398-AC126177.3</td>
<td>5.84</td>
<td>0.00066168</td>
</tr>
<tr>
<td>ENSG00000267779-LINC01533</td>
<td>5.01</td>
<td>0.00069983</td>
</tr>
<tr>
<td>ENSG00000286257-AL137100.3</td>
<td>5.61</td>
<td>0.00076289</td>
</tr>
<tr>
<td>ENSG00000246379-AC040168.1</td>
<td>4.79</td>
<td>0.0007725</td>
</tr>
<tr>
<td>ENSG00000205176-REXO1L1P</td>
<td>4.58</td>
<td>0.00154179</td>
</tr>
<tr>
<td><strong>Upregulated IncRNAs (24h)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSG00000280441-FP236383.1</td>
<td>3.71</td>
<td>3.48E-07</td>
</tr>
<tr>
<td>ENSG00000274333-CU633967.1</td>
<td>1.49</td>
<td>0.00085714</td>
</tr>
<tr>
<td>ENSG00000232354-VIPR1-AS1</td>
<td>1.13</td>
<td>0.00201162</td>
</tr>
<tr>
<td>ENSG00000226007-BX005266.2</td>
<td>1.32</td>
<td>0.00438355</td>
</tr>
<tr>
<td>ENSG00000255362-AP000785.2</td>
<td>3.91</td>
<td>0.01631123</td>
</tr>
<tr>
<td>ENSG00000277501-AC243571.2</td>
<td>2.14</td>
<td>0.01728598</td>
</tr>
<tr>
<td>ENSG00000232406-AL121895.1</td>
<td>0.71</td>
<td>0.01763396</td>
</tr>
<tr>
<td>ENSG00000176912-TYMSOS</td>
<td>0.62</td>
<td>0.0241523</td>
</tr>
<tr>
<td>ENSG00000258412-AL355102.1</td>
<td>0.67</td>
<td>0.03557555</td>
</tr>
</tbody>
</table>
Table 5.4 Top DE IncRNA that were downregulated in response to B[a]P in A549 cells after 6 and 24h of exposure.

<table>
<thead>
<tr>
<th>Transcript ID (Ensembl Gene ID)</th>
<th>Log₂ Fold Change</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Downregulated IncRNAs (6h)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSG00000099251-HSD17B7P2</td>
<td>-0.63</td>
<td>9.65E-06</td>
</tr>
<tr>
<td>ENSG00000224687-RASAL2-AS1</td>
<td>-0.66</td>
<td>0.02181896</td>
</tr>
<tr>
<td>ENSG00000269910-AL049840.2</td>
<td>-0.60</td>
<td>0.0230871</td>
</tr>
<tr>
<td>ENSG00000251148-AL158068.2</td>
<td>-4.10</td>
<td>0.03137112</td>
</tr>
<tr>
<td>ENSG00000258096-AC025031.2</td>
<td>-0.63</td>
<td>0.03507313</td>
</tr>
<tr>
<td>ENSG00000215859-PDZK1P1</td>
<td>-0.66</td>
<td>0.03752064</td>
</tr>
<tr>
<td>ENSG00000279529-AC008764.8</td>
<td>-0.62</td>
<td>0.03876248</td>
</tr>
<tr>
<td>ENSG00000225836-AL157400.1</td>
<td>-3.99</td>
<td>0.03965265</td>
</tr>
<tr>
<td><strong>Downregulated IncRNAs (24h)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSG00000235142-LINC02532</td>
<td>-1.11</td>
<td>0.00012001</td>
</tr>
<tr>
<td>ENSG00000235779-AC079779.3</td>
<td>-5.61</td>
<td>0.00016563</td>
</tr>
<tr>
<td>ENSG00000281641-SAMD12-AS1</td>
<td>-5.52</td>
<td>0.00022312</td>
</tr>
<tr>
<td>ENSG00000281383-FP671120.5</td>
<td>-3.40</td>
<td>0.00027595</td>
</tr>
<tr>
<td>ENSG00000235049-LINC00940</td>
<td>-5.36</td>
<td>0.0004044</td>
</tr>
<tr>
<td>ENSG00000228956-SATB1-AS1</td>
<td>-4.17</td>
<td>0.00043285</td>
</tr>
<tr>
<td>ENSG00000238266-LINC00707</td>
<td>-0.86</td>
<td>0.00044359</td>
</tr>
<tr>
<td>ENSG00000203987-AL772363.1</td>
<td>-5.32</td>
<td>0.00044624</td>
</tr>
<tr>
<td>ENSG00000254092-AC021613.1</td>
<td>-4.13</td>
<td>0.00063509</td>
</tr>
<tr>
<td>ENSG00000272168-CASC15</td>
<td>-0.71</td>
<td>0.00098923</td>
</tr>
</tbody>
</table>
5.6 Gene ontology (GO) Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of DR genes

To further characterize the DEGs in response to B[a]P, we performed gene ontology (GO) analysis. There were three different functional categories used according to the GO term enrichment analysis, including molecular function (MF), cellular component (CC), and biological process (BP). First, these analyses revealed that there were 18 molecular functions associated with the upregulated protein-coding genes after 6h of B[a]P exposure including ion binding, catalytic activity, small molecule binding as well as nucleoside phosphate purine nucleotide binding and purine nucleotide binding (Figure 5.8A). The number of upregulated protein-coding genes after 6h of B[a]P was significantly enriched in these 18 molecular functions (Figure 5.8B). Only 3 molecular functions were associated with the downregulated protein-coding genes after 24 of B[a]P; these were ion binding, cation binding, and metal-binding (Figure 5.9A). Additionally, the number of downregulated protein coding genes were significantly associated with those three molecular functions (Figure 5.9B). There was no significant association with differentially upregulated protein-coding genes after 24h with B[a]P or downregulated protein-coding genes after 6h with B[a]P.
A. Molecular functions

- GO:0005524 ATP binding
- GO:0032559 adenyl ribonucleotide binding
- GO:0030554 adenyl nucleotide binding
- GO:0042802 identical protein binding
- GO:0001882 nucleoside binding
- GO:0035639 purine ribonucleoside triphosphate binding
- GO:0032250 purine ribonucleoside binding
- GO:0032249 ribonucleoside binding
- GO:0001882 purine nucleoside binding
- GO:0043168 anion binding
- GO:0022555 purine ribonucleoside binding
- GO:0032553 ribonucleotide binding
- GO:0017076 purine nucleoside binding
- GO:1901265 nucleoside phosphate binding
- GO:0000166 nucleotide binding
- GO:0036094 small molecule binding
- GO:0003824 catalytic activity
- GO:0043167 ion binding

B. Number of genes

Figure 5.8 GO analysis for 6h treatment with B[a]P. (A) Molecular functions that were associated with upregulated protein-coding genes. (B) Number of upregulated protein-coding genes that were associated with molecular functions.
Figure 5.9 GO analysis for 24h exposure to B[a]P. (A) Molecular functions that were associated with downregulated protein-coding genes. (C) Number of downregulated protein-coding genes that were associated with molecular functions.
Finally, KEGG pathway analysis revealed 33 upregulated lncRNA after 6 hr of B[a]P are involved in the hsa01100 pathway, which is a collection of many metabolic pathways. There was no enriched KEGG pathways associated with up- or downregulated lncRNA after 24h of B[a]P. These data suggest that a potential and novel set of DEG in response to B[a]P might be associated with pathways and distinct molecular functions in A549 cells.
CHAPTER 6: DISCUSSION

The AhR is a cytosolic transcription factor known to be activated by a variety of ligands, including dioxin [1]. Since discovery of AhR, its mediation of the toxic effects of dioxin has been widely studied. However, the endogenous role of AhR is still under investigation [11]. The availability of genetically-engineered mice and new high-throughput technologies have provided insights into how AhR mediates both its toxic and physiological effects. AhR controlled physiological and regulatory processes include altered immunity, organ development, cell proliferation and survival, control over the cell cycle, and wound healing [62, 359-362]. In this regard, a number of non-canonical signaling mechanisms have been discovered for the AhR (e.g., protein-protein interaction and regulation of ncRNA). Some of these implicate crosstalk between AhR and other signaling pathways such as NF-κβ, the ER, and nuclear factor-erythroid 2-related factor 2 (NRF2) [117, 118, 121-123, 245]. AhR may be protective in several chronic inflammatory diseases including chronic rhinosinusitis, Crohn’s disease, multiple sclerosis ulcerative colitis and chronic obstructive pulmonary disease (COPD) [363-367]. We recently reported a protective role of AhR against cigarette smoke-induced apoptosis via miR-196a expression [88]. We hypothesized that AhR might regulate a wide range of physiological and biological processes by modulating other ncRNA, and in particular IncRNA.

In the present study, we used B[a]P as an AhR agonist to investigate the relationship between IncRNA expression and AhR activation. Previous observations found that some IncRNA regulate biological processes similar to AhR [316, 368] including MALAT1, linc000673, HOTTIP, NEAT1, H19, and SOX2OT [221, 316, 349, 353, 356, 368-371]. Therefore, we initially assessed if AhR regulates the expression of those IncRNAs in A549 cells exposed to B[a]P. Our data show
that these lncRNA are not induced by B[a]P and that presence or absence of AhR in A549 cells does not impact their expression levels. These data were first obtained by RT-qPCR followed by RNA-seq; both methods yielded similar results. These data were surprising for several reasons. First, lung cancer cell lines (e.g., A549, H1299, SK-MES-1, SPCA-1, 95-D, and NCI-H520 cell lines) show relatively high levels of NEAT1, which correlates with its ability to promote cell proliferation, invasion, and differentiation [316, 368]. Furthermore, evidence has shown a role of HOTTIP in cancer cell biology [369, 370, 372-374], including relatively high expression levels of HOTTIP in cancer cell lines [369, 370, 372, 374, 375]. Previous reports also demonstrated that SOX2OT upregulation was associated with A549 cell proliferation and migration [371, 376], that TCDD induces MALAT1 expression in pancreatic cancer cells and that knockdown of AhR or treating the cells an AhR antagonist inhibits MALAT1 expression [353], suggesting the importance of AhR in regulating this lncRNA in response to TCDD. In addition, our results on H19 are in contrast to a previous study showing induction of H19 expression in cells treated with 32 μM of B[a]P [349], a higher concentration than what was used in our study. Finally, our results showing a lack of regulation of Linc00673 by B[a]P are in contrast by a study by Wu et al. where there is induction of Linc00673 expression in response to B[a]P in A549 cells [221]. Thus, despite evidence in the literature showing regulation of these lncRNA by AhR and/or B[a]P, we did not show significant alteration in our study. Such discrepancy between our results and these could be due to the differences in the cell line and/or ligands used in the study, as well as the concentration of B[a]P (8 μM versus 32 μM). Another possible theory is related to differences in the experimental protocol such as the cell culture media. In the study conducted by Wu et al., A549 cells were cultured in Roswell Park Memorial Institute (RPMI) medium [221], while we cultured the cells in a DMEM. Even slight differences in experimental protocols may impact outcomes as a recent
report by Björnstedt et al. demonstrated that different culture conditions changed the phenotype of A549 cells [377]. While multiple experimental variables could explain differences between our results and those published in the literature, we show that these IncRNA are not controlled by the AhR.

However, we do not believe an experimental error contributed to the relative lack of induction of the above mentioned IncRNA, as our RNA-seq data show that well-characterized AhR target genes such as CYP1A1, TIPARP, and AhRR were upregulated in response to B[a]P. This indicates that AhR was activated by B[a]P and are consistent with published literature demonstrating these as AhR target genes [103, 109, 164, 378]. We also showed induction of CYP1A1 and CYP1B1 by qRT-PCR. CYP1A1 and CYP1B1 are detoxifying enzymes and a classical AhR target genes [103, 378, 379], having an established role in the regulation of xenobiotic metabolism, oxidative stress, apoptosis, and cell cycle control [192, 378]. AhRR and TIPARP are also AhR target genes and are involved in the negative regulation of AhR signaling, the net effect of which is reduced AhR activity and impaired CYP1A1 induction [109, 159, 162, 164]. Thus, the A549 cells used in our study responded as anticipated to B[a]P, a known AhR ligand.

Additional genes identified in our study as being regulated by B[a]P include aquaporin 3 (AQP3), BTG anti-proliferation factor 2 (BTG2), and doublecortin (DCX); these are also known to be regulated by AhR. AQP3, for instance, plays a critical role in HepG2 cell migration when exposed to different AhR ligands (TCDD, B[a]P, 3MC, and quercetine), and that absence of AhR reduces AQP3 expression. The AQP3 promoter has a potential XRE for AhR, suggesting direct transcriptional regulation [380]. BTG2 is another protein-coding RNA upregulated in A549 cells
exposed to B[a]P. Interestingly, BTG2 is an AhR target gene in WB-F344 cells exposed to TCDD [381]. BTG2 is a tumor suppressor that inhibits proliferation and induces apoptosis [381]. B[a]P can stimulate pathogenesis of HCC through the NF-kB pathway [382] and a recent study demonstrated the role of AhR in the progression HCC [383, 384]. It should be noted that activation of the AhR by its endogenous ligand ITE has inhibitory effects on proliferation and migration of HCC cells [385]. Thus, different AhR ligands can have very distinctive and often divergent outcomes [33, 212, 225, 386, 387]. It is interesting to speculate that B[a]P upregulation of BTG2 may be a mechanism through which the AhR contributes to the progression of different cancers. Delta like canonical Notch ligand 4 (DLL4) is one of the downregulated RNA we observed in response to B[a]P. DLL4 is a member of the Notch family, which play regulatory roles in the formation, development, and maintenance of vessels [388]. DLL4 expression in pregnant rats is reduced on day 4 gestation by B[a]P [389]. These and our results suggest a broader regulation of DLL4 by B[a]P in different cell types. Another example of a downregulated protein-coding RNA in response to B[a]P is DCX, which has a regulatory role in normal neuronal migration and proliferation [390]. Recently, FIZC was reported to induce DCX expression in mouse hippocampus, which improved long-term memory and learning skills, via AhR and Wnt/ß-catenin activation [391]. Our results also identified 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR) as a downregulated protein-coding RNA, which is also suppressed by TCDD [392]. LaPres et al. proposed that the severe liver damage induced by TCDD might be through AhR activation; however, this report did not show a direct relation between AhR and HMGCR [392]. Given that there is evidence supporting that the AhR plays an essential physiological role in liver development [141, 199, 359, 393], HMGCR downregulation by B[a]P might provide more broad implications for the AhR in other cell types including lung cancer cells. The majority of
protein-coding DEG are known genes and included lymphocyte cytosolic protein 2 (LCP2), EGF like, fibronectin type III and laminin G domains (EGFLAM), ciliary rootlet coiled-coil (CROCC2), XK related 6 (XKR6), myosin heavy chain 4 (MYH4), aquaporin 2 (AQP2), dehydrogenase/reductase 3 (DHRS3), caveolae associated protein 2 (CAVIN2), multiple EGF like domains 6 (MEGF6), intermediate filament family orphan 2 (IFFO2), stanniocalcin 1 (STC1), solute carrier organic anion transporter family member 4C1 (SLCO4C1), nuclear receptor subfamily 4 group A member 2 (NR4A2), lipase G (LIPG), and transmembrane protein 132D (TMEM132D)). These DE RNAs show regulation of various biological functions and disease processes. However, no reports to date have shown a relation between those DEGs and B[a]P or the AhR. Further investigation of these genes may reveal novel functions for the AhR.

Due to limited studies evaluating AhR regulation of IncRNA expression, we took a more comprehensive approach and investigated IncRNA expression via RNA-seq. These data show 811 differentially expressed IncRNA in response to B[a]P. The majority are novel IncRNA with no reported biological function; these include microRNA 4290 host gene (MIR4290HG), AC008969.1, long intergenic non-protein coding RNA 1533 (LINC01533), vasoactive intestinal peptide receptor 1 antisense RNA 1 (VIPR1-AS1), BX005266.2, RAS protein activator like 2 antisense RNA 1 (RASAL2-AS1), long intergenic non-protein coding RNA 2532 (LINC02532), and FP671120.5. Three of the DE IncRNAs (FP236383.1, HSD17B7P2, and AL049840.2) have been identified as DEGs several human diseases [394-396]. Interestingly, four DE LncRNA identified in our study- special AT-rich sequence binding protein 1 antisense RNA 1 (SATB1-AS1), thymidylate synthetase opposite strand (TYMSOS), sterile alpha motif domain containing 12 antisense RNA 1 (SAMD12-AS1), and hepatocellular carcinoma up-regulated long non-coding
RNA (HULC) - share functions similar to AhR, including regulation of cell proliferation. Numerous studies have investigated AhR expression and activation in regulating cell survival and proliferation [386, 387, 397] and some have shown that the AhR promotes cell survival and proliferation through a ligand-independent mechanism [397], whereby overexpression of the AhR was sufficient to stimulate proliferation of A549 cells [398]. Even in the absence of ligand, AhR in the NSCLC cell lines have been linked to alterations in cell function. For example, H1299 cells have higher invasion compared to A549 cells due to the cellular level of AhR, which is lower in H1299 cells than in A549 cells [397]. However, AhR activation also has effects on proliferation that are dependent on both the cell and ligand type [386, 387]. For example, in various NSCLC and HCC cell lines, AhR activation by TCDD promotes cell survival, proliferation, migration, and invasion via the activation AKT serine/threonine kinase (Akt) pathway. AhR activation upon B[a]P exposure can also promote proliferation of A549 cells [386] whereas AhR activation by ITE inhibits proliferation and migration of endometrial cancer cells [387]. However, the underlying mechanism of anti- versus pro-proliferative roles of AhR remains elusive. While it is possible that cell-specific differences can account for this, another possible mechanism could be through the regulation of IncRNA that may occur in a ligand-specific manner. Intriguingly, we showed that SATB1-AS1 is differentially regulated by B[a]P, in that was up- and downregulated IncRNA in response to B[a]P after 6h and 24h, respectively. Other studies have shown that SATB1-AS1 has an impact on cell proliferation. SATB1-AS1 is upregulated in many types of cancer including laryngeal squamous cell carcinoma (LSCC), acute myeloid leukemia (AML), and colorectal cancer [399-401]. In AML, SATB1-AS1 promotes proliferation and inhibits apoptosis by acting as a miRNA sponge for miR-580 [399]. Given that the AhR can modulate cellular proliferation in response to B[a]P A549 cells [386], B[a]P induction of SATB1-AS1 expression might provide
insights into the molecular mechanisms of AhR in regulating cell proliferation. Further exploration is needed to identify whether the presence or absence of AhR has a regulatory impact on SATB1-AS1 expression and consequent AhR function.

We also identified other lncRNA with important roles in cell proliferation and survival. TYMSOS is another lncRNA that was also upregulated by B[a]P in our study. TYMSOS can drive GC progression by acting as a ceRNA for miR-4739 [402]. TYMSOS has tumor-promoting role in GC cells by increasing cell proliferation, migration, invasion, and inhibiting apoptosis. Our findings show that another downregulated lncRNA is SAMD12-AS1, which also plays a role in cell proliferation. Sun et al. showed that upregulation of SAMD12-AS1 directly interacts with DNMT1, resulting in an increase in cell proliferation and migration via the suppression p53 pathway [403]. In vivo and in vitro studies revealed that activating the AhR pathway increase GC cell growth and invasion, while inhibiting AhR pathway results in a reduction in GC cell growth [89, 404, 405]. Given that AhR can promote the proliferation of GC cell lines in response to B[a]P [405-407], it is interesting to speculate that AhR regulation of TYMOS and SAMD12-AS1 may be an important overall mechanism in its ability to control various cellular processes including proliferation. Finally, our data also detected that HULC is another lncRNA upregulated by B[a]P. High expression of HULC promotes proliferation and inhibits apoptosis of NSCLC cell lines (NCI-H23 and NCI-H522) via activation of the Akt pathway [408]. Given that AhR can increase tumor cell growth and inhibit apoptosis through the Akt pathway [409-411], B[a]P induction of HULC expression may be one way in which AhR regulates cell proliferation.
B[a]P is a PAH that is produced during the combustion of organic material. As such, B[a]P is present in CS, diesel exhaust and air pollution. Many of these complex mixtures have been shown to have an effect on lncRNA expression. Tobacco, which results in the pathogenesis a variety of diseases [412], can also impact lncRNA expression [351, 412-416]. Cigarette smoke extract (CSE) is used as an in vitro model of tobacco exposure that contains B[a]P [413]. Qian et al. reported that CSE increases LCPAT1 in BEAS-2B cells, which was involved in DNA damage [413]. Upregulation of LCPAT1 is associated with promoting the proliferation of NSCLC cell lines (A549, H1299, and H1975) but not non-cancer cells (BEAS-2B) [417]. However, this study did not show that AhR was involved in regulating LCPAT1 upon CSE exposure, although CSE can activate the AhR [33]. It is therefore possible that AhR may regulate LCPAT1 expression in response to CSE. However, our data did not show a significant effect of B[a]P on LCPAT1 expression. This discrepancy could be due to the fact that CS contains thousands of other chemicals that may be contributing to the increase in LCPAT1. In addition, utilization of a chronic smoke-induced COPD mouse model has been used to evaluate lncRNA expression, including a study conducted by Lu et al. who detected hundreds of DE lncRNAs [351]. This study found that the most significant pathways associated with DE lncRNA were endoplasmic reticulum and oxidative stress pathways [351], pathways also controlled by the AhR in response to cigarette smoke [33, 418]. In addition, recently published data revealed the protective role of AhR against the pathogenesis of CS-induced COPD [367]. It is interesting to speculate that AhR regulation of lncRNA could be how the AhR protects against COPD development. Coal tar also contains B[a]P, and is an exposure that also affects lncRNA expression [419, 420]. For example, coal tar increases lncRNA ENST000001520 expression and promotes the proliferation of BEAS-2B cells [419]. However, the potential regulatory mechanism behind coal tar induction of ENST000001520
expression is still unknown. Altogether, these findings suggest that complex environmental toxicants containing B[a]P have significant effects on lncRNA expression.

GO analysis further revealed diverse molecular functions associated with DE protein-coding genes after B[a]P exposure. Furthermore, KEGG pathways analysis revealed an association between DE lncRNA after 6h of B[a]P treatment, including the hsa01100 pathway. The hsa01100 pathway is a combination of metabolic pathways, an example of which involves the cytochrome P450 enzymes. Previous studies reported that lncRNA significantly respond to xenobiotic exposure [421-423], and a recent study by Waxman et al. detected more than 1400 xenobiotic-responsive lncRNAs (xeno-lncs) using RNA-seq datasets from livers of rats exposed to 27 xenobiotics [421]. There were seven distinct mechanisms of action of each set of xenobiotics, one of which involved the AhR. Here, two AhR agonists, β-naphthoflavone and 3-methylcholanthrene, significantly upregulated 32 xeno-lncs [421]. In addition, two differentially expressed lncRNAs in response to TCDD were positively correlated with CYP1A1 expression [422]. Finally, the lncRNA HNF1α-AS1 and HNF4α-AS1 play a role in regulating the transcription of P450 genes although the exact mechanism remains unidentified [423]. Given that xenobiotic metabolism is one of the outcomes of AhR signaling, our findings that lncRNA expression is affected by B[a]P may yield novel insight into the mechanism through which AhR causes toxicity independent of CYP1A1 expression.

Limitations and future directions

Our work elucidates the relationship between AhR and lncRNA expression in response to the classic AhR ligand B[a]P using two different techniques- RT-qPCR and RNA-seq. These
techniques are not without their limitations. For example, RNA-seq results can be affected by the sample quality, sample preparation, read depth, and the number of biological replicates. One limitation of our experiment is related to the depth of sequencing. In our experiment, the read depth was 50 million reads, which could be insufficient in detecting genes with low expression [424]. Thus, deep sequencing of the transcriptome may be needed to detect lncRNA not in high abundance [424, 425] and higher depths increase the reliability of the data and the statistical power of the results [426, 427]. However, we validated our experimental approach using qRT-PCR to evaluate known genes that are increased upon AhR activation; similar results were obtained using RNA-seq. An important future experiment would be to study the AhR-dependent role in the regulation of lncRNA expression by performing RNA-seq analysis on A549WT and A549-AhRKO cells. Finally, it is well-established that the function of the AhR is cell-type specific and our results may not be reflective of lncRNA expression in primary cells. Finally, evaluation of the biological function of lncRNA identified in this study may reveal important insights into the role of the AhR and focus should be given on cellular functions such as apoptosis and proliferation.

There are limitations in the MTT assay, which was used to evaluate cell viability in this study. The MTT assay has a broad utility in evaluating drug cytotoxicity, proliferation, and cell viability [428, 429]. This assay depends on cellular mitochondrial activity to convert the tetrazole to formazan via NADPH [428, 429]. Generally, it is assumed that the cellular mitochondrial activity is proportionally correlated with the viable cell number, and thus increased mitochondria activity is reflected by the formazan concentration [428-430]. However, MTT reduction could be impacted by a number of factors, such as changes in metabolic activity. In addition, alterations in the physiological state of cells and culture conditions such as glucose
content of medium and pH can also influence results of the MTT assay [126, 429, 431, 432]. Another limitation is that the MTT reduction relies not only on the mitochondria but also on other non-mitochondrial sources such as the plasma membrane, endosome, and lysosome compartments [433-435]. Recent research has found that various chemical compounds (e.g., vitamin A, dithiothreitol, ascorbic acid, coenzyme A, flavonoids, genistein, and polyphenols) can interfere with the MTT assay [429, 436-440]. The ability of these compounds to interfere with the MTT assay can increase formazan production and background absorbance [429, 436-440]. These limitations can impact the interpretation of MTT results. Thus, our results showing that B[a]P increased cell viability beyond one hundred percent could be due to changes in metabolic activity due to a corresponding increase in mitochondrial activity. Another reason for this observation could be because of changes in the cellular physiological state caused by B[a]P. Nevertheless, using the MTT assay allowed us to determine a concentration of B[a]P that did not significantly affect viability, a concentration that still increased the expression of lncRNA.

Conclusions

Our results support that AhR activation in response to B[a]P affects lncRNA expression in A549 cells although further research is needed to evaluate AhR dependence of lncRNA expression changes elicited by B[a]P. Regardless, this study strongly supports the notion that AhR regulates lncRNA expression. We found that B[a]P significantly impacts lncRNA expression overall, although the AhR does not regulate the well-characterized lncRNA NEAT1, HOTTIP, SOX2OT, MALAT1, H19, and Linc00673. Further investigations on the effects of B[a]P on lncRNA expression caused by AhR ablation will help us to identify novel connections between B[a]P and lncRNAs and expand our understanding of the AhR physiological functions.
References


128


Appendix

DMSO_vs_BaP_results (6h)_DEG_.xls

DMSO_vs_BaP_results (24h)_DEG_.xls