

Copyright

by

Karen Larissa Pereira de Castro

2021

**The Dissertation Committee for Karen Larissa Pereira de Castro Certifies that this
is the approved version of the following dissertation:**

**Characterization of Aryl Hydrocarbon receptor and Kruppel-Like-
Factor-6 interaction**

Committee:



Cornelis Elferink (Aug 12, 2021 15:59 CDT)

Cornelis J. Elferink, Ph.D, Mentor



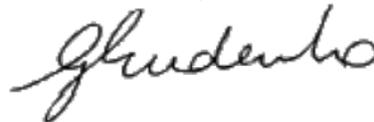
NESLIGHAN MARTINEZ (Aug 12, 2021 15:53 CDT)

Muge N. Kuyumcu-Martinez, Ph.D, Chair



Andrew Routh (Aug 13, 2021 10:56 CDT)

Andrew Routh, Ph.D



Gabrielle Rudenko Ph.D



Karl-Dimiter Bissig, MD, Ph.D

**Characterization of Aryl Hydrocarbon receptor and Kruppel-Like-
Factor-6 interaction**

by

Karen Larissa Pereira de Castro, M.S

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas Medical Branch

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas Medical Branch

July, 2021

Dedication

For my parents and my family that while distant are always close to my heart. To my husband who is always with me in this journey as we enjoy life and science together. To my friends and all the kindred spirits I met in this Country.

Acknowledgements

First, I would like to thank my family and friends for all the support throughout these years. Next, to Dr. Elferink for accepting to be my mentor, always encouraging me to be a better scientist. To all the members of the Elferink lab: Li Chen, the heart of the lab, Eric Wright, Dianne Yap, Aditya Joshi, and Nate Girer, who became a second family here in America. To Tracie Albritton and the BMB department for always being available and helpful. Finally, I would like to thank my committee members Muge N. Kuyumcu-Martinez, Andrew Routh, Gabrielle Rudenko and Karl-Dimiter Bissig for the support and guidance.

Characterization of Aryl Hydrocarbon receptor and Kruppel-Like-Factor-6 interaction

Publication No. _____

Karen Larissa Pereira de Castro, Ph.D

The University of Texas Medical Branch, 2021

Supervisor: Cornelis J. Elferink

The aryl hydrocarbon receptor (AhR) is a ligand activated transcription factor which regulates toxic response to environmental pollutants, including 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD). Upon ligand activation, the AhR translocates to the nucleus, dissociates from its chaperone complex, and dimerizes with the Aryl Hydrocarbon Receptor Nuclear Translocator (ARNT). The canonical AhR-ARNT complex binds to DNA regions known as Xenobiotic Response Elements (XRE), inducing the transcription of target genes. Our previous results demonstrated that the AhR is also involved in a novel protein-DNA complex with the Kruppel Like Factor 6 (KLF6), a zinc finger that is involved in processes such as angiogenesis, cell proliferation, and apoptosis. The AhR-KLF6 heterodimer functions differently from the ARNT-AhR complex by binding to a non-consensus XRE (NC-XRE) in a ligand-dependent manner. Using full length proteins as well as a series of point mutation and deletion proteins produced *in vitro*, we demonstrated by electrophoretic mobility shift assays (EMSAs) that the glutamine rich region (Q-rich) of AhR is crucial for protein-DNA binding. We identified the specific residues in the AhR Q-rich domain responsible for protein DNA-binding. We also identified the key amino acids of KLF6 and the regions of AhR that are responsible for

protein-protein interaction. We have applied ChIP-seq (Chromatin Immunoprecipitation–Sequencing) and RNA-sequencing to correlate DNA binding events with changes in transcriptomics in the liver of mice exposed to TCDD, linking AhR activation by TCDD with biological functions such as regulation of cell death, apoptotic process, cellular proliferation and response to xenobiotic stimulus.

TABLE OF CONTENTS

List of Tables	x
List of Figures	xi
List of Abbreviations	15
Chapter 1 Introduction	17
AHR Canonical Pathway	18
AhR Non-canonical pathway	19
AhR and dioxin toxicity	22
The AhR structure	24
AhR binding partners	26
Kruppel Like Factors	28
Kruppel like factors and liver	29
CPS1	32
The involvement of Plasminogen activator inhibitor-1 (pai-1) in the non- canonical pathway	34
AhR-KLF6 DNA interaction	35
AhR genomics and transcriptomics	35
KLF6 roles in genomics and transcriptomics	37
Chapter 2: Characterization of AhR and KLF6 interaction	39
INTRODUCTION	39
MATERIALS AND METHODS	40
Cell Culture	40
Nuclear extraction	40
AhR, KLF6, CPS1 constructs	41
Mutation constructs	42
Coupled <i>in vitro</i> transcription translation and western blots	46
Quantification of expressed proteins	46
Microplate Protein-Binding Assay	47
Electrophoretic Mobility Shift Assay	49
RESULTS	50

Protein expression.....	50
Characterization of the AhR-KLF6 protein-protein interaction	52
AhR-KLF6 DNA interaction	55
Characterization of the NC-XRE	56
The N-terminal region of KLF6 is necessary for DNA binding.....	58
The role of the AhR DNA-binding domain in the NC-XRE interaction	59
The Q-rich region of AhR is critical for NC-XRE	61
AhR residues between 160 and 223 are essential for KLF6 interaction..	62
AhR-KLF6-CPS1 DNA interaction.....	63
Characterization of AhR- CPS1 NC-XRE interaction.....	64
DISCUSSION.....	68
Chapter 3: A genome-wide functional assessment of AhR signaling.....	76
Introduction.....	76
ARNT inducible knockout mouse model	77
MATERIAL AND METHODS	79
Animals.....	80
Chromatin immunoprecipitation.....	80
Next generation sequencing.....	82
Chromatin immunoprecipitation sequencing	82
Transcription factor identification	83
Motif Mapping.....	83
Visualization of ChIP-seq data using Heatmaps.....	83
RNA extraction and RNA-seq	84
Semi-quantitative PCR.....	84
Quantitative PCR	85
Visualization of RNA-seq results using heat maps.....	85
RESULTS	86
RNA-sequencing for identification of AhR Target Genes	86
ChIP-sequencing for identification of AhR Target Genes.....	97
Motif mapping	102

Anchoring RNA-seq with ChIP-seq	103
DISCUSSION	106
Chapter 4: Dissertation conclusion	113
References	115
Appendices	133
Appendices I: Oligonucleotides	133
Appendices II: Human AhR binding partners	137
Appendices III: RNA seq ARNT flox upregulated genes	142
Appendices IV: RNA seq ARNT flox pathways analysis	150
Appendices V: RNA seq ARNT iKO upregulated genes	154
Appendices VI: RNA seq ARNT iKO genes pathways analysis	156
Appendices VIII: RNA seq ARNT iKO downregulated genes	162
Appendices IX: RNA seq ARNT flox downregulated genes pathway analysis	165
Appendices X: RNA seq ARNT iKO downregulated genes pathways	167
VITA	169

List of Tables

Table 1: Gene List pathway analysis by Metascape. Analysis of downregulated genes of ARNT Floxed animals treated with TCDD compared to TCDD treated ARNT iKO animals.	95
Table 2: Analysis of Enriched Transcription Binding Sites	101
Table 3: Possible transcription factors that may be associated with dioxin response	105

List of Figures

Figure 1: AhR domains.....	17
Figure 2: AhR canonical and non-canonical pathways.....	20
Figure 3: Mouse and human AhR domains.	24
Figure 4: The AhR Q-rich region is essential for nuclear localization.	26
Figure 5: KLF groups.....	29
Figure 6: KLF6 genomic DNA and splice isoforms.	31
Figure 7: AhR-CPS1 interaction.	33
Figure 8: Schematic diagram of the expression constructs used.	43
Figure 9 AhR and CPS1 clones. (A) AhR and (B) CPS1 internal deletions.	44
Figure 10: AhR and KLF6 alanine mutagenesis.....	45
Figure 11: Microplate Protein-Binding Assay schematics.	49
Figure 12: Immunoblots of AhR/KLF6 Protein expression.	51
Figure 13: Immunoblot of CPS1 protein expression.....	52
Figure 14: Detection of wild type AhR and AhR mutant proteins interaction with wild-type KLF6 by microplate binding assay.....	53

Figure 15: Detection of wild-type AhR interaction with wild type KLF6 and KLF6 mutants.....	54
Figure 16: AhR-KLF6 DNA binding.....	55
Figure 17: Characterization of the NC-XRE binding site.....	56
Figure 18 DNA binding analysis of MCF7 cells nuclear extracts to the NC-XRE. ...	57
Figure 19: Amino acids 29-30 and 33-34 of KLF6 are critical for protein-DNA binding.	59
Figure 20: Characterization of AhR DNA binding domain.....	60
Figure 21: The Q-Rich region of AhR is essential for NC-XRE binding.....	61
Figure 22: Residues 160-224 of AhR are essential for protein-protein interaction, consequently DNA bind.....	62
Figure 23: AhR-CPS1 binds to the NC-XRE. KLF6-CPS1 complex do not bind to the NC-XRE in the absence of AhR.	63
Figure 24: AhR-CPS1-KLF6 complex bind to the NC-XRE.	63
Figure 25: Characterization of the AhR-CPS1 interaction with NC-XRE.	65
Figure 26: AhR-CPS1 DNA interaction.	67
Figure 27: AhR-CPS1 DNA interaction.	67
Figure 28: AhR-KLF6 and CPS1 interaction summary.....	75

Figure 29: Schematic of inducible ARNT Knockout mice model.....	79
Figure 30: Semi-quantitative RT-PCR analysis of ARNT expression in Floxed animals and in induced knockout (iKO) animals.....	79
Figure 31: Principal component analysis of samples used for RNA-sequencing analysis.....	86
Figure 32: Effect of TCDD exposure on Cyp1a2, Cypbp1, Ugt1a2 and Notch1 expression.	87
Figure 33: Effect of TCDD exposure on PAI-1.....	88
Figure 34:Gene List pathway analysis by Metascape. Bar graph of enriched genes upregulated in ARNT floxed animals treated with TCDD for 2 hours.	88
Figure 35: Biological process analysis of upregulated genes of TCDD treated ARNT floxed animals.....	89
Figure 36:Gene List pathway analysis by Metascape. Bar graph of enriched genes of ARNT iKO animals treated with TCDD for 2 hours.....	89
Figure 37: Biological process analysis of upregulated genes of TCDD treated ARNT iKO animals.	89
Figure 38: Effect of TCDD exposure on ATF3, FosB and G6PC expression.....	91
Figure 39: Effect of TCDD exposure on liver Cyp1a1 expression.....	91
Figure 40: Effect of TCDD exposure on UGT1A6, HGF and Selenbp2 expression.....	93

Figure 41: Average RNA-seq heat map of significant gene expression changes between TCDD and vehicle treated Floxed and iKO animals.....	97
Figure 42: AhR and KLF6 bind to the PAI-1 promoter in a TCDD dependent manner.	98
Figure 43: ChIP-seq tracks for AhR occupancy of Vehicle and TCDD-treated wild type animals.	99
Figure 44: Overview of AhR binding sites.	100
Figure 45: Positional distribution of AhR DNA binding motifs.....	102
Figure 46: Motif discovery of chIP-seq regions that lack XRE motifs.	103
Figure 47: Comparison between ChIP-seq genes with a \log_2 ratio TCDD/vehicle ≥ 1 with floxed TCDD treated upregulated and downregulated genes according to RNA-seq.	104
Figure 48: Gene set enrichment analysis of upregulated genes of RNA-seq data of Floxed TCDD treated animals in comparison with functional binding site detected by ChIP-seq.....	104
Figure 49: ChIP-seq comparison with RNAseq. AhR enrichment 0-10kb from the TSS of genes that contain non-XREs and XREs compared with RNA- seq of upregulated genes upon TCDD treatment.	105

List of Abbreviations

AhR- Aryl Hydrocarbon Receptor

AhRR- Aryl-Hydrocarbon Receptor Repressor

ARNT iKO- ARNT inducible knockout

bHLH- PAS- basic helix-loop-helix (bHLH) Period/ARNT/Single minded

ChIP- Chromatin Immunoprecipitation.

CPS1- Carbamoyl Phosphate Synthetase I

EMSA- Electrophoretic Mobility Shift Assay

hAhR- Human AhR

HGF- Hepatocyte Growth Factor

HSP90- heat shock protein 90

KLF6- Kruppel Like Factor 6

mAhR- Mouse AhR

NC-XRE- Non Consensus Xenobiotic Response Element

NLS- Nuclear localization signal

p23- Prostaglandin E Synthase 3

p21^{cip1}- Cyclin-Dependent Kinase Inhibitor 1

PAI-1- (SERPINE 1) Plasminogen Activator Inhibitor-1

PAH- Polycyclic Aromatic Hydrocarbons

TAD: Transcription Activation Domain

TCDD- 2,3,7,8-Tetrachlorodibenzodioxin

TF- Transcription Factor

XAP2- AH Receptor-Interacting Protein

XRE- Xenobiotic Response Element.

Chapter 1 Introduction

The existence of the aryl hydrocarbon receptor (AhR) was first proposed by Poland and collaborators in 1976¹. AhR was identified as a ligand-activated transcription factor that upregulates the transcription of drug metabolizing enzymes². AhR can bind to different endogenous ligands, such as tetrapyrroles, kynurenine, and aradonic acid metabolites, and to various exogenous ligands such as polyphenols, polycyclic aromatic hydrocarbons (PAHs), and halogenated aromatic hydrocarbons (HAH). Among the HAHs 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) is the best characterized³⁻⁸.

The AhR belongs to the family of basic helix-loop-helix (bHLH) period/ARNT/single-minded (PAS) transcription factors. Mutagenic and deletion analysis were used to define the functions of each domain of AhR (Figure 1). These studies revealed that the basic domain is responsible for DNA binding; HLH, PAS A and PAS B as the regions responsible for ARNT dimerization; PAS B as the region where AhR ligands bind; and the C-terminal as the region containing a transcription activation domain (TAD), which contains an acidic, a Q-rich, and a P/S/T subdomains.⁹⁻¹¹

AhR is highly expressed in the fetal liver, lungs, kidneys, epithelial cells,

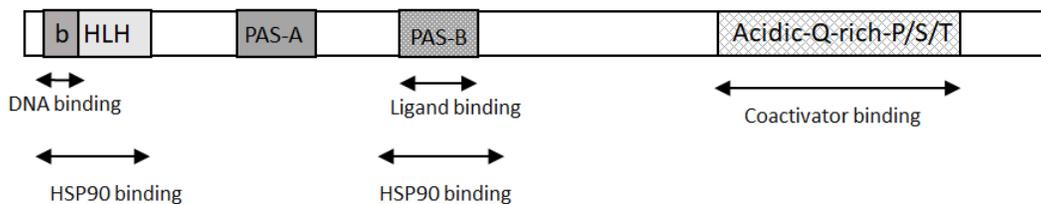


Figure 1: AhR domains.

esophagus, thymus and placenta and in the adult spleen, pancreas, lungs and liver. Although AhR was first described as a protein that evolved to facilitate the metabolism of xenobiotic elements, it plays essential roles in cellular adhesion, migration, differentiation

and proliferation, hematopoiesis, vascular development, neural system development, reproduction and circadian rhythm ^{3,4,7,9,12,13}.

The pathway that leads to AhR activation by exogenous ligands has been widely investigated. According to the canonical pathway, inactivated AhR is localized in the cytosol bound to the chaperones AH receptor-interacting protein (XAP2), heat shock protein 90 (HSP90), and prostaglandin E synthase 3 (p23). After ligand binding, AhR translocates to the nucleus forming a complex with the Aryl Hydrocarbon Receptor Nuclear Translocator (ARNT). The complex binds to the xenobiotic response element (XRE) regulating the expression of target genes such as cytochrome P450 1A1 (CYP1A1) and other phase I and II metabolic enzymes ^{5,6,14}.

Evidence indicates that besides ARNT, AhR can interact with other transcription factors¹⁵⁻²¹. AhR-ARNT complex interacts with Specificity Protein 1 (SP1) via the HLH-PAS domains of AhR and ARNT and a SP1 zinc finger. The AhR-ARNT-SP1 complex can bind to XRE adjacent to the GC motif that binds SP1^{15,21}. It was also demonstrated that AhR can interact with other proteins independently of ARNT. Upon TCDD induction, the retinoblastoma tumor suppressor protein (pRb) binds with the N-terminal and Q-rich region of AhR, not with ARNT, ^{20,22}. Data also show that the NF-kB subunits RelA and RelB interact with AhR. RelA-AhR complex can activate *c-myc* promoter, while interaction with RelB forms a complex that modulates the transcription of chemokines, by binding to a promoter region different than the XRE ^{16,23-25}.

AHR CANONICAL PATHWAY

The inactive AhR resides in the cytosol, forming a complex with a dimer of the chaperones HSP90 (a chaperone that stabilizes AhR conformation), co-chaperones p23 (maintains the HSP90 ATP-bound conformation) and XAP2 (hypothesized to block AhR nuclear localization signal). TCDD and other ligands enter the cell by diffusion, bind and induce AhR conformational changes that cause the exposure of AhR nuclear localization

signal (NLS), which will then facilitate the recognition of AhR by β -importins and promote nuclear translocation. Once in the nucleus, AhR dissociates from the chaperone complex and forms a heterodimer with ARNT. The complex binds to the DNA xenobiotic response elements (XRE) (defined by the core motif 5'GCGTG3'), induces co-activator recruitment, increases promoter accessibility to other transcription factors and induces the expression of target genes such as phase I xenobiotic metabolizing enzymes (CYP1A1, CYP1A2, CYPBP1), phase II xenobiotic metabolism enzymes (glutathione S-transferase (GST), aldehyde dehydrogenase 3a1), AhRR (Aryl-Hydrocarbon Receptor Repressor), and NAD(P)H-quinone oxidoreductase^{3,5,11,26-30} (Figure 2).

The AhRR shares amino acid sequence similarities with AhR from the N-terminal region to the end of the PAS-A domain. The differences lie in the C-terminal region and in the lack of a transcription activation domain, a ligand binding and a PAS-B domain, thus not being able to bind ligands. The C-terminal region contains a transrepression domain, where corepressors responsible for AhR negative feedback control bind. AhRR binds to ARNT and to the XRE, suppressing the transcription of AhR target genes. After transcriptional activation, AhR is exported from the nucleus and degraded by 26S proteasomes^{9,27,31-33}.

AHR NON-CANONICAL PATHWAY

The AhR non-canonical pathway was proposed based on the discoveries described above. Upon ligand binding, AhR translocates into the nucleus dissociating from the chaperone complex forming a DNA-protein complex with KLF6 at NC-XRE regions that are characterized by the 5'GGGA3' tetranucleotide repeat. AhR-KLF6 recruits Carbamoyl phosphate synthetase I (CPS1) and is responsible for homocitrullination of histone H1 at lysine 34. The pathway is responsible for the expression of target genes, such as cyclin-dependent kinase inhibitor 1 (p21^{cip1}) and plasminogen activator inhibitor 1 (PAI-1). (Figure 2)^{14,34-36}

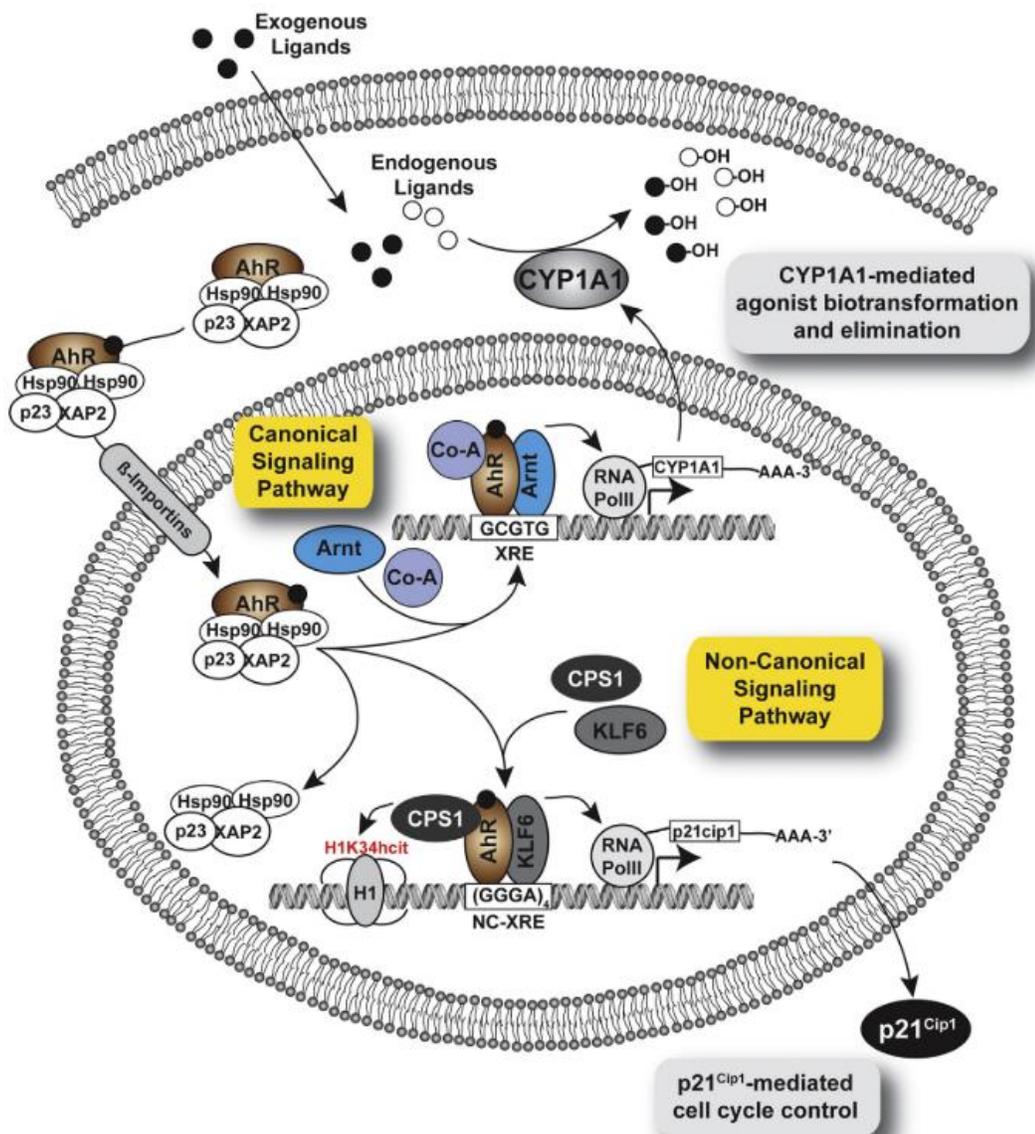


Figure 2: AhR canonical and non-canonical pathways. The inactive form of AhR resides in the cytosol bound to a chaperone complex (Hsp90, XAP2 and p23). According to the canonical pathway, AhR translocates into the nucleus dissociating from the chaperone complex upon endogenous or exogenous ligand binding forming a complex with ARNT and co-factors and then binding to the XRE DNA regions (core sequence of 5' GCGTG 3'). In the non-canonical pathway, AhR can also interact with KLF6, recruiting CPS1 as a co-factor, binding to the NC-XRE DNA sites defined by a 5'GGGA3' tetranucleotide repeat, and subsequently regulating the expression of target genes such as p21Cip1 (Reproduced under CC BY 4.0)¹⁴. Reproduced under CC BY-NC-ND.

Exposure to TCDD can lead to suppression of liver regeneration³⁷. Among the AhR target genes that can be responsible for the lack of regenerative response is the PAI-1, a regulator of the urokinase-type/tissue-type plasminogen activator (uPA/tPA) activities

³⁸. PAI-1 upregulation inhibits uPA expression, consequently disrupting the proteolytic activation of Hepatocyte growth factor (HGF)³⁹, a key factor in liver regeneration since it can trigger hepatocyte proliferation⁴⁰. Considering that PAI-1 promoter region lacks the classical XRE element, Huang and Elferink investigated and characterized the DNA region where AhR interacts, designating the region as *non-consensus XRE* (NC-XRE)³⁶.

Considering the homology observed between the Kruppel Like Factors (KLFs) DNA binding motif and the core motif of the NC-XRE (a tetranucleotide 5'GGGA3' repeat) and the high expression levels of Kruppel Like Factor 6 (KLF6) in the liver, Wilson and collaborators decided to investigate a novel protein interaction between the AhR and KLF6. By co-immunoprecipitation, Electrophoretic mobility shift assays (EMSAs), and chromatin immunoprecipitation, it was confirmed that upon ligand binding, AhR forms a complex with KLF6 in the nucleus and the heterodimer binds to the NC-XRE region⁴¹.

KLF6 is a tumor suppressor that can activate or suppresses gene expression and participates in several cellular processes, such as differentiation, proliferation and development^{42,43}. Growth suppression activity is linked to p53-independent transactivation of cyclin-dependent kinase inhibitor 1 (p21^{cip1})⁴⁴. Jackson and collaborators demonstrated that p21^{cip1} expression during liver regeneration is controlled by the AhR-KLF6 non-canonical pathway⁴⁵.

In an effort to better characterize the novel, non-canonical pathway, Joshi *et al.* used an NC-XRE probe to purify a TCDD inducible complex in mouse liver, identifying by mass spectrometry sequencing the Carbamoyl Phosphate Synthase 1 (CPS1) as cofactor of the AhR-KLF6 complex. Further investigation revealed that upon CPS1 recruitment, lysine 34 (K34) on histone H1 is homocitrullinated following AhR-KLF6 interaction with NC-XRE sites³⁴.

Histone H1 can restrain nucleosome mobility and impede chromatin access to interacting complexes. Post-translational modifications are linked with alterations in protein mobility⁴⁶, and it has been demonstrated that histone H1 K34 acetylation triggers

transcription by inducing H1 mobility and transcription factor recruitment ⁴⁷. It is known that K34 homocitrullination also increases H1 mobility, inducing a more open state that facilitates DNA-protein interactions or by modulating DNA interaction with regulatory proteins ¹⁴.

AHR AND DIOXIN TOXICITY

Exposure to environmental pollutants happens daily, through air, food and water, and AhR works as a sensor for many environmental toxicants⁹. AhR ligands include toxic, synthetic compounds such as polychlorinated biphenyls (PCBs), polycyclic aromatic hydrocarbons (PAHs), and halogenated aromatic hydrocarbon compounds (HAH) such as dibenzo-*p*-dioxins.

The existence of AhR was first identified in mouse liver in a binding assay using 2,3,7,8 -Tetrachlorodibenzo-*p*-dioxin (TCDD) ¹, a toxicant that belongs to the dioxin-like family, which includes polyhalogenated aromatic hydrocarbons, such as polychlorinated dibenzodioxins, polychlorinated biphenyls and polychlorinated dibenzofurans⁴. Dioxin structures and mechanisms of action are similar; the members of the family can bioaccumulate and are persistent pollutants for which biological responses are facilitated by their interaction with cellular proteins. These chemicals can accumulate in water and soil, can be transported through air and water, and continuously bioaccumulate through the food chain ^{48,49}.

TCDD is considered the prototypical dioxin as it is the most widely studied. TCDD, a human carcinogen, is not commercially produced, but can be generated as a by-product of paper bleaching, waste incineration, herbicide production, wood and fossil fuel combustion, volcanic eruptions and metal production⁵⁰. It is a persistent organic pollutant with a half-life estimated to be around eight years in humans, as observed in a long term study conducted on veterans of Operation Ranch Hand who were exposed to aerial release of agent orange in the Vietnam war. ^{48,51}

Lower dose exposures can activate the transcription of genes responsible for xenobiotic metabolism. Higher doses however can cause a variety of toxic effects such as birth defects, hepatotoxicity, cancer, immunotoxicity, weight loss, thymic involution, gastrointestinal hemorrhage and death ^{4,6,8,52,53}.

Several accidental exposures to TCDD have occurred during the production of chemicals such as pesticides. Most work exposures to TCDD occur during the production of trichlorophenol (TCP), pentachlorophenol (PCP) and 2,4,5 trichloroacetic acid. In 1976 a TCP reactor exploded in Seveso, Italy contaminating the surrounding area with TCDD. For the first time, levels of TCDD in human blood was measured early following exposure. The area of exposure was subdivided in three zones, A, B and R. A few weeks after the accident several residents developed chloracne, many plants and animals died, numerous residents presented with headaches, nausea and eye irritation, leading to the evacuation of zone A, the most proximal zone. Follow up studies revealed an increase in soft tissue sarcoma, hematopoietic neoplasms and gastrointestinal cancer ^{49,54,55}.

Studies with children exposed in the Seveso accident revealed high incidence of chloracne of residents of zone A, aberrations in tooth development, and a lower sperm concentration in adult males who were exposed with ages between 1 to 9 ⁵⁵.

A 25-year long-term study conducted by Consonni and collaborators revealed an increase in death rate linked to hematopoietic and lymphatic cancer in residents of zone A ⁵⁶. In the 35th year update the authors revealed that the mortality of lymphatic-hematopoietic cancer prevailed in zones A and B, was mostly among women ⁵⁷. The 25-year study also revealed an increased in fatal ischemic heart disease in men that lived in zone A, and an increased rate in female deaths linked to diabetes observed in all zones ⁵⁶.

To further understand the adverse effects cause by dioxin, studies of wildlife and laboratory animals exposure are of utmost importance⁵⁸. Research conducted in the Great Lakes area, a region polluted with dioxin and dioxin-like compounds, demonstrated the health impact of dioxin exposure in birds, reptiles, fish and mammals, which develop

reproductive impairment, impaired immunological function, endocrine disorders, altered glucose metabolism and so on⁵⁸⁻⁶⁰.

Dioxins are metabolic-resistant carcinogens that can cause a wide range of adverse effects. Although the AhR is essential to TCDD toxicity, the precise mechanism for AhR in TCDD toxicity has not been fully characterized. The non-canonical pathway discovery and characterization can generate more important information of TCDD toxicity at a molecular level.

THE AHR STRUCTURE

As previously stated, the structure of AhR has been widely studied and described. The AhR protein contains a bHLH domain in the N-terminal region, present in several transcriptional factors. The second domain contains two subdomains, PAS A and B. The C-terminal of the protein contains the transcription activation domain⁶¹⁻⁶³. hAhR and mAHR share a 86% of amino acid sequence similarity in the N-terminal region of the proteins, while the C-terminal region is 58% similar (Figure 3)⁶⁴.

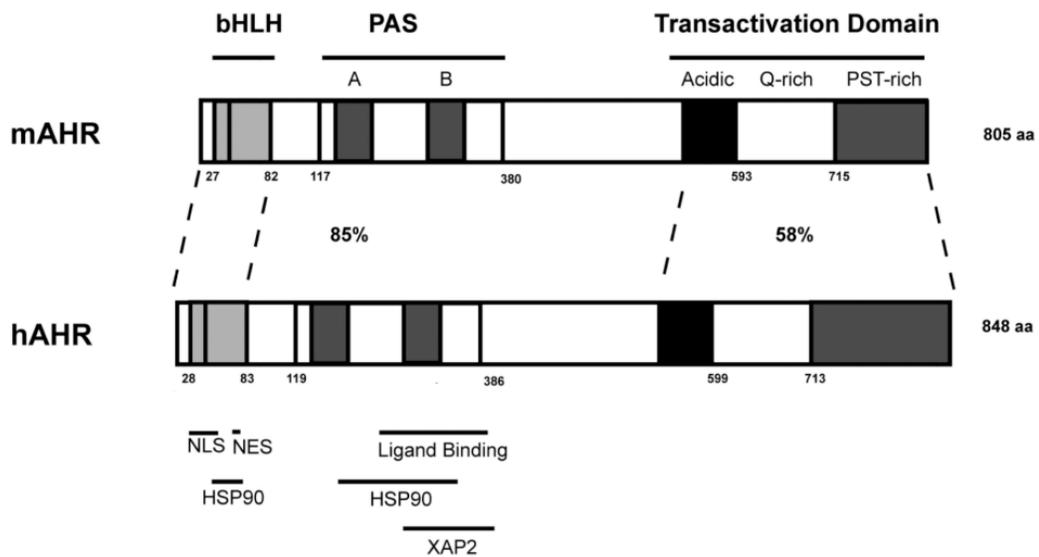


Figure 3: Mouse and human AhR domains. From Flaveny & Perdew 2009⁶⁴. Modified. Reproduced under CC BY-NC.

The characterization of mouse bHLH and PAS domains of AhR were first described by Fukunaga and coworkers (1995)⁶⁵ by deletion analysis. Ligand binding properties and ARNT interaction were evaluated using mutants synthesized *in vitro*. Deletion of AhR C-terminal region did not disrupt ligand binding, while the deletion of the PAS domain diminished ligand binding. A complete loss of binding was observed when the PAS A and B are absent. Binding analysis using full-length radiolabeled [³⁵S] ARNT and TCDD revealed that deletion of the basic region did not disrupt AhR-ARNT interaction, while deletion of bHLH (or individual helices), or PAS domains partially or completely disrupted interaction with ARNT. The authors also evaluated the protein complex XRE binding, concluding that an AhR protein lacking the basic region was not able to interact with DNA.

To examine ligand activation of AhR, Whitelaw *et al.* (1993)⁶⁶ mapped the AhR ligand binding region. By constructing chimeras containing the DNA binding domain of the glucocorticoid receptor (GR) and mouse AhR (mAHR) mutants lacking fragments of the bHLH domain, it was demonstrated that the ligand binding region of AhR is localized between amino acids 230 and 421 (a region harboring the PAS B domain and nearby amino acids), due to the inability of the mutant to form a complex with HSP90 and TCDD *in vitro*.

Different groups studied the C-terminal region of AhR. Rowlands *et al.*¹⁸ investigated the human (hAhR) transcription activation domain localization. Using AhR fusion proteins with a heterologous DNA binding domain they confirmed by β -galactosidase assay that deletion of 303 amino acids on the C-terminal region results in a >90% decrease in transactivation, while a fusion protein encoding the C-terminal 303 amino acids is able to induce the transcription of a reporter gene. Tkachenko and collaborators demonstrated that a hAhR protein lacking part of the Q-rich region, compared to the wild-type protein accumulates on the cytosol, and upon ligand activation the protein nuclear translocation is not stable, indicating that the region is important to inhibit protein export, and may be involved in other types of interactions(Figure 4)⁶⁷.

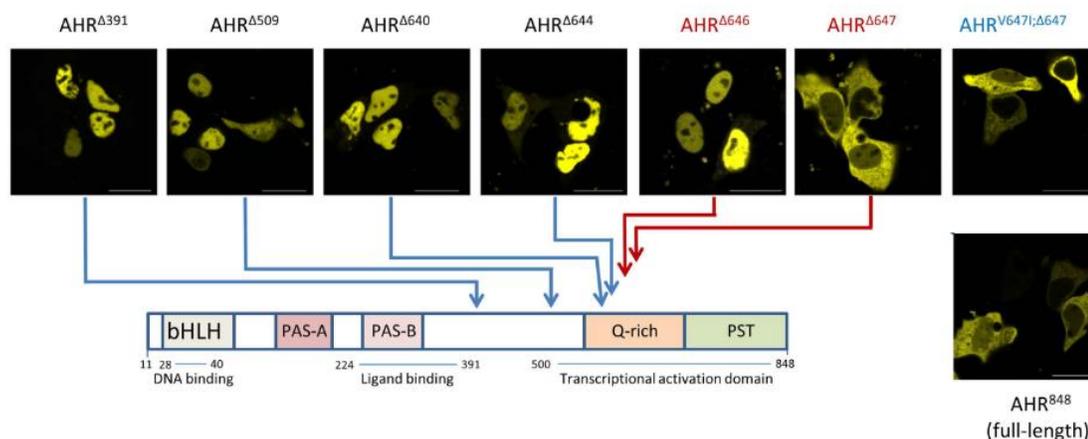


Figure 4: The AhR Q-rich region is essential for nuclear localization. Full length human AhR (AhR 848) resides on the cytosol. Deletions on the C terminal region of AhR compromised compartmentalization. Inclusion of the Q-rich region increased nuclear localization (in red). The valine 647 determines AhR localization, since full length AhR and AhR Δ 647 predominantly locate on the cytosol. From Tkachenko et al. 2016. Modified. Reproduced under CC BY 4.0) ⁶⁷.

According to published studies, the region of AhR responsible for XRE binding is located on the N-terminal region of the protein. Fukunaga *et al* (1996)⁶⁸ demonstrated by mutational analysis, alanine scanning, and EMSAs, that tyrosine 9 and arginine 14 of mAhR are required for DNA binding. Bungler and coworkers⁶⁹ generated mAhR mutants with amino substitutions from amino acids 37-39 and demonstrated that although the mutants were able to bind to the ligand 2-azido-3-[¹²⁵I]iodo-7,8-dibromodibenzo-p-dioxin, these were not able to bind to a XRE oligonucleotide in EMSA in the presence of wild-type ARNT and the agonist β -Naphthoflavone (β NF). According to the crystal structure of AhR and ARNT complex, PAS A and bHLH domains of AhR and ARNT bind to the XRE, AhR interacts with the DNA from amino acids 36 through 40 ^{70,71}.

AHR BINDING PARTNERS

According to Biogrid (2021)⁷², there is evidence of 76 interactors of AhR (Appendices III). The interactions with ARNT, HSP90, p23 and XAP2 have being

described by several groups, as well the interactions with estrogen receptor 1 (ESR1), retinoblastoma 1 (REL1), and nuclear receptor 1 and 2 (NCOA1 and NCOA2).

Upon TCDD exposure and in the absence of 17 β -estradiol, AhR recruits ESR1 to the XREs and away from ESR1 target genes, consequently disrupting estrogen signaling.⁷³ Pull down assays in TCDD treated MCF7 cells revealed that both AhR and ARNT interact with ESR1, and in the presence of 17 β -estradiol and cycloheximide, ESR1 can repress CYP1A1 and CYP1B1 upregulation⁷⁴.

The interaction of the AhR with retinoblastoma 1 (REL1) is linked to gene regulation and cell cycle⁷⁵. This interaction is responsible for maximal G1 arrest induced by TCDD in hepatoma cells⁷⁶. At least two AhR domains are essential for the interaction, an amino region with an LXCEX motif in the ligand binding domain and the Q-rich region²².

Nuclear receptor 1 (NCOA1) is a nuclear protein that interacts with nuclear receptors as a co-factor that stimulates transcription. The Q-rich domain of AhR interacts with NCOA1. The AhR transcription activation domain contains a flexible structure in the natural form, but in the presence of trimethylamine N-oxide (TMAO) and trifluoroethanol (TFE) adopts a folded conformation, indicating that this region is essential for proper protein folding and to induce transcription activation^{63,77}.

Nuclear receptor 2 (NCOA2) is a co-activator of nuclear and steroid receptor, that with NCOA1, regulates the energy balance between brown and white adipose tissue. In the presence of TCDD, this protein interacts with ARNT and AhR. NCOA1, NCOA2 and p160 binds to CYP1A1 promoter region, indicating these proteins act in gene regulation induced by TCDD.^{78,79}

Transcriptional factor 1 (Sp1) is a Kruppel like factor protein, that regulates gene expression of genes involved in immune responses, cell differentiation, apoptosis and growth. SP1 binding is identified in the promoter region of several XREs. SP1 interacts

with both AhR and ARNT (within the HLH-PAS domains), and the basal expression of SP1 target genes is dependent on the interaction with AhR-ARNT^{21,80,81}.

KRUPPEL LIKE FACTORS

The Kruppel Like Transcription Factors (KLFs) are zinc finger proteins that belong to the Sp1/KLF family of transcriptional factors. KLFs can act as regulators in cellular process like growth, migration, proliferation, differentiation apoptosis. The C-terminal region of the KLFs proteins consists of the highly conserved cysteine and histidine zinc fingers, while the N-terminal domain is responsible for protein-protein interaction⁸²⁻⁸⁴.

There are at least 17 members of the mammalian KLF family, and they can recognize GT and GC rich regions, activating or repressing transcription. The members of the family are divided into 3 groups (figure 5) based on their biological functions. Group 1 includes members that interact with transcriptional repressors and includes KLF3, KLF8 and KLF12. The second group consists of KLF1, KLF2, KLF4, KLF5, KLF6 and KLF7, proteins that are primarily activators, but may interact with repressors. The third group

includes KLF9, KLF10, KLF11, KLF13, KLF14 and KLF16, proteins that also act as transcriptional repressors ^{82,83,85}.

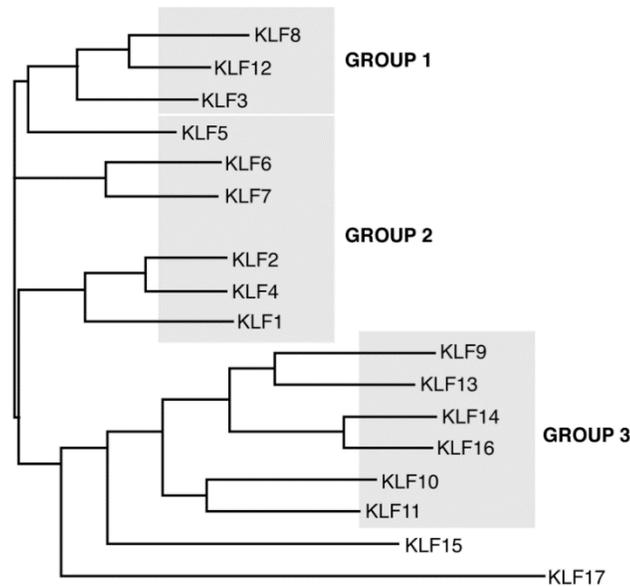


Figure 5: KLF groups. KLFs are divided in groups based on their homology and similarities on the N-terminal domain, that will induce the interactions with similar partners, leading to transcription activation or repression. From McConnell & Yang 2011 ⁸⁷ (Reproduced under CC BY 4.0).

Due to their ability to activate or repress a large gene repertoire, KLFs can regulate diverse processes such as cardiac system development, hematopoiesis, adipogenesis, formation of the epithelial barrier, development of the smooth and skeletal muscle, development of the cardiovascular system, lung formation, gluconeogenesis, intestinal development and so on. This vastness of functions is explained by the differences on the N-terminal domain that allow the interaction with different sets of proteins. Due to this variety of functions, some KLF can antagonize other functions. For example, while KLF4, 5 and 15 induce adipogenesis, while KLFs 2 and 3 suppress it. ^{86,87}

Kruppel like factors and liver

Among the Kruppel Like factors, the ones that have crucial roles in the liver are KLF1, KLF2, KLF4, KLF6, KLF11, KLF14, KLF15^{88,89}. Their functions are diverse including gluconeogenesis, apoptosis, oxidative stress, adipogenesis, cell differentiation and cycle progression⁹⁰.

In gluconeogenesis, amino acid catabolism and lipid metabolism regulation are controlled partly due to KLF15 action. KLF15 expression is upregulated by food deprivation and downregulated after food intake. Deletion of KLF15 leads to fasting hypoglycemia due to hepatic downregulation of gluconeogenic enzyme expression, deficient amino acid metabolism and improved hepatic insulin resistance provoked by a high fat diet^{87,89,91}.

The liver has the highest proportion of macrophages among the organs, where they participate in essential functions such as inflammation, tissue repair, metabolism and fibrosis⁹². Certain KLF members are also involved in macrophage activation. KLF4 plays a role in the induction of the M2 phenotype by inducing IL-4; KLF2 regulates nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) inhibiting the macrophage pro-inflammatory activation and KLF6 and NF- κ B can promote macrophage activation.^{89,93}

KLF6 is highly expressed in the liver, where it is involved in fetal liver development, glucose metabolism, regulation of hepatic fatty acid and it is also implicated in liver steatosis and fibrosis.^{89,94,95}

The expression of KLF6 in hepatocytes induces the activation of liver glucokinase that is linked to insulin resistance; and post translational activation of the nuclear receptor peroxisome proliferator activated receptor alpha (PPAR α) which will regulate the fatty acid and glucose metabolism⁸⁹. KLF6 is upregulated during liver steatosis and fibrosis by regulating the expression of fibrosis and steatosis related genes, such as transforming growth factor - β 1, collagen 1 and PPAR α ^{82,89,96,97}.

KLF6 (also known as Zinc-finger transcription factor 9 (ZF9), core promoter binding protein (CPBP) or G-box binding factor (GBF)) cDNA was first isolated in placental cells by using a promoter probe encoding a TATA box-less gene⁹⁸. KLF6 is

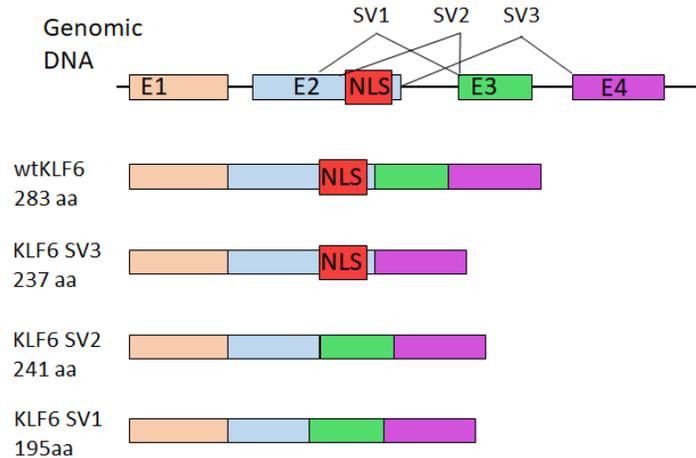


Figure 6: KLF6 genomic DNA and splice isoforms.

highly expressed on the placenta, as well in the hindgut, kidney and lung during development and in the adult liver, lung, heart, endothelial cells and intestines, being expressed in all other tissues at lower levels⁸⁷. The full-length transcript encodes a 42kDa protein with three C₂H₂ zinc fingers, 283 amino acids, 82 in the DNA binding domain and 201 amino acids in the transactivation domain. Three alternative splice forms have been isolated, and these encode the proteins KLF6-SV1 (195 aa), KLF6-SV2 (241 aa) and KLF6 SV3 (237 aa). (Figure 6)⁹⁹.

The human protein amino acid sequence indicates a structure similar to other transcription factors, including the three C-terminal zinc fingers, a serine/threonine rich domain in the central region of the protein and an acidic N-terminal region⁹⁹.

To better understand KLF6 physiological functions, Matsumoto & collaborators (2006)¹⁰⁰ generated a KLF6 knockout mouse model, and found that homozygous loss of KLF6 is embryonic lethal, and the lethality at day 12.5 is caused by reduced hematopoiesis, deficient yolk sac vascularization and impaired liver formation.¹⁰⁰

KLF6 acts as a transactivator of several target genes, such as cyclin-dependent kinase inhibitor 1 (p21^{Cip1})⁴², transforming growth factor beta 1 (TGF- β 1)^{101,102}, insulin-like growth factor I (IGF-I)¹⁰³, laminin α 1¹⁰⁴, nitric oxide synthase¹⁰⁵, keratin 12¹⁰⁶, acid ceramidase¹⁰⁷, keratin 4¹⁰⁸, endoglin¹⁰⁹, Leukotriene C4¹¹⁰, E-cadherin¹¹¹.

KLF6 can act as a suppressor of growth by inducing the expression of p21^{Cip1}, a regulator of cell cycle (G₁/S) progression, binding to its promoter, independently of p53⁴⁴. Narla *et al* (2007)⁴² generated a transgenic mice model capable of overexpressing human KLF6 on the liver. The transgenic animals exhibited lower levels of serum albumin, body weight and liver mass and an increased expression of p21^{Cip1} in the hepatocytes. These results demonstrated KLF6 is a key regulator of hepatocyte proliferation by the transcriptional regulation of p21^{Cip1}.

KLF6 stimulates pre-adipocyte differentiation by downregulating delta-like 1 (Dlk1), a transmembrane protein that acts as a cell growth regulator. According to experiments performed in 3T3-L1 cells, KLF6 recruits Histone Deacetylase 3 (HDAC3), and the complex binds to the promoter region of Dlk1, downregulating its expression, which will then stimulate pre adipocyte differentiation^{112,113}.

CPS1

CPS1 is a multidomain protein, with two distinct active centers¹¹⁴, that catalyzes the synthesis of carbamoyl phosphate from bicarbonate and ammonia^{115,116}. The three-step reaction consists of: the catalysis of ATP and bicarbonate which produces ADP and carboxy phosphate that will react with ammonia generating carbamic acid. The carbamic acid reacts with ATP creating carbamoyl phosphate and ADP¹¹⁵.

CPS1 is a protein with 1500 residues that after cleavage results in a 1463 polypeptide expressed predominantly in enterocytes and hepatocytes^{117,118}, divided into five major domains that are not fully characterized, since the crystal structure of the whole protein has not been determined and there was a lack of expression systems able to express

the protein^{115,117-119}. The domain composition may be similar to what is observed in *Escherichia coli* CPS. The N-terminal glutaminase-like domain does not use glutamine to produce ammonia, as observed in *E coli* protein, so the function of this region is not known; it is speculated it may be linked to protein activation and stabilization^{117,118}. The C-terminal moiety of the protein is divided in two 60 kilodaltons (kDa) subdomains, with a N-terminal 45 kDa phosphorylation domain (that phosphorylate bicarbonate), followed by a 15 kDa integrating domain, which function is not characterized yet, and a C-terminal 45 kDa region that phosphorylates carbamate and a 15 kDa catalytic and N-acetyl-L-glutamate (NAG) region, essential for protein activation^{118,120}.

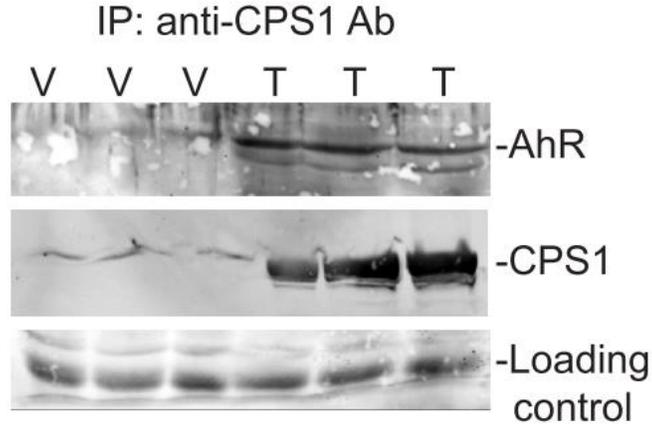


Figure 7: AhR-CPS1 interaction. Immunoprecipitation of mouse liver nuclear extracts of animals treated with vehicle (V) or TCDD (T) with CPS1 antibody. Lamin was used as a load control. AhR is detected in animals treated with TCDD. From Joshi *et al.* 2015³⁴. Reproduced under CC BY.

As stated previously, Joshi and coworkers (2015)³⁴ demonstrated that AhR interacts with CPS1. The nuclear fraction of mouse livers of animals treated with TCDD or vehicle was immunoprecipitated with AhR or CPS1 antibodies, and the results confirm the protein-protein interaction (Figure 7).

Chromatin immunoprecipitation (ChIP) performed with mouse livers using AhR, CPS1, KLF6, IgG (negative control) and histone H3 (positive control) antibodies followed

by polymerase chain reaction (PCR) using primers against PAI-1 and CYP1A1 promoters, harboring an NC-XRE and XRE respectively, demonstrated that CPS1, AhR and KLF6 are recruited to the PAI-1 promoter in the presence of TCDD. Neither CPS1 or KLF6 are recruited to the CYP1A1 promoter.

Further investigation using murine AML12 hepatoma cells and mouse primary hepatocytes revealed that CPS1 is essential for homocitrullination, a post-translational modification of lysine residues that reacted with isocyanic acid¹²¹, of a histone H1 lysine 34 (H1K34) involved in transcriptional regulation. This discovery identified a new epigenetic mark that could induce chromatin remodeling into a more accessible conformation.

THE INVOLVEMENT OF PLASMINOGEN ACTIVATOR INHIBITOR-1 (PAI-1) IN THE NON-CANONICAL PATHWAY

Aiming to map AhR target genes expressed in liver of adult male mice, Tijet & collaborators (2006)¹²² performed expression arrays to detect mRNA profiles of wild-type and knockout animals treated with TCDD for 19 hours. The transcript profiles revealed expression of AhR targets belonging to the classical pathway, such as Cyp1a1. Certain protease inhibitors were also identified, including Plasminogen activator inhibitor-1 (PAI-1).

Fibroblasts, immune cells, endometrium, peritoneum, adipocytes and liver cells produce PAI-1, a protein encoded by Serpine1. PAI-1 is involved in blood clotting by inhibiting the action of plasminogen activators, such as urokinase plasminogen activator (u-PA) and tissue type plasminogen activator (t-PA). Both of these activators can induce fibrinolysis that will culminate with the conversion of plasminogen to plasmin followed by fibrin proteolysis^{123,124}.

Considering that PAI-1 levels increase after partial liver transplantation in humans¹²⁵, it is a TCDD-responsive gene^{122,126-129}, and the promoter region lacks a recognizable

XRE¹²⁹, Huang and Elferink (2012)³⁶ decided to investigate the PAI-1 promoter region using a luciferase reporter system. They discovered that a 200 bp region lacking a XRE element showed AhR mediated transcription activation dependent on TCDD. Further investigations by EMSAs using mouse liver nuclear extracts revealed the region of the PAI-1 promoter that conferred protein-DNA binding, and that this region consists of a 5'GGGA3' tetranucleotide repeat.

ChIP experiments showed a TCDD-dependent association of AhR in the PAI-1 promoter, and this interaction is not observed when ARNT antibodies were used.

AHR-KLF6 DNA INTERACTION

While following up on Huang & Elferink's findings, Wilson *et al.* (2013)¹³⁰ observed homology between the DNA binding sequence of KLF6 and the NC-XRE which indicated KLF6 as a potential AhR binding partner.

To investigate AhR-KLF6 interaction, Wilson *et al.* (2013) performed co-immunoprecipitations, EMSAs, and ChIP assays. According to their major findings, AhR interacts with KLF6 and the complex binds to the NC-XRE in a TCDD dependent manner.

ChIP assays were used to evaluate the complex interaction of KLF6 with the NC-XRE *in vivo*. Using whole liver extracts of animals treated with vehicle or TCDD, ChIP was performed using antibodies against AhR and KLF6 proteins. Histone H3 and IgG were used as positive and negative controls, respectively. Based on their findings, both proteins are bound to the PAI-1 promoter at the same time. Since the interaction with KLF6 was not observable in ChIP experiments examining the Cyp1a1 promoter containing XREs, it was determined, that KLF6 is not involved in the canonical pathway.

AHR GENOMICS AND TRANSCRIPTOMICS

Alterations in gene expression observed due to AhR activation by ligands can induce pathological effects such as hepatotoxicity, developmental problems,

immunotoxicity, fibrosis¹³¹⁻¹³³ or physiological regulatory effects such as detoxification, cellular adhesion, heart development, hematopoiesis, female reproduction^{9,134-139}.

The use of chIP followed by microarrays or next generation sequencing contributed to the discovery of protein-DNA interactions by nuclear receptors and the location of regions where the DNA transcription factors bind. Similar strategies contribute to elucidating where AhR binds on the promoter region and its possible binding partners.¹³² Complementary transcriptomics analysis linked these DNA binding events to cellular functions in the presence or absence of ligands.³

Tijet and colleagues (2006)¹⁴⁰ mapped genes regulated by AhR using expression arrays to map genes of AhR null and wild-type mice. The ProbeSets expression was remarkably altered due to TCDD exposure in an AhR-dependent manner, and the upregulation of classical AhR target genes such as Cyp1a1, Cyp1b1, Cyp1a2 and NAD(P)H dehydrogenase [quinone] 1 (Nqo1) was observed. It was also observed that in the absence of an exogenous ligand, certain ProbeSets were affected, indicating that AhR plays important roles in normal physiology¹⁴⁰.

Dere and collaborators (2011)¹³² compared the transcriptomic profile of mice treated with TCDD with the gene enrichment observed in chIP-chip assays. They identified that 47.5% of TCDD responsive genes contain AhR enrichment in intragenic regions, (10 kb upstream of the transcription start site to the transcription end site). Around 70% of enriched genes at a 2-hour time point contained an XRE consensus motif, which suggests that the remaining genes could be secondary responses or be regulated through an NC-XRE dependent mechanism.

Prokopec *et al.* investigated the transcriptomic changes induced by TCDD in different mammalian species and tissues. It was observed that there was an upregulation of Cyp family across species and tissues as well as Aryl hydrocarbon Receptor Repressor (AhRR) and Nqo1, typical AhR-responsive genes. Although these AhR core genes were upregulated among different tissue and species, most of the genes believed to be activated

by AhR showed little overlap in the analysis, demonstrating the continued need for transcriptomics studies of changes induced by TCDD ¹⁴¹.

KLF6 ROLES IN GENOMICS AND TRANSCRIPTOMICS

The Kruppel Like Factors are transcription activators and/or repressors with divergent N-terminal domains and transcription regulatory motifs ^{84,142}; The remarkably conserved C-terminal regions possess three consecutive zinc fingers DNA-binding domain that can interact with CACC or GC boxes, located on the DNA of the responsive promoters ^{83,89,142,143}.

Kruppel like factors expression is altered in human cancers ^{83,88,144} and they can act as oncogenes or tumor suppressors, depending on the type of the tumor ^{83,145,146}. Since KLF6 is expressed in diverse tissues, its functions are altered in bone, breast, head and neck, liver, lung, ovary, pituitary, prostate and uterus cancers ^{142,146}.

KLF6 expression is decreased in gastric cancer ^{88,147}, and this downregulation can result in decreased cell differentiation, since KLF6 regulates the expression of Cyclin Dependent Kinase Inhibitor 1A (CDKN1A) and the MYC proto-oncogene bHLH transcription factor (MYC) ¹⁴⁸.

Li *et al.* ¹¹² demonstrated by EMSAs and ChIP the importance of KLF6 in adipocyte differentiation. According to their research, KLF6 transactivation domain recruits Histone Deacetylase 3 (HDAC3) and the complex binds to the promoter region of Delta Like Non-Canonical Notch Ligand 1 (Dlk1), repressing its activity, which culminates with adipocyte differentiation. The lack of KLF6, achieved by siRNA transfection on 3T3-L1 cells resulted in impaired cell differentiation, demonstrating the importance of KLF6 in the maturation of pre-adipocytes.

The levels of an alternatively spliced form of KLF6 that lacks a nuclear localization region are increased in pancreatic tumor, and the protein accumulates in the cytosol ¹⁴⁹. Diaferia & collaborators ¹⁵⁰ analyzed the genetic expression of human pancreatic ductal

adenocarcinoma using different pancreatic cell lines. To evaluate the genomic occupancy of certain transcription factors (TF), including KLF6, CFPAC-1 cells were subjected to ChIP-seq. Interestingly, almost 30% of the peaks identified included AhR target genes observed in ChIP-seq performed in livers of animals treated with TCDD for two hours, such as CYP1B1, PAI-1, AHRR, CDKN1A, KLF9¹⁵¹, which could indicate that AhR and KLF6 regulate common pathways.

Chapter 2: Characterization of AhR and KLF6 interaction

INTRODUCTION

AhR is a member of the PER-ARNT-SIM superfamily of proteins, acting as a sensor of endogenous and exogenous signals. Activation of AhR by endogenous or exogenous ligands triggers its translocation to the nucleus, where it binds to ARNT and the heterodimer complex binds to XRE regions in the DNA^{9-11,152}.

While evaluating AhR target genes, several research groups demonstrated that AhR controls gene expression of targets that lack XREs by interacting with different transcription factors. KLF6 was characterized as one of the novel AhR binding partners. The AhR-KLF6 complex binds to a non-consensus XRE (NC-XRE), that consists of a 5'GGGA' tetranucleotide repeat that is found in the PAI-1 promoter. The AhR-KLF6 complex recruits CPS1 which induces homocitrullination of lysine 34 of histone H1, and subsequently may induce a more relaxed chromatin conformation facilitating transcription^{14,34,35,152,153}.

KLF6 is a zinc finger transcription factor involved in cell proliferation, differentiation, growth, inflammation and apoptosis. It is the best characterized KLFs found in the liver, being essential for liver development and glucose and fatty acid metabolism inducing activation of glucokinase and peroxisome proliferator-activated receptor α (PPAR α), a key regulator of lipid metabolism^{89,90,154,155}.

In order to increase our understanding of the novel AhR non-canonical pathway, EMSAs using AhR, KLF6, and CPS1 mutant proteins produced *in vitro* were performed to characterize the protein complex binding to the NC-XRE. To determine the regions and amino acids responsible for AhR and KLF6 interaction, deletion and alanine mutation

constructs encoding human proteins were produced and tested using microplate binding assays.

MATERIALS AND METHODS

Cell Culture

MCF7 cells were cultivated following standard protocol¹⁵⁶. Briefly, cells were cultivated in 75cm² flasks in Dulbecco's Modified Eagle Medium with 10% of fetal bovine serum. For sub culturing, media was removed, the flask was washed with PBS and incubated with 2mL of trypsin-EDTA until the cell layer was dispersed. 8mL of growth media was added and the suspension was centrifuged at 180x G for 5 minutes. The supernatant was discarded, and 10 mL of fresh media was added. Cells were counted in a hemocytometer and 5.0 x 10⁶ cells were plated in 150mm dishes.

Nuclear extraction

For nuclear extraction, cells were cultivated in 150mm dishes until 80% confluency. Cells were treated with 6nM TCDD or DMSO as control for 1 hour at 37° C. Nuclear extraction at 0°C was performed after treatment. Nuclear and cytoplasmic extraction was performed based on the methodology proposed by Miller *et al.* (1983)¹⁵⁷. After treatment, dishes were washed twice with HEPES (10mM, pH 7.5) and incubated for 15 minutes. Cells were scrapped with MDH buffer (3mM MgCl₂, 1mM DTT, 25mM HEPES, pH 7.5) and homogenized with 15 strokes on a tight fitting dounce homogenizer. The homogenate was centrifuged at 1000x G for 2 minutes to extract the crude nuclei. The crude nuclei was homogenized in MDH buffer containing 0.5% triton X, and centrifuged for 2 minutes at 1000x G. The pellet was washed with MDH buffer and lysed with 0.4M KCL, 1mM DTT, 25mM HEPES (pH 7.5). The lysate was centrifuged at 105,000x G for 60 minutes at 4°C. Protein concentration was measured by Lowry DC Protein Assay

(Biorad, Hercules, CA) following manufacturer instructions and 10% glycerol was added to the samples. For EMSA experiments, 10 µg of nuclear proteins were used.

AhR, KLF6, CPS1 constructs

Human AhR, CPS1, and KLF6 full length cDNA were obtained from the Mammalian Gene collection (MGC) Horizon Discovery. The bacterial stocks provided by the manufacturer were cultivated on LB plates with ampicillin (for AhR and KLF6 vectors and kanamycin for CPS1). Plasmid DNA was purified using Qiagen QIAprep® Miniprep (Hilden, Germany) according to manufacturer instructions. The plasmids were used as a template to amplify by PCR ORFs for the mature AhR, KLF6 and CPS1 (lacking the signal peptide). Two step PCR was performed using Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA), following the manufacturer's protocol. Samples were prepared to be cloned into the pT7CFE1-Chis vector using the MscI (for AhR and KLF6), NdeI (for CPS1) and XhoI (for AhR and KLF6) restriction sites. AhR constructs were designed to express a HA (YPYDVPDYA) tag in the C-terminal region, KLF6 constructs to express a FLAG (DYKDDDDK) tag in the C-terminal region and CPS1 to express the His tag present in the vector located after the XhoI restriction site. PCR reactions were purified using Qiagen QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). Purified PCR products and pT7CFE1-Chis vector (Thermo Fisher Scientific, Waltham, MA) were digested overnight at 37 °C with the appropriate restriction enzymes. The following day, the DNA was purified in agarose gels and was extracted using the QIAquick Gel Extraction Kit according to manufacturer instructions. After purification, ligation was performed overnight at 16°C with T4 DNA ligase (New England Biolabs, Ipswich, MA) and the required insert to vector ratio was calculated by NEBcalculator. After ligation, DNA was transformed in NEB® 5-alpha Competent *E. coli* (High Efficiency) (New England Biolabs, Ipswich, MA) and plated in LB agar plates with ampicillin. Plates were incubated overnight at 37°C. Single colonies were selected, and

grown in LB media with ampicillin. Plasmid DNA was purified using Qiagen QIAprep® Miniprep (Hilden, Germany). Constructs were sequence verified using specific targeted primers at the UTMB Molecular Genomics Core.

Mutation constructs

Deletion of N-terminal regions of KLF6 (Δ 15, Δ 27, Δ 37) (Figure 8B) and CPS1 (Δ 135) (Figure 8C) and C-terminal region of AhR (1-348, 1-386, 1-424,1-490, 1-540, 1-599, 1-604, 1-610, 1-620, 1-640) (Figure 8A) was performed by PCR on the wild-type construct with specific primer sequences (Appendices I). PCR reactions, purification, digestion, gel extraction, ligation, transformation and extraction were performed as described before.

Internal deletions (AHR Δ b, Δ 17-39, Δ 40-78, Δ 275-327, Δ 40-274, Δ 40-223, Δ 121-223, Δ 160-223 and CPS1 Δ 369-508) (Figure 9) and alanine mutagenesis (KLF6 29-30, 33-34, AhR 13-17, 37-39, 13-17 & 37-39, 600-601, 602-603, 600-604) (Figure 10) were performed using specific primers (Appendices I) and the Q5® Site-Directed Mutagenesis Kit (New England Biolabs, Ipswich, MA). PCR reactions, purification, digestion, gel extraction, ligation, transformation and extraction were performed as described before.

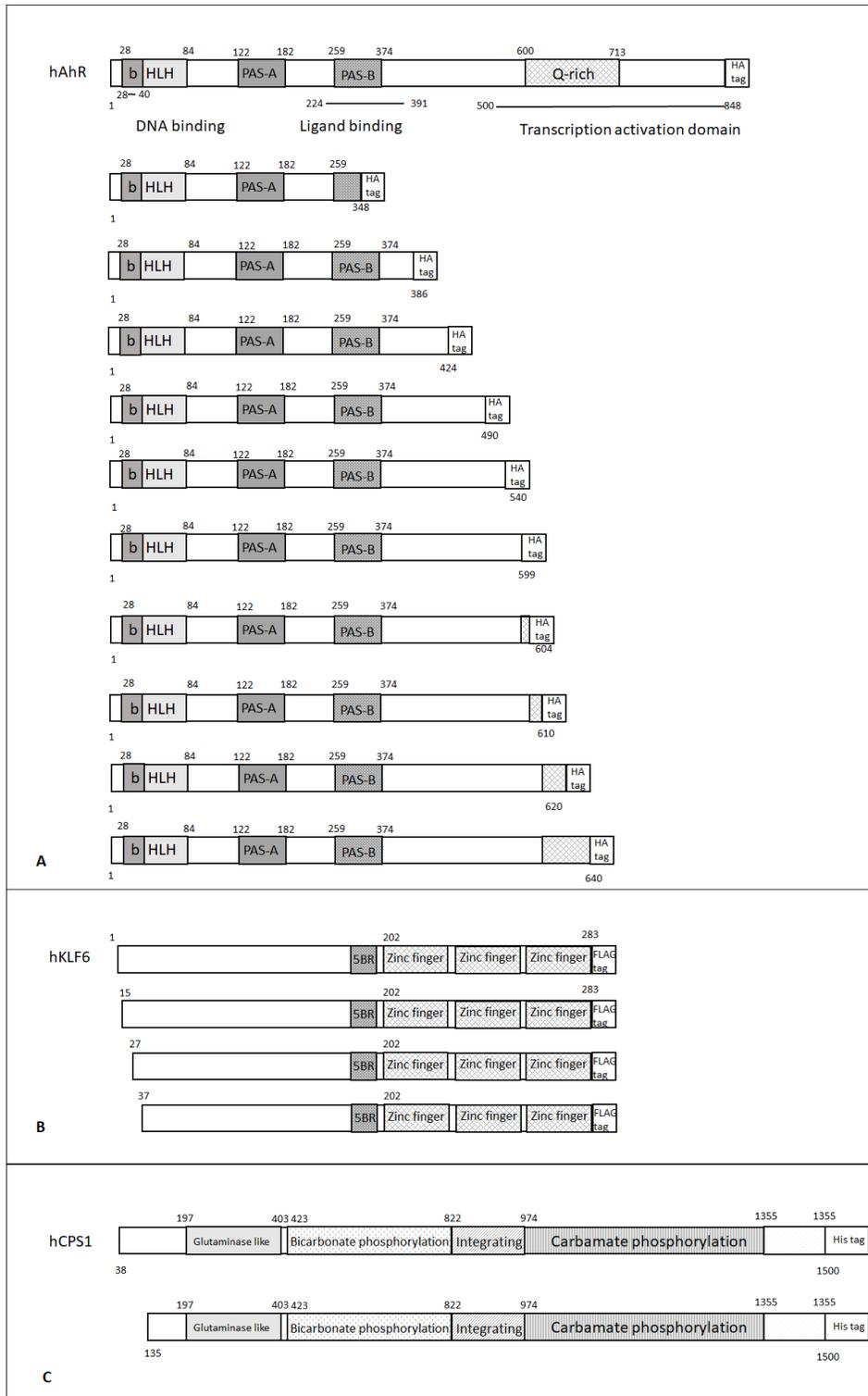


Figure 8: Schematic diagram of the expression constructs used. Full length cDNA encoding human AhR, KLF6 and CPS1 were cloned into the pTF7CFE1 C-his vector. (A) Wild-type AhR and C-terminal deletions, (B) Wild-type KLF6 and N-terminal deletions, (C) Wild-type CPS1 and N-terminal deletion.

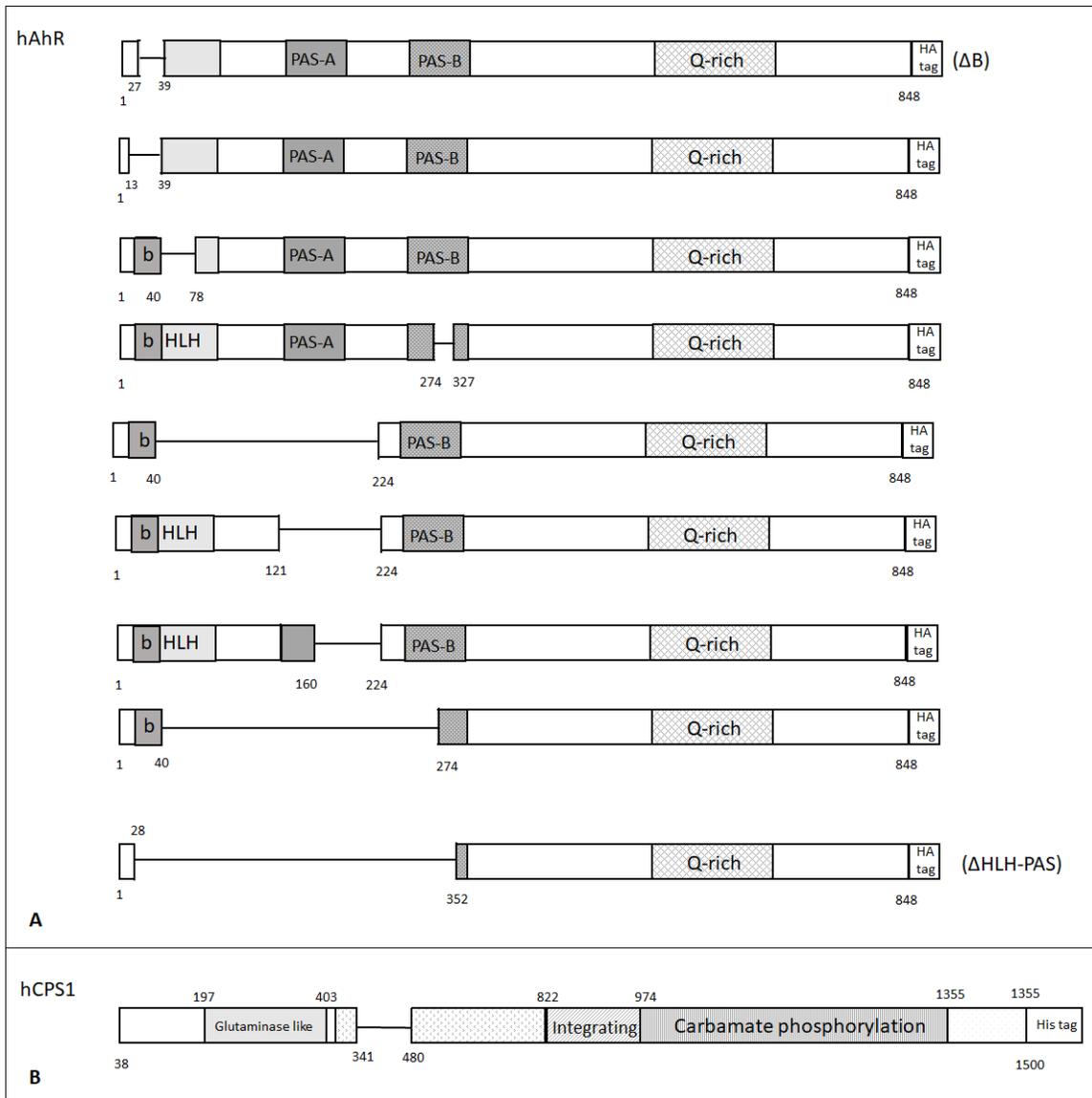


Figure 9 AhR and CPS1 clones. (A) AhR and (B) CPS1 internal deletions.

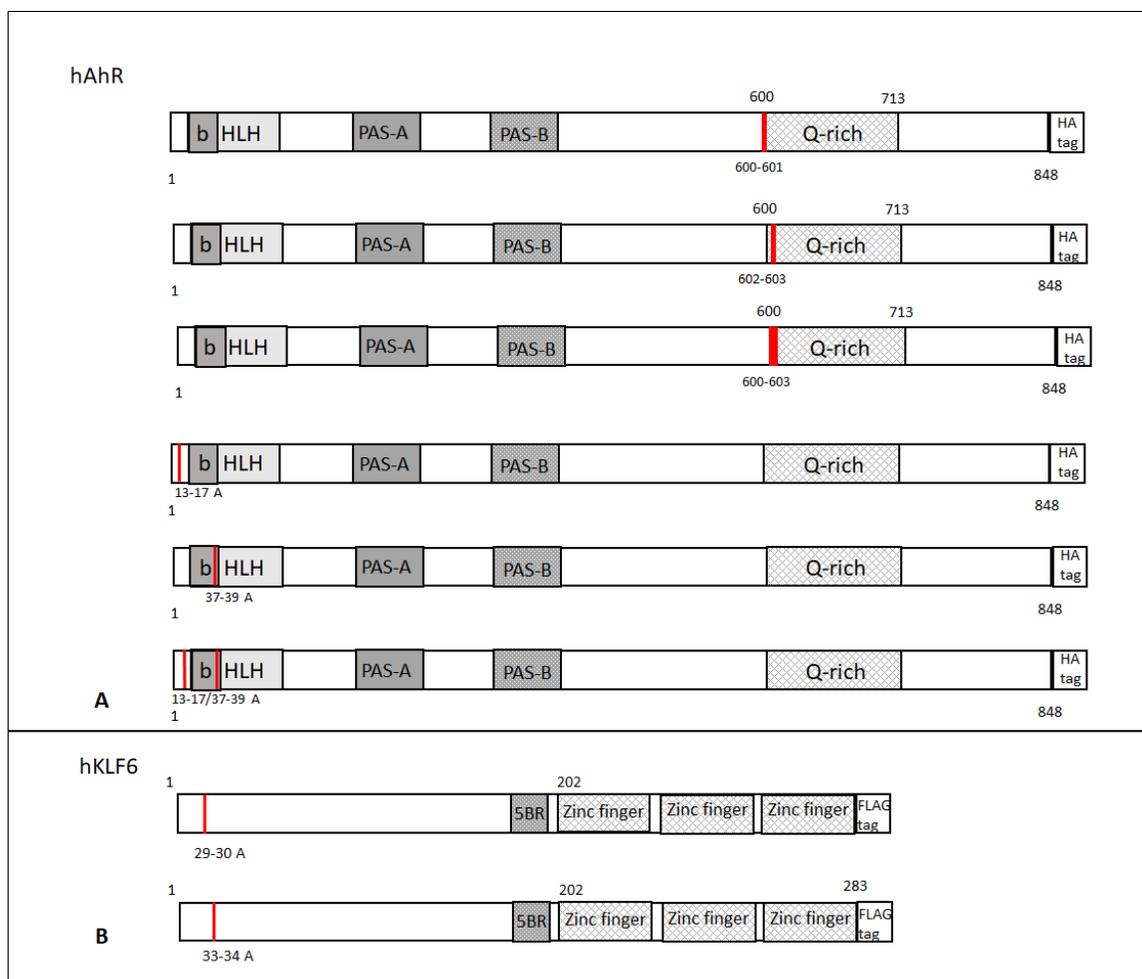


Figure 10: AhR and KLF6 alanine mutagenesis. Alanine substitutions were design to evaluate AhR (A) protein-DNA interactions and KLF6 (B) protein-protein interactions. Antibodies

Anti-AhR (rabbit polyclonal) antibody was obtained from Thermo (Waltham, MA) (catalog AV31635). Kruppel like factor 6 antibody (rabbit polyclonal) was purchased from Active Motif (Carlsbad, CA) (catalog number 61243). Anti-HA tag antibody (rabbit monoclonal) was acquired from Cell Signaling (Danvers, MA) (catalog 3956S), anti-HA tag antibody HRP conjugated,(goat polyclonal) was purchased from GenScript (Piscataway, NJ) (Catalog A00169), anti-Flag antibody HRP conjugated (Mouse monoclonal) (catalog A01428) from GenScript (Piscataway, NJ) (catalog A01428), anti-Flag antibody (Mouse monoclonal) (catalog A00187) and anti-His tag antibody (rabbit polyclonal) (Catalog A00174) from GenScript (Piscataway, NJ), anti-CPS1 rabbit

polyclonal HRP conjugated was acquired from Cusabio (Houston, TX) and anti CPS1 mouse monoclonal (catalog 376190) from Santa Cruz (Dallas, TX).

Coupled *in vitro* transcription translation and western blots

Initial efforts to express full-length proteins using the reticulocyte transcription-translation system proved unsuccessful. This system was not able to express all the AhR constructs or CPS1, which expression *in vitro* was successfully achieved only in baculovirus-insect cell expression system so far^{117,158}. To overcome the protein expression challenges, the 1-Step Human Coupled IVT Kit – DNA (Thermofisher, Waltham, MA) was utilized, following manufacturer instructions. 1 µg of each plasmid was used for each 25 µl reaction. The AhR and KLF6 reactions were incubated for 3 hours at 30 °C while CPS1 reactions were incubated for 5 hours. Total protein concentration was measured by DC™ Protein Assay (Bio-Rad, Hercules, CA). 2µL of each reaction was diluted in SDS sample buffer in reducing conditions and separated by polyacrylamide gel electrophoresis. The proteins were transferred to PVDF (Millipore-Sigma, Burlington, MA) membranes by a semi-dry electrophoretic transfer (Bio Rad, Hercules, CA).

After transferring, membranes were blocked in a Tris Buffered Saline-Tween 20 (TBS-T) solution of 5% dry milk and probed with specific primary antibodies. Proteins were detected using species-specific fluorescent secondary antibodies (GE Healthcare, Chicago, IL) and images were captured by a Typhoon Trio imager (GE Healthcare, Chicago, IL).

Quantification of expressed proteins

To quantify the expression of each AhR and KLF6 mutant protein compared with the wild-type expression, Sandwich Enzyme-Linked ImmunoSorbent Assay (ELISA) were used¹⁵⁹.

Solid 96-well polystyrene microplate (MilliporeSigma Burlington, MA) were coated overnight at 4°C with specific antibodies in carbonate buffer pH9.6. (α -HA or α -FLAG tag 1:500). After blocking (5% dry milk in PBS) for 3 hours and washing with Tris-buffered saline with 0.05% tween 20 (TBS-T) (three times), 25 μ g of each protein reaction diluted in PBS were added in each well (in triplicates). After 90 minutes incubation at room temperature with orbital shaking, plates were washed and incubated with detection antibody (α AhR or α KLF6 1:1000) for 1 hour. Next, plates were incubated with species-specific horseradish conjugated secondary antibody diluted in PBS for 1 hour. After washing with TBS-T and PBS (3 times TBS-T, one time PBS), QuantaRed™ Enhanced Chemifluorescent HRP Substrate (Thermo Fisher Scientific, Waltham, MA) was added, and the plate fluorescence was measured using Glomax Explorer automated plate reader (Promega, Madison, WI). The ratio of the fluorescent units from each recombinant protein was normalized to the wild-type protein. For protein binding assays, equivalent concentration of each mutant compared with 25 μ g of wild-type were utilized.

Microplate Protein-Binding Assay

To evaluate protein-protein interactions, solid phase protein binding assay were performed as described previously¹⁶⁰⁻¹⁶³, with some modifications (Figure 11). 96-well solid polystyrene microplates (Corning, MilliporeSigma Burlington, MA) were coated overnight with antibodies against one of the proteins of interest (AhR or HA tag antibodies; KLF6 or FLAG tag antibodies) in carbonate buffer pH 9.6. On the following day, plates were washed with TBS-T and blocked with 5% milk in PBS for 3 hours. Meanwhile, the target proteins were incubated with 6 nM of TCDD or DMSO as control in PBS at 30°C for one hour. After the incubation, 100 μ l of the protein complex reaction were added to each well, stirred on a shaker for 90min at room temperature and subsequently probed with an antibody conjugated with horseradish peroxidase for 1 hour. To detect the signal, QuantaRed™ Enhanced Chemifluorescent HRP Substrate (Thermo Fisher

Scientific, Waltham, MA) was added to the wells and neutralized with QuantaRed Stable Peroxide Solution after approximately 8 minutes of incubation. Fluorescence signal was measured with Glomax automated plate reader (Promega, Madison, WI).

To analyze the interaction between AhR and respective mutants, and KLF6, plates were coated with 1:500 α FLAG antibody (Genscript, Piscataway, NJ). Each plate contained negative control (wild-type AhR without KLF6) and blank wells. Wild-type AhR and KLF6 concentration was 25 μ g while the mutants concentration was dependent on the values observed during the protein quantification step discussed on the previous section. For protein detection, HA-tagged HRP-conjugated antibody was used.

To evaluate the interaction between KLF6, KLF6 mutants, and AhR, plates were coated with 1:500 α HA tag antibody (Cell Signaling, Danvers, MA). Each plate contained negative control (wild-type KLF6 without AhR) and blank wells. Wild-type KLF6 and AhR concentration used was 25 μ g while the amount of mutant protein used was dependent

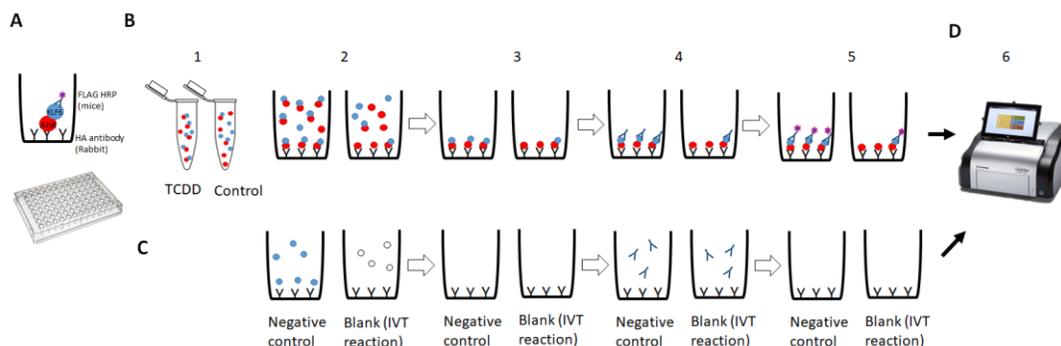


Figure 11: Microplate Protein-Binding Assay schematics. General assay design: (A) 96 well plates were coated with an antibody against one of the proteins of interest (HA for AhR detection). A second HRP conjugated antibody against the partner protein is added (FLAG for KLF6 detection). If a complex is formed, it can be detected by fluorescence. (B) Assays steps, with not filled arrows representing plate washing. (B1) *In vitro* produced protein diluted in PBS were treated with TCDD or DMSO as control. (B2) After incubation, samples were added in plates pre-coated with an antibody against one of the proteins. (B3) Plates were washed and incubated (B4) with HRP conjugated secondary antibody targeting the second protein of interest. (B5) QuantaRed™ Enhanced Chemifluorescent HRP Substrate was used for antibody detection. (C) For negative control, the target protein in PBS is incubated in the plate coated against the partner protein. Blank signal was obtained in wells were IVT lysate reaction was incubated. (D) Fluorescence signal was measured by Glomax Explorer plate reader. For each assay, reads were normalized to the blank signal and if a signal below the blank was detected, it was considered that no interaction occurred).

on the values observed during the protein quantification step discussed on the previous section. For protein detection, FLAG-tagged HRP-conjugated antibody was used.

Electrophoretic Mobility Shift Assay

For Electrophoretic Mobility Shift Assay (EMSA) 0.5 μl of AhR and KLF6 and 1 μl of CPS1 protein expressed *in vitro* were used in each reaction as described previously¹³⁰. The recombinant proteins were produced with the same concentration of plasmid (1 μg of plasmid per 25 μL reaction) for each assay and the products were analyzed by western blot to ensure similar expression was used on each assay.

For double stranded DNA oligonucleotides synthesis primer annealing, 1 nmol/ μl of each primer (primer sequences are available in appendices I) was diluted in nuclease-free water and placed on a 95 °C for five minutes and cooled down at room temperature. 100ng/ μl of the annealed primers were mixed with nuclease-free water, dGTP, dCTP and

dTTP. ^{32}P -ATP was added to the mix with Klenox buffer, klenow enzyme and ddH₂O. The mixture was incubated in Tris-EDTA buffer (TE buffer, 10 mM Tris, 1 mM EDTA, pH8.0) and purified with phenol:chloroform NaCl extraction. On the following day, after centrifugation the samples were added to G25 columns (GE healthcare, Chicago, IL) to remove the unincorporated radiolabeled ^{32}P -dATP. After spinning the radiolabeled DNA was collected.

The *in vitro* produced proteins (0.5 μl of AhR and KLF6 and 1 μl of CPS1) were incubated with 6nM of TCDD or DMSO as control for 15 minutes in HEDG buffer (1mM EDTA, 1mM DTT in 25mM Hepes with 10% glycerol, pH 7.4) at 37 °C. Samples were then incubated for 10 minutes at room temperature, with 200ng/ μl of poly deoxyinosinic-deoxy- cytidylic acid (PolydI·dC), 0.1M of DTT and 3mM of KCl. After incubation, 300 ng of ^{32}P -NC-XRE was added to the solution and the mix was incubated 10 minutes at room temperature.

Samples were electrophoresed under nondenaturing conditions on polyacrylamide gel (6%) in TAE buffer (40 mM tris, 1mM EDTA and 20mM acetic acid), exposed to phosphoscreens (GE Healthcare Life Sciences) overnight and imaged using Typhoon Trio Imager (GE Healthcare). Band intensities were quantified using ImageJ software ¹⁶⁴.

RESULTS

Protein expression

To study protein-protein and protein-DNA interactions, 26 AhR, 6 KLF6 (Figure 12) and 3 CPS1 (Figure 13) constructs were used. Proteins were successfully expressed *in vitro* using the 1-Step Human Coupled IVT Kit – DNA (Thermofisher, Waltham, MA)

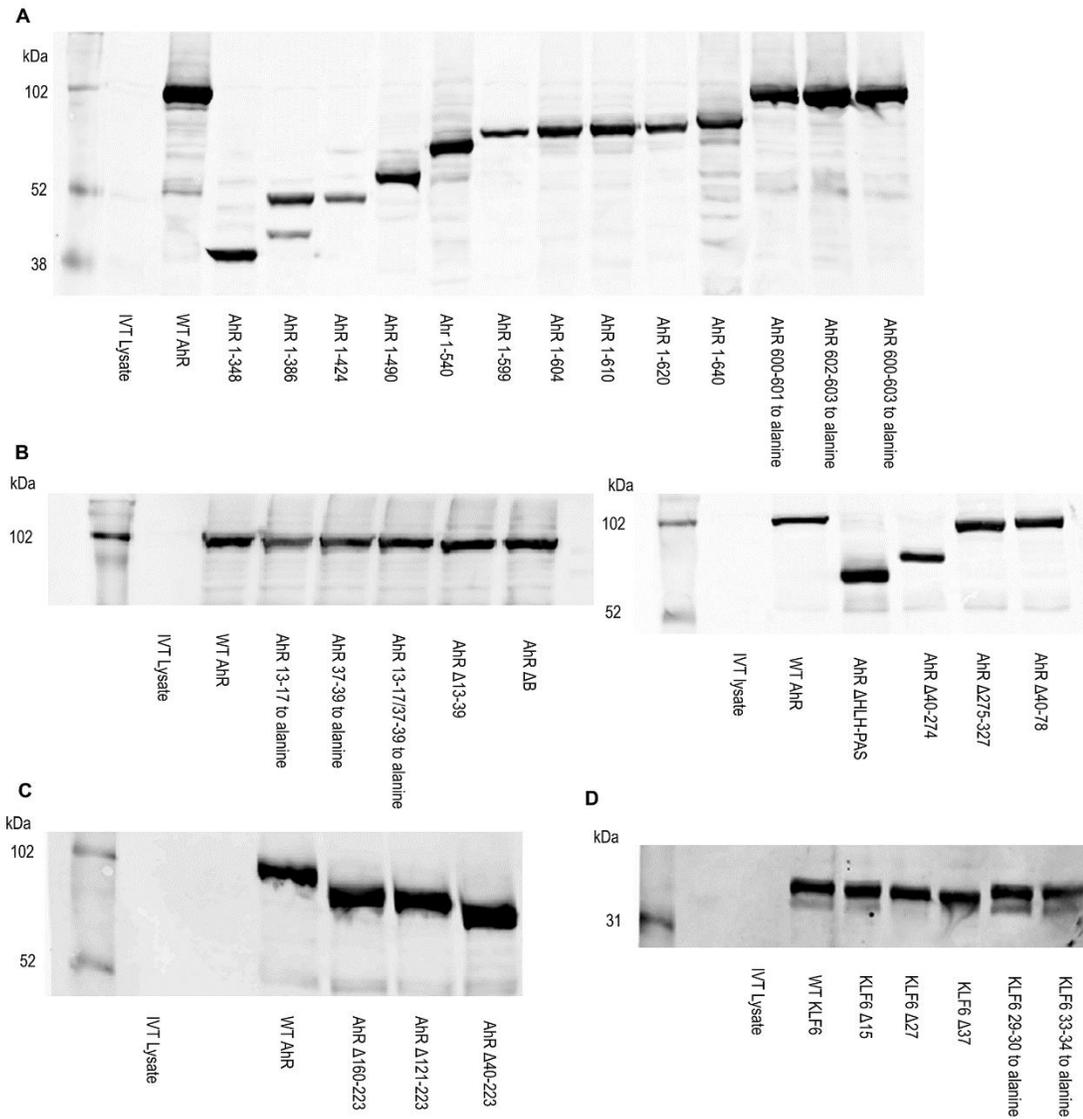


Figure 12: Immunoblots of AhR/KLF6 Protein expression. Constructs encoding full length proteins, deletions, truncations and alanine substitutions were used for protein expression with the transcription translation 1 Step Human Coupled IVT Kit. Western blots with AhR recombinant proteins were incubated with anti-HA tag as a primary antibody and probed with Cy5 fluorescence antibody. KLF6 western blots were incubated with anti-Flag tag as primary antibody and . (A) Expression of recombinant WT AhR, C-terminal deletions and Alanine substitution of glutamines present on the Q-rich region.(B,C) Alanine substitutions and deletions on the DNA binding region of AhR and internal deletions on the HLH-PAS region.(D) Expression of wild-type, deletion constructs and alanine substitutions of KLF6.

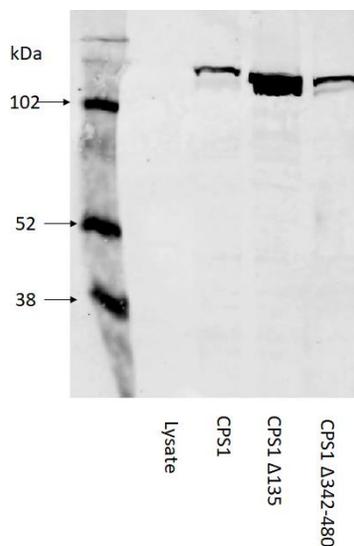


Figure 13: Immunoblot of CPS1 protein expression. Western blot using the construct encoding the full-length protein, N-terminal deletion and internal deletion.

Characterization of the AhR-KLF6 protein-protein interaction

Studies carried out by our group revealed the importance of C-terminal domain of AhR and N-terminal of KLF6 in the complex formation¹³⁰. To determine the exact region responsible for interaction on each protein using a quantitative approach, a Microplate Protein-Binding Assay was applied. In this assay, an antibody against one of the proteins of interest is non-covalently bound to a polystyrene microplate, and a mix of the binding proteins is incubated in a physiological solution. Binding is detected by using an antibody against the second protein, similarly to an enzyme-linked immunosorbent assay (ELISA) methodology^{160,162,163,165}.

To investigate the AhR-KLF6 protein-protein and protein-DNA interaction, AhR and KLF6 deletion and site-directed mutants were constructed. The *in vitro* transcription translation was achieved successfully as confirmed by Western Blot (Figure 13).

To examine the region of AhR responsible for KLF6 binding, 19 AhR mutants were used. The results demonstrated the mutant protein lacking amino acids 160-223 failed to form a complex with wild-type KLF6, indicating that this region is essential for protein-protein interaction (Figure 14). The absence of the basic helix-loop-helix, PAS A and PAS B domain also compromised complex formation, since the ligand-binding domain is absent (Figure 14).

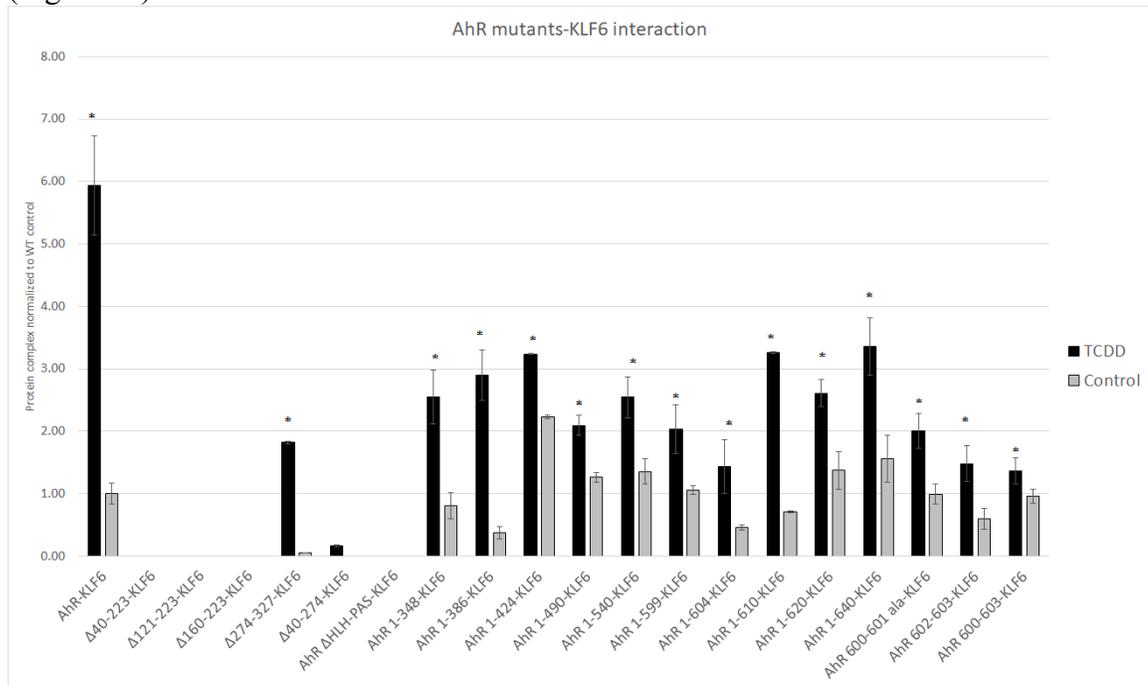


Figure 14: Detection of wild type AhR and AhR mutant proteins interaction with wild-type KLF6 by microplate binding assay. Error bars represent the standard error between the experiments (N=3). Residues between 160-223 of AhR are essential for interaction with KLF6.

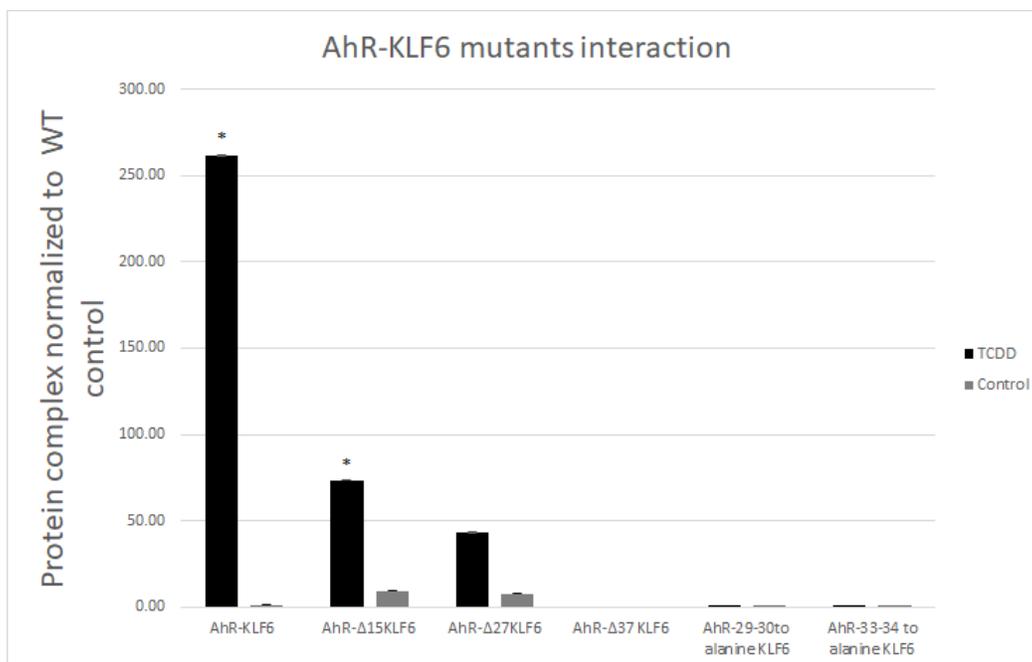


Figure 15: Detection of wild-type AhR interaction with wild type KLF6 and KLF6 mutants. Error bars represent the standard error between the experiments (N=3). Deletion of the first 27 amino acids of KLF6 compromise interaction, while deletion of the first 37 amino acids completely disrupt complex formation. Alanine mutagenesis demonstrated amino acids 29-30 and 33-34 are essential for binding.

The microplate protein-binding assay also demonstrated that the removal of the first 37 amino acids on the N-terminal region of KLF6 disrupts protein-protein interaction (Figure 15), according to the results observed on the microplate binding assay. To pinpoint the exact amino acids responsible for the complex formation, 2 mutants harboring alanine substitutions were designed: substitutions targeted the glutamic acids (29-30) and glutamines (33-34). The results demonstrated that the glutamine substitution reduced the protein interaction, while the absence of glutamic acids completely compromised the protein interaction.

AhR-KLF6 DNA interaction

AhR	+	+	+	+	-	-	+	+	-	-
KLF6	+	+	+	+	+	+	-	-	-	-
³² P-NC-XRE	+	+	+	+	+	+	+	+	+	+
Cold NC-XRE	+	+	-	-	-	-	-	-	-	-
TCDD	+	-	+	-	+	-	+	-	+	-

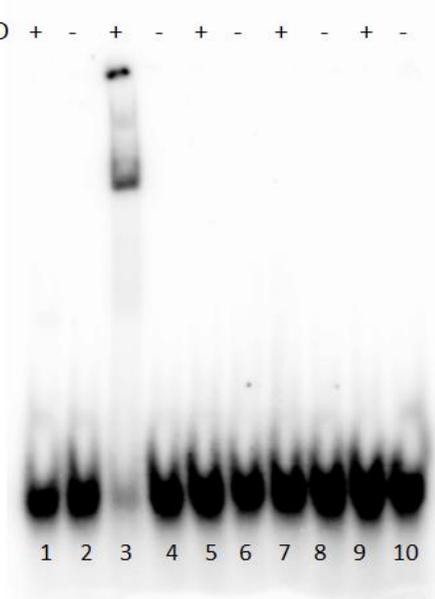


Figure 16: AhR-KLF6 DNA binding. EMSA performed with proteins produced *in vitro* and incubated with labeled (lanes 3-10) or unlabeled (lanes 1,2) NC-XRE. The presence of both proteins (lanes 3-4) and TCDD is required for complex formation).

Human AhR and KLF6 were expressed *in vitro* and the NC-XRE oligonucleotide sequence was radiolabeled with ³²P. To evaluate the viability of the proteins for EMSAs, equal volumes of wild-type proteins (0.5µl) were incubated with TCDD or vehicle and labeled and unlabeled NC-XRE. The data revealed that, as expected, the presence of both proteins and TCDD is necessary for the complex to bind to DNA (Figure 16). The lack of AhR or KLF6 prevented the complex formation, demonstrating the presence of both proteins is required.

Characterization of the NC-XRE

Previous studies developed in our laboratory demonstrated that the NC-XRE binding by the AhR complex occurs at the second tetranucleotide 5'-GGGA-3' repeat.

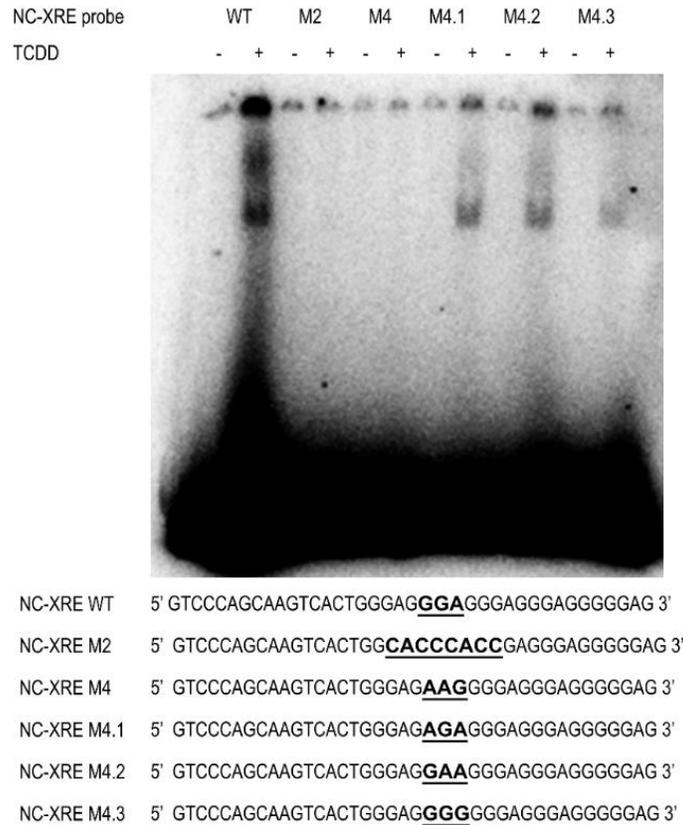


Figure 17: Characterization of the NC-XRE binding site. Wild-type AhR and KLF6 human proteins produced *in vitro* were treated with TCDD or DMSO (as control) and EMSA was performed with ³²P-labeled NC-XRE WT or mutated oligonucleotide probes (M2, M4, M4.1, M4.2 and M4.3)

When the second 5'-GGGA-3' repeat was mutated to a 5'-GAAG-3' motif (M4), proteins from nuclear extracts of animals treated with TCDD or vehicle failed to recognize the DNA probe³⁶. In order to investigate protein-DNA formation, we designed NC-XRE probes with mutations in the second 5'-GGGA-3' repeat (Figure 17). We also tested a mutated probe designed from Huang *et al* (2012)³⁶, M2 which mutates the first three 5'-GGGA-3' repeats. As expected according to what was observed with murine proteins, proteins failed to form

a complex with the M2 and M4 probes. Single point mutations in the NC-XRE oligonucleotide targeting the second 5'-GGGA-3' repeat did not prevent the complex formation, although the signals observed were weaker when compared to the wild-type probe. M4.1 and M4.2 exhibited binding decrease of approximately 59% while M4.3 was reduced 75 %, which suggests that a purine substitution may not completely disrupt protein-DNA complex formation.

Mathews & Lo (2012)¹³³ performed ChIP-seq in MCF-7 cells treated with TCDD,

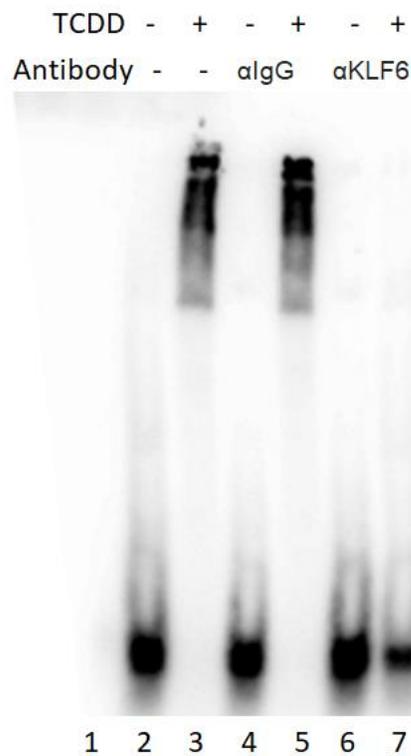


Figure 18 DNA binding analysis of MCF7 cells nuclear extracts to the NC-XRE. Nuclear extracts were prepared from vehicle- and TCDD-treated MCF7 cells, and were pre-incubated with either IgG (negative control) (Lanes 4-5) or αKLF6 antibodies (Lanes 6-7) prior to EMSA with an ³²P radiolabeled NC-XRE probe.

demonstrating that Kruppel Like Factors binding sites are highly enriched in AhR/ARNT

cobound regions ¹⁶⁶. Knockdown of AhR in this cell line abolished dioxin inducible repression of estrogen-dependent genes, while the same is not observed in cells lacking ARNT, suggesting that repression of estrogen repressor occurs independently of ARNT ¹⁶⁶. EMSA performed with nuclear extracts of TCDD treated MCF-7 cells demonstrated XRE binding¹⁶⁷. To investigate NC-XRE binding, MCF-7 cells were exposed to 6nM of TCDD for 1 hour and the nuclear fraction was extracted. NC-XRE binding was observed in treated cells, and addition of KLF6 antibody to the sample abolished complex formation (Figure 18), indicating existence of AhR non-canonical pathway in human cells.

The N-terminal region of KLF6 is necessary for DNA binding

Constructs encoding full length human KLF6 and mutants were used for *in vitro* protein expression, to evaluate protein-protein and protein –DNA interactions. Microplate Protein-Binding Assays demonstrated that the lack of the first 27 amino acids did not compromised the complex formation, in contrast to previous findings by Wilson and collaborators ¹³⁰. Quantitative assessment revealed a 23.63% reduced binding with the complex comprising the KLF6 Δ 27 mutant.

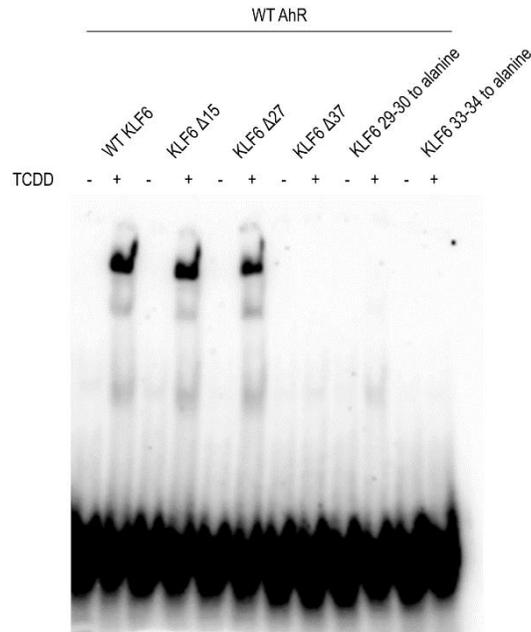


Figure 19: Amino acids 29-30 and 33-34 of KLF6 are critical for protein-DNA binding. The N terminal region of KLF6 is essential for DNA binding. The lack of the first 37 amino acids of KLF6 abolishes AhR-KLF6 DNA interaction.

Deletion of the first 37 amino acids fully compromised protein-DNA complex formation (Figure 19). These results complement the protein-protein interaction findings, where the substitution of amino acids 29-30 or 33-34 to alanine, disrupted NC-XRE binding.

The role of the AhR DNA-binding domain in the NC-XRE interaction

Previous research demonstrated the basic region of mAhR (amino acids 27-39) is essential for XRE binding. Although the lack of this region did not prevent dimerization with ARNT, the complex failed to bind the DNA. Additionally, the region located between amino acids 13-17 is also required for XRE binding. Deletion of the HLH region located between amino acids 40-79 reduced heterodimerization with ARNT and abolished DNA binding^{65,69,168,169}. More recent research determined the crystal structure of the AhR-

ARNT-XRE complex and revealed that the AhR interacts with the XRE motifs via Ser 36, His 39 and Arg 40^{70,71}.

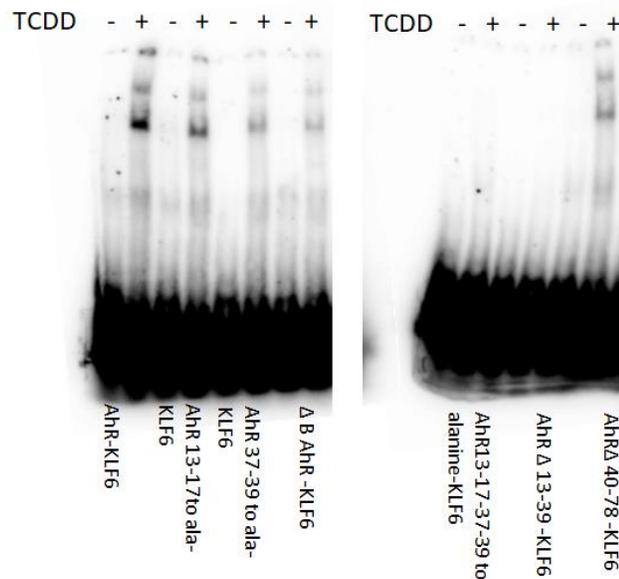


Figure 20: Characterization of AhR DNA binding domain. Alanine substitutions on each the AhR DNA binding region (aa 13-17/37-39), deletion of the basic domain (27-39) or HLH 40-79) did not prevent complex DNA bind. Alanine substitution on both DNA regions (13-17 and 37-39) or deletion of the region (13-39) abolished complex-DNA binding.

Aiming to investigate if the regions responsible for XRE binding also impact NC-XRE complex formation a series of hAhR mutants were generated: AhR mutants where amino acids 13-17 (RKRRK) and/or 37-39 (KRH) were substituted to alanines, a ΔB clone (amino acids 27-39 were deleted), and a clone where the HLH domain was deleted. The proteins generated were tested in EMSAs. Results showed that the alanine substitution of amino acids 13-17, 37-39, deletion of amino acids 27-39 (ΔB) or deletion of the HLH (Δ40-78), reduced but did not abolish complex-DNA binding. Complete lack of binding to NC-XRE was observed when both amino acids 13-17 and 37-39 were substituted to alanine or

when the region between 13-39 was deleted (Figure 20). These results demonstrate that these basic residues in the AhR interact with DNA in both XRE and NC-XRE

The Q-rich region of AhR is critical for NC-XRE

Our laboratory previously demonstrated that deletion of the C-terminal region of AhR disrupts protein-DNA binding to the NC-XRE³⁵. To determine the exact region responsible for complex-DNA binding, a series of AhR mutants with progressively larger C-terminal deletions were generated and tested in EMSAs. The results demonstrated that proteins encoding only the first 604 amino acids incubated with wild-type KLF6 fail to bind the NC-XRE (Figure 21). Protein DNA interactions are observed with AhR protein encoding the first 610 amino acids, which demonstrate the importance of a glutamine-rich amino acid region spanning residues 600-610 (⁶⁰⁰QQQSLALNS⁶¹⁰S) of the AhR. To

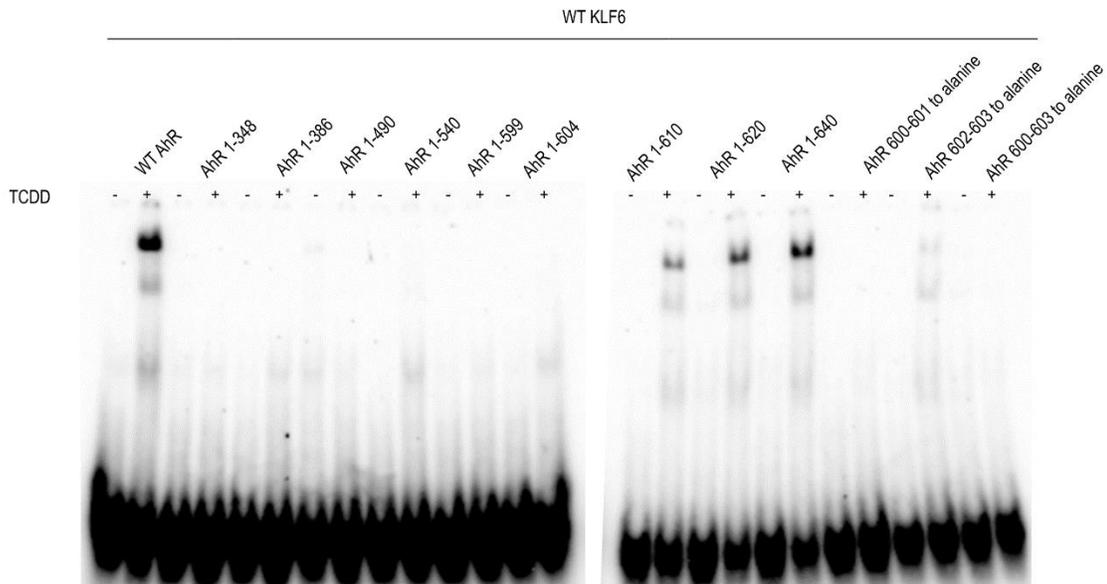


Figure 21: The Q-Rich region of AhR is essential for NC-XRE binding. EMSA using AhR C-terminal deletions, and alanine mutations on the Q-rich region. The Q-rich region 604-610 and neighboring amino acids are essential for NC-XRE binding.

pinpoint the key amino acids, we constructed three distinct mutants with alanine substitutions targeting the amino acids 600-603. Alanine substitutions of amino acids 600-

601 completely abolished AhR DNA binding while substitutions of residues 602-603 severely compromised DNA binding to the NC-XRE. Accordingly, substituting all four glutamines to alanines also completely abolished DNA binding.

AhR residues between 160 and 223 are essential for KLF6 interaction

To examine the role of AhR in NC-XRE binding we tested clones encoding AhR Δ 40-224, AhR Δ 40-274, AhR Δ 121-223, AhR Δ 160-223, AhR Δ 275-327, AhR Δ HLH-PAS, in the presence of wild-type KLF6 (Figure 22). These constructs were previously used in the Microplate Protein-Binding Assay.

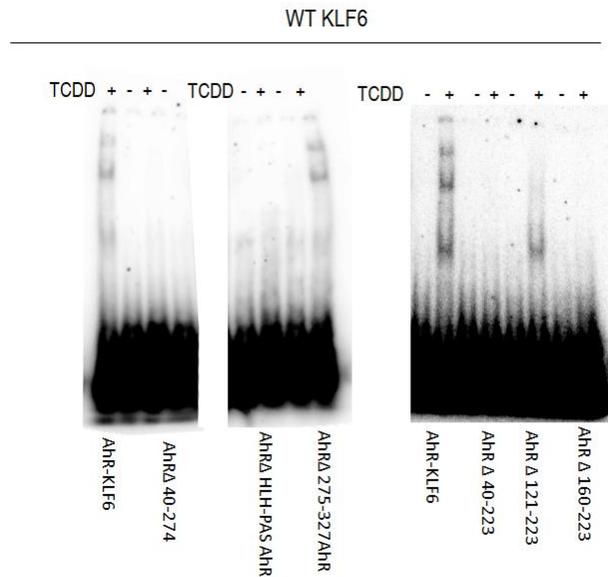


Figure 22: Residues 160-224 of AhR are essential for protein-protein interaction, consequently DNA bind.

Absence of the HLH and PAS region disrupts complex formation and consequently DNA binding. Deletion of this region also compromised the AhR interaction of Sp1⁸⁰, a member of the Sp/KLF family.

EMSA also confirmed that the absence of amino acids 160-223 disrupted DNA binding. Since the protein interaction was not observed in the Microplate Protein-Binding Assay (Figure 14), DNA binding accordingly also affected.

AhR-KLF6-CPS1 DNA interaction

CPS1 is a gene that codes a 1500 residues enzyme expressed in hepatocytes and enterocytes¹¹⁸. In an initial attempt to characterize the domains responsible for AhR NC-

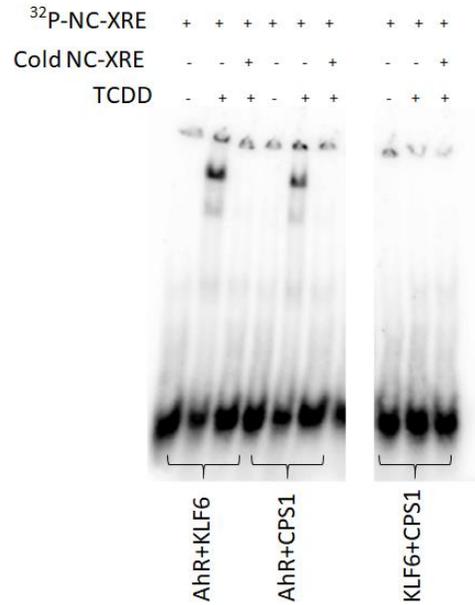


Figure 23: AhR-CPS1 binds to the NC-XRE. KLF6-CPS1 complex do not bind to the NC-XRE in the absence of AhR.

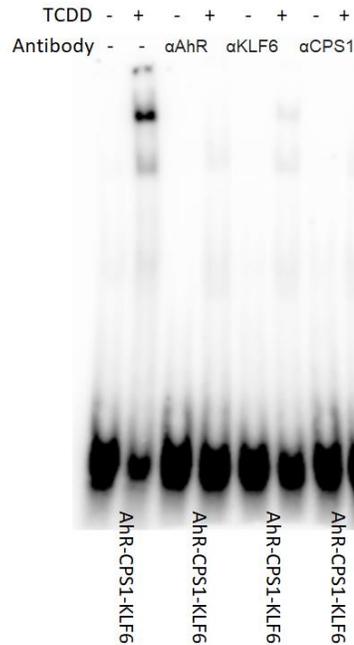


Figure 24: AhR-CPS1-KLF6 complex bind to the NC-XRE.

XRE interaction, a construct encoding the mature CPS1 protein without the signal peptide (first 38 amino acids) was built and successfully expressed *in vitro*.

Human AhR, KLF6, and CPS1 proteins were expressed and were incubated with TCDD and subjected to EMSA using the NC-XRE as a probe.

Complex formation is observed when AhR is incubated with either KLF6 or CPS1, but not when CPS1 is incubated with KLF6 in the absence of AhR (Figure 23). DNA binding is also observed if the three proteins are incubated together (Figure 24). The protein DNA complex is almost eliminated with the addition of antibodies as competitor against each of the proteins (Figure 25). These data imply that CPS1 interacts exclusively with the AhR and all three components comprise the protein-DNA complex.

Characterization of AhR- CPS1 NC-XRE interaction

To investigate the importance of the second GGGGA repeat in the NC-XRE motif for AhR-CPS1 DNA binding, wild-type AhR and CPS1 proteins were incubated with the mutated oligonucleotides M2, M4, M4.1, M4.2 and M4.3. As expected, and as previously observed with AhR-KLF6 complex, the AhR and CPS1 proteins form a complex with the

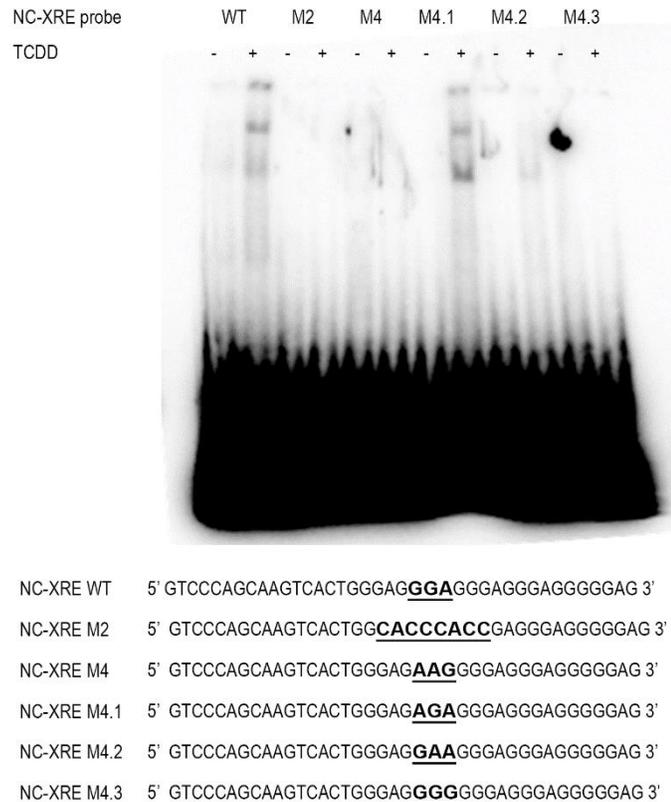


Figure 25: Characterization of the AhR-CPS1 interaction with NC-XRE. Wild-type AhR and CPS1 proteins expressed *in vitro* were treated with TCDD or DMSO (as control) and EMSA was performed with ³²P-labeled wild-type NC-XRE or mutant oligonucleotides.

wild-type oligonucleotide. The complex failed to bind the M2 and M4 mutant oligonucleotides, akin to the finding with the AhR-KLF6 complex. The heterodimer however, failed to interact to the M4.2 and M4.3 indicating that mutation of the second guanine nucleotide to an adenosine did not disrupt the interaction substantively, but

substitutions of the third and fourth nucleotides largely abolished complex formation (Figure 25).

AhR- CPS1 DNA interaction

To further investigate the role of CPS1 in the non-canonical pathway, a series of AhR deletion and site-directed mutant constructs were analyzed. Progressive deletions of the AhR C-terminus did not markedly affect DNA complex formation (Figure 26), in contrast to our findings with the AhR-KLF6 DNA interaction. There appeared to be a slight increase in protein-DNA binding (based on signal strength) with mutants encompassing amino acids 620-640, a region immediately C-terminal to the receptor's Q-rich region.

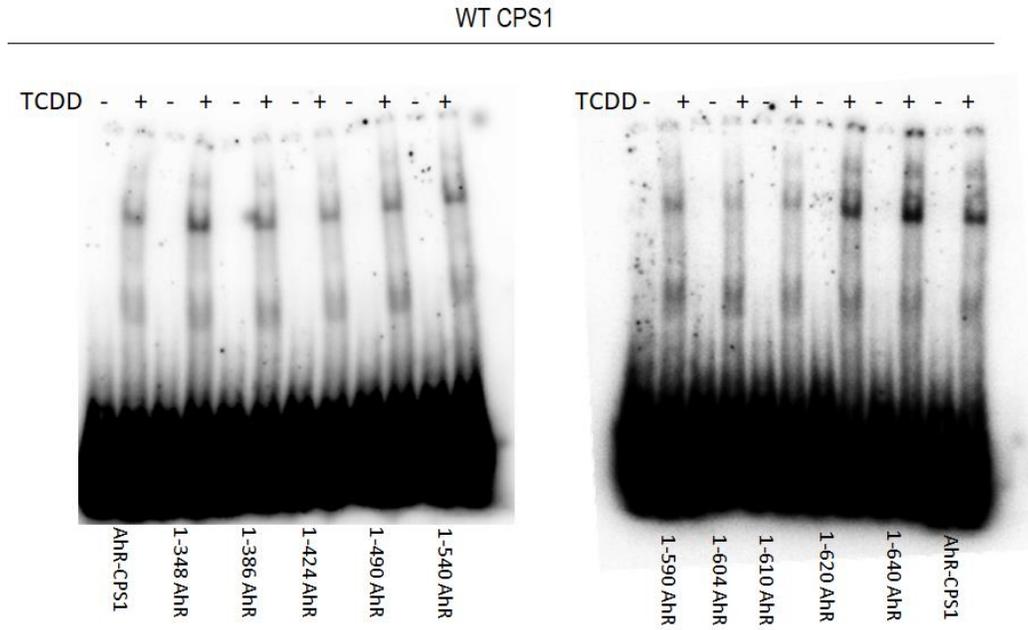


Figure 26: AhR-CPS1 DNA interaction. AhR mutants with C-terminal deletions were tested with wild-type CPS1 to evaluate NC-XRE binding.

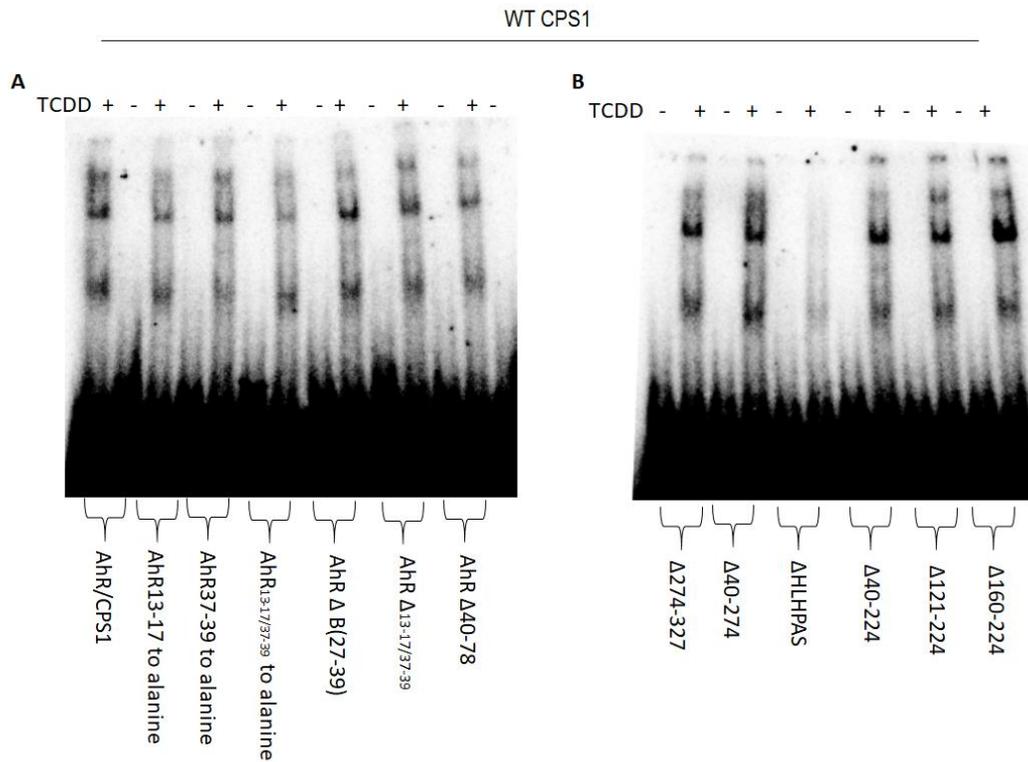


Figure 27: AhR-CPS1 DNA interaction. AhR mutants with alanine substitutions and deletions on the DNA binding region (A) and internal deletions were tested in EMSAs with wild-type CPS1

AhR constructs harboring alanine substitutions in the DNA binding domain, and internal deletions targeting the same region were also tested in EMSAs with CPS1 protein (Figure 27). Formation of protein-DNA complexes in the EMSA is observed despite the lack of a functional AhR DNA binding region in this series of AhR mutants. These data stand in contrast to what was observed with the AhR-KLF6 interaction. It is noteworthy that the AhR-CPS1 protein complex failed to bind to the NC-XRE when the AhR protein lacked the HLH-PAS region. Deletion of this region (Δ 40-78) also abolished the AhR-KLF6 protein-protein interaction, and may indicate the importance of this region in these interactions or simply reflect a deleterious effect on protein folding compromising function.

DISCUSSION

The main goal of this work was to characterize AhR complex formation at the NC-XRE. Previously, our laboratory demonstrated that responsiveness to TCDD could be independent of the canonical XRE. Huang and Elferink (2012)³⁶ discovered and characterized a Non-Consensus XRE (NC-XRE) in the PAI-1 promoter region. The core motif consisted of a 5'GGGA3' tetranucleotide repeat that allows DNA binding independent of ARNT. Wilson *et al*³⁵(2013) determined that KLF6 is an AhR binding partner to the NC-XRE. Joshi and collaborators studies showed that CPS1 is recruited as a cofactor that induces the homocitrullination of the lysine 34 of histone H1³⁴.

According to our findings, the presence of the three proteins results in DNA binding: addition of specific antibodies against each of the three components completely disrupts complex-DNA binding according to what is observed in EMSAs (Figure 24). Corroborating this finding, addition of KLF6 antibody in nuclear extracts from pretreated MCF-7 human cell line prior to an EMSA assay, prevented complex formation with the NC-XRE (Figure 18). Moreover, AhR is the central factor in that interaction, since

incubation of CPS1 and KLF6 in the absence of AhR did not induce complex formation (Figure 23). The interaction between AhR-KLF6 or AhR-CPS1 is observed in EMSAs and in the microplate binding assays occurred despite the absence of the respective other proteins. Since the proteins were generated *in vitro* using HeLa cell lysates, it is formally possible that the lysate contributed the missing protein.

Previous findings demonstrated that the second 5'-GGGA-3' repeat of the NC-XRE confers protein binding according to EMSAs performed with mouse liver extracts³⁶. In the present study, our results indicated that the second repeat is also essential for human AhR-KLF6 binding to the NC-XRE. Adenosine and guanine substitutions (5'-GGGA-3' to 5'-GAGA-3', 5'-GGAA-3' or 5'-GGGG-3') in the binding motif weakened protein-DNA interactions, especially with the 5'-GGGG-3' sequence. The AhR-CPS1 interaction with the NC-XRE was also investigated, and the results demonstrated that binding persisted only when the 5'-GGGA-3' was mutated to 5'-GAGA-3', but was fully disrupted if the motif was mutated to 5'-GGAA-3' or 5'-GGGG-3'.

Protein-nucleic acid interaction can lead to DNA distortion thus modulating the protein-DNA interface. Indeed the AhR-ARNT complex bound to the XRE bends the DNA at the site of interaction¹⁷⁰. Conversely, single alterations in the target DNA can alter the interaction dynamics. DNA-binding proteins can also bind DNA in different modes, with different motif preferences and with different specificity if a multi-protein complex is formed^{171,172}. Analysis of protein-DNA structures available in the Protein Data Bank (PDB) revealed the ability of adenine to form contacts with asparagines and glutamines, whereas guanines preferentially interact with arginines, supporting the idea of specificity in residue recognition by the target DNA. Analysis of DNA guanine+adenine strand combinations with circular dichroism demonstrated that the guanine content of repeats can change DNA conformation^{173,174}. Circular dichroism analysis revealed that (5'-GGGA-3')₈ repeats tend to form hairpins that are stabilized by the formation of G/A or A/A pairs¹⁷⁴. This finding suggests that mutations in the NC-XRE core motif sequence may disturb

protein-DNA interactions by radically altering the secondary structure (i.e., topology) of the DNA binding site ^{175,176}.

Although the absence of 3D models makes it difficult to fully understand the AhR's conformational flexibility during DNA binding, research on other members of the bHLH-PAS family contributes to our understanding of AhR interactions. The bHLH-PAS domains are highly conserved among species. The basic region is responsible for DNA interaction and the HLH domain, representing two amphipathic helices connected by a linker region is responsible for protein dimerization ¹⁷⁷⁻¹⁷⁹.

The PAS-A domain is essential for binding partner selection and specificity of target gene activation and the PAS-B domain is a signal sensor that regulates protein conformational changes ¹⁸⁰⁻¹⁸³.

The AhR nuclear localization signal (amino acids 13-39) overlaps with the DNA binding region. The basic amino acids and linker present in this region are essential for nuclear localization and DNA binding activity, and may be crucial to define an appropriate conformation for binding partners interaction and protein-DNA binding¹⁶⁹.

The AhR-KLF6 complex fails to bind DNA when both of the receptor's DNA binding regions 13-17 and 37-39 are mutated to alanine or deleted (Δ 13-39), or if the glutamines at position 600-603 are mutated to alanine. However, if either N-terminal sequence—amino acids 13-17 or 37-39—is left intact, NC-XRE binding is preserved. Likewise, deletion of the HLH domain (Δ 40-78), essential for AhR-ARNT complex formation at the XRE ⁶⁵, did not prevent the NC-XRE interaction.

Both amino acid regions 13-17 and 37-39 in the AhR are essential for XRE binding ^{69,184} and substitution of amino acids 37-39 generates an AhR-ARNT complex that is incapable of binding the XRE ⁶⁹. Hence, while the basic residues appear to function in AhR binding to both XRE and NC-XRE, distinct differences have been observed. Interestingly, in contrast to the AhR-ARNT complex, alanine substitutions of only one of the DNA binding regions (13-17 or 37-39) is insufficient to disrupt AhR-KLF6 DNA binding,

indicating that the two different protein complexes may have different nucleotide recognition sequences and rely on distinct protein-DNA interfaces^{185,186}. Nevertheless, these basic residues appear to function in AhR-DNA binding to both the XRE and NC-XRE.

Lack of AhR binding to the NC-XRE is also observed following the deletion of amino acids 600-610, encompassing a Q-Rich region in the AhR. Berg & Von Hippel (1989) proposed that certain proteins possess residues outside the DNA-binding domain that can facilitate DNA interactions^{185,187}. In yeast, genes regulated by transcription factors with glutamine-rich regions display higher levels of expression and mutational variability. The number of glutamines present in the transcription factor can alter its stability and interactions, which impact the expression of target genes¹⁸⁸. The Q-rich region is likely essential for stable AhR interactions with other proteins in the nucleus, although the precise molecular events remain unknown⁶⁷.

Studies with the bacterial σ^{54} transcriptional factor Q-Rich domain demonstrated that glutamine substitutions result in impaired transcription of target genes, since deletions or substitutions of glutamine cause a loss of ability to melt DNA¹⁸⁹. The glutamine-rich region Q2 of the cAMP Response Element-Binding Protein (CREB) stabilizes the protein interaction with DNA, which may indicate that interaction of transcription activators with Q-Rich transcription factors may favor DNA binding by nucleotide stabilization¹⁹⁰. Investigation of the glutamine-rich region of Sp1 suggested that this region is responsible for transcription activation, since a Sp1 mutant lacking a DNA-binding domain and conserved Q-Rich domain interacts with DNA-bound Sp1, superactivating transcription¹⁹¹. It is also important to mention that amino acids other than the glutamines present in a Q-Rich regions are important for functional specificity¹⁹². Collectively, these findings establish that the Q-Rich domain in the AhR is critical for AhR-KLF6-NC-XRE interaction *in vitro*. Our findings suggest that this region may stabilize the protein-protein interaction and thus facilitate protein-DNA binding, since microplate binding assays revealed a

diminished protein-protein interaction when the glutamines at position 600-603 were substituted to alanine. This results in a complex unable to stably bind DNA, as observed in EMSAs.

The AhR-CPS1-DNA interaction was also investigated. Interestingly there is variability between AhR-CPS1-NC-XRE binding and AhR-KLF6-NC-XRE binding. As previously stated, when mutated oligonucleotides were tested in EMSAs, binding is observed only with the M4.1 mutant oligonucleotide—where the second 5'-GGGA-3' repeat was mutated to 5'-GAGA-3'. NC-XRE binding by the AhR complex is still observed with the CPS1 Δ 135 or CPS1 Δ 341-480 deletion mutants when incubated with wild-type AhR. The binding of AhR mutants with wild-type CPS1 to the NC-XRE was also evaluated. Surprisingly protein-DNA binding was observed between all but the Δ HLLH-PAS (Δ 40-78) AhR mutant. Conceivably, the AhR-CPS1 interactions depend on residues in this deleted region, or this deletion may have disrupted normal protein folding rendering the receptor non-functional.

Deletion or substitution of AhR basic residues spanning the 13-17 or 37-39 sequences separately did not prevent DNA binding to the NC-XRE. Different groups studied the AhR DNA binding region responsible for XRE binding. Fukunaga & Hankinson (1996)⁶⁸ analyzed XRE binding using mouse AhR deletion mutants and ARNT in EMSAs, demonstrating that alanine substitution of tyrosine 9 or arginine 14 (positions 10 and 13 on human proteins, respectively) results in loss of XRE binding. Substitutions of amino acids 12-16 compromised or completely disrupted XRE binding by the AhR-ARNT heterodimer. Structural analysis revealed that AhR interacts with DNA via serine 36, histidine 39 and arginine 40, lysine 62 and lysine 65^{70,71}. Considering the sequence difference between the XRE core motif (5'-GCGTG-3') and the NC-XRE motif, it is possible that other residues present in the N-terminal region of AhR are responsible for recognizing NC-XRE nucleotides.

Knowledge gaps regarding the protein-protein interactions still remain. Although some proteins have distinct binding interfaces for each binding partner, others can bind different partners through the same or overlapping contact points, indicating that some binding sites are more promiscuous by interacting with multiple partners, albeit with different affinities^{193,194}. Proteins can bind to DNA by different modes with preferences and specificity dependent of the composition of multi-protein complexes¹⁷¹, including the binding of cofactors that alter DNA recognition properties¹⁷². If CPS1 acts as a non-DNA binding cofactor in the AhR's recruitment to the NC-XRE, it may do so as a cooperative cofactor, increasing specificity, affinity, and inducing maximal gene expression^{171,195}.

To investigate the regions responsible for AhR-KLF6 protein-protein interaction, deletion mutants and alanine substitutions were examined. Our findings demonstrate that amino acids 160-223 of the AhR and amino acids 27-37 in KLF6 are essential for complex formation. Although AhR deletion constructs expressing proteins from amino acids 348 through 604 fails to bind to DNA, protein-protein interactions were observed. The lack of the HLH and PAS domain compromised dimerization with KLF6. The essential region for AhR interaction with KLF6 is located in the PAS domain. This domain is subdivided into PAS-A and PAS-B highly conserved regions, connected by a poorly conserved linker. Previous studies determined that the PAS-A domain is crucial for ARNT dimerization and target gene activation^{32,183,196}. Our findings demonstrated that this region is also essential for heterodimerization with KLF6. Conceivably, this would predict that ARNT protein and KLF6 binding to the AhR might be mutually exclusive and could explain why the ARNT protein is not detected in the NC-XRE bound complex, and KLF6 is absent from the XRE bound complex.

Removal of the first 37 amino acids in KLF6 precludes an AhR protein interaction, and consequently DNA binding. The key amino acids of KLF6 responsible for protein-protein interaction are the glutamic acids (29-30) and glutamines (33-34). The N-terminal region of KLF6 is rich in acidic residues, consistent with the presence of a transcription

activation domain^{90,197}. Zhang *et al.* mapped KLF6 domains, indicating that the N-terminal region is required for interaction with p65, anchoring the protein to the promoter region of target genes¹⁹⁸. The acidic residues present in the N-terminal region of KLF1 may be responsible for protein-protein interactions¹⁹⁹. Point mutations in KLF4 glutamates present in the acidic domain resulted in lack of reporter expression²⁰⁰. Further investigation demonstrated that this region is responsible for intermolecular interactions that will contribute to transactivation. Two glutamines present in the N-terminal region of KLF6 also proved to be essential for protein interaction. Glutamine, a polar neutral amino acid, can be involved in hydrogen bond formation in a protein active site, and this interaction is important for structural stabilization^{201,202}. Glutamine dimers present in a protein domain can form backbone to side chain and backbone to backbone interactions with binding partners²⁰³.

Whilst we are cognizant of the fact that introducing deletions or mutation in the sequences of the studied proteins may alter protein structure, folding and physicochemical properties, the lack of computational tools that can infer the effects of these alterations,^{204,205} and the lack of crystal structures to verify the conformational integrity of the mutants, justifies the use of *in vitro* generated proteins is justified. Substitution and deletion mutagenesis are important strategies to study protein function^{206–208} and the use of these methods was essential to characterize this AhR non-canonical pathway. Taking into consideration that deletion mutagenesis can alter the structure of the target protein and that alanine substitutions do not induce steric or electrostatic effects on uncharged residues and are a powerful tool to examine the functions of charged residues^{209–211}, a future aim could encompass substitution analyses to pinpoint AhR residues responsible for KLF6 and CPS1 binding.

In an effort to expose potential physiologic relevance, a search of AhR and KLF6 mutations associated with cancer was performed using the online platform cBioPortal for cancer genomics²¹². Mutations in the regions associated with KLF6 interactions were

detected. AhR E169Q is linked to bladder urothelial carcinoma, R172C to rectal adenocarcinoma, W176L to cutaneous melanoma, P214L to bladder urothelial carcinoma, R223C is observed in uterine endometrioid carcinoma, cutaneous melanoma, rectal adenocarcinoma, and A606V was detected in uterine endometrioid carcinoma. Three KLF6 mutations are described between amino acids 27-37: W32 deletion was observed in lung squamous cell carcinoma, Q34H in cutaneous melanoma, and a splice mutation on amino acid 35 correlated with bladder urothelial carcinoma. Future studies could examine whether any or all of these mutations affect AhR-KLF6 interaction as the basis for these tumors.

Protein interactions regulate cellular, metabolic process and signaling pathways, playing essential roles in organism function and in the balance between health and disease. Protein-protein interactions studies are essential for the understanding of protein function and how mutations could impact the binding, biological functions, and gene regulation^{213,214}. An evaluation of these interactions will provide valuable insights into AhR biology and modulation. Figure 28 summarizes the major finds regarding this chapter.

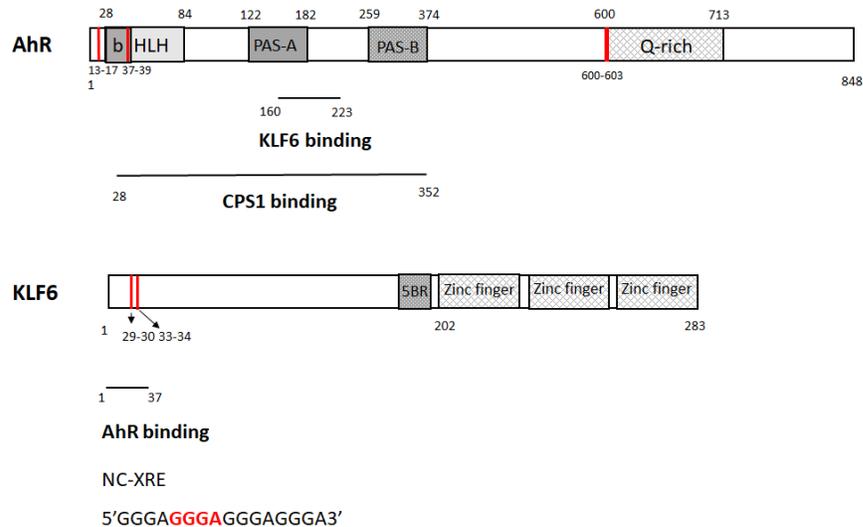


Figure 28: AhR-KLF6 and CPS1 interaction summary. Our results demonstrated that the absence of part of the PAS A domain and linker region (amino acids 160-223) compromised AhR-KLF6 interaction. AhR interaction with CPS1 occur within the bHLH and PAS domains. Amino acids 13-17, 37-39 and the glutamines 600-603 are essential for DNA binding. KLF6 N terminal region interacts with AhR by EE 29-30 and QQ 33-34. The second GGGA repeat of the NC-XRE is essential for protein-DNA interaction; however, purines substitution did not prevent complex formation.

Chapter 3: A genome-wide functional assessment of AhR signaling

INTRODUCTION

According to the canonical pathway, upon ligand activation, the AhR translocates to the nucleus where it binds with ARNT, and the complex binds to the XREs, with the core sequence 5'GCGTG3'. However, several groups observed in their studies that the promoter region and the transcription start site of certain target genes do not contain a classical XRE ¹⁵³ indicating that AhR may interact with other recognition sequences and act through distinct pathways ^{5,14}.

Using EMSAs and ChIP, Voguel *et al* (2007) demonstrated that RelB, a subunit of nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), physically interacts with AhR upon TCDD or forskolin activation, and the complex binds to a RelB/AhR responsive element, in the promoter region of interleukin 8 (IL-8) gene, activating gene transcription ¹⁶.

Kimura and collaborators (2009) ²¹⁵ investigated the involvement of AhR in the negative regulation of the lipopolysaccharide (LPS) response pathway. According to immunoprecipitation studies performed with macrophages stimulated with LPS, AhR forms a complex with Signal transducer and activator of transcription 1 (Stat1) and NF-κB. The complex formation results in the inhibition of the promoter activity of IL-6. Their findings elucidated the crucial role of AhR/Stat1 complex in the downregulation of LPS signaling pathway.

ChIP-chip and gene expression analysis of microarrays performed on liver tissue obtained from mice treated with TCDD for 2 or 24 hours revealed that around 50% of AhR-enriched regions do not contain a XRE core sequence. Around 32% of TCDD-activated genes identified by microarray studies contain a XRE binding site. *De novo* motif analysis identified certain sequences that are known binding elements for to other transcription

factors (TF) binding motifs. The existence of binding motifs different from the core motif 5'-GCGTG-3' suggested that AhR could interact with different binding partners, acting through different DNA regions.¹³²

Lo & Matthews (2012)¹³³ mapped AhR-ARNT binding sites by using ChIP-seq in MCF-7 cells treated with TCDD. Although they detected overlap between AhR and ARNT binding regions, only 60% of the identified sequences corresponded to a XRE, indicating that this heterodimer might interact with different DNA motifs or each protein is capable of acting independently at non XRE sites.

Analyses of published gene expression studies and ChIP-on-chip data of female mice gavaged with TCDD revealed that differentially expressed genes are also regulated by other transcription factors, including KLF4²¹⁶. Motif analysis of promoter regions associated with AhR regulated genes in human cells revealed GC-rich motifs recognized as binding sites of SP1 and KLF12²¹⁷.

Collectively these studies demonstrated that the AhR may interact with different partners and bind to motifs other than XREs to control gene expression. Studying alternative AhR interactions can help to fill some gaps in the knowledge about the roles in AhR in dioxin toxicity.

ARNT inducible knockout mouse model

To investigate the role of a gene in a specific cell or organ, genetic strategies can be developed to introduce known mutations in the mouse genome. The Cre recombinase is a widely used system, based on the manipulation of a 34 base-pair (bp) DNA site named loxP by the use of a recombinase Cre (cyclization recombination). To achieve a gene knockout, the Cre/lox is directed to the gene of interest, flanking an essential exon with loxP sites. The Cre is then “delivered”, excising the target DNA contained between by loxP sites, generating a null allele²¹⁸. Transgenic animals that have Cre driven by the serum

albumin (*Alb*) gene promoter are a useful model to study gene excision specific to hepatocytes, since the *alb* gene is expressed only in these cells of the liver.²¹⁹ .

To better understand how ARNT is involved in AhR biology, Bradfield *et al.* (2010)²²⁰ developed a hepatocyte-specific ARNT null allele, to overcome the embryonic lethality of global ARNT full knockout observed in other studies.

To generate a conditional knockout mouse model, ARNT^{fx/fx} animals (mice with loxP sites flanking exon 6 of ARNT) were crossed with Cre^{alb} animals expressing Cre recombinase in hepatocytes. The offspring animals (Arnt^{fx/+} Cre^{alb}) were crossed to ARNT^{fx/fx} to generate hepatocyte-specific conditional ARNT-KO animals. PCR was used to evaluate if the animals resulting for this crossing lost exon 6 in the hepatocytes.

Analysis of the organs of conditioned knockout animals demonstrated that the livers have normal morphology and normal ductus venosus closure during development. The animals were also resistant to acute exposure to TCDD, not presenting hepatic injury, inflammation, degeneration or notably hepatomegaly. Certain AhR classical target genes such as Cyp1a1, Cyp1a2, Cyp1b1 and AhRR were not upregulated, indicating the essential role of ARNT in AhR mediated xenobiotic metabolism.

In order to determine the roles of AhR in the non-canonical pathway we decided to use the same floxed ARNT mouse model crossed with a tamoxifen inducible Alb-Cre (CreERT²). This system contains the human estrogen receptor ligand-binding domain fused with the Cre recombinase, expressed under the control of the albumin promoter. The resulting protein stays inactive in the cytosol, turning active after injection of the synthetic estrogen receptor ligand 4-hydroxytamoxifen (OHT)²¹⁸

The timed knockout of ARNT is achieved by administration by administration of tamoxifen as demonstrated in figure 29. The RNA-seq and RT-PCR (figure 30) demonstrated that tamoxifen treatment resulted in ARNT deletion.

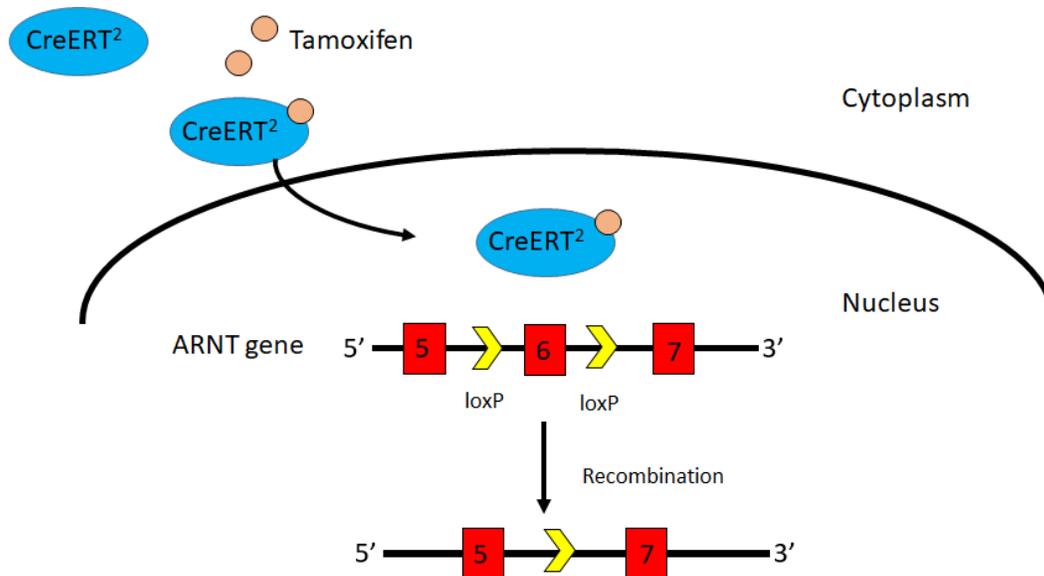


Figure 29: Schematic of inducible ARNT Knockout mice model. CreERT2 remains inactive on the cytosol of hepatocytes due to the albumin expression. After tamoxifen treatment, the chimera translocates to the nucleus, where the Cre recombinase cleaves the loxP sites flanking ARNT exon 6, which will result in the knockout of ARNT in the liver parenchyma. The timed knockout of ARNT is achieved by administration of tamoxifen as demonstrated. The treatment resulted in a truncated ARNT product, representing the exon 6 deletion.

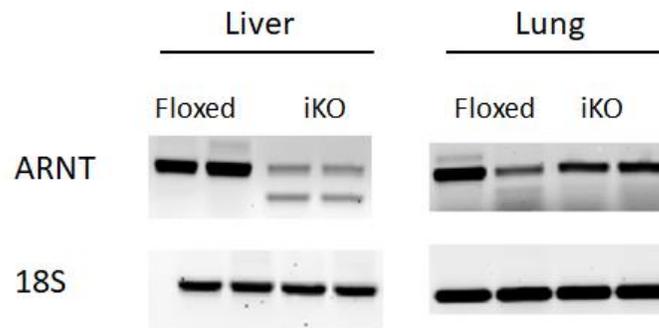


Figure 30: Semi-quantitative RT-PCR analysis of ARNT expression in Floxed animals and in induced knockout (iKO) animals. ARNT expression in mRNA isolated from livers and lungs of ARNT floxed or ARNT iKO mice treated with tamoxifen (via oral gavage every 24 hours 3 days). Note that the excision of loxP sites on the liver of iKO animals generated a truncated ARNT product, while the lung of these animals generated a full length product.

MATERIAL AND METHODS

Animals

All procedures used in the experiments were approved in accordance with the guidelines set by the Institutional Animal Care and Use Committee at the University of Texas Medical Branch at Galveston.

The experiments were performed with 8-10 weeks old ARNT^{fx/fx} and ARNT inducible knockout (ARNT^{IKO}) animals. All animals were treated by oral gavage with 80 mg/Kg tamoxifen for three consecutive days and the experiments were performed a week after the first tamoxifen treatment. Liver specific knockout of ARNT was confirmed by RT-PCR (Figure 31).

For ChIP, ChIP-seq and RNA-seq, 20µg/kg of TCDD in peanut oil or with a corresponding volume of peanut oil (Vehicle) were administered via oral gavage 2 hours prior to sacrifice of the mice by isoflurane anesthesia followed by cervical dislocation.

Chromatin immunoprecipitation

ChIP was used to validate the antibodies used for ChIP-sequencing. The assays were performed using whole liver from the animals as described previously^{34,36,130,221}.

After TCDD or vehicle 2-hour treatment, ARNT floxed and knockout animals were sacrificed, their livers quickly removed and placed on phosphate buffered saline (PBS). The gallbladder was removed and discarded, and the liver was minced with a razor blade. For chromatin crosslinking, the samples were incubated for 10 minutes in 2.5% formaldehyde 37% at room temperature in rotation. Crosslinking was stopped by adding 200 nM glycine solution. The samples were centrifuged at 3200xG, 4°C for 5 minutes. The supernatant was removed, and samples were resuspended in 5mL of ice-cold PBS and transferred to a dounce homogenizer, and tissue homogenized with 7 strokes with a tight-fitting pestle. The homogenate was centrifuged at 3200xG at 4°C. After discarding the supernatant, samples were resuspended in 2mL of cell lysis buffer (NaCl 150nM, 25 nM Tris, 5mM EDTA, 1% TritonX100, 0.1% Sodium dodecyl sulfate, 0.5% deoxycholate)

with 1:100 protease inhibitor cocktail. Samples were placed on ice for 15 minutes, following a 5-minute centrifugation at 3000xG at 4°C. The supernatant was discarded and the samples were processed using the ChIP-IT Express Enzymatic Kit (Active Motif) following the manufacturer protocol with modifications. Briefly, the pellet was resuspended in ChIP lysis buffer with protease inhibitor cocktail, phenylmethylsulfonyl fluoride (PMSF) and placed in ice for 30 minutes. After incubation, samples were homogenized with 50 strokes using a tight fit pestle for nuclei release. The nuclei were centrifuged for 10 minutes at 5000xG at 4°C. The pellet was resuspended in digestion buffer (provided by the kit) with protease inhibitor cocktail and incubated for 5 minutes at 37°C. Enzymatic shearing cocktail (1:100 in 50% glycerol solution) was added on the pre-warmed samples, that were then incubated for 15 minutes at 37°C. After vortexing the samples, the reaction was stopped with ethylenediaminetetraacetic acid (EDTA) and the samples were incubated on ice for 10 minutes. The sheared samples were centrifuged at 18000xg and the sheared chromatin was obtained in the supernatant. An aliquot was obtained to be used as input DNA. Samples were incubated overnight at 4°C with rotation with Protein G magnetic beads, buffers from the manufacturer kit and antibodies. For immunoprecipitation, the following antibodies were used: AHR Monoclonal Antibody (RPT9) (Thermo Fisher Scientific, Waltham, MA), KLF6 (pAb) (Active Motif), H3 (positive control) and IgG (negative control). The immunoprecipitated samples were collected following the instructions provided by the manufacturer protocol, washed, eluted and reversed crosslinked with the solutions provided by the kit, and incubated at 95°C for 15 minutes, treated with proteinase K, and incubated for 1 hour at 37°C. After incubation, proteinase K stop solution was added and the DNA was precipitated with phenol/chloroform extraction.

The samples were PCR amplified with primers specific for the XRE present in the CYP1A1 promoter (Forward 5'-CGGGACATCACAGACAGCCTCATT-3'/ reverse 5'-CCTGCCACTGGTTCACAAAGACAC-3') and the NC-XRE on the PAI-1 (Forward 5'-

ACGTTGTGGAAGTCCCTAC-3'/ Reverse (5'-GCCAGGGTTGCACTAAACAT-3') and visualized after separation on 5% polyacrylamide gel stained with SYBR green (Thermo Fisher Scientific, Waltham, MA) for 30 minutes and imaged with Typhoon Trio.

Next generation sequencing

Illumina next generation technology generate clusters by the immobilization of the templates in a flow cell followed by a solid phase amplification, creating copies of each template in close proximity. The next step is sequencing by synthesis, where four fluorescent label nucleotides are used to sequence the clusters in the flow cell. One single label deoxynucleotide triphosphate is added on each sequencing cycle, where it serves as a polymerization terminator, so after each incorporation the fluorescent dye is imaged and the base is identified, subsequently being cleaved to allow the incorporation of the following nucleotide²²².

The NextSeq500 employs a novel technology that requires only two channels. In this system, thymine is labeled green, cytosine is labeled red, guanine is unlabeled and adenine is a mixture of the red and green labels. Contrary to the four channel strategy employed by HiSeq and MiSeq, the NextSeq platform needs only two images for detection, which reduces time and increases the reading capacity, while error rates are similar to other platforms²²³.

Chromatin immunoprecipitation sequencing

ChIP-seq protocol was similar as described for ChIP. After the liver samples were processed and lysed as previously described, the generated pellet was resuspended in 1mL of lysis buffer supplemented with protease inhibitor cocktails and PMSF. The solution was dounce homogenized with 10 strokes for nuclei release. Samples were centrifuged for 10 min at 1500xg at 4°C. The supernatant was discarded, and the pellet was resuspended in shearing buffer supplemented with protease inhibitor cocktail and PMSF). DNA was

sheared with Covaris ultrasonicator. Sheared chromatin was centrifuged for 10 minutes at 15000xg at 4°C and the supernatant was used for immunoprecipitation with AHR Monoclonal Antibody (RPT9) (Thermo Fisher Scientific, Waltham, MA), KLF6 (pAb) (Active Motif), following the same steps described on the previous section. Immunoprecipitated DNA samples were submitted to the UTMB Next Generation Sequencing Core Facility (Galveston, Texas) for library preparation and sequencing.

Liver samples of 6 wild-type female C57BL/6J mice treated with TCDD or vehicle (3 in each group) were submitted to Active Motif (Carlsbad, CA) for ChIP-seq. Samples were precipitated with AhR Enzo BML-SA210 (Farmingdale, NY antibody) and the 75 single end reads generated by the Illumina NextSeq 500 were mapped to the mouse (mm10) genome. MACS 2.1.0 was used for peak calling, with a cutoff p-value of 1e-7 for narrow peaks and 1e-1 for broad peaks. To compare the metrics between the three samples of each group, the overlap intervals were grouped in merged regions.

Transcription factor identification

To identify transcription factors with AhR binding regions²²⁴ TFcheckpoint was used. A list of genes with a Log2Ratio TCDD/Vehicle ≥ 4.5 identified by ChIP-seq was uploaded into the search database.

Motif Mapping

Motif mapping was performed with Meme suit FIMO tool²²⁵. DNA sequences were obtained by Cisgenome²²⁶ utilizing the data obtained by AhR ChIP-seq with merged regions of peaks present in triplicate and loaded on Meme suit server. Motif mapping was performed using the XRE sequence 5'GCGTG3' (P-value ≤ 0.000709).

Visualization of ChIP-seq data using Heatmaps

For visualization of binding signals between TCDD treated and control samples close to promoter regions, Easeq²²⁷ was used. Binding distribution +/- 5kb from the TSS were clustered using k-means algorithm (with 10 clusters as default). The method read densities of each sample around the TSS of the reference coordinate obtained in the ChIP-seq analysis. The Y-axis represents individual positions while the X-axis corresponds to the metrics of the overlapping intervals of the 3 samples of each group (merged regions).

RNA extraction and RNA-seq

Twelve samples were submitted for next generation sequencing. 6 ARNT^{fx/fx} (ARNT floxed) female animals and 6 ARNT^{iKO} (ARNT iKO) female animals were treated with (N=3 for each group) TCDD or vehicle two hours prior liver extraction.

The livers of the treated mice were placed in Trizol (Thermo Fisher, catalog number 15596026) and homogenized using a polytron homogenizer (Kinematica AG, Lucerne, Switzerland). The total RNA was extracted as described by Rio *et al.*²²⁸ Samples were submitted to the UTMB genomics core facility (University of Texas, Medical Branch, Galveston, TX), where they performed quality control using Qubit fluorometer. The RNA was reversed transcribed into cDNA, fragmented and sequenced using NextSeq550, with paired end reads of 75 nucleotides, and a total of 400 million reads for all twelve samples.

Samples were processed by the UTMB genomics core facility. Quality control of the data were obtained from the Sequencer output reports and processed by FastQC²²⁹. The data were aligned to the mouse reference genome with STAR²³⁰ software. The program provided raw read counts of each transcript. Differential expression analysis to determine fold change in expression was performed with DESeq2²³¹.

Semi-quantitative PCR

For semi-quantitative PCR, RNA was extracted as described in the previous section. qRT-PCR iScriptTM gDNA Clear cDNA Synthesis Kit (Biorad, Hercules, CA)

was used according to the manufacturer instructions for cDNA synthesis using 1 µg RNA. 2 µl of cDNA reaction was used as template for the PCR reactions. 10x PCR buffer (Millipore-Sigma, Burlington, MA) was diluted in nuclease-free water (Thermo Fisher Scientific, Waltham, MA), forward and reverse primers were added at a 0.2 µM final concentration, and 10 U/ per reaction of Taq polymerase (Millipore-Sigma, Burlington, MA) was added on each reaction. Cycling parameters included a hot start at 94°C for 3 minutes, primer annealing temperature was set according to the estimates obtained from Neb Tm Calculator ²³² and set for 30 seconds, then 72°C for 90 seconds . Target genes were amplified for 30 cycles. Final extension was set at 72°C for 10 minutes. Reaction products were loaded in 0.8% agarose ethidium bromide gels, and then visualized on Typhoon Trio (GE Healthcare, Chicago, IL).

Quantitative PCR

Quantitative real time PCR was performed using The CFX96 Touch System (Biorad, Hercules, CA) and cDNA was produced as described in the previous section. SsoAdvanced™ Universal SYBR® Green Supermix (Biorad, Hercules, CA) was diluted in nuclease-free water (Thermo Fisher Scientific, Waltham, MA), forward and reverse primers were added at a final dilution of 0.45 µM each. Cycling parameters were initial denaturation at 98°C for 30 seconds, denaturation 95°C for 10 seconds and annealing, extension, and fluorescence read at 60°C for 30 seconds. Target genes were amplified for 39 cycles. PCR products were subjected to melt curve analyses. Gene expression was determined by analyzing data using ΔC_t method to adjust for expression of 18s and $2^{-\Delta\Delta C_t}$ to adjust the expression of treated versus control samples.

Visualization of RNA-seq results using heat maps

To visualize RNA-seq results expression.-based heat maps with comparisons between upregulated and downregulated genes in floxed treated animals, upregulated and

downregulated genes in iKO treated animals and upregulated and downregulated genes in Floxed treated animals compared to iKO treated animals, heatmapper was used ²³³.

RESULTS

RNA-sequencing for identification of AhR Target Genes

As discussed previously, the AhR regulates transcription of several genes. This process involves at least two different protein partners (ARNT and KLF6) and the formation of distinct AhR-ARNT-XRE or AhR-KLF6-NC-XRE complexes. This study employed NGS technologies to identify novel AhR target genes regulated by the non-canonical pathway. To focus the interest on the NC-XRE pathway we used an ARNT floxed mouse that expresses normal ARNT levels in the liver and a liver-specific inducible ARNT conditional knockout model (ARNT iKO), which contain LoxP sites flanking

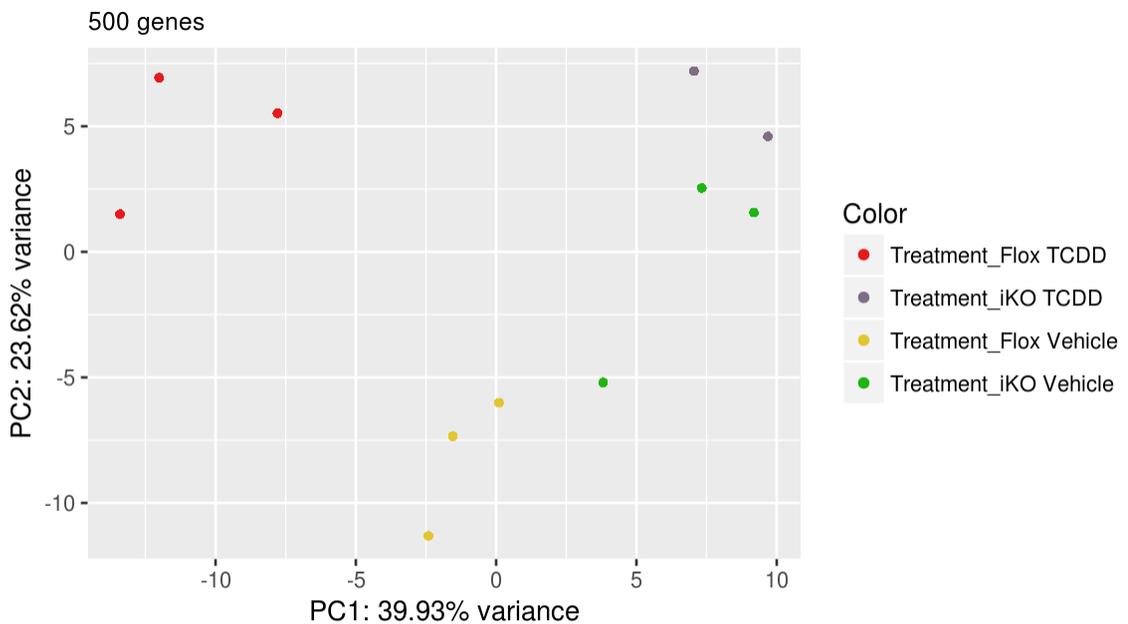


Figure 31: Principal component analysis of samples used for RNA-sequencing analysis. The floxed TCDD treated samples cluster together. Two of the iKO vehicle treated samples cluster with the iKO TCDD treated samples.

ARNT exon 6 and expresses Cre recombinase under the control of the albumin promoter that is activated by tamoxifen treatment.

We treated the mice (N=3 per genotype and per treatment) with vehicle or TCDD for 2 hours via oral gavage. The livers were extracted, the RNA was purified and submitted for RNA sequencing. For sample analysis, a Principal Scale Plot was generated²³⁴ by PCAGO (Figure 31). One of the iKO TCDD treated samples was an outlier, therefore, removed from the analysis.

The RNA sequencing analyses of differential expression between floxed animals treated with TCDD or vehicle identified 199 upregulated genes ($\log_2\text{foldchange} \geq 1$, $\text{FDR} \leq 0.1$, $p \text{ value} < 0.05$) (Appendices III) while the same analysis of iKO animal data revealed 34 genes (Appendices IV).

Classical AhR target genes Cyp1a2, Cyp1b1 and UDP glucuronosyltransferase 1-2 (Ugt1a2) upregulation in Floxed animals was confirmed by qRT-PCR (Figure 32). Cyp1a2 was also upregulated with lower $2^{(-\Delta\Delta Ct)}$ in iKO animals, while levels of Cyp1b1 were similar to floxed animals. Cyp1b1 is not expressed in hepatocytes²³⁵, the cell target for deletion of ARNT. KLF6 and Nuclear Protein 1, Transcriptional Regulator (Nupr1) are

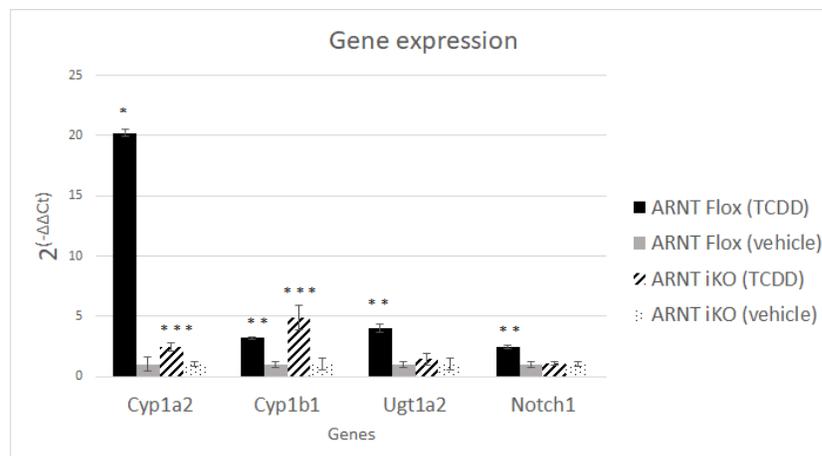


Figure 32: Effect of TCDD exposure on Cyp1a2, Cyp1b1, Ugt1a2 and Notch1 expression. qRT-PCR quantification in mRNA isolated from floxed or iKO mice treated with TCDD or vehicle for 2 h, n=3. * $p < 0.0001$, ** $p < 0.01$, *** $p < 0.05$. The values are presented as mean \pm SEM.

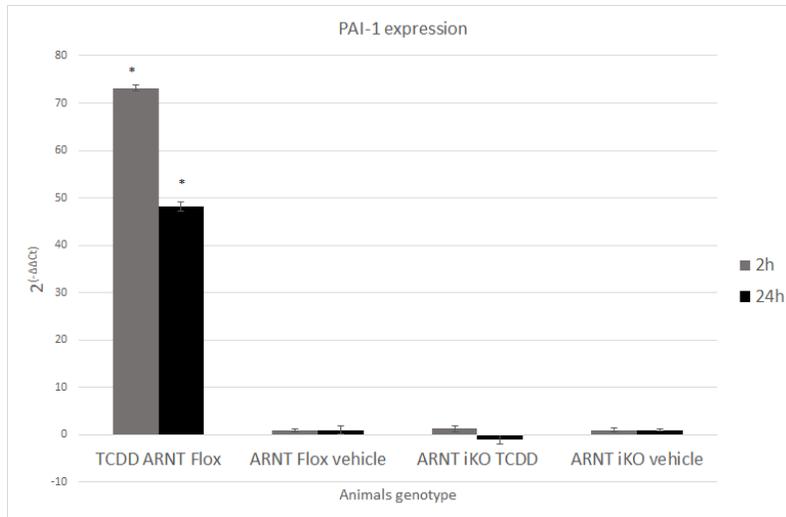


Figure 33: Effect of TCDD exposure on PAI-1. qRT-PCR quantification of mRNA isolated from flox or iKO mice treated with TCDD or vehicle for 2 and 24 hours, n=3. *p<0.0001. The values are presented as mean ± SD.

downregulated in Cyp1b1 knockout animals feed with a low fat diet, which can lead to a decrease in stress and inflammation signaling²³⁵.

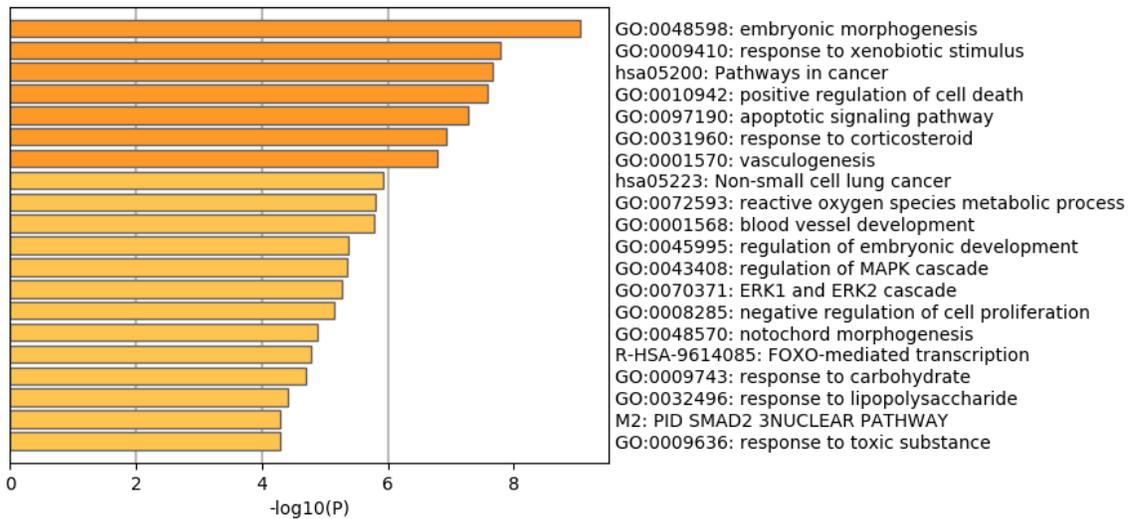


Figure 34:Gene List pathway analysis by Metascape. Bar graph of enriched genes upregulated in ARNT floxed animals treated with TCDD for 2 hours.

KEGG 2019 mouse) include metabolism of xenobiotics by cytochrome P450, breast cancer, steroid hormone biosynthesis, basal cell carcinoma, chemical carcinogenesis, pathways in cancer and hepatocellular carcinoma . Metascape²³⁷ was utilized to convert the list of genes into human species. The findings are summarized in Figure 34. The diseases²³⁸ the genes may be linked include hepatocarcinogenesis, anaplasia, fatty liver and steatohepatitis.

According to the analysis performed by GSEA (Figure 35), the biological process that correlate with the dataset are: positive regulation of cell death, apoptotic process, , embryonic development and morphogenesis, epithelium development, regulation of intracellular transduction and response to lipids. These results are consistent with previously described AhR mechanisms of action^{7,9,30,239-242}.

The top signaling networks associated with ARNT iKO animals (By KEGG 2019 mouse) are metabolism of xenobiotics by cytochrome P450, ECM-receptor interaction, steroid hormone biosynthesis, retinol metabolism, and chemical carcinogenesis. Metascape pathway analysis was utilized to convert the gene list into human orthologs (Figure 36). Diseases associated with these genes include systemic scleroderma, fibrosis, enteritis, T

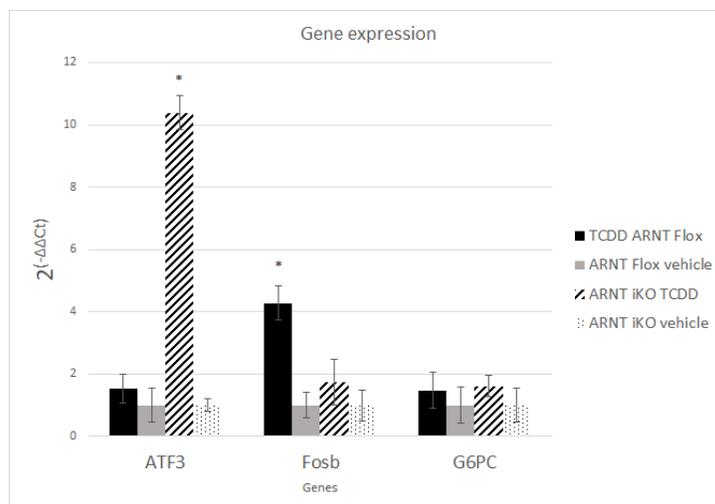


Figure 38: Effect of TCDD exposure on ATF3, FosB and G6PC expression. qRT-PCR quantification in mRNA isolated from floxed or iKO mice treated with TCDD or vehicle for 2, n=3. *p<0.0001; **p<0.01. The values are presented as mean ± SEM.

cell lymphoma. Ontology analyses revealed molecular functions associated with the list of genes including integrin binding, C-C chemokine binding, ATP transmembrane transporter activity, G-protein coupled chemoattractant receptor activity, and fibroblast growth factor

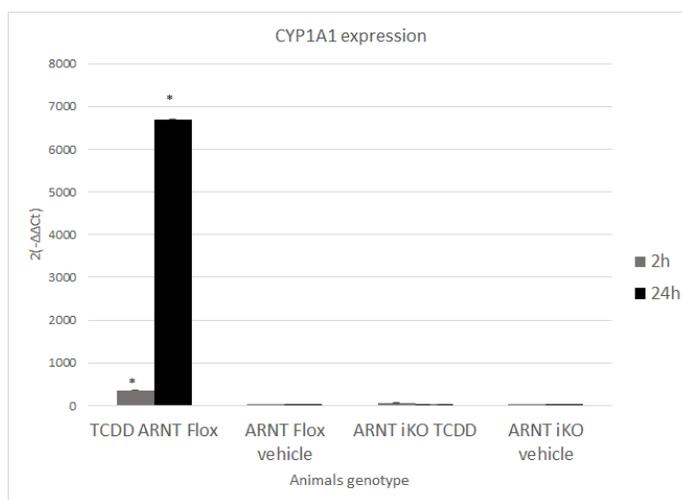


Figure 39: Effect of TCDD exposure on liver Cyp1a1 expression. qRT-PCR quantification of mRNA isolated from floxed or iKO mice treated with TCDD or vehicle for 2 or 24 hours, n=3. *p<0.0001. The values are presented as mean ± SE. Lack of ARNT drastically reduced.

binding. Biological processes linked to the genes are response to mechanical stimulus, cell locomotion, response to xenobiotic stimulus, response to food, positive regulation of locomotion and cell population proliferation (Figure 38). These findings are consistent with what is described in the aspects of AhR biology^{134,243-248}.

Comparisons between TCDD treated floxed and iKO animals revealed only 4 upregulated genes: Cyclic AMP-dependent transcription factor ATF-3 (*Atf3*), *Cyp1a1*, G0/G1 Switch Regulatory Protein 3 (*FosB*), glucose-6-phosphatase catalytic subunit (*G6pc*) (Figure 36). qRT-PCR confirmed CYP1A1 upregulation 2 and 24 hours after treatment, although inducible knockout animal expression was drastically reduced when compared to floxed animals (Figure 38). The residual CYP1A1 expression seen in ARNT iKO mice is attributed to the continued ARNT expression in the non-parenchymal liver cells still responsive to TCDD. qRT-PCR on ATF3, FosB and G6pc expression were also evaluated. Although G6pc expression was not altered upon TCDD treatment in the two genotypes, FosB was upregulated in floxed animals, while ATF3 was upregulated in iKO animals. These data suggest that ARNT contributes to ATF3 and FosB expression by suppressing the former and activating the latter.

FosB is one of the members of the Fos family proteins that interacts with JUN proteins forming the complex AP-1, regulating cell proliferation and differentiation²⁴⁹. The FosB promoter contains two XRE binding regions, but *in vitro* binding by the AhR-ARNT heterodimer in EMSAs could not be demonstrated with oligonucleotides containing XRE in pretreated Hepa-1 cells^{11,250}. ATF3, a stress induction transcription factor, is induced in the early phases of inflammatory response, and its deficiency leads to increase in liver damage due to reperfusion injury and hepatocellular apoptosis²⁵¹.

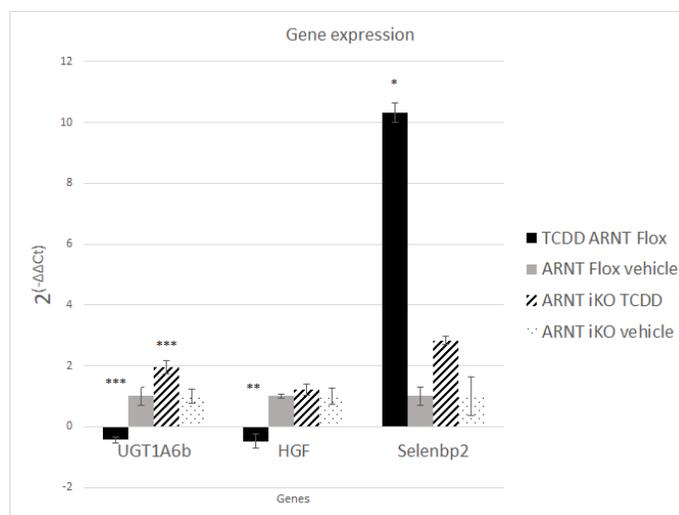


Figure 40: Effect of TCDD exposure on UGT1A6, HGF and Selenbp2 expression. qRT-PCR quantification in mRNA isolated from floxed or iKO mice treated with TCDD or vehicle for 2, n=3. * $p < 0.0001$; ** $p < 0.01$, *** $p < 0.05$. The values are presented as mean \pm SEM.

Within the RNA-seq datasets 30 genes that were detected to be upregulated only in the TCDD treated iKO animals, qRT-PCR validation demonstrated upregulation of UGT1A6b and Selenbp2 in iKO animals (Figure 40). Upregulation of Selenbp2 is also observed in floxed animals.

UGT1A6b is a UDP-glucuronosyltransferase that transforms small lipophilic molecules in metabolites²⁵². Selenium- and acetaminophen-binding protein (Selenbp2) may be linked to detoxification of xenobiotics. It was already demonstrated in the literature that exposure to TCDD lead to upregulation of expression of UGT1A6b in rats liver²⁵³⁻²⁵⁵.

Although HGF is not upregulated in iKO animals, it is downregulated in floxed animals. iKO animals did not have an increase of PAI-1 levels upon TCDD treatment (results shown on the next topic). PAI-1 interacts with urokinase-type plasminogen activator, staling the activation of HGF^{5,256}.

Some of the 30 genes initially identified as upregulated in iKO mice by RNA seq that were analyzed by qRT-PCR, and their induction could not be validated. These genes include: HTRA1, PTP4A1, Cyp2c55, Itih5, STON1, GLIS3, GSTM2 and BTG2.

RNA-seq analysis of downregulated genes ($\log_2\text{foldchange} \leq -1$, $\text{FDR} \leq 0.1$, $p \text{ value} < 0.05$) in floxed animals treated with TCDD or vehicle identified 40 genes (Appendix VI) while 43 genes (Appendix VII) were identified in iKO. The signaling pathways identified in the list of genes of floxed animals include signaling pathways regulating pluripotency, steroid hormone biosynthesis, TGF-beta signaling pathway, primary bile acid biosynthesis (Appendix VIII). Among the pathways associated with downregulated genes in iKO animals (Appendix IX) are highlighted chemical carcinogenesis, metabolism of xenobiotics by cytochrome P450, bile secretion, retinol metabolism, steroid hormone biosynthesis.

The data obtained from TCDD treated floxed and iKO animals was also compared. Analysis revealed 256 upregulated genes ($\log_2\text{foldchange} \geq 1$, $\text{FDR} \leq 0.1$, $p \text{ value} < 0.05$) and 136 downregulated genes ($\log_2\text{foldchange} \leq -1$, $\text{FDR} \leq 0.1$, $p \text{ value} < 0.05$). The list of genes was analyzed in Enrich. Pathway analysis revealed that upregulated genes are related to chemical carcinogenesis, cytochrome p450 metabolism of xenobiotics, retinol metabolism, steroid hormone biosynthesis, ascorbate and aldarate metabolism, drug metabolism, pentose and glucuronate interconversions, porphyrin and chlorophyll metabolism, linoleic acid metabolism. Diseases associated with those genes include chronic glomerulonephritis, muscular dystrophy, endocrine gland carcinoma, chemical burns.

Pathway analysis of genes in floxed animals treated with TCDD that were downregulated when compared to treated iKO animals are related to cell cycle, p53 signaling, cellular senescence, cytokine-cytokine receptor interaction, ECM-receptor interaction PPAR signaling pathway and steroid hormone biosynthesis. From the 134

genes, 114 had human homologs and were analyzed by Metascape²³⁷. The top signaling pathways are highlighted in table 1.

Table 1: Gene List pathway analysis by Metascape. Analysis of downregulated genes of ARNT Floxed animals treated with TCDD compared to TCDD treated ARNT iKO animals.

^a number of genes in the list. ^b percentage of all the list of genes. ^c p value in log base 10. ^d multi test adjusted p value in log base 10.

GO	Category	Description	Count ^a	% ^b	Log10(P) ^c	Log10(q) ^d
GO:1903047	GO Biological Processes	mitotic cell cycle process	40	33.33	-32.79	-28.43
GO:0010564	GO Biological Processes	regulation of cell cycle process	27	22.5	-18.54	-15.22
GO:0008608	GO Biological Processes	attachment of spindle microtubules to kinetochore	10	8.33	-15.65	-12.51
M129	Canonical Pathways	PID PLK1 PATHWAY	10	8.33	-14.32	-11.25
M176	Canonical Pathways	PID FOXM1 PATHWAY	9	7.5	-13.08	-10.12
R-HSA-983189	Reactome Gene Sets	Kinesins	9	7.5	-11.44	-8.74
GO:0000910	GO Biological Processes	Cytokinesis	12	10	-10.93	-8.26
M14	Canonical Pathways	PID AURORA B PATHWAY	7	5.83	-9.53	-6.91
GO:0140013	GO Biological Processes	meiotic nuclear division	9	7.5	-6.98	-4.66
WP2882	WikiPathways	Nuclear receptors meta-pathway	11	9.17	-6.85	-4.54
WP2361	WikiPathways	Gastric cancer network 1	5	4.17	-6.85	-4.54

GO:0000912	GO Biological Processes	assembly of actomyosin apparatus involved in cytokinesis	3	2.5	-5.83	-3.58
GO:0051383	GO Biological Processes	kinetochore organization	4	3.33	-5.67	-3.43
WP2874	WikiPathways	Liver X receptor pathway	3	2.5	-5.05	-2.84
M242	Canonical Pathways	PID AURORA A PATHWAY	4	3.33	-5.04	-2.83
GO:0097435	GO Biological Processes	supramolecular fiber organization	13	10.83	-4.73	-2.53
GO:0000212	GO Biological Processes	meiotic spindle organization	3	2.5	-4.58	-2.4
WP4240	WikiPathways	Regulation of sister chromatid separation at the metaphase-anaphase transition	3	2.5	-4.39	-2.23
GO:0022412	GO Biological Processes	cellular process involved in reproduction in multicellular organism	9	7.5	-4.39	-2.23
GO:0007096	GO Biological Processes	regulation of exit from mitosis	3	2.5	-4.31	-2.17

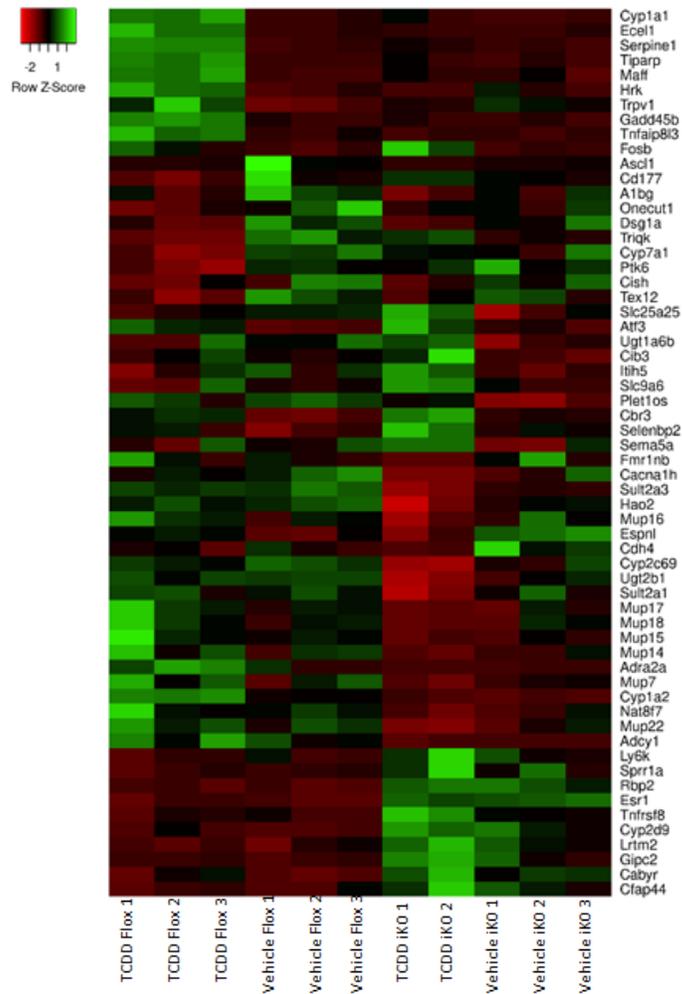


Figure 41: Average RNA-seq heat map of significant gene expression changes between TCDD and vehicle treated Floxed and iKO animals.

A heat map with the top 10 genes of each of the following comparisons: TCDD floxed x vehicle floxed upregulated and downregulated genes, TCDD iKO x vehicle iKO upregulated and downregulated genes. TCDD floxed x iKO TCDD upregulated and downregulated genes (Figure 41).

ChIP-sequencing for identification of AhR Target Genes

The next goal of this project was to perform ChIP-seq on DNA immunoprecipitated with AhR and KLF6 antibodies from ARNT flox and ARNT iKO animals. To analyze immunoprecipitated samples, RT-PCR was used to assess IP enrichment, using CYP1A1 and PAI-1 primers specific to XRE and NC-XRE promoter regions, respectively. Results confirm that upon TCDD exposure, AhR and ARNT bind to the XRE containing promoter

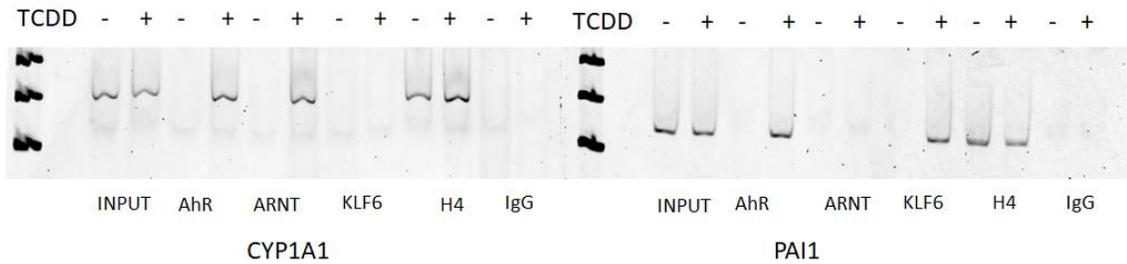


Figure 42: AhR and KLF6 bind to the PAI-1 promoter in a TCDD dependent manner.

region of Cyp1a1, while AhR and KLF6 binds to PAI-1 (Figure 42). The same material was submitted for next generation sequencing. Peak analysis of the sequenced samples however did not find known AhR binding sites in the promoter region of AhR targets such as Cyp1a1, PAI-1, AhRR, indicating that with current conditions our protocol was not sufficiently to identify binding motifs. Taking this into consideration, we are currently using Active Motif Inc (Carlsbad, California) ChIP-sequencing (ChIP-seq) services to conduct these studies in lieu of in-house ChIP-seq to mitigate trouble-shooting requirements.

Since previous studies have already identified some of the AhR direct target genes to validate the ChIP-seq experiment performed by Active Motif, peak analysis of AhR classical target genes (Figure 43) such as Cyp1a1, AhRR, Cyp1a2, Cyp1b1, Notch1, UGT1A2 was performed to validate the experiment results.

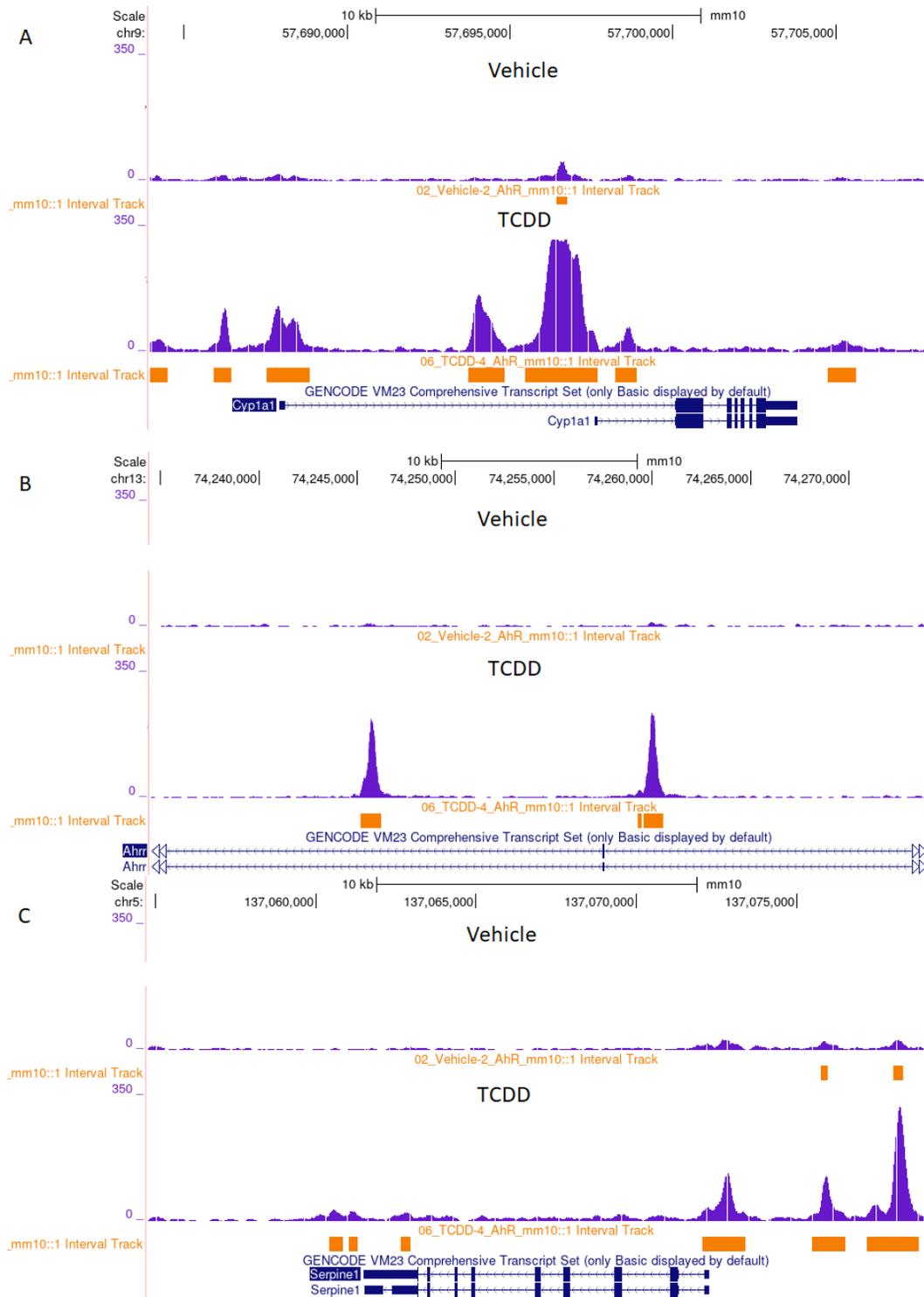


Figure 43: ChIP-seq tracks for AhR occupancy of Vehicle and TCDD-treated wild type animals. Representative tracks are displayed for Cyp1a1 (A), AhRR (B), PAI-1 (Serpine1).

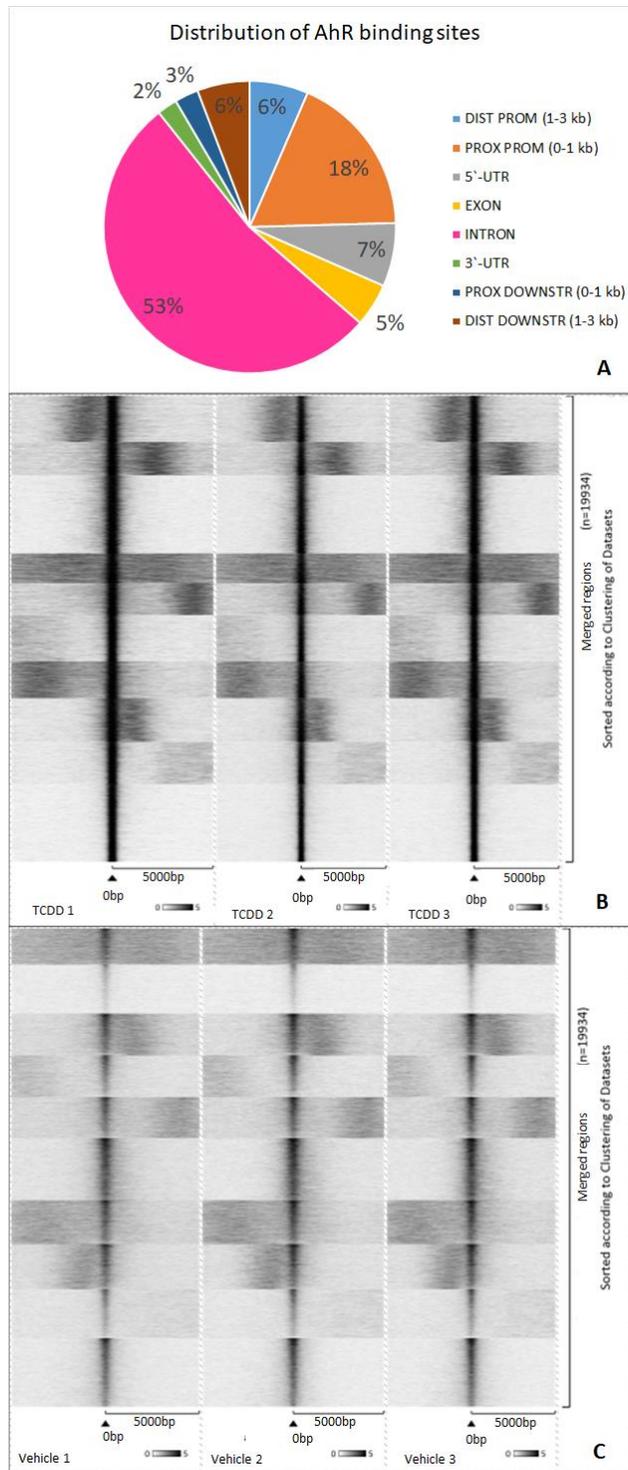


Figure 44: Overview of AhR binding sites. The genomic locations of peaks detected by chIP-seq was categorized in 8 classes (A). (B,C) Visualization of the results was achieved through a heat map. (B) TCDD treated samples, (C) Vehicle treated samples.

revealed majority of binding events occur in the genes intronic regions (Figure 44). Profile plots of the ChIP-seq signals showed that majority of AhR binding events on each gene occurred predominantly on promoter regions.

TFcheckpoint²²⁴ was used to identify the top 20 (Table 2) transcription factors with AhR DNA binding sites identified by ChIP-seq. Notable transcription factors which AhR is bound include: TCF4, a Wnt response element²⁵⁷; Nuclear Factor Of Activated T Cells 1 (Nfatc1), that is known to induce cell migration upon TCDD exposure and to act as a tumor suppressor in hepatocellular carcinoma^{258,259}; AhRR which represses AhR transcription²¹⁷; Nuclear Factor Of Activated T Cells 2 (Nfatc2) an inducer of the expression of cytokines in T-cells²⁶⁰; CAMP Responsive Element Binding Protein 3 Like 1 (Creb3l1) an inhibitor of cell cycle progression, upregulated in TCDD treated MCF-7 cells lacking AhR^{261,262}; Paired related homeobox protein 1 (Prrx1), a transcription co-activator that modulates platelet-derived growth factor that is upregulated during liver injury²⁶³.

Table 2: Analysis of Enriched Transcription Binding Sites

Gene_symbol	Gene_Name	Synonym
TCF4	transcription factor 4	E2-2 ITF-2 ITF2 PTHS SEF-2 SEF2 SEF2-1 SEF2-1A SEF2-1B TCF-4 BHLHB19
FOXP1	forkhead box P1	12CC4 QRF1 HFKH1B
NFATC1	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1	NF-ATC NFAT2 NFATC
ZFX3	zinc finger homeobox 3	ATBF1 ATBT ZNF927
AHRR	aryl-hydrocarbon receptor repressor	AHH AHHR BHLHE77
CREB3L1	cAMP responsive element binding protein 3-like 1	OASIS
ESRRB	estrogen-related receptor beta	DFNB35 ERR2 ERRB ESRL2 NR3B2
NFATC2	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2	NFAT1 NFATP
MECOM	MDS1 and EVI1 complex locus	AML1-EVI-1 EVI1 MDS1 MDS1-EVI1 PRDM3

MBD2	methyl-CpG binding domain protein 2	DMTASE NY-CO-41
KDM3A	lysine (K)-specific demethylase 3A	JHDM2A JHMD2A JMJD1 JMJD1A TSGA
WWOX	WW domain containing oxidoreductase	D16S432E FOR FRA16D HHCMA56 PRO0128 SDR41C1 WOX1
PRDM2	PR domain containing 2, with ZNF domain	HUMHOXY1 KMT8 MTB-ZF RIZ RIZ1 RIZ2
L3MBTL4	l(3)mbt-like 4 (Drosophila)	HST1031
GTF2A1	general transcription factor IIA, 1, 19/37kDa	TF2A1 TFIIA TFIIA-42 TFIIAL
APC	adenomatous polyposis coli	BTPS2 DP2 DP2.5 DP3 GS PPP1R46
PRRX1	paired related homeobox 1	AGOTC PHOX1 PMX1 PRX-1 PRX1
KAT6B	K(lysine) acetyltransferase 6B	GTPTS MORF MOZ2 MYST4 QKF QUERKOPF
NCOA2	nuclear receptor coactivator 2	GRIP1 KAT13C NCOA-2 SRC2 TIF2 BHLHE75

MOTIF MAPPING

Motif mapping using the ChIP-seq data detected that approximately 58% of the

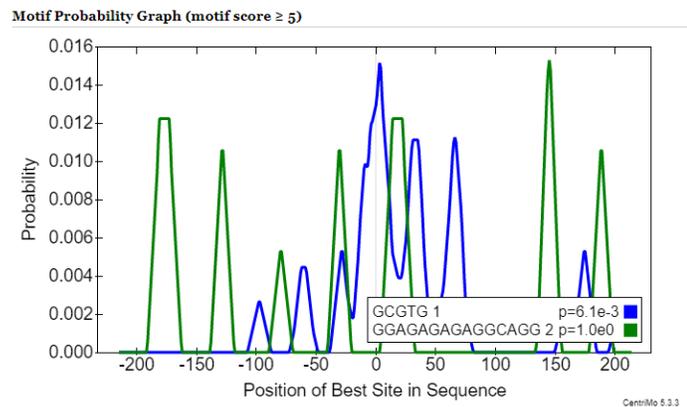


Figure 45: Positional distribution of AhR DNA binding motifs.

peaks contained 5'-GCGTG-3' motifs. Centrimo²⁶⁴ analysis was applied to investigate if AhR binding occurs near the center of the ChIP-seq peaks, which is usually observed if ChIP-seq experiment is successful. Analysis revealed that the probability of XRE's to be centrally enriched is 1.4%, an expected percentage in a motif that represents a transcription

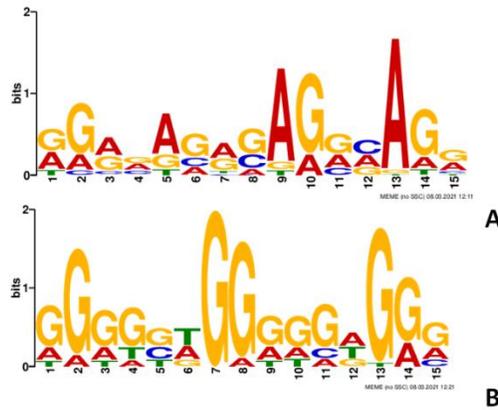


Figure 46: Motif discovery of ChIP-seq regions that lack XRE motifs. First (A) and fourth (B) motif detected by MEME motif mapping.

factor binding site²⁶⁵. 25% of the motifs are located within 145 bp from the peak center (Figure 46). Meme²⁶⁶ motif discovery was used to discover motifs in the regions that did not contain XRE's revealing the motif 5'-GGAGAGAGAGGCAGG-3' (e value 7.5e-058) (Figure 47). Comparison with the distribution of AhR binding between XRE non-XRE motifs showed a broader distribution of non-XRE binding events (Figure 46). Tomtom²⁶⁷ analysis was applied to compare the non-XRE motif with known motifs, revealing similarities with several TF binding sites, including znf263, SP2, SP3, SP4, KLF5, and KLF16.

Anchoring RNA-seq with ChIP-seq

Among the ChIP-seq peaks detected with anti-AhR antibody using a log₂ratio TCDD/vehicle ≥ 1 cut-off in TCDD treated samples, 134 of the 199 genes were also upregulated in the RNA-seq data obtained using ARNT floxed animals (Figure 47). Gene Set enrichment analysis²⁶⁸ linked this gene set with regulation of cell death, apoptotic process, embryonic morphogenesis, cell population proliferation, response to oxygen

The 87 TCDD responsive genes that did not feature an AhR peak in the ChIP-seq were analyzed by Binding Analysis for Regulation of Transcription (BART)²⁶⁹ to identify other putative transcription factors that may mediate dioxin response (Table 3).

Table 3: Possible transcription factors that may be associated with dioxin response

TF	statistic	pvalue	zscore	max_auc	re_rank	irwin_hall_pvalue
RXRA	5.514	1.75E-08	5.346	0.73	0.026	8.17E-05
FOXA2	4.415	5.04E-06	5.406	0.746	0.027	9.31E-05
HNF4A	8.234	9.07E-17	4.289	0.758	0.036	2.04E-04
NR3C1	6.4	7.79E-11	4.066	0.765	0.043	3.49E-04
ONECUT1	4.511	3.23E-06	7.705	0.649	0.044	3.78E-04
ISL1	6.516	3.61E-11	6.569	0.628	0.047	4.77E-04
FOXA1	5.355	4.29E-08	4.569	0.695	0.049	5.31E-04
AR	4.925	4.23E-07	5.841	0.635	0.051	6.11E-04
PTF1A	4.781	8.73E-07	4.963	0.643	0.06	9.52E-04

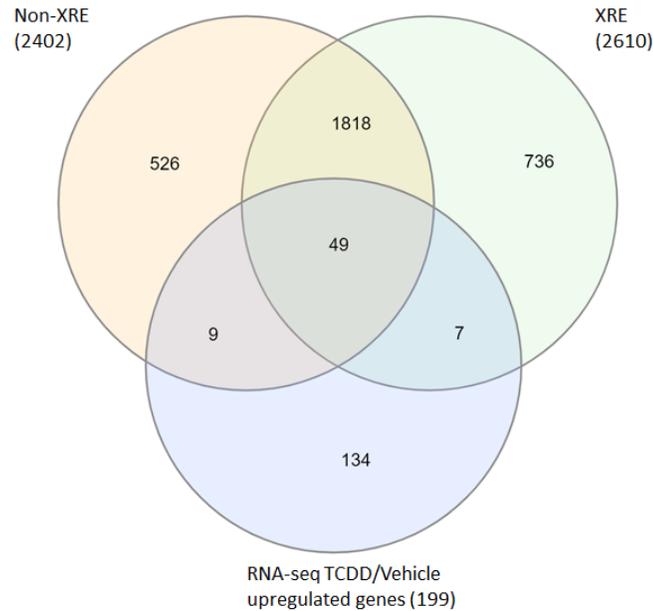


Figure 49: ChIP-seq comparison with RNAseq. AhR enrichment 0-10kb from the TSS of genes that contain non-XREs and XREs compared with RNA-seq of upregulated genes upon TCDD treatment.

Motif analysis revealed the 65 of genes upregulated in the RNA-seq dataset upon TCDD exposure contained XREs on the promoter regions. 49 of those genes however, also contained a distinct motif mapped with the highest score (5'-GGAGAGAGAGGCAGG-3'). Nine genes contained only Non-XREs motifs on the mapped region (figure 49). Those genes were *Ptges*, *Pmm1*, *Lekr1*, *Cxcl11*, *Adgrg3*, *Amot*, *G6pc*, *Per2* and *Dusp1*.

DISCUSSION

Different research groups have tried to identify the effects of exposure to environmental contaminants such as dioxin on liver damage. TCDD exposure leads to AhR activation, resulting in altered gene expression. Recent studies identified KLF6 as a new AhR binding partner. This complex binds to a novel DNA sequence, the NC-XRE, regulating novel AhR target genes. The main goal of this work was to better characterize AhR pathways by leveraging next generation sequencing. To achieve this aim we used an inducible liver-specific ARNT knockout mouse model. To unmask ARNT independent AhR activities, floxed and knockout animals were treated with TCDD or vehicle for 2 hours and their livers were prepared for next generation sequencing.

The analysis of the RNA-seq of treated floxed animals by Enrich revealed that most of the results are complement previous work regarding the AhR, demonstrating that the receptor plays crucial roles in xenobiotic metabolism, cell cycle regulation, liver regeneration, and homeostasis^{9,270-272}.

RNA-seq identified 199 upregulated genes upon TCDD treatment in ARNT floxed animals. Several of the upregulated genes have being previously identified as AhR target genes, including *CYP1A1*, *CYP1A2*, *CYP1B1*, UDP Glucuronosyltransferases, *CDKN1A*, *Gadd45b*, *Nfe2l2*, *Notch1*, and *Selenbp1*^{11,45,271,273-275}.

There were four common upregulated genes between floxed and iKO animals: *Cyp1a1*, *ATF3*, *FosB* and *G6P6*. Although *Cyp1a1* is upregulated in iKO treated animals,

a massive decrease in expression is observed in the iKO mice when comparing TCDD treated floxed with iKO animals. The liver is composed of different cell populations, with hepatocytes by far being the major cell type, with the non-parenchymal cells consisting of Kupffer-cells, endothelial and stellate cells, collectively representing only about 20% of the total liver cell population. Immunocytochemical analysis of mouse liver demonstrated that the cellular composition is similar to other mammals, with hepatocytes being the largest population, followed by Kupffer cells ($\approx 35\%$ of the hepatocyte population), stellate cells and squamous endothelial cells²⁷⁶. Since the conditional knockout mouse model utilized is hepatocyte specific, the presence of other cells populations may explain the observed residual Cyp1a1 and Cyp1b1 expression in ARNT knockout animals.

ATF3, a transcription factor involved in stress response, is upregulated by KLF6 in prostate cancer cells and after 72 hours in the liver of mice exposed to TCDD^{142,251,277-279}. The upregulation was confirmed by qRT-PCR in iKO animals, albeit similar induction was not detected in floxed mice treated with TCDD for 2 hours. ChIP-seq on wild-type animals identified the binding of the AhR both upstream and within ATF3 gene. Peaks in activating other members of the Activating Transcription Factors (ATF4, ATF6, ATF6b, ATF7 and ATF7ip) were also detected by ChIP-seq, highlighting that the AhR regulates the expression of homeostasis maintenance genes.

Glucose-6-Phosphatase Catalytic Subunit (G6PC) was also identified as a common upregulated gene between the floxed and iKO animals. However, qRT-PCR did not statistically detect differences between treated and control animals. Forgacs *et al.* (2012) investigated metabolic changes induced by TCDD in mouse and rat liver. Ovariectomized mice exposed to TCDD presented downregulated levels of G6PC 72 hours after treatment²⁷⁸. ChIP-seq with AhR antibody on wild-type animals, identified peaks on TCDD treated samples upstream and in the gene. This indicates that the levels of G6PC may be time- and dose-dependent, and that this gene is indeed a dioxin responsive gene.

FosB, a member of the Fos family proteins, interacts with JUN proteins forming the AP-1 complex regulates cell proliferation and differentiation^{280 281}. Upregulation was observed by qPCR only in TCDD-treated floxed animals. ChIP-seq analysis revealed AhR binding sites located both upstream and downstream of this gene. This find suggest that FosB may be an ARNT dependent gene since upregulation was not observed in iKO animals.

One of the goals of this work was to correlate RNA-sequencing with ChIP-sequencing targeting the AhR-KLF6 complex to analyze NC-XRE-regulated genes that change expression following TCDD treatment. As our attempt to perform ChIP-seq with ARNT floxed and ARNT iKO animals yielded ambiguous results, we elected to compare the RNA-seq data (of ARNT flox and ARNT iKO animals) with ChIP-seq from wild-type animals treated with TCDD for 2 hours.

Comparison between the RNA-seq data from floxed mice with ChIP-seq data revealed that the majority ($\approx 66\%$) of the genes upregulated upon TCDD treatment contain XRE motifs. 92% of these genes also contain peaks with the core motif identifying non-XRE regions (5'-GGAGAGAGAGGCAGG-3'). Approximately 33% of the upregulated genes of Floxed treated animal's possessed AhR enrichment 10 kb upstream of the TSS associated with an NC-XRE or XRE core. Nine genes where binding events occurred upstream of the TSS contained non-XREs motifs only. Those genes were Ptges, Pmm1, Lckr1, Cxcl11, Adgrg3, Amot, G6pc, Per2 and Dusp1.

PAI-1 is among of the genes that besides having peaks containing XRE motifs, it also contains peaks with the mapped motif 5'RGRVAGRSAGR VAGR3' (R=purine base, V=C, G or A, S=C or G), with the region mapped to 5'-AGGGAGGGAGGGAGG-3'. Interestingly, PAI-1 is not upregulated in ARNT iKO TCDD-treated animals, even though AhR binding to the NC-XRE appears to be ARNT independent. This may indicate that the lack of ARNT interferes with AhR activity elsewhere on the PAI-1 promoter suggestive of co-regulation between AhR canonical and non-canonical pathways.

Among the list of 199 upregulated and 40 downregulated transcripts upon TCDD exposure identified by RNA-seq of ARNT floxed animals, 87 of the corresponding genes did not have AhR bound to DNA according to ChIP-seq results. Since our studies focused on rapid (2 hours) transcriptional response, the lack of AhR-DNA binding is less likely indicative of secondary (indirect) transcriptional events, and instead supports direct AhR processes not detected by the pull-down antibody, possibly due to other as yet uncharacterized protein interactions.

The AhR-ARNT heterodimer, and AhR interactions with other partners such as KLF6 may act in a synergistic manner, to regulate the expression of dioxin responsive genes. A transcription factor may bind different sites of the genome due to interaction with different partners, favoring the binding in cells where a particular partner is expressed. Cooperative binding may occur if different complexes have a preference to bind DNA in certain conformations, so the energy spent by one protein to bend the DNA will not be spent by another when a different protein binds.²⁸²⁻²⁸⁴

The AhR canonical pathway is a complex mechanism not yet fully understood. It is proposed that ligand binding induces a conformational change in the PAS-B, to expose the AhR nuclear localization signal normally masked by hsp90 allowing the complex to translocate into the nucleus²⁸⁵. Upon ligand binding, the PAS-A domain also becomes more exposed, which facilitates the interaction with ARNT upon nuclear entry. ARNT was first described as an AhR nuclear translocation protein, but later it was demonstrated that it forms a complex with AhR subsequent to its entry into the nucleus²⁸⁶. ARNT binds to the AhR:hsp90 complex, which induces the release of the chaperone molecules and the formation of an active AhR-ARNT heterodimer²⁸⁷. It is unclear if this ARNT function is needed to render the AhR accessible to other proteins such as KLF6. Protein-protein interactions of multi-specific proteins depend on surface properties, net charge, and cellular concentration²⁸⁸. Regulation of transcription involves the formation of transcription factor complexes, and the kinetics of the process depends on the whole nuclear environment. The

alteration of any factor that affects protein diffusion also impacts reaction dynamics.²⁸⁹ Woolf & Linderman Monte Carlo simulations of membrane proteins indicated that binding is a dynamic process, favoring the formation of clusters that will facilitate the crosstalk between different partners. Alterations in dimerization of one species may alter the whole system²⁹⁰. It is possible that the absence of ARNT abolished AhR interaction with other partners, including KLF6.

AhR degradation is an important mechanism that prevents prolonged AhR activity. AhR proteolysis after ligand activation was observed in different cell lines. The conformational changes induced by ligand binding do not affect AhR stability, but does affect nuclear export of the activated receptor.^{27,291} It was previously reported that AhR nuclear export signal is localized within the region of ARNT interaction, which presumably blocks nuclear export²⁹². Robert & Whitelaw (1999)²⁹¹ demonstrated that the presence of ARNT is not essential for AhR stability, since AhR rapid turnover of the activated AhR occurs in both murine hepatoma Hepa1c1c7 cells (that contain ARNT) and in mutant Hepa1c4 cells that lack ARNT.

To better understand the roles of AhR in TCDD toxicity, hepatic RNA sequencing in liver-specific ARNT knockout animals was performed. qRT-PCR demonstrated decreased expression of AhR targets CYP1A1, Ugt1a2 and Notch1. Cyp1b1 was upregulated both in floxed and knockout animals. Interestingly, qRT-PCR analysis of genes that were upregulated in iKO TCDD treated animals compared with floxed treated animals (Inf2, Cyp2d9, Cyp7a1), revealed that the basal levels of these genes are similar in vehicle or TCDD-treated animals. Elevated basal levels of CYP1B1 were observed in treated ARNT deficient Hepa1 variant LA2. The increased expression of CYP1B1 suggests an inhibitory role of ARNT in the expression of certain AhR target genes. Transient transfection reporter studies in LA2 cells with a construct containing four XRE binding sites from the CYP1B1 promoter resulted in increased reporter gene expression in LA2

cells compared to wild-type cells, suggesting that ARNT suppresses CYP1B1 promoter activity in Hepa cells²⁹³.

The results obtained in this work showed upregulation in CYP1B1 levels in TCDD treated ARNT iKO animals, while CYP2D9 was unaffected. ChIP-seq analysis demonstrated AhR binding on the promoter region of CYP1B1 and CYP2D9, while binding in CYP2D9 was inconsistent, being observed in only one TCDD treated animal. Although a ChIP-seq analysis of iKO animals would be necessary to evaluate AhR DNA binding in ARNT knockout animals, it is interesting to point out that while CYP1B1 contained XRE binding sites commensurate with peak locations, CYP2D9 lacked XREs in the AhR binding sites. These results suggest that AhR and other transcription factors may be more effective in activating certain genes in the absence of ARNT.

Motif analysis revealed that approximately 41% of the peaks detected in ChIP-seq of WT animals, do not contain XRE motifs. Motif discovery mapping of the regions lacking XREs detected consensus motif 5'-RGRVAGRSAGR VAGR-3', with the best match 5'-GGAGAGAGAGGCAGG-3'. Around 30% of genes with a Log2Ratio TCDD/Vehicle \geq 1 contained the proposed motif, demonstrating that other transcription factors besides ARNT dimerize with the AhR and bind DNA. Considering EMSA assays demonstrated that AhR-KLF6 and AhR-CPS1 complexes bound to mutated NC-XRE sequences, ChIP peaks binding to non-XRE activated genes detected through motif mapping could seek to establish the binding of AhR, KLF6 and CPS1 at these sites. A distribution analysis of the AhR binding sites defined as XRE and non-XRE peaks revealed a higher probability of narrow central enrichment of XRE binding. Non-XRE motifs do not have a central enrichment, and instead exhibit a very broad binding distribution, which may connote a more diverse range of AhR interactions with other DNA-binding proteins^{265,294}.

While more research is needed to increase our understanding of the AhR canonical and non-canonical pathways, the RNA-seq and ChIP-seq analyses provide insights into

AhR functions in the liver in the presence and absence of ARNT, and how DNA binding correlates with changes in the transcriptome. Further studies with KLF6 knockout animals and ChIP-seq performed with a KLF6 antibody would provide valuable information about AhR-KLF6 DNA co-binding and how this protein complex modulates changes in gene expression due to dioxin exposure.

Chapter 4: Dissertation conclusion

As stated in the previous sections, a novel AhR pathway triggered by the interaction between AhR and KLF6 upon dioxin exposure was discovered and involves an interaction with a non-consensus XRE. Upon ligand binding, the AhR translocates into the nucleus, dissociates from the chaperone complex rendering the receptor competent to bind with KLF6. This heterodimer binds to NC-XREs associated with and regulate the transcription of target genes.

We were able to identify key peptides regions in both the AhR and KLF6 responsible for protein interaction and NC-XRE binding. Characterization of the NC-XRE binding sequence by EMSA revealed that single adenosine and guanine substitutions in the second 5'-GGGA-3' repeat comprising the PAI-1 NC-XRE tetranucleotide repeat did not abolish protein-DNA interactions, suggesting that the AhR-KLF6 DNA binding may tolerate substantial sequence variability.

ChIP-seq performed in TCDD-treated MCF-7 cells revealed that approximately 63% of AhR binding sites lack consensus XREs, and *de novo* motif analysis revealed GC boxes (a characteristic binding site of KLFs) as possible alternative AhR binding sites. Overlap between AhR and ARNT binding sites of 65% was observed^{295,296}. Meta-analysis on the promoter region of dioxin targets in cultured human cells also identified GC-rich motifs²¹⁷. This indicates that the AhR can interact with DNA independently of ARNT.

Interestingly, motif analysis in ChIP-seq studies on WT animals, revealed that more than 40% of the peaks detected in TCDD-treated mice did not contain XRE regions, and among those regions, the motifs detected are similar to other Kruppel like factors recognition motifs. Future ChIP-seq investigations performed using anti-KLF6 antibodies in mice treated with TCDD may be needed to better characterize genome-wide AhR-KLF6 DNA binding precisely fine-tune the functional NC-XRE sequence.

To anchor changes in the transcriptome with DNA binding events, ChIP-seq was compared to RNA-seq. Among the 199 upregulated genes, 138 ($\approx 69\%$) of the genes contained AhR DNA binding sites. Approximately 40% of these genes contained peaks located upstream of the transcription start site, $\approx 46\%$ contained peaks within the structural gene, and 24% of the peaks occurred downstream of the gene. These findings indicate that AhR binding events in the coding region and downstream in the transcription termination site may play an underappreciated role in AhR biology.

The observation that 127 of 199 of TCDD upregulated genes contained motifs similar to Kruppel like factors recognition motifs (5'-GGAGAGAGAGGCAGG-3' mapped motif) and that PAI-1 expression is not upregulated in iKO animals exposed to dioxin, suggest that NC-XRE and XRE sites may act cooperatively in regulating AhR target gene expression, and conspire to prevent sustained AhR signaling activity.

The discovery and investigation of the non-canonical AhR pathway yields new information on how the AhR regulates transcriptional activity, although more investigation is necessary to fully understand NC-XRE signaling. Future analyses of the AhR-KLF6 complex will shed much needed additional light on AhR receptor biology and how this important pathway plays a role in physiological and pathophysiological processes.

References

1. Poland, A., Glover, E. & Kende, A. Stereospecific, high affinity binding of 2,3,7,8-tetrachlorodibenzo-p-dioxin by hepatic cytosol. *J. Biol. Chem.* **251**, 4936–4946 (1976).
2. Greenlee, F. & Poland, A. Nuclear uptake of 2,3,7,8-tetrachlorodibenzo-p-dioxin in C57BL/6J and DBA/2J mice. Role of the hepatic cytosol receptor protein. *J. Biol. Chem.* **254**, 9814–9821 (1979).
3. Esser, C. & Rannug, A. The Aryl Hydrocarbon Receptor in Barrier Organ Physiology, Immunology, and Toxicology. *Pharmacol. Rev.* **67**, 259–279 (2015).
4. Sorg, O. AhR signalling and dioxin toxicity. *Toxicol. Lett.* **230**, 225–233 (2014).
5. Jackson, D. P., Joshi, a. D. & Elferink, C. J. Ah receptor pathway intricacies; signaling through diverse protein partners and DNA-motifs. *Toxicol. Res.* **4**, 1143–1158 (2015).
6. Mulero-Navarro, S. & Fernandez-Salguero, P. M. New Trends in Aryl Hydrocarbon Receptor Biology. *Front. Cell Dev. Biol.* **4**, 1–14 (2016).
7. Yin, J. *et al.* Role of AhR in positive regulation of cell proliferation and survival. *Cell Prolif.* **49**, 554–560 (2016).
8. Stejskalova, L., Dvorak, Z. & Pavek, P. Endogenous and Exogenous Ligands of Aryl Hydrocarbon Receptor: Current State of Art. *Curr. Drug Metab.* **12**, 198–212 (2011).
9. Larigot, L., Juricek, L., Dairou, J. & Coumoul, X. AhR signaling pathways and regulatory functions. *Biochim. Open* **7**, 1–9 (2018).
10. Corrada, D., Soshilov, A. A., Denison, M. S. & Bonati, L. Deciphering Dimerization Modes of PAS Domains: Computational and Experimental Analyses of the AhR:ARNT Complex Reveal New Insights Into the Mechanisms of AhR Transformation. *PLoS Comput. Biol.* **12**, (2016).
11. Beischlag, T. V, Morales, J. L., Hollingshead, B. D. & Perdew, G. H. The Aryl Hydrocarbon Receptor Complex and the Control of Gene Expression. *Crit Rev Eukaryot Gene Expr* **18**, 207–250 (2008).
12. Yi, T. *et al.* Aryl Hydrocarbon Receptor : A New Player of Pathogenesis and Therapy in Cardiovascular Diseases. *Biomed Res. Int.* **2018**, 1–11 (2018).
13. Shimba, S. & Watabe, Y. Crosstalk between the AHR signaling pathway and circadian rhythm. *Biochem. Pharmacol.* **77**, 560–565 (2009).
14. Wright, E. J., Castro, K. P. De, Joshi, A. D. & Elferink, C. J. Canonical and non-canonical aryl hydrocarbon receptor signaling pathways. *Curr. Opin. Toxicol.* **2**, 87–92 (2017).
15. Gene, C., Kobayashi, A., Sogawa, K. & Fujii-kuriyama, Y. Nucleic Acids , Protein Synthesis , and Molecular Genetics : Cooperative Interaction between AhR · Arnt and Sp1 for the Drug-inducible Expression Cooperative Interaction between AhR ½ Arnt and Sp1 for the Drug-inducible Expression of CYP1A1 Gene *. *Biochemistry* **271**, 12310–12316 (1996).
16. Vogel, C. F. A. *et al.* RelB, a new partner of aryl hydrocarbon receptor-mediated

- transcription. *Mol. Endocrinol.* **21**, 2941–2955 (2007).
17. Wu, Y., Baumgarten, S. C., Zhou, P. & Stocco, C. Testosterone-Dependent Interaction between Androgen Receptor and Aryl Hydrocarbon Receptor Induces Liver Receptor Homolog 1 Expression in Rat Granulosa Cells. *Mol. Cell. Biol.* **33**, 2817–2828 (2013).
 18. Rowlands, J. C., McEwan, I. J. & Gustafsson, J. a. Trans-activation by the human aryl hydrocarbon receptor and aryl hydrocarbon receptor nuclear translocator proteins: direct interactions with basal transcription factors. *Mol. Pharmacol.* **50**, 538–548 (1996).
 19. Dougherty, E. D. & Pollenz, R. Analysis of Ah Receptor-ARNT and Ah Receptor-ARNT2 complexes in vitro and in cell culture. *Toxicol Sci* **23**, 1–7 (2008).
 20. Ge, N. L. & Elferink, C. J. A direct interaction between the aryl hydrocarbon receptor and retinoblastoma protein: Linking dioxin signaling to the cell cycle. *J. Biol. Chem.* **273**, 22708–22713 (1998).
 21. Wang, F., Wang, W. & Safe, S. Regulation of constitutive gene expression through interactions of Sp1 protein with the nuclear aryl hydrocarbon receptor complex. *Biochemistry* **38**, 11490–11500 (1999).
 22. Puga, A. *et al.* Aromatic hydrocarbon receptor interaction with the retinoblastoma protein potentiates repression of E2F-dependent transcription and cell cycle arrest. *J. Biol. Chem.* **275**, 2943–2950 (2000).
 23. Kim, D. W. *et al.* The RelA NF- κ B subunit and the aryl hydrocarbon receptor (AhR) cooperate to transactivate the c- myc promoter in mammary cells. 5498–5506 (2000).
 24. Ruby, C. E., Leid, M. & Kerkvliet, N. I. 2,3,7,8-Tetrachlorodibenzo-p-dioxin Suppresses Tumor Necrosis Factor- α and Anti-CD40-Induced Activation of NF- κ B/Rel in Dendritic Cells: p50 Homodimer Activation Is Not Affected. *Mol. Phar* **62**, 722–728 (2002).
 25. Vogel, C., Sciallo, E. & Matsumura, F. Involvement of RelB in aryl hydrocarbon receptor-mediated induction of chemokines. *Biochem Biophys Res Commun* **363**, 722–726 (2007).
 26. Gutiérrez-Vázquez, C. & Quintana, F. J. Regulation of the Immune Response by the Aryl Hydrocarbon Receptor. *Immunity* **48**, 19–33 (2018).
 27. Pollenz, R. S. The mechanism of AH receptor protein down-regulation (degradation) and its impact on AH receptor-mediated gene regulation. *Chem. Biol. Interact.* **141**, 41–61 (2002).
 28. Seok, S. H. *et al.* Structural hierarchy controlling dimerization and target DNA recognition in the AHR transcriptional complex. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5431–5436 (2017).
 29. Petrusis, J. R., Kusnadi, A., Ramadoss, P., Hollingshead, B. & Perdew, G. H. The hsp90 Co-chaperone XAP2 Alters Importin β Recognition of the Bipartite Nuclear Localization Signal of the Ah Receptor and Represses Transcriptional Activity. *J. Biol. Chem.* **278**, 2677–2685 (2003).
 30. Denison, M. S., Soshilov, A. A., He, G., Degroot, D. E. & Zhao, B. Exactly the same but different: Promiscuity and diversity in the molecular mechanisms of action of the aryl hydrocarbon (dioxin) receptor. *Toxicol. Sci.* **124**, 1–22 (2011).
 31. Pecanova, L. & Farkas, R. Multiple functions and essential roles of nuclear

- receptor coactivators of bHLH-PAS family. *Endocr. Regul.* **50**, 165–181 (2016).
32. Kewley, R. J., Whitelaw, M. L. & Chapman-Smith, A. The mammalian basic helix-loop-helix/PAS family of transcriptional regulators. *Int. J. Biochem. Cell Biol.* **36**, 189–204 (2004).
 33. Ma, Q., Baldwin, K. T. & Virginia, W. Hydrocarbon Receptor (AhR) by the Ubiquitin-Proteasome Pathway. *J. Biol. Chem.* **275**, 8432–8438 (2000).
 34. Joshi, A. D., Mustafa, M. G., Lichti, C. F. & Elferink, C. J. Homocitrullination is a Novel Histone H1 Epigenetic Mark Dependent on Aryl Hydrocarbon Receptor Recruitment of Carbamoyl Phosphate Synthase 1. *J. Biol. Chem.* (2015). doi:10.1074/jbc.M115.678144
 35. Wilson, S. R., Joshi, A. D. & Elferink, C. J. The Tumor Suppressor Kruppel-Like Factor 6 Is a Novel Aryl Hydrocarbon Receptor DNA Binding Partner. *J. Pharmacol. Exp. Ther.* **345**, 419–429 (2013).
 36. Huang, G. & Elferink, C. J. A novel nonconsensus xenobiotic response element capable of mediating aryl hydrocarbon receptor-dependent gene expression. *Mol. Pharmacol.* **81**, 338–47 (2012).
 37. Mitchell, K. a, Lockhart, C. a, Huang, G. & Elferink, C. J. Sustained aryl hydrocarbon receptor activity attenuates liver regeneration. *Mol. Pharmacol.* **70**, 163–170 (2006).
 38. Ghosh, A. K. & Vaughan, D. E. PAI-1 in tissue fibrosis. *J. Cell. Physiol.* **227**, 493–507 (2012).
 39. Coudriet, G. M. *et al.* A Noncanonical Role for Plasminogen Activator Inhibitor Type 1 in Obesity-Induced Diabetes. *Am. J. Pathol.* **189**, 1413–1422 (2019).
 40. Nejak-Bowen, K., Orr, A., Bowen, W. C. & Michalopoulos, G. K. Conditional Genetic Elimination of Hepatocyte Growth Factor in Mice Compromises Liver Regeneration after Partial Hepatectomy. *PLoS One* **8**, (2013).
 41. Wilson, S. R., Joshi, A. D. & Elferink, C. J. The Tumor Suppressor Kruppel-Like Factor 6 Is a Novel Aryl Hydrocarbon Receptor DNA Binding Partner. *J. Pharmacol. Exp. Ther.* **345**, 419–429 (2013).
 42. Narla, G. *et al.* In vivo regulation of p21 by the Kruppel-like factor 6 tumor-suppressor gene in mouse liver and human hepatocellular carcinoma. *Oncogene* **26**, 4428–4434 (2007).
 43. Lang, U. *et al.* GSK3 β phosphorylation of the KLF6 tumor suppressor promotes its transactivation of p21. *Oncogene* **32**, 4557–4564 (2014).
 44. Narla, G. *et al.* KLF6, a candidate tumor suppressor gene mutated in prostate cancer. *Science (80-.)*. **294**, 2563–2566 (2001).
 45. Jackson, D. P. & Mitchell, K. A. Ah Receptor – Mediated Suppression of Liver Regeneration through NC- XRE – Driven p21 Cip1 Expression. *Mol. Pharmacol.* 533–541 (2013).
 46. Catez, F., Tetsuya, U. & Bustin, M. Determinants of histone H1 mobility and chromatin binding in living cells. *Nat Struct Mol Biol* **13**, 1–7 (2006).
 47. Kamieniarz, K. *et al.* A dual role of linker histone transcriptional activation. *Genes Dev.* **26**, 797–802 (2012).
 48. Mandal, P. K. Dioxin: A review of its environmental effects and its aryl hydrocarbon receptor biology. *J. Comp. Physiol. B Biochem. Syst. Environ. Physiol.* **175**, 221–230 (2005).

49. Pelclová, D. *et al.* Adverse health effects in humans exposed to 2,3,7,8-tetrachlorodibenzo-p- dioxin (TCDD). *Rev. Environ. Health* **21**, 119–138 (2006).
50. National Toxicology Program, D. of H. and H. S. 2,3,7,8-Tetrachlorodibenzo-p-dioxin. *Rep. Carcinog.* (2009).
51. Michalek, J. E. *et al.* Pharmacokinetics of TCDD in veterans of operation ranch hand: 10-year follow-up. *J. Toxicol. Environ. Heal. - Part A* **47**, 209–220 (1996).
52. Wu, L. & Whitlock, J. P. Mechanism of dioxin action: Ah receptor-mediated increase in promoter accessibility in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 4811–4815 (1992).
53. Kerkvliet, N. I. TCDD: An Environmental Immunotoxicant Reveals a Novel Pathway of Immunoregulation--A 30-Year Odyssey. *Toxicol. Pathol.* **40**, 138–142 (2012).
54. Landi, M. T. *et al.* 2,3,7,8-Tetrachlorodibenzo-p-dioxin plasma levels in Seveso 20 years after the accident. *Environ. Health Perspect.* **106**, 273–277 (1998).
55. Eskenazi, B. *et al.* The Seveso accident: A look at 40 years of health research and beyond. *Environ. Int.* **121**, 71–84 (2018).
56. Consonni, D. *et al.* Mortality in a population exposed to dioxin after the Seveso, Italy, accident in 1976: 25 Years of follow-up. *Am. J. Epidemiol.* **167**, 847–858 (2008).
57. Pesatori, A. C. *et al.* Cohort study of the population exposed to dioxin after the Seveso, Italy accident: Cancer incidence results, 1977-2012. *ISEE Conf. Abstr.* **1**, 10–11 (2016).
58. White, S. & Birnbaum, L. An Overview of the Effects of Dioxins and Dioxin-like Compounds on Vertebrates, as Documented in Human and Ecological Epidemiology. *J Env. Sci Heal. C Env. Carcinog Ecotoxicol* **27**, 197–211 (2009).
59. Fox, G. A. Wildlife As Sentinels of Human Health Effects in the Great Lakes–St. Lawrence Basin. *Environ. Health Perspect.* **109**, 853–861 (2001).
60. Gandhi, N. *et al.* Dioxins in Great Lakes fish: Past, present and implications for future monitoring. *Chemosphere* **222**, 479–488 (2019).
61. Yamaguchi, Y. & Kuo, M. T. Functional analysis of aryl hydrocarbon receptor nuclear translocator interactions with aryl hydrocarbon receptor in the yeast two-hybrid system. *Biochem. Pharmacol.* **50**, 1295–1302 (1995).
62. Tsuji, N. *et al.* The activation mechanism of the aryl hydrocarbon receptor (AhR) by molecular chaperone HSP90. *FEBS Open Bio* **4**, 796–803 (2014).
63. Watt, K., Jess, T. J., Kelly, S. M., Price, N. C. & McEwan, I. J. Induced α -helix structure in the aryl hydrocarbon receptor transactivation domain modulates protein - Protein interactions. *Biochemistry* **44**, 734–743 (2005).
64. Flaveny, C. & Perdew, G. H. Transgenic Humanized AHR Mouse Reveals Differences between Human and Mouse AHR Ligand Selectivity. *Mol. Cell. Pharmacol.* **1**, 119–123 (2009).
65. Fukunaga, B. N., Probst, M. R., Reisz-Porszasz, S. & Hankinson, O. Identification of functional domains of the aryl hydrocarbon receptor. *J Biol Chem* **270**, 29270–29278 (1995).
66. Whitelaw, M. L., Göttlicher, M., Gustafsson, J. A. & Poellinger, L. Definition of a novel ligand binding domain of a nuclear bHLH receptor: co-localization of ligand and hsp90 binding activities within the regulable inactivation domain of the dioxin

- receptor. *EMBO J.* **12**, 4169–4179 (1993).
67. Tkachenko, A. *et al.* The Q-rich/PST domain of the AHR regulates both ligand-induced nuclear transport and nucleocytoplasmic shuttling. *Sci. Rep.* **6**, 1–11 (2016).
 68. Fukunaga, B. N. & Hankinson, O. Identification of a novel domain in the aryl hydrocarbon receptor required for DNA binding. *J Biol.Chem* **271**, 3743–3749 (1996).
 69. Bunger, M. K. *et al.* Resistance to 2,3,7,8-tetrachlorodibenzo-p-dioxin toxicity and abnormal liver development in mice carrying a mutation in the nuclear localization sequence of the aryl hydrocarbon receptor. *J. Biol. Chem.* **278**, 17767–17774 (2003).
 70. Schulte, K. W., Green, E., Wilz, A., Platten, M. & Daumke, O. Structural Basis for Aryl Hydrocarbon Receptor-Mediated Gene Activation. *Structure* **25**, 1025-1033.e3 (2017).
 71. Seok, S. *et al.* Structural hierarchy controlling dimerization and target DNA recognition in the AHR transcriptional complex. **114**, 5431–5436 (2017).
 72. AHR Result Summary | BioGRID. Available at: <https://thebiogrid.org/106699/summary/homo-sapiens/ahr.html>. (Accessed: 1st August 2021)
 73. Matthews, J. & Gustafsson, J.-A. Estrogen receptor and aryl hydrocarbon receptor signaling pathways. *Nucl. Recept. Signal.* **4**, e016 (2006).
 74. Beischlag, T. V. & Perdew, G. H. ER α -AHR-ARNT protein-protein interactions mediate estradiol-dependent transrepression of dioxin-inducible gene transcription. *J. Biol. Chem.* **280**, 21607–21611 (2005).
 75. Tappenden, D. M., Hwang, H. J., Yang, L., Thomas, R. S. & Lapres, J. J. The aryl-hydrocarbon receptor protein interaction network (AHR-PIN) as identified by tandem affinity purification (TAP) and mass spectrometry. *J. Toxicol.* **2013**, (2013).
 76. Elferink, C. J., Ge, N. L. & Levine, A. Maximal aryl hydrocarbon receptor activity depends on an interaction with the retinoblastoma protein. *Mol. Pharmacol.* **59**, 664–673 (2001).
 77. NCOA1 - Nuclear receptor coactivator 1 - Homo sapiens (Human) - NCOA1 gene & protein. Available at: <https://www.uniprot.org/uniprot/Q15788>. (Accessed: 2nd August 2021)
 78. NCOA2 - Nuclear receptor coactivator 2 - Homo sapiens (Human) - NCOA2 gene & protein. Available at: <https://www.uniprot.org/uniprot/Q15596>. (Accessed: 2nd August 2021)
 79. Beischlag, T. V. *et al.* Recruitment of the NCoA/SRC-1/p160 Family of Transcriptional Coactivators by the Aryl Hydrocarbon Receptor/Aryl Hydrocarbon Receptor Nuclear Translocator Complex. *Mol. Cell. Biol.* **22**, 4319–4333 (2002).
 80. Gene, C., Kobayashi, A., Sogawa, K. & Fujii-kuriyama, Y. Cooperative Interaction between AhR · Arnt and Sp1 for the Drug-inducible Expression of CYP1A1 Gene *. *Biochemistry* **271**, 12310–12316 (1996).
 81. SP1 - Transcription factor Sp1 - Homo sapiens (Human) - SP1 gene & protein. Available at: <https://www.uniprot.org/uniprot/P08047>. (Accessed: 2nd August 2021)

82. Pollak, N. M., Hoffman, M., Goldberg, I. J. & Drosatos, K. Krüppel-Like Factors Crippling and uncropping metabolic pathways. *JACC Basic to Transl. Sci.* **3**, 132–156 (2018).
83. Nagai, R., Friedman, S.L., Kasuga, M. *The Biology of Kruppel like Factors.* (Springer US, 2009).
84. Lomberk, G. & Urrutia, R. The family feud: Turning off Sp1 by Sp1-like KLF proteins. *Biochem. J.* **392**, 1–11 (2005).
85. Lu, X. J., Shi, Y., Chen, J. L. & Ma, S. Kruppel-like factors in hepatocellular carcinoma. *Tumor Biol.* **36**, 533–541 (2015).
86. Swamynathan, S. K. Krüppel-like factors: three fingers in control. *Hum. Genomics* **4**, 263–270 (2010).
87. McConnell, B. B. & Yang, V. Mammalian Krüppel-Like Factors in Health and Diseases. *Physiol Rev* **23**, 1–7 (2011).
88. Kim, C. K., He, P., Bialkowska, A. B. & Yang, V. W. SP and KLF Transcription Factors in Digestive Physiology and Diseases. *Gastroenterology* **152**, 1845–1875 (2017).
89. Oishi, Y. & Manabe, I. Krüppel-Like Factors in Metabolic Homeostasis and Cardiometabolic Disease. *Front. Cardiovasc. Med.* **5**, 1–14 (2018).
90. Goutham, N. & Friedman, S. L. Krüppel-like Factors and the Liver. in *The Biology of Krüppel-like Factors* 1–258 (2009). doi:10.1007/978-4-431-87775-2
91. Gray, S. *et al.* Regulation of Gluconeogenesis by Krüppel-like Factor 15. *Cell Metab* **5**, 305–312 (2007).
92. Krenkel, O. & Tacke, F. Liver macrophages in tissue homeostasis and disease. *Nat. Rev. Immunol.* **17**, 306–321 (2017).
93. Jha, P. & Das, H. KLF2 in regulation of NF- κ B-mediated immune cell function and inflammation. *Int. J. Mol. Sci.* **18**, (2017).
94. Sydor, S. *et al.* Krüppel-like factor 6 is a transcriptional activator of autophagy in acute liver injury. *Sci. Rep.* 1–12 (2017). doi:10.1038/s41598-017-08680-w
95. Nakamura, H., Chiambaretta, F., Sugar, J., Sapin, V. & Yue, B. Y. J. T. Developmentally regulated expression of KLF6 in the mouse cornea and lens. *Investig. Ophthalmol. Vis. Sci.* **45**, 4327–4332 (2004).
96. Miele, L. *et al.* The Kruppel-Like Factor 6 Genotype Is Associated With Fibrosis in Nonalcoholic Fatty Liver Disease. *Gastroenterology* **135**, 282–291 (2010).
97. Ray, K. KLF6 activates PPAR α signalling in hepatic steatosis. *Nat. Rev. Gastroenterol. Hepatol.* **10**, 128 (2013).
98. Koritschoner, N. P. *et al.* A novel human zinc finger protein that interacts with the core promoter element of a TATA box-less gene. *J. Biol. Chem.* **272**, 9573–9580 (1997).
99. Andreoli, V., Gehrau, R. C. & Bocco, J. L. Biology of Kruppel-Like Factor 6 Transcriptional Regulator in Cell Life and Death. *IUBMB Life* **62**, 896–905 (2010).
100. Matsumoto, N. *et al.* Developmental regulation of yolk sac hematopoiesis by Krüppel-like factor 6. *Blood* **107**, 1357–1365 (2006).
101. Mgbemena, V., Segovia, J., Chang, T. & Bose, S. Krüppel-like factor 6 regulates transforming growth factor- gene expression during human respiratory syncytial virus infection. *Virology* **8**, 1–6 (2011).
102. Kojima, S. *et al.* Transcriptional activation of urokinase by the Kruppel-like factor

- Zf9/COPEB activates latent TGF- β 1 in vascular endothelial cells. *Blood* **95**, 1309–1316 (2000).
103. Rubinstein, M. *et al.* Transcriptional activation of the insulin-like growth factor I receptor gene by the Kruppel-like factor 6 (KLF6) tumor suppressor protein: Potential interactions between KLF6 and p53. *Endocrinology* **145**, 3769–3777 (2004).
 104. Niimi, T., Hayashi, Y., Sekiguchi, K. & Kitagawa, Y. The Sp family of transcription factors regulates the human laminin α 1 gene in JAR choriocarcinoma cells. *Biochim. Biophys. Acta - Gene Struct. Expr.* **1759**, 573–579 (2006).
 105. Warke, V. G. *et al.* Transcriptional activation of the human inducible nitric-oxide synthase promoter by Krüppel-like factor 6. *J. Biol. Chem.* **278**, 14812–14819 (2003).
 106. Chiambaretta, F. *et al.* Regulation of corneal keratin-12 gene expression by the human Krüppel-like transcription factor 6. *Investig. Ophthalmol. Vis. Sci.* **43**, 3422–3429 (2002).
 107. Park, J. H. *et al.* KLF6 is one transcription factor involved in regulating acid ceramidase gene expression. *Biochim. Biophys. Acta - Gene Struct. Expr.* **1732**, 82–87 (2005).
 108. Okano, J. I. *et al.* The Kruppel-like transcriptional factors Zf9 and GKLF coactivate the human keratin 4 promoter and physically interact. *FEBS Lett.* **473**, 95–100 (2000).
 109. Botella, L. M. *et al.* Transcriptional activation of endoglin and transforming growth factor- β signaling components by cooperative interaction between Sp1 and KLF6 : their potential role in the response to vascular injury. *Homeostasis, Thromb. Vasc. Biol.* **100**, 4001–4010 (2002).
 110. Zhao, J. L., Austen, K. F. & Lam, B. K. Cell-specific transcription of leukotriene C4 synthase involves a Kruppel-like transcription factor and Sp1. *J. Biol. Chem.* **275**, 8903–8910 (2000).
 111. DiFeo, A. *et al.* E-cadherin is a novel transcriptional target of the KLF6 tumor suppressor. *Oncogene* **25**, 6026–6031 (2006).
 112. Li, D. *et al.* Krüppel-like factor-6 promotes preadipocyte differentiation through histone deacetylase 3-dependent repression of DLK1. *J. Biol. Chem.* **280**, 26941–26952 (2005).
 113. Smas, C. M. & Sul, H. S. Pref-1, a protein containing EGF-like repeats, inhibits adipocyte differentiation. *Cell* **73**, 725–734 (1993).
 114. Nyunoya, H., Brogfie, K. E., Widgren, E. E. & Lusty, J. Characterization and Derivation of the Gene Coding for Mitochondrial Carbamyl Phosphate Synthetase I of Rat. *J. Biol. Chem.* **260**, 9346–9356 (1985).
 115. Rubio, V. Structure-function studies in carbamoyl phosphate synthetases. *Biochem. Soc. Trans* **21**, (1993).
 116. Simmer, J. P., Kelly, R. E., Rinker, A. G., Scully, J. L. & Evans, D. R. Mammalian Carbamyl Phosphate Synthetase (CPS). *J. BIOLOCKAL Chem.* **265**, 10395–10402 (1990).
 117. Yefimenko, I. *et al.* Understanding Carbamoyl-phosphate Synthetase I (CPS1) Deficiency by Using Expression Studies and Structure- Based Analysis. *Hum. Mutat.* **31**, (2010).

118. Díez-Fernández, C., Hu, L., Cervera, J., Häberle, J. & Rubio, V. Understanding carbamoyl phosphate synthetase (CPS1) deficiency by using the recombinantly purified human enzyme: Effects of CPS1 mutations that concentrate in a central domain of unknown function. *Mol. Genet. Metab.* **112**, 123–132 (2014).
119. Rubio, V. & Cervera, J. The carbamoyl-phosphate synthase family and carbamate kinase: Structure-function studies. *Biochem. Soc. Trans.* **23**, 879–883 (1995).
120. De Cima, S. *et al.* Structure of human carbamoyl phosphate synthetase: Deciphering the on/off switch of human ureagenesis. *Sci. Rep.* **5**, 1–15 (2015).
121. Badar, A., Arif, Z. & Alam, K. Role of Carbamylated Biomolecules in Human Diseases. *IUBMB Life* **70**, 267–275 (2018).
122. Tijet, N., Boutros, P. C., Moffat, I. D., Okey, A. B. & Tuomisto, J. Aryl Hydrocarbon Receptor Regulates Distinct Dioxin- Dependent and Dioxin-Independent Gene Batteries. *Mol. Pharmacol.* **69**, 140–153 (2006).
123. Carmona, P., Mendez, N., Ili, C. G. & Brebi, P. The Role of Clock Genes in Fibrinolysis Regulation: Circadian Disturbance and Its Effect on Fibrinolytic Activity. *Front. Physiol.* **11**, 1–7 (2020).
124. Cesari, M., Pahor, M. & Incalzi, R. A. Plasminogen activator inhibitor-1 (PAI-1): A key factor linking fibrinolysis and age-related subclinical and clinical conditions. *Cardiovasc. Ther.* **28**, 72–91 (2010).
125. Sato, Y. *et al.* Coagulation and Fibrinolytic Systems During Liver Regeneration in the Early Period after Adult Living Related Partial Liver Transplantation. *Transplant. Proc.* **40**, 2501–2502 (2008).
126. Puga, A., Maier, A. & Medvedovic, M. The transcriptional signature of dioxin in human hepatoma HepG2 cells. *Biochem. Pharmacol.* **60**, 1129–1142 (2000).
127. Frueh, F. W., Hayashibara, K. C., Brown, P. O. & Whitlock, J. P. Use of cDNA microarrays to analyze dioxin-induced changes in human liver gene expression. *Toxicol. Lett.* **122**, 189–203 (2001).
128. Boverhof, D. R. *et al.* Temporal and dose-dependent hepatic gene expression patterns in mice provide new insights into TCDD-mediated hepatotoxicity. *Toxicol. Sci.* **85**, 1048–1063 (2005).
129. Son, D. S. & Rozman, K. K. 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) induces plasminogen activator inhibitor-1 through an aryl hydrocarbon receptor-mediated pathway in mouse hepatoma cell lines. *Arch. Toxicol.* **76**, 404–413 (2002).
130. Wilson, S., Joshi, A. & Elferink, C. The tumor suppressor Kruppel-like factor 6 is a novel aryl hydrocarbon receptor DNA binding partner. *J. Pharmacol. Exp. Ther.* **345**, 419–429 (2013).
131. Pombo, M., Lamé, M. W., Walker, N. J., Huynh, D. H. & Tablin, F. TCDD and omeprazole prime platelets through the aryl hydrocarbon receptor (AhR) non-genomic pathway. *Toxicol. Lett.* **235**, 28–36 (2015).
132. Dere, E., Lo, R., Celius, T., Matthews, J. & Zacharewski, T. R. Integration of Genome-Wide Computation DRE Search, AhR CHIP-chip and Gene Expression Analyses of TCDD-Elicited Responses in the Mouse Liver. *BMC Genomics* **12**, 365 (2011).
133. Lo, R. & Matthews, J. High-resolution genome-wide Mapping of AHR and ARNT binding sites by CHIP-Seq. *Toxicol. Sci.* **130**, 349–361 (2012).
134. Kung, T., Murphy, K. A. & White, L. A. The aryl hydrocarbon receptor (AhR)

- pathway as a regulatory pathway for cell adhesion and matrix metabolism. *Biochem. Pharmacol.* **77**, 536–546 (2009).
135. Fletcher, N. *et al.* 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) alters the mRNA expression of critical genes associated with cholesterol metabolism, bile acid biosynthesis, and bile transport in rat liver: A microarray study. *Toxicol. Appl. Pharmacol.* **207**, 1–24 (2005).
 136. Carreira, V. S. *et al.* Ah receptor signaling controls the expression of cardiac development and homeostasis genes. *Toxicol. Sci.* **147**, 425–435 (2015).
 137. Smith, B. W. *et al.* The aryl hydrocarbon receptor directs hematopoietic progenitor cell expansion and differentiation. *Hematop. stem cells* **122**, 376–385 (2013).
 138. Juricek, L. & Coumoul, X. The aryl hydrocarbon receptor and the nervous system. *Int. J. Mol. Sci.* **19**, (2018).
 139. Ochoa, I. H., Karman, B. N. & Flaws, J. A. The Role of the Aryl Hydrocarbon Receptor in the Female Reproductive System. *Biochem Pharmacol.* **23**, 1–7 (2005).
 140. Tijet, N. *et al.* Aryl hydrocarbon receptor regulates distinct dioxin-dependent and dioxin-independent gene batteries. *Mol. Pharmacol.* **69**, 140–153 (2006).
 141. Prokopec, S. D. *et al.* Compendium of TCDD-mediated transcriptomic response datasets in mammalian model systems. *BMC Genomics* **18**, 1–8 (2017).
 142. DiFeo, A., Martignetti, J. A. & Narla, G. The role of KLF6 and its splice variants in cancer therapy. *Drug Resist. Updat.* **12**, 1–7 (2009).
 143. Rodriguez, E., Aburjania, N., Priedigkeit, N. M., Difeo, A. & Martignetti, J. A. Nucleo-cytoplasmic localization domains regulate kruppel-like factor 6 (KLF6) protein stability and tumor suppressor function. *PLoS One* **5**, 1–13 (2010).
 144. Presnell, J. S., Schnitzler, C. E. & Browne, W. E. KLF/SP transcription factor family evolution: Expansion, diversification, and innovation in eukaryotes. *Genome Biol. Evol.* **7**, 2289–2309 (2015).
 145. Lu, X. J., Shi, Y., Chen, J. L. & Ma, S. Kruppel-like factors in hepatocellular carcinoma. *Tumor Biol.* **36**, 533–541 (2015).
 146. Tetreault, M., Yang, Y. & Katz, J. P. Krüppel-like factors in cancer. *Nat. Publ. Gr.* **13**, 701–713 (2013).
 147. Cho, Y. G. *et al.* Genetic alterations of the KLF6 gene in gastric cancer. *Oncogene* **24**, 4588–4590 (2005).
 148. Sangodkar, J. *et al.* Functional role of the KLF6 tumour suppressor gene in gastric cancer. *Eur. J. Cancer* **45**, 666–676 (2009).
 149. Hartel, M. *et al.* Increased alternative splicing of the KLF6 tumor suppressor gene correlates with prognosis and tumor grade in patients with pancreatic cancer. *Eur J Cancer* **44**, 1895–1903 (2008).
 150. Diaferia, G. R. *et al.* Dissection of transcriptional and cis -regulatory control of differentiation in human pancreatic cancer. *EMBO J.* **35**, 595–617 (2016).
 151. Nault, R. *et al.* Pyruvate kinase isoform switching and hepatic metabolic reprogramming by the environmental contaminant 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Toxicol. Sci.* **149**, 358–371 (2016).
 152. Rothhammer, V. & Quintana, F. J. The aryl hydrocarbon receptor: an environmental sensor integrating immune responses in health and disease. *Nat. Rev. Immunol.* **19**, 184–197 (2019).

153. Pohjanvirta, R. & Viluksela, M. Novel Aspects of Toxicity Mechanisms of Dioxins and Related Compounds. *Int. J. Mol. Sci.* **21**, 4–9 (2020).
154. Sabatino, M. E. *et al.* Krüppel-Like Factor 6 Is Required for Oxidative and Oncogene-Induced Cellular Senescence. *Front. Cell Dev. Biol.* **7**, 1–11 (2019).
155. Bechmann, L. P. *et al.* Post-transcriptional activation of PPAR alpha by KLF6 in hepatic steatosis. *J Hepatol.* **58**, 1000–1006 (2014).
156. ATCC. MCF7 ATCC® HTB-22™. Available at: <https://www.atcc.org/products/all/htb-22.aspx>. (Accessed: 6th October 2020)
157. Miller, A. G., Israel, D. & Whitlock Jr., J. P. Biochemical and genetic analysis of variant mouse hepatoma cells defective in the induction of Benzo(a)pyrene-metabolizing enzyme activity. *J. Biol. Chem.* **258**, 3523–3527 (1983).
158. Diez-fernandez, C. *et al.* Molecular Characterization of Carbamoyl-Phosphate Synthetase (CPS1) Deficiency Using Human Recombinant CPS1 as a Key Tool. *Hum. Mutat.* **34**, 1149–1159 (2013).
159. He, J. *Practical Guide to ELISA Development. The Immunoassay Handbook* (Elsevier, 2013). doi:10.1016/b978-0-08-097037-0.00025-7
160. Biesiadecki, B. J. & Jin, J. A High-Throughput Solid-Phase Microplate Protein-Binding Assay to Investigate Interactions between Myofilament Proteins. *J. Biomed. Biotechnol.* **2011**, 1–8 (2011).
161. Zappelli, E., Daniele, S., Abbracchio, M. P., Martini, C. & Trincavelli, M. L. A Rapid and Efficient Immunoenzymatic Assay to Detect Receptor Protein Interactions : G Protein-Coupled Receptors. *Int. J. Mol. Sci.* **15**, 6252–6264 (2014).
162. Graham, N. A., Pope, M. D., Rimchala, T., Huang, B. K. & Asthagiri, A. R. A Microtiter Assay for Quantifying Protein-Protein Interactions Associated with Cell-Cell Adhesion. *J. Biomol. Screen.* **12**(5); **12**, 683–693 (2007).
163. Weng, Z. & Zhao, Q. Utilizing ELISA to Monitor Protein-Protein Interaction. in *Protein-Protein Interactions: Methods and Applications* (eds. Meyerkord, C. L. & Fu, H.) 341–352 (Springer New York, 2015). doi:10.1007/978-1-4939-2425-7_21
164. ImageJ. Available at: <https://imagej.nih.gov/ij/index.html>. (Accessed: 20th July 2020)
165. Thompson, P. M., Beck, M. R. & Campbell, S. L. *Protein-protein interaction analysis by nuclear magnetic resonance spectroscopy. Methods in molecular biology (Clifton, N.J.)* **1278**, (2015).
166. Labrecque, M. P. *et al.* Distinct roles for Aryl hydrocarbon receptor nuclear translocator and Ah receptor in estrogen-mediated signaling in human cancer cell lines. *PLoS One* **7**, 1–10 (2012).
167. Dorh O.; Vogel C., Abel, J. Different Response of 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD)-Sensitive Genes in Human Breast Cancer MCF-7 and MDA-MB 231 Cells. *Arch. Biochem. Biophys.* **321**, 405–412 (1995).
168. Bunger, M. K. *et al.* Abnormal liver development and resistance to 2,3,7,8-tetrachlorodibenzo-p-dioxin toxicity in mice carrying a mutation in the DNA-Binding domain of the aryl hydrocarbon receptor. *Toxicol. Sci.* **106**, 83–92 (2008).
169. Ikuta, T., Eguchi, H., Tachibana, T., Yoneda, Y. & Kawajiri, K. Nuclear localization and export signals of the human aryl hydrocarbon receptor. *J. Biol. Chem.* **273**, 2895–2904 (1998).
170. Elferink, C. J. & Whitlock, J. P. 2,3,7,8-Tetrachlorodibenzo-p-dioxin-inducible,

- Ah receptor-mediated bending of enhancer DNA. *J. Biol. Chem.* **265**, 5718–5721 (1990).
171. Siggers, T. & Gordân, R. Protein-DNA binding: Complexities and multi-protein codes. *Nucleic Acids Res.* **42**, 2099–2111 (2014).
 172. nd Kelsey C. MartinMhatre V. Ho, Ji-Ann Lee, A. Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* **23**, 1–7 (2012).
 173. Kejnovská, I., Kypr, J. & Vorlíčková, M. Circular dichroism spectroscopy of conformers of (guanine + adenine) repeat strands of DNA. *Chirality* **15**, 584–592 (2003).
 174. Huertas, D. & Azorín, F. Structural polymorphism of homopurine DNA sequences. d(GGA)(n) and d(GGGA)(n) repeats form intramolecular hairpins stabilized by different base- pairing interactions. *Biochemistry* **35**, 13125–13135 (1996).
 175. Jakubec, D., Laskowski, R. A. & Vondrasek, J. Sequence-specific recognition of DNA by proteins: Binding motifs discovered using a novel statistical/computational analysis. *PLoS One* **11**, 1–24 (2016).
 176. Chou, S. H., Chin, K. H. & Wang, A. H. J. Unusual DNA duplex and hairpin motifs. *Nucleic Acids Res.* **31**, 2461–2474 (2003).
 177. Murre, C., McCaw, P. S. & Baltimore, D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* **56**, 777–783 (1989).
 178. Kato, G. J. & Dang, C. V. Function of the c-Myc oncoprotein. *FASEB J.* **6**, 3065–3072 (1992).
 179. Peyton, M. *et al.* BETA3, a novel helix-loop-helix protein, can act as a negative regulator of BETA2 and MyoD-responsive genes. *Mol. Cell. Biol.* **16**, 626–633 (1996).
 180. Chen, Y. Q., Sengchanthalangsy, L. L., Hackett, A. & Ghosh, G. NF- κ B p65 (RelA) homodimer uses distinct mechanisms to recognize DNA targets. *Structure* **8**, 419–428 (2000).
 181. Dobbins, S. E., Lesk, V. I. & Sternberg, M. J. E. Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 10390–10395 (2008).
 182. Bonati, L., Corrada, D., Tagliabue, S. G. & Motta, S. Molecular modeling of the AhR structure and interactions can shed light on ligand-dependent activation and transformation mechanisms. *Curr Opin Toxicol.* **2**, 42–49 (2018).
 183. Kolonko, M. & Greb-Markiewicz, B. BHLH–PAS proteins: Their structure and intrinsic disorder. *Int. J. Mol. Sci.* **20**, (2019).
 184. Fujita, H. *et al.* Characterization of the aryl hydrocarbon receptor repressor gene and association of its Pro185Ala polymorphism with micropenis. *Teratology* **65**, 10–18 (2002).
 185. Yesudhas, D., Battoo, M., Anwar, M. A., Panneerselvam, S. & Choi, S. Proteins recognizing DNA: Structural uniqueness and versatility of DNA-binding domains in stem cell transcription factors. *Genes (Basel)*. **8**, (2017).
 186. Corsi, F., Lavery, R., Laine, E. & Carbone, A. Multiple protein-DNA interfaces unravelled by evolutionary information, physico-chemical and geometrical properties. *PLoS Comput. Biol.* **16**, e1007624 (2020).

187. Von Hippel, P. H. & Berg, O. G. Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675–678 (1989).
188. Gemayel, R. *et al.* Variable Glutamine-Rich Repeats Modulate Transcription Factor Activity. *Mol. Cell* **59**, 615–627 (2015).
189. Hsieh, M., Tintut, Y. & Gralla, J. D. Functional roles for the glutamines within the glutamine-rich region of the transcription factor σ 54. *J. Biol. Chem.* **269**, 373–378 (1994).
190. Mayr, B. M., Guzman, E. & Montminy, M. Glutamine Rich and Basic Region / Leucine Zipper (bZIP) Domains Stabilize cAMP-response Element-binding Protein (CREB) Binding to Chromatin. *J. Biol. Chem.* **280**, 15103–15110 (2005).
191. Courey, A. J., Holtzman, D. A., Jackson, S. P. & Tjian, R. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. *Cell* **59**, 827–836 (1989).
192. Sen, A., Hsieh, W. C. & Aguilar, R. C. The Information Content of Glutamine-Rich Sequences Define Protein Functional Characteristics. *Proc. IEEE* **105**, 385–393 (2017).
193. Keskin, O. & Nussinov, R. Similar Binding Sites and Different Partners: Implications to Shared Proteins in Cellular Pathways. *Structure* **15**, 341–354 (2007).
194. Dequeker, C., Laine, E. & Carbone, A. Decrypting protein surfaces by combining evolution, geometry, and molecular docking. *Proteins Struct. Funct. Bioinforma.* **87**, 952–965 (2019).
195. Vinet, L. & Zhedanov, A. *Protein Dimerization and Oligomerization in Biology. Journal of Physics A: Mathematical and Theoretical* **44**, (2011).
196. Crews, S. T. Control of cell lineage-specific development and transcription by bHLH- PAS proteins. *Genes Dev.* **12**, 607–620 (1998).
197. Fischer, E. a, Verpont, M. C., Garrett-Sinha, L. a, Ronco, P. M. & Rossert, J. a. Klf6 is a zinc finger protein expressed in a cell-specific manner during kidney development. *J. Am. Soc. Nephrol.* **12**, 726–35 (2001).
198. Zhang, Y. *et al.* Krüppel-like factor 6 is a co-activator of nf-kb that mediates p65-dependent transcription of selected downstream genes. *J. Biol. Chem.* **289**, 12876–12885 (2014).
199. Chen, X. & Bieker, J. J. Erythroid Kruppel-like factor (EKLF) contains a multifunctional transcriptional activation domain important for inter- and intramolecular interactions. *EMBO J.* **15**, 5888–5896 (1996).
200. Geiman, D. E., Ton-That, M., Johnson, J. M. & Yang, V. W. Transactivation and growth suppression by the gut-enriched Kruppel-like factor (Kruppel-like factor 4) are dependent on acidic amino acid residues and protein-protein interaction. *Nucleic Acids Res.* **28**, 1106–1113 (2000).
201. Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **285**, 1735–1747 (1999).
202. Vasudev, P. G., Banerjee, M., Ramakrishnan, C. & Balaram, P. Asparagine and glutamine differ in their propensities to form specific side chain-backbone hydrogen bonded motifs in proteins. *Proteins Struct. Funct. Bioinforma.* **80**, 991–1002 (2012).

203. Rhys, N. H., Soper, A. K. & Dougan, L. The hydrogen-bonding ability of the amino acid glutamine revealed by neutron diffraction experiments. *J. Phys. Chem. B* **116**, 13308–13319 (2012).
204. Banerjee, A., Levy, Y. & Mitra, P. Analyzing Change in Protein Stability Associated with Single Point Deletions in a Newly Defined Protein Structure Database. *J. Proteome Res.* **18**, 1402–1410 (2019).
205. Nisthal, A., Wang, C. Y., Ary, M. L. & Mayo, S. L. Protein stability engineering insights revealed by domain-wide comprehensive mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 16367–16377 (2019).
206. Shortle, D. & Sondek, J. The emerging role of insertions and deletions in protein engineering. *Curr. Opin. Biotechnol.* **6**, 387–393 (1995).
207. Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A. & Sidhu, S. S. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8950–8954 (2000).
208. Shortle, D. Probing the determinants of protein folding and stability with amino acid substitutions. *J. Biol. Chem.* **264**, 5315–5318 (1989).
209. Raghunathan, G., Soundrarajan, N., Sokalingam, S., Yun, H. & Lee, S. G. Deletional Protein Engineering Based on Stable Fold. *PLoS One* **7**, (2012).
210. Cunningham, B. C. & Wells, J. A. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science (80-)*. **244**, 1081–1085 (1989).
211. Edelheit, O., Hanukoglu, I., Dascal, N. & Hanukoglu, A. Identification of the roles of conserved charged residues in the extra cellular domain of an epithelial sodium channel (ENaC) subunit by alanine mutagenesis. *Am. J. Physiol. - Ren. Physiol.* **300**, 887–897 (2011).
212. cBioPortal for Cancer Genomics. Available at: <http://www.cbioportal.org/>. (Accessed: 16th April 2021)
213. Gonzalez, M. W. & Kann, M. G. Chapter 4: Protein Interactions and Disease. *PLoS Comput. Biol.* **8**, (2012).
214. Rao, V. S., Srinivas, K., Sujini, G. N. & Kumar, G. N. S. Protein-Protein Interaction Detection: Methods and Analysis. *Int. J. Proteomics* **2014**, 1–12 (2014).
215. Kimura, A. *et al.* Aryl hydrocarbon receptor in combination with Stat1 regulates LPS-induced inflammatory responses. *J. Exp. Med.* **206**, 2027–2035 (2009).
216. Josyula, N. *et al.* Gene co-regulation and co-expression in the aryl hydrocarbon receptor-mediated transcriptional regulatory network in the mouse liver. *Arch. Toxicol.* **94**, 113–126 (2020).
217. Oshchepkova, E., Sizentsova, Y., Wiebe, D., Mironova, V. & Kolchanov, N. Meta-analysis of transcriptome data detected new potential players in response to dioxin exposure in humans. *Int. J. Mol. Sci.* **21**, 1–20 (2020).
218. Feil, S., Valtcheva, N. & Feil, R. *Inducible cre mice. Methods in Molecular Biology* **530**, (humana press, 2009).
219. Activity, C. R. E., Fetal, I. N., For, U. & Studies, D. Cre activity in fetal albCre mouse hepatocytes: utility for developmental studies. *Genesis* **47**, 789–792 (2010).
220. Nukaya, M., Walisser, J. A., Moran, S. M., Kennedy, G. D. & Bradfield, C. A. Aryl Hydrocarbon Receptor Nuclear Translocator in Hepatocytes Is Required for

- Aryl Hydrocarbon Receptor – Mediated Adaptive and Toxic Responses in Liver. *Toxicol. Sci.* **118**, 554–563 (2010).
221. Nguyen, T. T., Cho, K., Stratton, S. A. & Barton, M. C. Transcription Factor Interactions and Chromatin Modifications Associated with p53-Mediated, Developmental Repression of the Alpha-Fetoprotein Gene. *Mol. Cell. Biol.* **25**, 2147–2157 (2005).
 222. Illumina Inc. Illumina Sequencing Technology. *Oct.* 23 (2013).
 223. Reuter, J. A., Damek, S. & Snyder, M. High-Throughput Sequencing Technologies. *Mol Cell* **58**, (2015).
 224. Chawla, K., Tripathi, S., Thommesen, L., Lægreid, A. & Kuiper, M. TFcheckpoint: A curated compendium of specific DNA-binding RNA polymerase II transcription factors. *Bioinformatics* **29**, 2519–2520 (2013).
 225. Introduction - MEME Suite. Available at: <http://meme-suite.org/>. (Accessed: 26th January 2021)
 226. Ji, H., Jiang, H., Ma, W. & Wong, W. H. Using CisGenome to Analyze ChIP-chip and ChIP-seq Data Hongkai. *Curr Protoc Bioinforma.* 1–63 (2011). doi:10.1002/0471250953.bi0213s33.Using
 227. EaSeq – Interactive ChIP-seq analysis software. Available at: <https://easeq.net/>. (Accessed: 8th August 2021)
 228. Rio, D. C., Ares, M., Hannon, G. J. & Nilsen, T. W. Purification of RNA using TRIzol (TRI Reagent). *Cold Spring Harb. Protoc.* **5**, 1–4 (2010).
 229. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. (Accessed: 9th June 2020)
 230. Alignment with STAR | Introduction to RNA-Seq using high-performance computing. Available at: https://hbctraining.github.io/Intro-to-rnaseq-hpc-O2/lessons/03_alignment.html. (Accessed: 9th June 2020)
 231. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, (2014).
 232. New England Biolabs. NEB Tm Calculator. Available at: <https://tmcalculator.neb.com/#!/main>. (Accessed: 11th October 2020)
 233. Babicki, S. *et al.* Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res.* **44**, W147–W153 (2016).
 234. PCAGO. Available at: <https://pcago.bioinf.uni-jena.de/>. (Accessed: 30th November 2020)
 235. Bushkofskya, J. R., Maguire, M., Larsen, M. C., Fong, Y. H. & Jefcoate, C. R. Cyp1b1 affects external control of mouse hepatocytes, fatty acid homeostasis and signaling involving HNF4 α and PPAR α . *Arch Biochem Biophys* **597**, 30–47 (2016).
 236. Enrichr. Available at: <https://amp.pharm.mssm.edu/Enrichr/enrich>. (Accessed: 10th July 2020)
 237. Metascape. Available at: <https://metascape.org/gp/index.html#/main/step1>. (Accessed: 25th October 2020)
 238. Piñero, J. *et al.* DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **45**, D833–D839

- (2017).
239. Gialitakis, M. *et al.* Activation of the Aryl Hydrocarbon Receptor Interferes with Early Embryonic Development. *Stem Cell Reports* **9**, 1377–1386 (2017).
 240. Metidji, A. *et al.* The Environmental Sensor AHR Protects from Inflammatory Damage by Maintaining Intestinal Stem Cell Homeostasis and Barrier Integrity. *Immunity* **49**, 353–362.e5 (2018).
 241. Delescluse, C., Lemaire, G., De Sousa, G. & Rahmani, R. Is CYP1A1 induction always related to AHR signaling pathway? *Toxicology* **153**, 73–82 (2000).
 242. Fader, K. A. *et al.* 2,3,7,8-tetrachlorodibenzo-p-dioxin alters lipid metabolism and depletes immune cell populations in the Jejunum of C57BL/6 mice. *Toxicol. Sci.* **148**, 567–580 (2015).
 243. Bradshaw, T. D. & Bell, D. R. Relevance of the aryl hydrocarbon receptor (AhR) for clinical toxicology. *Clin. Toxicol.* **47**, 632–642 (2009).
 244. Andreola, F. *et al.* Accumulation and Reduced Retinoic Acid Metabolism. *Cancer Res.* **57**, 2835–2838 (1997).
 245. Lahoti, T. S. *et al.* Aryl hydrocarbon receptor activation synergistically induces lipopolysaccharide-mediated expression of proinflammatory chemokine (c-c motif) ligand 20. *Toxicol. Sci.* **148**, 229–240 (2015).
 246. Nakata, A. *et al.* G-protein signalling negatively regulates the stability of aryl hydrocarbon receptor. *EMBO Rep.* **10**, 622–628 (2009).
 247. Brawner, K. M. *et al.* Depletion of dietary aryl hydrocarbon receptor ligands alters microbiota composition and function. *Sci. Rep.* **9**, 1–12 (2019).
 248. Tsai, C. H. *et al.* The inhibition of lung cancer cell migration by AhR-regulated autophagy. *Sci. Rep.* **7**, 1–14 (2017).
 249. FOSB Gene - GeneCards | FOSB Protein | FOSB Antibody. Available at: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=FOSB>. (Accessed: 18th October 2020)
 250. Hoffer, A., Chang, C. Y. & Puga, A. Dioxin induces transcription of fos and jun genes by Ah receptor-dependent and -independent pathways. *Toxicol. Appl. Pharmacol.* **141**, 238–247 (1996).
 251. Zhu, Q. *et al.* Loss of ATF3 exacerbates liver damage through the activation of mTOR/p70S6K/ HIF-1 α signaling pathway in liver inflammatory injury. *Cell Death Dis.* **9**, 1–13 (2018).
 252. UGT1A6 UDP glucuronosyltransferase family 1 member A6 [Homo sapiens (human)] - Gene - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=54578>. (Accessed: 23rd October 2020)
 253. Selenbp2 - Selenium-binding protein 2 - Mus musculus (Mouse) - Selenbp2 gene & protein. Available at: <https://www.uniprot.org/uniprot/Q63836>. (Accessed: 23rd October 2020)
 254. Chu, R. *et al.* Protein Profiling of Mouse Livers with Peroxisome Proliferator-Activated Receptor α Activation. *Mol. Cell. Biol.* **24**, 6288–6297 (2004).
 255. Pastorelli, R. *et al.* Differential Expression Profiling of the Hepatic Proteome in a Rat Model of Dioxin Resistance. *Mol. Cell. Proteomics* **5**, 882–894 (2006).
 256. Shimizu, M. *et al.* Mechanism of retarded liver regeneration in plasminogen activator — Deficient mice: Impaired activation of hepatocyte growth factor after

- fas-mediated massive hepatic apoptosis. *Hepatology* **33**, 569–576 (2001).
257. Wang, Q. *et al.* Ah receptor activation by dioxin disrupts activin, BMP, and WNT signals during the early differentiation of mouse embryonic stem cells and inhibits cardiomyocyte functions. *Toxicol. Sci.* **149**, 346–357 (2016).
 258. Seifert, A., Rau, S., Küllertz, G., Fischer, B. & Santos, A. N. TCDD induces cell migration via NFATc1/ATX-signaling in MCF-7 cells. *Toxicol. Lett.* **184**, 26–32 (2009).
 259. Xu, S. *et al.* NFATc1 is a tumor suppressor in hepatocellular carcinoma and induces tumor cell apoptosis by activating the FasL-mediated extrinsic signaling pathway. *Cancer Med.* **7**, 4701–4717 (2018).
 260. Nfatc2 - Nuclear factor of activated T-cells, cytoplasmic 2 - Mus musculus (Mouse) - Nfatc2 gene & protein. Available at: <https://www.uniprot.org/uniprot/Q60591>. (Accessed: 17th April 2021)
 261. Salisbury, T. B. *et al.* Endogenous aryl hydrocarbon receptor promotes basal and inducible expression of tumor necrosis factor target genes in MCF-7 cancer cells. *Biochem Pharmacol* **91**, 390–399 (2014).
 262. Creb3l1 - Cyclic AMP-responsive element-binding protein 3-like protein 1 - Mus musculus (Mouse) - Creb3l1 gene & protein. Available at: <https://www.uniprot.org/uniprot/Q9Z125>. (Accessed: 17th April 2021)
 263. Gong, J. *et al.* Paired related homeobox protein 1 regulates PDGF-induced chemotaxis of hepatic stellate cells in liver fibrosis. *Lab. Investig.* **97**, 1020–1032 (2017).
 264. CentriMo Output Formats - MEME Suite. Available at: <https://meme-suite.org/meme/doc/centrimo-output-format.html>. (Accessed: 6th March 2021)
 265. Bailey, T. L. & MacHanick, P. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res.* **40**, 1–10 (2012).
 266. MEME - MEME Suite. Available at: <https://meme-suite.org/meme/doc/meme.html>. (Accessed: 9th March 2021)
 267. Tomtom - MEME Suite. Available at: <https://meme-suite.org/meme/doc/tomtom.html>. (Accessed: 9th March 2021)
 268. GSEA | MSigDB | Investigate Gene Sets. Available at: <http://www.gsea-msigdb.org/gsea/msigdb/annotate.jsp>. (Accessed: 29th March 2021)
 269. BART web. Available at: <http://bartweb.org/>. (Accessed: 15th April 2021)
 270. Harper, P. A., Riddick, D. S. & Okey, A. B. Regulating the regulator: Factors that control levels and activity of the aryl hydrocarbon receptor. *Biochem. Pharmacol.* **72**, 267–279 (2006).
 271. Boverhof, D. R. *et al.* Temporal and dose-dependent hepatic gene expression patterns in mice provide new insights into TCDD-mediated hepatotoxicity. *Toxicol. Sci.* **85**, 1048–1063 (2005).
 272. Moreno-Marin, N. *et al.* Dioxin Receptor Adjusts Liver Regeneration After Acute Toxic Injury and Protects Against Liver Carcinogenesis. *Sci. Rep.* **7**, 1–12 (2017).
 273. Lu, P. *et al.* Constitutive Activation of the Human Aryl Hydrocarbon Receptor in Mice Promotes Hepatocarcinogenesis Independent of Its Coactivator Gadd45b. *Toxicol. Sci.* **167**, 581–592 (2019).
 274. Tonelli, C., Chio, I. I. C. & Tuveson, D. A. Transcriptional Regulation by Nrf2. *Antioxidants Redox Signal.* **29**, 1727–1745 (2018).

275. Tsujimoto, S. *et al.* Selenium-binding protein 1: Its physiological function, dependence on aryl hydrocarbon receptors, and role in wasting syndrome by 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Biochim. Biophys. Acta - Gen. Subj.* **1830**, 3616–3624 (2013).
276. Baratta, J. L. *et al.* Cellular organization of normal mouse liver: A histological, quantitative immunocytochemical, and fine structural analysis. *Histochem. Cell Biol.* **131**, 713–726 (2009).
277. Huang, X., Li, X. & Guo, B. KLF6 induces apoptosis in prostate cancer cells through up-regulation of ATF3. *J. Biol. Chem.* **283**, 29795–29801 (2008).
278. Forgacs, A. L. *et al.* Comparative metabolomic and genomic analyses of TCDD-elicited metabolic disruption in mouse and rat liver. *Toxicol. Sci.* **125**, 41–55 (2012).
279. Yoon, C. Y. *et al.* Gene expression profile by 2,3,7,8-tetrachlorodibenzo-p-dioxin in the liver of wild-type (AhR^{+/+}) and aryl hydrocarbon receptor-deficient (AhR^{-/-}) mice. *J. Vet. Med. Sci.* **68**, 663–668 (2006).
280. FOSB FosB proto-oncogene, AP-1 transcription factor subunit [Homo sapiens (human)] - Gene - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/gene/2354>. (Accessed: 26th October 2020)
281. Casado, F. L., Singh, K. P. & Gasiewicz, T. A. Aryl hydrocarbon receptor activation in hematopoietic stem/progenitor cells alters cell function and pathway-specific gene modulation reflecting changes in cellular trafficking and migration. *Mol. Pharmacol.* **80**, 673–682 (2011).
282. Vandell, J., Cassan, O., Lèbre, S., Lecellier, C. H. & Bréhélin, L. Probing transcription factor combinatorics in different promoter classes and in enhancers. *bioRxiv* 1–19 (2017). doi:10.1101/197418
283. Guo, Y. & Gifford, D. K. Modular combinatorial binding among human trans-acting factors reveals direct and indirect factor binding. *BMC Genomics* **18**, 1–16 (2017).
284. Morgunova, E. & Taipale, J. Structural perspective of cooperative transcription factor binding. *Curr. Opin. Struct. Biol.* **47**, 1–8 (2017).
285. Soshilov, A. A., Motta, S., Bonati, L. & Denison, M. S. Transitional states in ligand-dependent transformation of the aryl hydrocarbon receptor into its DNA-binding form. *Int. J. Mol. Sci.* **21**, (2020).
286. Hankinson, O. The Aryl Hydrocarbon receptor complex. *Annu. Rev. Pharmacol. Toxicol.* **35**, 307–340 (1995).
287. Soshilov, A. & Denison, M. S. Role of the Per/Arnt/Sim domains in ligand-dependent transformation of the aryl hydrocarbon receptor. *J. Biol. Chem.* **283**, 32995–33005 (2008).
288. Cohen-Khait, R., Dym, O., Hamer-Rogotner, S. & Schreiber, G. Promiscuous Protein Binding as a Function of Protein Stability. *Structure* **25**, 1867-1874.e3 (2017).
289. Woringer, M. & Darzacq, X. Protein motion in the nucleus: From anomalous diffusion to weak interactions. *arXiv* **0**, 945–956 (2018).
290. Woolf, P. J. & Linderman, J. J. Self organization of membrane proteins via dimerization. *Biophys. Chem.* **104**, 217–227 (2003).
291. Roberts, B. J. & Whitelaw, M. L. Degradation of the basic helix-loop-helix/Per-

- ARNT-Sim homology domain dioxin receptor via the ubiquitin/proteasome pathway. *J. Biol. Chem.* **274**, 36351–36356 (1999).
292. Davarinos, N. A. & Pollenz, R. S. Aryl hydrocarbon receptor imported into the nucleus following ligand binding is rapidly degraded via the cytoplasmic proteasome following nuclear export. *J. Biol. Chem.* **274**, 28708–28715 (1999).
293. Eltom, S. E., Zhang, L. & Jefcoate, C. R. Regulation of cytochrome P-450 (CYP) 1B1 in mouse hepa-1 variant cell lines: A possible role for aryl hydrocarbon receptor nuclear translocator (ARNT) as a suppressor of CYP1B1 gene expression. *Mol. Pharmacol.* **55**, 594–604 (1999).
294. Gordân, R., Hartemink, A. J. & Bulyk, M. L. Distinguishing direct versus indirect transcription factor-DNA interactions. *Genome Res.* **19**, 2090–2100 (2009).
295. Yang, S. Y., Ahmed, S., Satheesh, S. V. & Matthews, J. Genome-wide mapping and analysis of aryl hydrocarbon receptor (AHR)- and aryl hydrocarbon receptor repressor (AHRR)-binding sites in human breast cancer cells. *Arch. Toxicol.* **92**, 225–240 (2018).
296. Dionyssiou, M. G. *et al.* Krüppel-like factor 6 (KLF6) promotes cell proliferation in skeletal myoblasts in response to TGF β / Smad3 signaling. **6**, 1–9 (2013).

Appendices

APPENDICES I: OLIGONUCLEOTIDES

Primers encoding for wild-type and deletion constructs were designed as followed

KLF6 WT F 5'ACGTGCTCCCATGTGCAGCATCTTCCAGGAGCTCCAGATC'3

KLF6 WT R 5'CCCGCTCGAGTCAGAGGTGCCTCTTCATGTGCAGGGCCAGG'3

KLF6 Δ 15 F

5'TGCACGAGACCGGCTACTTCTCGGCGCTGCCGTCTCTGGAGGAGTA'3

KLF6 F Δ 27 F

5'TGGAGGAGTACTGGCAACAGACCTGCCTAGAGCTGGAACGTTACCT3'

KLF6 F Δ 37 F

5'GAGCTGGAACGTTACCTCCAGAGCGAGCCCTGCTATGTTTCAGCCTCAGAA
AT'3

AhR WT F 5'ACAGCAGCAGCGCCAACATCACCTACGCCAGTCGCAAGCGG 3'

AhR WT R

5'CCCGCTCGAGTTAAGCGTAATCTGGAACATCGTATGGGTACAGGAATCCAC
TGGATGTCAAAT 3'

AhR R 1-348 R

5'5'CCCGCTCGAGTTAAGCGTAATCTGGAACATCGTATGGGTACATGCCACTT
TCTCCAGTCT 3'

AhR R 1-386 R

5'CCCGCTCGAGTTAAGCGTAATCTGGAACATCGTATGGGTATAGTGGTCTCT
GAGTTACAA 3'

AhR R 1-424 R

5'CCCGCTCGAGTTAAGCGTAATCTGGAACATCGTATGGGTACATAGATTTCGT
TGAAAAAGT 3'

AhR R 1-490 R

5'CCCGCTCGAGTTAAGCGTAATCTGGAACATCGTATGGGTACATTATGGCAG
GAAAAGGGT 3'

AhR R 1-540 R

5'CCCGCTCGAGTTAAGCGTAATCTGGAACATCGTATGGGTAGTACAAGTCAC
TGTTTTT 3'

AhR R 1-599 R

5'CCCGCTCGAGTTAAGCGTAATCTGGAACATCGTATGGGTAATAATCTGAAG
GTATGAAGGG 3'

AhR R 1-604 R

5'CCCGCTCGAGTTAAGCGTAATCTGGAACATCGTATGGGTAGGACTGTTGCT
GTTGATAATC 3'

AhR R 1-610 R

5'CCCGCTCGAGTTAAGCGTAATCTGGAACATCGTATGGGTAGCTTGAGTTCA
GAGCCAAGGA 3'

AhR R 1-620 R

5'CCCGCTCGAGTTAAGCGTAATCTGGAACATCGTATGGGTATTCTAGATGTA
GGTGTTCCT 3'

AhR R 1-640 R

5'CCCGCTCGAGTTAAGCGTAATCTGGAACATCGTATGGGTACTGACACAGCT
GTTGCTGTG 3'

AhR Δb F 5'AGAGACCGACTTAATACAG 3'

AhR Δb R 5'GATTGGCTTTACTGTTTT 3'

hAhR ΔHLH/PAS F 5'CTTCTTACAAAAACAACCGATG 3'

hAhR ΔHLH/PAS R 5'AGCTGGGATTGGCTTTAC 3'

hAhR Δ 275-327 F 5'AGCTGATATGCTTTATTGTGC 3'

hAhR Δ 275-327 R 5'GGTGGCTGAAGTGGAGTA 3'

AhR Δ 40-78 F 5'GCCAAGAGCTTCTTTGATG 3'

AhR Δ 40-78 R 5'ATGCCGCTTGGGAAGGATT 3'

AhR Δ40-274 F 5'CCATCCATACTTGAAATCCG 3'

AhR Δ40-274 R 5'ATGCCGCTTGGGAAGGATT 3'

AhR Δ40-223 F 5'CTAAGGTGTCTGCTGGATAATTC 3'

AhR Δ40-223 R 5'ATGCCGCTTGGGAAGGATT 3'

AhR Δ121-223 F 5'CTAAGGTGTCTGCTGGATAATTC 3'

AhR Δ121-223 R 5' CTGTAATAAGAATTCTCCTTCTTG 3'

AhR Δ160-223 F 5' CTAAGGTGTCTGCTGGATAATTC 3'

AhR Δ160-223 R 5' ATATACACTCTGATGTATGACATC 3'

Primers designed with alanine substitutions were designed as following:

KLF6 29-30 to ala F 5' GCCGTCTCTGGCGGCGTACTGGCAAC 3'

KLF6 29-30 to ala R 5' AGCGCCGAGAAGTAGCCG 3'

KLF6 33-34 to ala F

5' GGAGTACTGGGCAGCGACCTGCCTAGAGCTGGAACGTTAC 3'

KLF6 33-34 to ala R 5' TCCAGAGACGGCAGCGCC 3'

AhR 600-601 to ala F

5' AAGTAAGTCTCCCTTCATACCTTCAGATTATGCAGCGCAACAGTCCTTGGCT
3'

AhR 600-601 to ala R 5'-

AGCCAAGGACTGTTGCGCTGCATAATCTGAAGGTATGAAGGGAGACTTACTT-
3'

AhR 602-603 to ala F 5'-

TCATACCTTCAGATTATCAACAGGCAGCGTCCTTGGCTCTGAACTCAAGCT-3'

AhR 602-603 to ala R 5'-

AGCTTGAGTTCAGAGCCAAGGACGCTGCCTGTTGATAATCTGAAGGTATGA-3'

AHR 600-603 to ala F

5' AAGTAAGTCTCCCTTCATACCTTCAGATTATGCAGCGGCAGCGTCCTTGGCT
3'

AHR 600-603 R

5' AGCCAAGGACGCTGCCGCTGCATAATCTGAAGGTATGAAGGGAGACTTACT
T 3'

hAhR 13-17 to Ala F 5' GGCGGCGCCGGTGCAGAAAACAGTAAA 3'

hAhR 13-17 to Ala R 5' GCCGCGGCACTGGCGTAGGTGATGTT 3'

hAhR 37-39 to Ala F 5' GGCTAGAGACCGACTTAATACAG 3'

hAhR 37-39 to Ala R 5' GCCGCGGAAGGATTTGACTTGATTC 3'

Oligonucleotides for gel shift assays

NC XRE WT F 5' GTCCCAGCAAGTCACTGGGAGGGAGGGAGGGAGGGGGAG
3'

NC XRE WT R 5' CTCCCCCTCCCTCCCTCCCTCCCAGTGACTTGCTGGGAC

NC XRE M4.1 F 5'GTCCCAGCAAGTCACTGGGAGAGAGGGAGGGAGGGGGAG

NC XRE M4.1 R 5'CTCCCCCTCCCTCCCTCTCTCCCAGTGACTTGCTGGGAC

NC XRE M4.2 F5'GTCCCAGCAAGTCACTGGGAGGAAGGGAGGGAGGGGGAG

NC XRE M4.2 R 5'CTCCCCCTCCCTCCCTTCCCTCCCAGTGACTTGCTGGGAC

NC XRE M4.3 F GTCCCAGCAAGTCACTGGGAGGGGGGGAGGGAGGGGGAG

NC XRE M4.3 R 5'CTCCCCCTCCCTCCCCCCTCCCAGTGACTTGCTGGGAC

APPENDICES II: HUMAN AHR BINDING PARTNERS

Interactor	Organism	Aliases	Description	Evidence
ARNT	H. sapiens	HIF1B, TANGO, bHLHe2, HIF1BETA, HIF-1beta, HIF1-beta, HIF-1-beta	aryl hydrocarbon receptor nuclear translocator	29
AIP	H. sapiens	ARA9, XAP2, XAP-2, FKBP16, FKBP37, SMTPHN	aryl hydrocarbon receptor interacting protein	10
ESR1	H. sapiens	ER, ESR, Era, ESRA, ESTRR, NR3A1, RP1-130E4.1	estrogen receptor 1	9
RB1	H. sapiens	RB, pRb, OSRC, pp110, p105-Rb, PPP1R130, RP11-174I10.1	retinoblastoma 1	9
HSP90AA1	H. sapiens	EL52, LAP2, HSPN, Hsp90, HSPC1, LAP-2, HSP86, HSPCA, Hsp89, HSP90N, ... more	heat shock protein 90kDa alpha (cytosolic), class A member 1	6
NCOA1	H. sapiens	SRC1, RIP160, KAT13A, bHLHe74, F-SRC-1, bHLHe42	nuclear receptor coactivator 1	4
NCOR2	H. sapiens	TRAC, SMRT, TRAC1, CTG26, SMRTE, TNRC14, N-CoR2, TRAC-1, SMAP270, SMRTE-tau	nuclear receptor corepressor 2	4
CCNT1	H. sapiens	CCNT, HIVE1, CYCT1	cyclin T1	3
CTNNB1	H. sapiens	CTNNB, MRD19, armadillo, OK/SW-cl.35	catenin (cadherin-associated protein), beta 1, 88kDa	3
CUL4B	H. sapiens	SFM2, MRXSC, MRXS15, CUL-4B, MRXHF2	cullin 4B	3
DDB1	H. sapiens	XPE, XPCE, DDBA, XAP1, XPE-BF, UV-DDB1	damage-specific DNA binding protein 1, 127kDa	3
EP300	H. sapiens	p300, RSTS2, KAT3B, RP1-85F18.1	E1A binding protein p300 nuclear receptor	3
NRIP1	H. sapiens	RIP140	interacting protein 1	3
ARNTL	H. sapiens	TIC, JAP3, MOP3, BMAL1, PASD3, bHLHe5, BMAL1c	aryl hydrocarbon receptor nuclear translocator-like	2
BRCA1	H. sapiens	IRIS, PSCP, FANCS, RNF53, BRCC1, PNCA4, BRCAI, PPP1R53, BROVCA1	breast cancer 1, early onset	2
NCOA2	H. sapiens	SRC2, TIF2, GRIP1, KAT13C, NCoA-2, bHLHe75	nuclear receptor coactivator 2	2
NEDD8	H. sapiens	NEDD-8	neural precursor cell expressed, developmentally down-regulated 8	2

NR2F1	H. sapiens	EAR3, BBOAS, NR2F2, SVP44, EAR-3, BBSOAS, ERBAL3, TFCOUP1, TCFCOUP1, COUP- TFI	nuclear receptor subfamily 2, group F, member 1 v-rel avian reticuloendotheliosis viral oncogene homolog A	2
RELA	H. sapiens	p65, NFKB3 CHIP, UBOX1, SCAR16, NY-CO- 7, SDCCAG7, HSPABP2, LA16c- 313D11.6	STIP1 homology and U- box containing protein 1, E3 ubiquitin protein ligase	2
STUB1	H. sapiens	SAZD, UTP13	transducin (beta)-like 3	2
TBL3	H. sapiens	HDL4, SCA17, TFIID, GTF2D, GTF2D1, RP1-191N21.3	TATA box binding protein	2
TBP	H. sapiens	emb, CRM1, exp1	exportin 1	2
XPO1	H. sapiens	bHLHe76 KD, TFM, AIS, SBMA, DHTR, NR3C4, HYS1, SMAX1, HUMARA, RP11-383C12.1	aryl hydrocarbon receptor	1
AHR	H. sapiens	WEDAS, bHLHe1	androgen receptor aryl-hydrocarbon receptor nuclear translocator 2	1
AR	H. sapiens	BAG-5	BCL2-associated athanogene 5	1
ARNT2	H. sapiens	JAB1, CSN5, SGN5, MOV-34	COP9 signalosome subunit 5	1
BAG5	H. sapiens	CP1B, GLC3A, CYPIB1, P4501B1 DAP-3, MRPS29, bMRP-10, MRP-S29	cytochrome P450, family 1, subfamily B, polypeptide 1	1
CYP1B1	H. sapiens	RBP3, RBBP3, RBAP1, E2F-1	death associated protein 3	1
DAP3	H. sapiens	PHD3, HIFPH3, HIFP4H3 DRAL, AAG11, SLIM3, FHL-2, SLIM-3	E2F transcription factor 1 egl-9 family hypoxia- inducible factor 3	1
E2F1	H. sapiens	HNF3A, TCF3A	four and a half LIM domains 2	1
EGLN3	H. sapiens	G13	forkhead box A1 guanine nucleotide binding protein (G protein), alpha 13	1
FHL2	H. sapiens	BTF4, RAP74, TF2F1, TFIIF	general transcription factor IIF, polypeptide 1, 74kDa	1
FOXA1	H. sapiens	BTF4, RAP30, TFIIF, TF2F2 LAP1, NIP71, LAP-1, HSC70, HSC71, HSC54, HSP71, HSP73, HEL-33, HSPA10, ... more	general transcription factor IIF, polypeptide 2, 30kDa	1
GNA13	H. sapiens		heat shock 70kDa protein 8	1
GTF2F1	H. sapiens			
GTF2F2	H. sapiens			
HSPA8	H. sapiens			

ICAM1	H. sapiens	BB2, CD54, P3.58 ND1, NS-1, NS1BP, FLARA3, KLHL39, NS1-BP, HSPC068,	intercellular adhesion molecule 1	1
IVNS1ABP	H. sapiens	RP4-635A23.5	influenza virus NS1A binding protein	1
KIAA1429	H. sapiens	MSTP054, fSAP121	KIAA1429	1
KIAA1683	H. sapiens	-	KIAA1683	1
KIF14	H. sapiens	-	kinesin family member 14 v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog	1
MAF	H. sapiens	CCA4, c-MAF PBP, CRSP1, RB18A, TRIP2, PPARBP, TRAP220, PPARGBP,	mediator complex subunit 1	1
MED1	H. sapiens	DRIP230, DRIP205, CRSP200 MP25, MAGP2, MAGP-2,	microfibrillar associated protein 5	1
MFAP5	H. sapiens	MFAP-5 TLDC4, ESNA1, ERAP140, NCOA7-AS, Nbla00052, Nbla10993, dJ187J11.3, RP1-	nuclear receptor coactivator 7	1
NCOA7	H. sapiens	187J11.1	nuclear factor, erythroid 2-like 2	1
NFE2L2	H. sapiens	NRF2 ORF1ab, R1AB_SARS2, ORF1ab-nsp16, PRO_0000449633, SARS-CoV2 nsp16, SARS-CoV-2 nsp16,	2-O-methyltransferase; Non-structural protein 16	1
NSP16	SARS-CoV-2	GU280_gp01_nsp16, 2-O- methyltransferase ORF1ab, R1AB_SARS2, ORF1ab-nsp7, SARS-CoV2 nsp7, PRO_0000449625, GU280_gp01_nsp7, SARS-	Non-structural protein 7	1
NSP7	SARS-CoV-2	CoV-2 nsp7 X2, 3b, ns3b, ORF3b, NS3B_CVHSA, SARS-CoV-3b, PRO_0000106132, SARS-CoV	Non-structural protein 3b; ns3b; Accessory protein 3b; Protein X2	1
ORF3B	SARS-CoV	ORF3b, SARS-CoV ... more	proteasome (prosome, macropain) 26S subunit, ATPase, 2	1
PSMC2	H. sapiens	S7, MSS1, Nbla10058	proteasome (prosome, macropain) 26S subunit, ATPase, 3	1
PSMC3	H. sapiens	TBP1	proteasome (prosome, macropain) 26S subunit, ATPase, 4	1
PSMC4	H. sapiens	S6, TBP7, RPT3, TBP-7, MIP224		1

PSMD3	H. sapiens	S3, P58, RPN3, TSTA2	proteasome (prosome, macropain) 26S subunit, non-ATPase, 3	1
PSMD4	H. sapiens	AF, S5A, ASF, AF-1, MCB1, Rpn10, pUB-R5, RP11-126K1.1	proteasome (prosome, macropain) 26S subunit, non-ATPase, 4	1
PSMD5	H. sapiens	S5B, RP11-27I1.5	proteasome (prosome, macropain) 26S subunit, non-ATPase, 5	1
PTGES3	H. sapiens	P23, TEBP, cPGES	prostaglandin E synthase 3 (cytosolic)	1
RAB14	H. sapiens	FBP, RAB-14, RP11-165P4.4	RAB14, member RAS oncogene family	1
RBX1	H. sapiens	ROC1, RNF75, BA554C12.1, RP11-554C12.1	ring-box 1, E3 ubiquitin protein ligase	1
RNF123	H. sapiens	KPC1, FP1477	ring finger protein 123	1
SIRT6	H. sapiens	SIR2L6	sirtuin 6	1
SLC7A5	H. sapiens	E16, LAT1, CD98, MPE16, 4F2LC, hLAT1, D16S469E	solute carrier family 7 (amino acid transporter light chain, L system), member 5	1
SMAD4	H. sapiens	JIP, DPC4, MYHRS, MADH4	SMAD family member 4 SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	1
SMARCA4	H. sapiens	BRG1, SWI2, SNF2, RTPS2, MRD16, SNF2L4, SNF2LB, BAF190, hSNF2b, BAF190A	SMARCA4	1
SP1	H. sapiens	-	Sp1 transcription factor suppressor of variegation	1
SUV39H1	H. sapiens	MG44, KMT1A, SUV39H, H3-K9-HMTase 1	3-9 homolog 1 (Drosophila)	1
TAF4	H. sapiens	TAF4A, TAF2C, TAF2C1, TAFII135, TAFII130, RP5-1107C24.1	TAF4 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 135kDa	1
TAF6	H. sapiens	TAF2E, TAFII80, TAFII70, TAFII85, TAFII-70, TAFII-80, MGC:8964, TAF(II)70, TAF(II)80	TAF6 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 80kDa	1
TAF9	H. sapiens	TAF2G, AD-004, TAFII31, TAFII32, MGC:5067, TAFII-31, TAFII-32, TAFIID32, STAF31/32	TAF9 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 32kDa	1
TRIM25	H. sapiens	EFP, Z147, ZNF147, RNF147	tripartite motif containing 25	1
TRIM66	H. sapiens	TIF1D, C11orf29, TIF1DELTA	tripartite motif containing 66	1

TSPYL6	H. sapiens	-	TSPY-like 6 ubiquitin carboxyl- terminal esterase L3	1
UCHL3	H. sapiens	UCH-L3, RP11-173B14.3 HEL113, CTRCT30, RP11-	(ubiquitin thiolesterase)	1
VIM	H. sapiens	124N14.1	vimentin	1
FASN	H. sapiens	FAS, OA-519, SDR27X1	fatty acid synthase	1

APPENDICES III: RNA SEQ ARNT FLOX UPREGULATED GENES

a	base	log2Fold			pval		gene_na	
	Mean	Change	lfcSE	stat	ue	padj	me	gene_type
ENSMUSG000	1108		0.90	10.1	2.15	2.69		
00032315.6	8.84	9.223818	4992	9216	E-24	E-21	Cyp1a1	protein_coding
ENSMUSG000	17.31		1.27	6.57	4.82	1.23		
00026247.13	118	8.402286	7664	6288	E-11	E-08	Ecel1	protein_coding
ENSMUSG000	3181.		0.38	16.4	5.96	4.85		
00037411.10	107	6.344857	5222	7064	E-61	E-57	Serpine1	protein_coding
ENSMUSG000	21.68		0.77	6.85	6.94	2.09	Gm2691	
00097834.2	936	5.298798	2553	8813	E-12	E-09	1	lncRNA
ENSMUSG000	3160.		0.43	11.1	6.74	1.1E		
00034640.9	623	4.830021	2976	554	E-29	-25	Tiparp	protein_coding
ENSMUSG000	928.4		0.52	7.23	4.64	1.62		
00042622.14	722	3.777569	21	533	E-13	E-10	Maff	protein_coding
ENSMUSG000	10.84		0.77	4.66	3.06	0.00		
00046607.6	904	3.599911	1424	6579	E-06	03	Hrk	protein_coding
ENSMUSG000	10.51		0.79	4.44	8.59	0.00		
00005952.15	355	3.546576	7014	9832	E-06	0746	Trpv1	protein_coding
ENSMUSG000	66.03		0.37	9.41	4.53	3.68	4931440	transcribed_process
00074580.11	375	3.537268	5523	9583	E-21	E-18	P22Rik	ed_pseudogene
ENSMUSG000	28.72		1.05	3.33	0.00	0.03	Gm1080	
00075014.1	578	3.528123	8898	1882	0863	2248	0	protein_coding
ENSMUSG000	740.5		0.22	15.7	4.29	1.75		
00015312.9	407	3.526574	3488	7968	E-56	E-52	Gadd45b	protein_coding
ENSMUSG000	113.9		0.35	9.62	6.31	5.7E		
00074345.4	554	3.449859	8452	4334	E-22	-19	Tnfaip8l3	protein_coding
ENSMUSG000	111.8		0.95	3.24	0.00	0.04		
00003545.3	039	3.090603	2057	6237	1169	012	Fosb	protein_coding
ENSMUSG000	15.51		0.62	4.72	2.3E	0.00	Gm1611	
00090217.1	375	2.964137	7351	4847	-06	0236	6	lncRNA
ENSMUSG000	15.54		0.56	5.12	2.99	3.6E		
00025094.8	35	2.881364	2294	4296	E-07	-05	Slc18a2	protein_coding
ENSMUSG000	2498.		0.21	13.0	3.3E	7.66		
00021453.2	031	2.851535	7676	9989	-39	E-36	Gadd45g	protein_coding
ENSMUSG000	116.0		0.31	8.89	6.04	3.64		
00042417.5	617	2.791604	3971	1291	E-19	E-16	Ccno	protein_coding
ENSMUSG000	8.612		0.69	3.90	9.24	0.00	9530046	
00086047.1	015	2.719219	5507	9694	E-05	5329	B11Rik	lncRNA
ENSMUSG000	3749		0.16	16.1	9.02	4.89		
00032310.4	9.37	2.708797	758	6421	E-59	E-55	Cyp1a2	protein_coding
ENSMUSG000	260.5		0.27	9.68	3.43	3.28		
00043421.8	201	2.68417	7092	6945	E-22	E-19	Hilpda	protein_coding
ENSMUSG000	31.22		0.41	6.30	2.97	6.28		
00001103.7	502	2.643698	9621	0204	E-10	E-08	Sebox	protein_coding
ENSMUSG000	8.541		0.75	3.42	0.00	0.02	Gm2941	
00100320.1	359	2.602984	9516	716	061	431	7	lncRNA
ENSMUSG000	62.26		0.30	8.29	1.11	5.48	Tnfrsf10	
00022074.6	094	2.520198	3923	2232	E-16	E-14	b	protein_coding
ENSMUSG000	157.3		0.31	7.74	9.2E	3.84	Gm4400	
00107624.1	604	2.462132	7701	9852	-15	E-12	5	TEC

ENSMUSG000	9.615		0.71	3.40	0.00	0.02	A230103	
00087026.7	406	2.447919	8247	8186	0654	5875	J11Rik	lncRNA
ENSMUSG000	145.9		0.54	4.44	8.99	0.00		
00030827.5	609	2.431685	7652	0198	E-06	0769	Fgf21	protein_coding
ENSMUSG000	546.1		0.44	5.44	5.13	7.25		
00059970.7	996	2.428746	5907	6757	E-08	E-06	Hspa2	protein_coding
ENSMUSG000	811.5		0.20	11.9	4.41	7.97		
00032500.10	329	2.424854	2373	8211	E-33	E-30	Dclk3	protein_coding
ENSMUSG000	9.669		0.69	3.45	0.00	0.02	Gm1536	unprocessed_pseudogene
00081488.1	577	2.398546	3556	8329	0544	2265	8	
ENSMUSG000	61.14		0.32	7.33	2.18	8.07	1200007	
00087684.1	418	2.38496	5055	7097	E-13	E-11	C13Rik	lncRNA
ENSMUSG000	29.16		0.45	5.17	2.27	2.82		
00050957.4	151	2.356162	5242	5629	E-07	E-05	Insl6	protein_coding
ENSMUSG000	59.85		0.49	4.71	2.36	0.00		
00038295.14	246	2.320273	1618	967	E-06	024	Atg9b	protein_coding
ENSMUSG000	8.412		0.77	2.97	0.00	0.07		
00090171.1	931	2.306096	6304	0611	2972	7084	Ugt1a2	protein_coding
ENSMUSG000	116.2		0.33	6.82	8.98	2.56		
00038155.11	549	2.276389	3688	1917	E-12	E-09	Gstp2	protein_coding
ENSMUSG000	1573.		0.23	9.51	1.88	1.61		
00021281.15	795	2.249683	6518	1673	E-21	E-18	Tnfaip2	protein_coding
ENSMUSG000	111.0		0.28	7.83	4.8E	2.11		
00024087.4	951	2.244002	6516	2027	-15	E-12	Cyp1b1	protein_coding
ENSMUSG000	8.489		0.74	2.95	0.00	0.07		
00050737.13	933	2.193198	2359	4365	3133	9363	Ptges	protein_coding
ENSMUSG000	22.03		0.46	4.65	3.17	0.00		
00074472.10	207	2.189382	9874	9513	E-06	0308	Zfp872	protein_coding
ENSMUSG000	40.69		0.47	4.57	4.84	0.00	Tmem26	
00032313.11	253	2.188987	882	1624	E-06	0447	6	protein_coding
ENSMUSG000	20.78		0.47	4.58	4.59	0.00	Gm4398	
00108238.1	188	2.155356	0327	2679	E-06	0427	4	TEC
ENSMUSG000	324.4		0.21	10.0	1.18	1.38		
00022474.15	322	2.149102	4376	249	E-23	E-20	Pmm1	protein_coding
ENSMUSG000	336.2		0.50	4.20	2.64	0.00		
00025582.4	087	2.102025	0205	2324	E-05	1918	Nptx1	protein_coding
ENSMUSG000	21.33		0.50	4.14	3.44	0.00		
00074579.14	951	2.092964	5251	2426	E-05	2348	Lekr1	protein_coding
ENSMUSG000	226.2		0.26	7.76	8.24	3.53		
00003541.6	122	2.06902	6497	3764	E-15	E-12	Ier3	protein_coding
ENSMUSG000	18.44		0.60	3.42	0.00	0.02	AY51293	
00066158.7	002	2.0681	3095	9145	0605	4193	1	lncRNA
ENSMUSG000	89.39		0.34	5.98	2.23	4.12		
00042035.11	445	2.064633	5243	024	E-09	E-07	Igsf3	protein_coding
ENSMUSG000	53.53		0.36	5.60	2.08	3.14	Gm1940	
00106229.1	911	2.056598	6922	5002	E-08	E-06	9	lncRNA
ENSMUSG000	17.94		0.61	3.32	0.00	0.03		
00033717.6	645	2.055301	7493	8459	0873	2497	Adra2a	protein_coding
ENSMUSG000	1643.		0.23	8.76	1.79	1.01		
00033060.16	35	2.048111	355	9484	E-18	E-15	Lmo7	protein_coding
ENSMUSG000	229.4		0.34	6.00	1.89	3.58		
00026628.13	088	2.044175	0318	6668	E-09	E-07	Atf3	protein_coding
ENSMUSG000	6322.		0.13	15.4	6.04	1.97		
00015839.6	705	2.03157	1372	6428	E-54	E-50	Nfe2l2	protein_coding
ENSMUSG000	25.35		0.50	3.96	7.21	0.00		
00072919.4	631	2.014453	7505	9324	E-05	4406	Noxred1	protein_coding

ENSMUSG000	23.26		0.44	4.50	6.54	0.00	Gm4399	
00107835.1	118	1.989501	1323	8038	E-06	0585	9	TEC
ENSMUSG000	1575.		0.13	14.1	1.06	2.87		
00026923.15	2	1.966704	8598	8994	E-45	E-42	Notch1	protein_coding
ENSMUSG000	20.90		0.50	3.82	0.00	0.00	9630050	
00115275.1	406	1.941523	8132	0899	0133	733	E16Rik	TEC
ENSMUSG000	18.12		0.61	3.16	0.00	0.04		
00044309.8	573	1.940083	2751	6186	1545	9638	Apol7c	protein_coding
ENSMUSG000	42.56		0.34	5.57	2.47	3.65	Gm1699	
00097069.1	123	1.927596	573	5443	E-08	E-06	8	lncRNA
ENSMUSG000	93.26		0.38	4.90	9.19	0.00		
00044337.5	501	1.877504	2527	8161	E-07	0102	Ackr3	protein_coding
ENSMUSG000	2268.		0.47	3.97	7.11	0.00		
00027954.9	26	1.875308	2074	249	E-05	4381	Efna1	protein_coding
ENSMUSG000	274.8		0.19	9.33	9.84	7.62		
00031609.3	102	1.859821	9173	7732	E-21	E-18	Sap30	protein_coding
ENSMUSG000	71.27		0.28	6.56	5.04	1.24		
00060314.13	001	1.857087	2669	983	E-11	E-08	Zfp941	protein_coding
ENSMUSG000	27.44		0.44	4.18	2.88	0.00		
00015957.12	785	1.84785	1798	2568	E-05	2038	Wnt11	protein_coding
ENSMUSG000	82.08		0.34	5.27	1.32	1.73	Fam222	
00041930.7	497	1.820855	5086	6525	E-07	E-05	a	protein_coding
ENSMUSG000	1257.		0.17	10.4	1.81	2.45		
00037621.8	977	1.793765	1982	2998	E-25	E-22	Atoh8	protein_coding
ENSMUSG000	723.9		0.20	8.78	1.56	9.08		
00086040.8	409	1.793048	4103	5012	E-18	E-16	Wipf3	protein_coding
ENSMUSG000	17.45		0.52	3.38	0.00	0.02		
00010044.12	365	1.775268	3716	9755	07	7195	Zmynd10	protein_coding
ENSMUSG000	75.82		0.31	5.65	1.6E	2.48		
00020175.15	739	1.769626	3183	0446	-08	E-06	Rab36	protein_coding
ENSMUSG000	63.48		0.29	5.93	3.01	5.49		
00050914.16	424	1.745298	4252	1306	E-09	E-07	Ankrd37	protein_coding
ENSMUSG000	1157.		0.19	8.98	2.49	1.56		
00054545.17	271	1.737957	3337	9267	E-19	E-16	Ugt1a6a	protein_coding
ENSMUSG000	9946.		0.47	3.67	0.00	0.01		
00038393.14	981	1.736144	2973	0703	0242	1777	Txnip	protein_coding
ENSMUSG000	12.30		0.60	2.85	0.00	0.09	4930532	
00097261.2	842	1.731431	5435	9815	4239	668	G15Rik	lncRNA
ENSMUSG000	9827.		0.23	7.26	3.66	1.32		
00021127.7	009	1.721509	6878	748	E-13	E-10	Zfp361l	protein_coding
ENSMUSG000	80.36		0.38	4.43	9.06	0.00		
00000320.10	815	1.719727	7463	8428	E-06	0772	Alox12	protein_coding
ENSMUSG000	1380		0.25	6.67	2.55	6.79		
00024052.17	5.69	1.686874	2878	0693	E-11	E-09	Lpin2	protein_coding
ENSMUSG000	1480		0.13	12.6	1.36	2.77	Selenbp	
00068874.13	6.17	1.667877	201	3447	E-36	E-33	1	protein_coding
ENSMUSG000	14.45		0.53	3.09	0.00	0.05		
00041794.13	069	1.654829	5432	0645	1997	916	Myrip	protein_coding
ENSMUSG000	1214.		0.51	3.19	0.00	0.04		
00023067.14	545	1.648508	6502	1677	1414	647	Cdkn1a	protein_coding
ENSMUSG000	44.57		0.52	3.12	0.00	0.05	E430021	
00105519.1	494	1.643965	6189	4286	1782	4381	H15Rik	TEC
ENSMUSG000	555.8		0.24	6.79	1.09	3.07		polymorphic_pseudogene
00060183.6	645	1.641207	1587	3429	E-11	E-09	Cxcl11	
ENSMUSG000	16.73		0.57	2.87	0.00	0.09		
00026955.13	532	1.641049	0079	8636	3994	3253	Sapcd2	protein_coding

ENSMUSG000	360.1		0.26	6.22	4.77	9.82			
00025902.13	682	1.636258	2788	6529	E-10	E-08	Sox17	protein_coding	
ENSMUSG000	1985.		0.09	17.6	1.29	2.1E			
00027253.15	73	1.631646	2515	3659	E-69	-65	Lrp4	protein_coding	
ENSMUSG000	19.74		0.48	3.36	0.00	0.02	Gm3836		
00103621.1	782	1.621348	1496	7316	0759	9043	6	TEC	
ENSMUSG000	811.1		0.22	7.13	9.82	3.33			
00047109.13	311	1.604479	4939	2961	E-13	E-10	Cldn14	protein_coding	
ENSMUSG000	109.7		0.27	5.74	9.2E	1.5E	Gm1365		
00086813.1	194	1.599254	838	4863	-09	-06	7	lncRNA	
ENSMUSG000	2050.		0.49	3.23	0.00	0.04			
00009633.3	51	1.594866	3567	1304	1232	1673	G0s2	protein_coding	
ENSMUSG000	17.00		0.50	3.14	0.00	0.05	Gm4331		
00105896.1	628	1.582494	2452	9543	1635	1462	1	TEC	
ENSMUSG000	262.7		0.23	6.67	2.45	6.65	Cyp21a2	unprocessed_pseudogene	
00092471.1	924	1.574708	5869	6211	E-11	E-09	-ps		
ENSMUSG000	34.38		0.43	3.57	0.00	0.01			
00022388.14	03	1.569379	9143	3727	0352	5848	Ttll8	protein_coding	
ENSMUSG000	106.4		0.38	4.12	3.79	0.00			
00021789.14	02	1.568823	0775	0079	E-05	2535	Sftpa1	protein_coding	
ENSMUSG000	15.29		0.50	3.06	0.00	0.06	Gm3663		
00104214.1	898	1.558859	8315	6719	2164	2624	8	TEC	
ENSMUSG000	4626.		0.21	7.23	4.68	1.62			
00035828.11	893	1.539183	2766	415	E-13	E-10	Pim3	protein_coding	
ENSMUSG000	76.97		0.39	3.92	8.8E	0.00			
00022468.12	479	1.534523	1323	1377	-05	5113	Endou	protein_coding	
ENSMUSG000	572.2		0.17	8.58	9E-	4.57			
00060470.7	969	1.523049	7387	6043	18	E-15	Adgrg3	protein_coding	
ENSMUSG000	21.14		0.50	2.99	0.00	0.07			
00050671.12	939	1.521191	7093	9828	2701	2371	lsm2	protein_coding	
ENSMUSG000	166.9		0.26	5.78	7.19	1.22			
00047875.6	812	1.514647	1762	6343	E-09	E-06	Gpr157	protein_coding	
ENSMUSG000	1000.		0.28	5.27	1.3E	1.73			
00040093.16	019	1.511028	6265	8432	-07	E-05	Bmf	protein_coding	
ENSMUSG000	856.5		0.16	9.14	6.09	4.12			
00038957.13	312	1.495091	3526	2807	E-20	E-17	Edc3	protein_coding	
ENSMUSG000	68.21		0.34	4.30	1.7E	0.00	Gm4554		
00110397.1	974	1.486173	5492	1613	-05	1326	0	TEC	
ENSMUSG000	21.00		0.45	3.24	0.00	0.04			
00019301.5	13	1.481697	6885	3039	1183	0486	Hsd17b1	protein_coding	
ENSMUSG000	40.67		0.43	3.43	0.00	0.02	F830115		
00104576.1	078	1.480456	1296	2578	0598	4006	B05Rik	TEC	
ENSMUSG000	55.30		0.36	4.02	5.58	0.00	5330406		
00104030.1	777	1.471884	5243	9876	E-05	3531	M23Rik	TEC	
ENSMUSG000	123.6		0.29	4.95	7.37	8.33	Gm3012		
00112433.1	491	1.468817	6657	1223	E-07	E-05	2	lncRNA	
ENSMUSG000	20.89		0.50	2.88	0.00	0.09	Gm2674		
00097783.2	685	1.468463	8496	7858	3879	1365	7	lncRNA	
ENSMUSG000	248.2		0.34	4.24	2.18	0.00			
00006777.7	093	1.460366	4005	5182	E-05	1637	Krt23	protein_coding	
ENSMUSG000	72.24		0.30	4.81	1.44	0.00			
00002020.15	956	1.457621	2475	8986	E-06	0156	Ltbp2	protein_coding	
ENSMUSG000	15.17		0.48	3.00	0.00	0.07	Gm4442		
00108206.1	815	1.455486	5133	0181	2698	2371	7	lncRNA	
ENSMUSG000	1399.		0.14	9.70	2.82	2.87			
00029650.10	402	1.448474	9222	6835	E-22	E-19	Slc46a3	protein_coding	

ENSMUSG000	1615.		0.22	6.44	1.18	2.77			
00003134.10	008	1.429245	1851	2358	E-10	E-08	Tbc1d8	protein_coding	
ENSMUSG000	432.9		0.22	6.20	5.62	1.13			
00036492.12	416	1.421207	9198	0795	E-10	E-07	Rnf39	protein_coding	
ENSMUSG000	17.91		0.49	2.86	0.00	0.09	Gm2021		
00110218.1	748	1.414058	2933	8662	4122	4949	9	protein_coding	
ENSMUSG000	72.28		0.26	5.38	7.06	9.81	2500002		
00096917.2	326	1.40974	1563	9683	E-08	E-06	B13Rik	lncRNA	
ENSMUSG000	24.85		0.41	3.41	0.00	0.02			
00031896.7	141	1.407229	2587	0747	0648	5696	Ctrl	protein_coding	
ENSMUSG000	176.1		0.32	4.27	1.91	0.00			
00008153.12	476	1.397626	6961	4596	E-05	1483	Clstn3	protein_coding	
ENSMUSG000	1018.		0.12	10.6	1.99	2.95			
00037674.15	97	1.377213	9467	3759	E-26	E-23	Rfx7	protein_coding	
ENSMUSG000	856.4		0.20	6.62	3.37	8.85			
00044469.8	76	1.373663	7212	9266	E-11	E-09	Tnfaip81	protein_coding	
ENSMUSG000	495.9		0.23	5.89	3.71	6.62			
00041688.16	414	1.370932	2499	6497	E-09	E-07	Amot	protein_coding	
ENSMUSG000	19.14		0.42	3.23	0.00	0.04			
00045725.5	725	1.368906	3192	4715	1218	1339	Prr15	protein_coding	
ENSMUSG000	20.84		0.41	3.28	0.00	0.03		processed_pseudogene	
00102269.1	428	1.367059	6542	1924	1031	6291	Gm7357		
ENSMUSG000	1336.		0.46	2.91	0.00	0.08			
00053113.3	288	1.366116	8348	6885	3535	6068	Socs3	protein_coding	
ENSMUSG000	1177		0.26	5.21	1.86	2.34			
00078650.2	7.34	1.362803	1424	2992	E-07	E-05	G6pc	protein_coding	
ENSMUSG000	2540.		0.13	9.77	1.38	1.5E			
00018387.12	732	1.361912	9263	9416	E-22	-19	Shroom1	protein_coding	
ENSMUSG000	50.27		0.33	4.06	4.72	0.00	Gm1175		
00087404.1	628	1.356321	3325	9063	E-05	3058	2	lncRNA	
ENSMUSG000	236.8		0.24	5.54	2.97	4.35			
00012123.17	435	1.343205	2316	3197	E-08	E-06	Crybg2	protein_coding	
ENSMUSG000	1924.		0.20	6.61	3.65	9.43			
00055866.9	479	1.336825	2015	7448	E-11	E-09	Per2	protein_coding	
ENSMUSG000	51.74		0.33	3.92	8.63	0.00			
00098678.1	398	1.332031	9264	6238	E-05	5066	Mroh6	protein_coding	
ENSMUSG000	1005.		0.42	3.12	0.00	0.05			
00053846.5	833	1.326314	4307	5836	1773	4197	Lipg	protein_coding	
ENSMUSG000	2378.		0.22	5.77	7.86	1.32			
00048756.11	435	1.311889	7304	1504	E-09	E-06	Foxo3	protein_coding	
ENSMUSG000	3172.		0.14	8.64	5.61	2.94			
00027397.14	828	1.295118	9893	0313	E-18	E-15	Slc20a1	protein_coding	
ENSMUSG000	880.4		0.15	8.15	3.5E	1.67			
00008167.14	675	1.292909	8548	4696	-16	E-13	Fbxw9	protein_coding	
ENSMUSG000	62.58		0.27	4.74	2.09	0.00			
00055150.15	571	1.284324	0719	4129	E-06	0216	Zfp78	protein_coding	
ENSMUSG000	46.76		0.34	3.68	0.00	0.01	D630024		
00085772.7	099	1.283036	7816	8836	0225	1135	D03Rik	lncRNA	
ENSMUSG000	905.1		0.44	2.86	0.00	0.09			
00002831.14	557	1.282552	8355	0573	4229	6599	Plin4	protein_coding	
ENSMUSG000	8978.		0.18	7.08	1.37	4.53			
00044783.16	311	1.279164	0484	7428	E-12	E-10	Hjrp	protein_coding	
ENSMUSG000	1798.		0.16	7.56	4.02	1.59			
00033446.7	789	1.269721	7944	0401	E-14	E-11	Lpar6	protein_coding	
ENSMUSG000	47.19		0.41	3.06	0.00	0.06			
00062991.9	315	1.264786	2099	913	2147	2454	Nrg1	protein_coding	

ENSMUSG000	25.59		0.40	3.07	0.00	0.06	Gm3832	
00103632.1	164	1.259143	8915	9233	2075	114	9	TEC
ENSMUSG000	163.0		0.22	5.70	1.14	1.81		
00062861.8	095	1.258311	045	7908	E-08	E-06	Zfp28	protein_coding
ENSMUSG000	33.49		0.37	3.35	0.00	0.02		
00023979.14	007	1.254008	337	8618	0783	9763	Guca1b	protein_coding
ENSMUSG000	1947.		0.37	3.31	0.00	0.03		
00024190.7	732	1.254	8419	3786	092	3864	Dusp1	protein_coding
ENSMUSG000	879.3		0.26	4.71	2.44	0.00		
00047180.8	617	1.251446	5526	308	E-06	0245	Neurl3	protein_coding
ENSMUSG000	306.1		0.23	5.40	6.49	9.09		
00020926.16	749	1.248418	098	488	E-08	E-06	Adam11	protein_coding
ENSMUSG000	1186.		0.17	7.02	2.13	6.91		
00006445.3	61	1.243687	7012	6021	E-12	E-10	Epha2	protein_coding
ENSMUSG000	45.19		0.34	3.58	0.00	0.01	Mrgprb1	unprocessed_pseudogene
00099889.1	625	1.241176	6479	2252	0341	5604	1-ps	
ENSMUSG000	67.12		0.28	4.24	2.15	0.00	Dnase1l	
00024136.14	64	1.218291	6753	8579	E-05	162	2	protein_coding
ENSMUSG000	453.7		0.19	6.32	2.57	5.55	4930503	
00044906.6	88	1.207765	1027	2486	E-10	E-08	L19Rik	protein_coding
ENSMUSG000	635.7		0.17	6.99	2.66	8.48	Ppp1r15	
00040435.12	947	1.19938	147	469	E-12	E-10	a	protein_coding
ENSMUSG000	261.1		0.16	7.36	1.8E	6.81		
00026637.13	635	1.198818	282	2857	-13	E-11	Traf5	protein_coding
ENSMUSG000	74.79		0.27	4.40	1.07	0.00	Gm4572	unprocessed_pseudogene
00110659.1	71	1.19589	1647	2377	E-05	0898	7	
ENSMUSG000	27.46		0.35	3.29	0.00	0.03	3010003	
00085355.2	239	1.181654	8392	7099	0977	5242	L21Rik	lncRNA
ENSMUSG000	123.4		0.25	4.58	4.54	0.00		
00024049.15	189	1.177484	6807	5094	E-06	0427	Myom1	protein_coding
ENSMUSG000	44.80		0.30	3.82	0.00	0.00	Gm1045	
00073000.3	336	1.174353	6707	8911	0129	7119	1	lncRNA
ENSMUSG000	1405		0.40	2.87	0.00	0.09		
00092274.3	7.92	1.162142	3744	8409	3997	3253	Neat1	lncRNA
ENSMUSG000	987.5		0.12	9.05	1.42	9.26		
00032624.16	697	1.157517	7895	051	E-19	E-17	Eml4	protein_coding
ENSMUSG000	45.85		0.33	3.48	0.00	0.02		
00028977.17	78	1.156853	1538	9348	0484	0229	Casz1	protein_coding
ENSMUSG000	45.29		0.30	3.78	0.00	0.00		
00074682.4	323	1.156031	5532	3663	0155	8159	Zcchc3	protein_coding
ENSMUSG000	25.34		0.40	2.85	0.00	0.09		
00020389.19	09	1.153142	352	7706	4267	7052	Cdkl3	protein_coding
ENSMUSG000	96.02		0.27	4.24	2.19	0.00		
00037617.12	696	1.15232	1505	4187	E-05	1637	Spag1	protein_coding
ENSMUSG000	43.00		0.29	3.91	9.17	0.00		
00090124.7	893	1.140096	1466	1587	E-05	5306	Ugt1a7c	protein_coding
ENSMUSG000	82.45		0.33	3.35	0.00	0.02		
00031530.6	235	1.138	8789	9022	0782	9763	Dusp4	protein_coding
ENSMUSG000	69.24		0.34	3.22	0.00	0.04	Gm1044	
00072963.2	235	1.123563	8351	5377	1258	2183	7	lncRNA
ENSMUSG000	508.7		0.16	6.91	4.65	1.43		
00074221.12	284	1.12235	2288	5795	E-12	E-09	Zfp568	protein_coding
ENSMUSG000	51.76		0.30	3.58	0.00	0.01	9130230	
00092626.1	109	1.108443	9568	0616	0343	5614	N09Rik	lncRNA
ENSMUSG000	1076.		0.14	7.65	1.95	7.94		
00041075.8	957	1.106392	4556	3731	E-14	E-12	Fzd7	protein_coding

ENSMUSG000	273.5		0.24	4.56	4.89	0.00	AI46322	
00114828.1	268	1.104479	1721	9238	E-06	045	9	lncRNA
ENSMUSG000	1150.		0.11	9.28	1.62	1.2E		
00046982.11	981	1.104093	8913	4867	E-20	-17	Tshz1	protein_coding
ENSMUSG000	5536.		0.21	5.04	4.58	5.39		
00019232.14	788	1.103679	8842	327	E-07	E-05	Etnppl	protein_coding
ENSMUSG000	353.9		0.20	5.26	1.37	1.79	Gm4331	
00104528.1	596	1.102952	9337	8783	E-07	E-05	4	TEC
ENSMUSG000	35.92		0.35	3.06	0.00	0.06		
00015812.8	981	1.09611	7622	4993	2177	2874	Gnrh1	protein_coding
ENSMUSG000	100.2		0.27	3.96	7.28	0.00	Gm1793	processed_pseudogene
00106664.1	729	1.093012	5534	6884	E-05	4435	6	
ENSMUSG000	2817.		0.26	4.11	3.84	0.00		
00023087.15	896	1.092642	5401	6954	E-05	2559	Noct	protein_coding
ENSMUSG000	1786.		0.15	6.98	2.88	9.02		
00039745.8	52	1.09176	6339	3306	E-12	E-10	Htati2	protein_coding
ENSMUSG000	3710.		0.17	6.32	2.6E	5.55		
00029201.14	043	1.090756	2556	1158	-10	E-08	Ugdh	protein_coding
ENSMUSG000	1429.		0.25	4.22	2.35	0.00		
00022528.8	488	1.088781	7452	9068	E-05	1734	Hes1	protein_coding
ENSMUSG000	64.95		0.32	3.31	0.00	0.03		
00021057.15	81	1.085264	7757	1187	0929	4026	Akap5	protein_coding
ENSMUSG000	54.18		0.25	4.20	2.57	0.00	B230217	
00097785.2	459	1.084766	7752	8571	E-05	1874	O12Rik	lncRNA
ENSMUSG000	74.05		0.27	3.94	8.03	0.00	1300014	
00112545.1	29	1.083197	4687	3393	E-05	4839	J16Rik	lncRNA
ENSMUSG000	1452.		0.23	4.65	3.18	0.00		
00053964.17	976	1.081964	2245	8718	E-06	0308	Lgals4	protein_coding
ENSMUSG000	1491.		0.19	5.62	1.88	2.89		unprocessed_pseudogene
00090555.1	601	1.080755	2223	2394	E-08	E-06	Gm8893	
ENSMUSG000	533.0		0.17	6.09	1.08	2.1E		
00054717.7	457	1.074076	6172	6732	E-09	-07	Hmgb2	protein_coding
ENSMUSG000	82.70		0.28	3.75	0.00	0.00		
00078651.8	363	1.061558	25	7722	0171	8937	Aoc2	protein_coding
ENSMUSG000	625.7		0.13	7.85	3.87	1.75		
00010067.13	13	1.05888	4733	9104	E-15	E-12	Rassf1	protein_coding
ENSMUSG000	453.3		0.18	5.76	8.07	1.34		
00000142.15	475	1.05866	3575	6908	E-09	E-06	Axin2	protein_coding
ENSMUSG000	317.8		0.25	4.15	3.18	0.00	A930033	
00090622.1	566	1.056497	3969	9943	E-05	2222	H14Rik	protein_coding
ENSMUSG000	896.4		0.11	9.16	4.88	3.45		
00029708.9	635	1.052259	4793	6579	E-20	E-17	Gcc1	protein_coding
ENSMUSG000	2813.		0.18	5.57	2.47	3.65		
00024580.8	73	1.050471	8416	5287	E-08	E-06	Grpel2	protein_coding
ENSMUSG000	2053.		0.17	6.12	8.97	1.76		
00040170.13	025	1.049211	125	6767	E-10	E-07	Fmo2	protein_coding
ENSMUSG000	115.2		0.30	3.44	0.00	0.02		
00052415.5	256	1.046956	3932	4702	0572	3184	Tchh	protein_coding
ENSMUSG000	119.8		0.26	3.95	7.53	0.00	Gm3709	
00103183.1	138	1.04528	4026	9004	E-05	4567	0	TEC
ENSMUSG000	1258.		0.18	5.61	1.97	2.99		
00030107.10	911	1.040321	5278	4912	E-08	E-06	Usp18	protein_coding
ENSMUSG000	68.90		0.28	3.66	0.00	0.01		
00039099.7	25	1.039928	3846	3713	0249	2031	Wdr93	protein_coding
ENSMUSG000	2004		0.15	6.84	7.43	2.2E		
00017009.3	0.25	1.038488	1624	9095	E-12	-09	Sdc4	protein_coding

ENSMUSG000	34.58		0.35	2.90	0.00	0.08			processed_pseudog
00113423.1	584	1.031309	5619	0035	3731	9732	Gm3791		ene
ENSMUSG000	163.8		0.20	4.93	8.13	9.11	Tmem12		
00050854.9	158	1.026585	8135	2301	E-07	E-05	5		protein_coding
ENSMUSG000	118.0		0.29	3.46	0.00	0.02	4930404		
00113630.1	32	1.024102	5279	8252	0524	1677	H11Rik		lncRNA
ENSMUSG000	546.3		0.16	6.38	1.69	3.87			
00000282.12	061	1.022538	0095	7079	E-10	E-08	Mnt		protein_coding
ENSMUSG000	430.6		0.21	4.75	1.96	0.00			
00027381.16	539	1.012276	2761	7809	E-06	0207	Bcl2l11		protein_coding
ENSMUSG000	47.68		0.30	3.32	0.00	0.03			
00068263.11	143	1.009887	3376	8826	0872	2497	Efcc1		protein_coding
ENSMUSG000	48.33		0.35	2.84	0.00	0.09			
00065147.1	651	1.007124	3847	6212	4424	9309	Snora31		snoRNA

APPENDICES IV: RNA SEQ ARNT FLOX PATHWAYS ANALYSIS

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Metabolism of xenobiotics by cytochrome P450	0.000004118	0.001248	10.66	132.18
2	Ascorbate and aldarate metabolism	0.0001395	0.005283	14.89	132.18
3	Basal cell carcinoma	0.00003815	0.002890	9.57	97.38
4	Pentose and glucuronate interconversions	0.0003491	0.009615	11.82	94.12
5	Non-small cell lung cancer	0.00004977	0.002513	9.14	90.53
6	Steroid hormone biosynthesis	0.00003004	0.003034	7.90	82.31
7	Chemical carcinogenesis	0.00004280	0.002594	7.48	75.28
8	Endometrial cancer	0.0002770	0.009326	8.66	70.97
9	Breast cancer	0.00001682	0.002549	6.15	67.64
10	Caffeine metabolism	0.05824	0.5190	16.75	47.62
11	Thyroid cancer	0.005878	0.08906	8.15	41.86
12	Hepatocellular carcinoma	0.0003216	0.009745	4.70	37.81
13	Drug metabolism	0.0009851	0.02296	5.29	36.62
14	Colorectal cancer	0.001863	0.04031	5.71	35.89
15	Porphyryn and chlorophyll metabolism	0.007834	0.1130	7.35	35.66
16	Retinol metabolism	0.002159	0.04088	5.52	33.90
17	Gastric cancer	0.0007711	0.01947	4.69	33.62
18	p53 signaling pathway	0.005497	0.09253	5.66	29.46
19	Melanoma	0.005776	0.09211	5.58	28.78
20	FoxO signaling pathway	0.002088	0.04218	4.57	28.19
21	Pathways in cancer	0.00009483	0.004105	3.01	27.84
22	Tryptophan metabolism	0.01208	0.1465	6.28	27.74
23	Ovarian steroidogenesis	0.01916	0.2073	5.29	20.92

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
24	Hippo signaling pathway	0.005223	0.09309	3.79	19.93
25	Small cell lung cancer	0.01346	0.1569	4.37	18.82
26	Apoptosis	0.01350	0.1515	3.56	15.34
27	Cellular senescence	0.01064	0.1344	3.26	14.81
28	Epstein-Barr virus infection	0.008152	0.1123	3.07	14.78
29	Bladder cancer	0.06272	0.5279	4.90	13.58
30	Glioma	0.03887	0.3926	4.02	13.06
31	Pancreatic cancer	0.03887	0.3799	4.02	13.06
32	MAPK signaling pathway	0.009412	0.1240	2.73	12.76
33	Chronic myeloid leukemia	0.04018	0.3805	3.97	12.75
34	Cocaine addiction	0.08252	0.6411	4.19	10.45
35	Amino sugar and nucleotide sugar metabolism	0.08549	0.6476	4.10	10.09
36	Notch signaling pathway	0.08549	0.6318	4.10	10.09
37	Fluid shear stress and atherosclerosis	0.05477	0.5029	2.81	8.17
38	PI3K-Akt signaling pathway	0.02687	0.2808	2.25	8.15
39	Phenylalanine metabolism	0.2056	1.000	4.37	6.91
40	Glycerolipid metabolism	0.1233	0.8691	3.30	6.90
41	Cushing syndrome	0.07462	0.6111	2.53	6.56
42	Mitophagy	0.1300	0.8752	3.19	6.51
43	Wnt signaling pathway	0.07597	0.6058	2.51	6.48
44	MicroRNAs in cancer	0.06245	0.5406	2.15	5.95
45	Amphetamine addiction	0.1470	0.9277	2.96	5.67
46	Maturity onset diabetes of the young	0.2368	1.000	3.72	5.36
47	Phototransduction	0.2368	1.000	3.72	5.36
48	Adipocytokine signaling pathway	0.1573	0.9348	2.83	5.24
49	Cell cycle	0.1243	0.8562	2.45	5.11
50	Prolactin signaling pathway	0.1608	0.9372	2.79	5.10

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
51	Transcriptional misregulation in cancer	0.1102	0.7952	2.20	4.84
52	Circadian rhythm	0.2593	1.000	3.35	4.52
53	Serotonergic synapse	0.1446	0.9528	2.28	4.42
54	Signaling pathways regulating pluripotency of stem cells	0.1564	0.9476	2.20	4.08
55	beta-Alanine metabolism	0.2740	1.000	3.14	4.07
56	Galactose metabolism	0.2740	1.000	3.14	4.07
57	Starch and sucrose metabolism	0.2813	1.000	3.05	3.86
58	Proteoglycans in cancer	0.1447	0.9326	1.98	3.83
59	ErbB signaling pathway	0.2037	1.000	2.39	3.81
60	Prion diseases	0.2884	1.000	2.96	3.68
61	Fructose and mannose metabolism	0.2955	1.000	2.87	3.50
62	Arachidonic acid metabolism	0.2219	1.000	2.26	3.40
63	IL-17 signaling pathway	0.2292	1.000	2.21	3.25
64	Human papillomavirus infection	0.1507	0.9316	1.68	3.17
65	Hepatitis C	0.2136	1.000	1.88	2.91
66	Glycerophospholipid metabolism	0.2513	1.000	2.07	2.86
67	Protein processing in endoplasmic reticulum	0.2214	1.000	1.85	2.79
68	Tyrosine metabolism	0.3299	1.000	2.51	2.79
69	Glycine, serine and threonine metabolism	0.3299	1.000	2.51	2.79
70	Ras signaling pathway	0.2029	1.000	1.73	2.75
71	Melanogenesis	0.2624	1.000	2.01	2.69
72	Influenza A	0.2345	1.000	1.79	2.60
73	Longevity regulating pathway	0.2698	1.000	1.97	2.58
74	NF-kappa B signaling pathway	0.2698	1.000	1.97	2.58
75	HIF-1 signaling pathway	0.2771	1.000	1.93	2.48

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
76	Carbohydrate digestion and absorption	0.3498	1.000	2.34	2.46
77	TNF signaling pathway	0.2992	1.000	1.83	2.20
78	Insulin resistance	0.2992	1.000	1.83	2.20
79	Type II diabetes mellitus	0.3816	1.000	2.09	2.02
80	Cholesterol metabolism	0.3877	1.000	2.05	1.94
81	Linoleic acid metabolism	0.3938	1.000	2.01	1.87
82	Herpes simplex virus 1 infection	0.2628	1.000	1.39	1.86
83	Fanconi anemia pathway	0.3999	1.000	1.97	1.81
84	AMPK signaling pathway	0.3574	1.000	1.60	1.64
85	Inflammatory mediator regulation of TRP channels	0.3610	1.000	1.58	1.61
86	Osteoclast differentiation	0.3646	1.000	1.57	1.58
87	Rap1 signaling pathway	0.3450	1.000	1.44	1.54
88	Estrogen signaling pathway	0.3859	1.000	1.50	1.43
89	Legionellosis	0.4406	1.000	1.73	1.42
90	Insulin signaling pathway	0.4034	1.000	1.45	1.31
91	Glutathione metabolism	0.4732	1.000	1.57	1.17
92	Non-alcoholic fatty liver disease (NAFLD)	0.4445	1.000	1.33	1.08
93	Glycolysis / Gluconeogenesis	0.4889	1.000	1.50	1.07
94	Renal cell carcinoma	0.4940	1.000	1.48	1.04
95	mTOR signaling pathway	0.4545	1.000	1.31	1.03
96	Acute myeloid leukemia	0.4990	1.000	1.46	1.01
97	Adherens junction	0.5139	1.000	1.40	0.93
98	Neuroactive ligand-receptor interaction	0.4566	1.000	1.16	0.91
99	JAK-STAT signaling pathway	0.4871	1.000	1.23	0.88

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
100	Tight junction	0.4967	1.000	1.20	0.84

APPENDICES V: RNA SEQ ARNT iKO UPREGULATED GENES

gene_id	base Mean	log2FoldC hange	lfcSE	stat	pvalu e	padj	gene_na me	gene_type
ENSMUSG00000	6681.		0.241	6.240	4.36E	1.59E		
026819.15	338	1.504117	019	644	-10	-06	Slc25a25	protein_coding
ENSMUSG00000	229.4		0.380	6.180	6.39E	1.86E		
026628.13	088	2.348858	038	596	-10	-06	Atf3	protein_coding
ENSMUSG00000	453.4		0.201	5.497	3.84E	5.09E		
090145.7	322	1.105174	017	913	-08	-05	Ugt1a6b	protein_coding
ENSMUSG00000	340.7		0.330	5.204	1.95E	0.000		
074240.5	069	1.717631	047	195	-07	218	Cib3	protein_coding
ENSMUSG00000	11088		1.057	4.840	1.29E	0.000		
032315.6	.84	5.118722	395	882	-06	897	Cyp1a1	protein_coding
ENSMUSG00000	111.8		1.067	4.622	3.79E	0.002		
003545.3	039	4.933125	168	633	-06	124	Fosb	protein_coding
ENSMUSG00000	157.9		0.273	4.421	9.82E	0.004	Gstm2-	processed_pse
107369.1	809	1.210026	699	017	-06	474	ps1	udogene
ENSMUSG00000	788.4		0.246	4.371	1.23E	0.004		
025780.7	662	1.077286	425	655	-05	897	Itih5	protein_coding
ENSMUSG00000	648.3		0.251	4.297	1.73E	0.006		
060681.15	327	1.082717	926	752	-05	287	Slc9a6	protein_coding
ENSMUSG00000	39.87		0.464	4.275	1.9E-	0.006		
101304.1	011	1.984035	005	897	05	767	Plet1os	lncRNA
ENSMUSG00000	62.22		0.265	4.269	1.96E	0.006		
022947.8	243	1.134718	803	019	-05	813	Cbr3	protein_coding
ENSMUSG00000	147.3		0.353	4.262	2.02E	0.006		
068877.12	837	1.506513	403	878	-05	84	Selenbp2	protein_coding
ENSMUSG00000	75.30		0.249	4.184	2.86E	0.008		
022231.10	837	1.045326	835	064	-05	695	Sema5a	protein_coding
ENSMUSG00000	485.5		0.284	4.152	3.29E	0.009		
019997.11	645	1.182042	641	75	-05	576	Ccn2	protein_coding
ENSMUSG00000	104.6		0.320	4.138	3.49E	0.009	5330417	
040412.16	277	1.327142	651	897	-05	597	C22Rik	protein_coding
ENSMUSG00000	66.95		0.370	4.111	3.94E	0.010		
044229.9	874	1.522956	456	036	-05	631	Nxpe4	protein_coding
ENSMUSG00000	57.80		0.331	4.058	4.94E	0.012		
002808.7	141	1.344906	391	363	-05	004	Epdr1	protein_coding
ENSMUSG00000	38.58		0.561	4.019	5.82E	0.013		
061322.15	057	2.257762	642	934	-05	685	Dnaic1	protein_coding
ENSMUSG00000	93.18		0.320	4.001	6.29E	0.014		
025002.5	946	1.281676	275	799	-05	043	Cyp2c55	protein_coding

ENSMUSG00000	68.06		0.323	3.923	8.73E	0.018			
114253.1	121	1.269706	627	366	-05	181	Gm47798	lncRNA	
ENSMUSG00000	11777		0.301	3.805	0.000	0.024			
078650.2	.34	1.149067	923	829	141	818	G6pc	protein_coding	
ENSMUSG00000	104.7		0.358	3.796	0.000	0.025			
040522.5	073	1.362234	852	091	147	205	Tlr8	protein_coding	
ENSMUSG00000	214.9		0.298	3.679	0.000	0.034			
028864.7	669	1.097507	301	19	234	102	Hgf	protein_coding	
ENSMUSG00000	3257.		0.279	3.586	0.000	0.043			
026043.18	473	1.002598	526	777	335	182	Col3a1	protein_coding	
ENSMUSG00000	83.15		0.292	3.562	0.000	0.045			
024511.15	864	1.040943	218	219	368	809	Rab27b	protein_coding	
ENSMUSG00000	59.96		0.386	3.554	0.000	0.046			
025321.14	935	1.374587	716	51	379	774	Itgb8	protein_coding	
ENSMUSG00000	236.4		0.313	3.532	0.000	0.049			
006205.13	209	1.108485	785	63	411	561	Htra1	protein_coding	
ENSMUSG00000	113.8		0.333	3.502	0.000	0.052			
033855.15	82	1.16742	277	856	46	826	Ston1	protein_coding	
ENSMUSG00000	78.20		0.553	3.470	0.000	0.057			
026064.16	506	1.921414	613	682	519	142	Ptp4a1	protein_coding	
ENSMUSG00000	210.6		0.311	3.427	0.000	0.062			
040152.8	402	1.067537	469	427	609	871	Thbs1	protein_coding	
ENSMUSG00000	61.26		0.376	3.406	0.000	0.066			
026479.13	863	1.283628	829	397	658	167	Lamc2	protein_coding	
ENSMUSG00000	28.77		0.390	3.337	0.000	0.076			
025804.5	743	1.304761	921	658	845	922	Ccr1	protein_coding	
ENSMUSG00000	3504.		0.680	3.307	0.000	0.080			
020423.6	418	2.250439	492	077	943	385	Btg2	protein_coding	
ENSMUSG00000	46.20		0.371	3.292	0.000	0.082			
052942.13	626	1.221786	127	096	994	352	Glis3	protein_coding	

APPENDICES VI: RNA SEQ ARNT iKO GENES PATHWAYS ANALYSIS

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Metabolism of xenobiotics by cytochrome P450	0.0001909	0.05784	26.74	228.98
2	ECM-receptor interaction	0.0003758	0.02847	21.26	167.68
3	Steroid hormone biosynthesis	0.0004613	0.02795	19.83	152.31
4	Retinol metabolism	0.0004923	0.02486	19.39	147.70
5	Chemical carcinogenesis	0.0005413	0.02343	18.77	141.20
6	Malaria	0.003138	0.1188	24.01	138.40
7	Focal adhesion	0.0003491	0.03526	11.82	94.12
8	Ascorbate and aldarate metabolism	0.04493	1.000	21.79	67.60
9	PI3K-Akt signaling pathway	0.0003203	0.04852	8.24	66.29
10	Arachidonic acid metabolism	0.01001	0.3372	13.22	60.86
11	Galactose metabolism	0.05303	1.000	18.38	53.99
12	Starch and sucrose metabolism	0.05464	1.000	17.83	51.82
13	Pentose and glucuronate interconversions	0.05625	1.000	17.30	49.79
14	Amoebiasis	0.01398	0.4235	11.10	47.39
15	Bladder cancer	0.06745	1.000	14.35	38.69
16	Porphyrin and chlorophyll metabolism	0.06745	1.000	14.35	38.69
17	Carbohydrate digestion and absorption	0.07062	1.000	13.68	36.26
18	Tryptophan metabolism	0.07851	1.000	12.25	31.18
19	Cocaine addiction	0.07851	1.000	12.25	31.18
20	Linoleic acid metabolism	0.08165	1.000	11.76	29.47
21	Ovarian steroidogenesis	0.09255	1.000	10.32	24.56
22	Glycolysis / Gluconeogenesis	0.1079	1.000	8.78	19.55
23	Amphetamine addiction	0.1094	1.000	8.65	19.14

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
24	Renal cell carcinoma	0.1094	1.000	8.65	19.14
25	Human papillomavirus infection	0.02292	0.6313	4.90	18.51
26	Adipocytokine signaling pathway	0.1140	1.000	8.29	17.99
27	p53 signaling pathway	0.1140	1.000	8.29	17.99
28	Proteoglycans in cancer	0.04651	1.000	5.80	17.78
29	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	0.1155	1.000	8.17	17.64
30	Melanoma	0.1155	1.000	8.17	17.64
31	Rap1 signaling pathway	0.04900	1.000	5.63	16.98
32	Cardiac muscle contraction	0.1245	1.000	7.54	15.71
33	RNA degradation	0.1320	1.000	7.09	14.35
34	Hypertrophic cardiomyopathy (HCM)	0.1364	1.000	6.84	13.63
35	Dilated cardiomyopathy (DCM)	0.1423	1.000	6.54	12.75
36	Protein digestion and absorption	0.1423	1.000	6.54	12.75
37	IL-17 signaling pathway	0.1437	1.000	6.46	12.54
38	TGF-beta signaling pathway	0.1437	1.000	6.46	12.54
39	Small cell lung cancer	0.1452	1.000	6.39	12.34
40	Toll-like receptor signaling pathway	0.1554	1.000	5.94	11.06
41	AGE-RAGE signaling pathway in diabetic complications	0.1583	1.000	5.82	10.74
42	Glucagon signaling pathway	0.1597	1.000	5.77	10.58
43	Pancreatic secretion	0.1640	1.000	5.60	10.13
44	Toxoplasmosis	0.1683	1.000	5.45	9.71
45	Insulin resistance	0.1711	1.000	5.35	9.44
46	Drug metabolism	0.1768	1.000	5.16	8.94
47	Platelet activation	0.1921	1.000	4.71	7.76
48	AMPK signaling pathway	0.1935	1.000	4.67	7.67

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
49	Inflammatory mediator regulation of TRP channels	0.1949	1.000	4.63	7.57
50	Osteoclast differentiation	0.1963	1.000	4.60	7.48
51	Relaxin signaling pathway	0.2004	1.000	4.49	7.22
52	FoxO signaling pathway	0.2017	1.000	4.46	7.13
53	Serotonergic synapse	0.2017	1.000	4.46	7.13
54	Apelin signaling pathway	0.2099	1.000	4.26	6.65
55	Insulin signaling pathway	0.2113	1.000	4.23	6.58
56	Gastric cancer	0.2260	1.000	3.92	5.83
57	Hippo signaling pathway	0.2378	1.000	3.70	5.31
58	Cell adhesion molecules (CAMs)	0.2521	1.000	3.46	4.77
59	Hepatocellular carcinoma	0.2534	1.000	3.44	4.72
60	Axon guidance	0.2648	1.000	3.27	4.34
61	Phagosome	0.2648	1.000	3.27	4.34
62	Huntington disease	0.2798	1.000	3.06	3.90
63	Chemokine signaling pathway	0.2860	1.000	2.99	3.74
64	Alcoholism	0.2884	1.000	2.96	3.68
65	Pathways in cancer	0.2305	1.000	2.20	3.23
66	Kaposi sarcoma-associated herpesvirus infection	0.3089	1.000	2.72	3.20
67	Regulation of actin cytoskeleton	0.3101	1.000	2.71	3.17
68	Ras signaling pathway	0.3288	1.000	2.52	2.81
69	Human cytomegalovirus infection	0.3538	1.000	2.31	2.40
70	MicroRNAs in cancer	0.3821	1.000	2.09	2.01
71	Cytokine-cytokine receptor interaction	0.3938	1.000	2.01	1.88
72	MAPK signaling pathway	0.3958	1.000	2.00	1.85

Appendices VII: RNA seq ARNT flox downregulated genes

gene_id	base Mean	log2Fold Change	lfcSE	stat	pvalu e	padj	gene_na me	gene_type
ENSMUSG00000020052.9	12.33 523	-5.30574	1.163 279	4.56 102	5.09 E-06	0.000 465	Ascl1	protein_coding
ENSMUSG00000052212.6	7.902 163	-2.79287	0.902 18	3.09 569	0.001 964	0.058 59	Cd177	protein_coding
ENSMUSG00000022347.8	2149. 401	-2.64001	0.916 824	2.87 952	0.003 983	0.093 253	A1bg	protein_coding
ENSMUSG00000108145.1	8.885 334	-2.46431	0.726 869	3.39 031	0.000 698	0.027 195	Gm38811	lncRNA
ENSMUSG00000043013.10	191.9 677	-2.21054	0.340 352	6.49 486	8.31 E-11	2.02 E-08	Onecut1	protein_coding
ENSMUSG00000069441.3	12.70 627	-1.81691	0.607 6	2.99 03	0.002 787	0.073 816	Dsg1a	protein_coding
ENSMUSG00000055963.12	10.03 007	-1.71349	0.602 954	2.84 182	0.004 486	0.099 926	Triqk	protein_coding
ENSMUSG00000028240.2	2476. 157	-1.66744	0.262 809	6.34 467	2.23 E-10	4.97 E-08	Cyp7a1	protein_coding
ENSMUSG00000106303.1	94.31 898	-1.65235	0.420 589	3.92 865	8.54 E-05	0.005 052	Gm7652	unprocessed_pseudogene
ENSMUSG00000038751.5	41.94 226	-1.55657	0.442 881	3.51 466	0.000 44	0.018 843	Ptk6	protein_coding
ENSMUSG00000106040.1	253.7 03	-1.52601	0.299 261	5.09 925	3.41 E-07	4.05 E-05	Cyp3a63-ps	unprocessed_pseudogene
ENSMUSG00000032578.7	953.5 237	-1.5174	0.493 54	3.07 452	0.002 108	0.061 626	Cish	protein_coding
ENSMUSG00000032065.6	22.19 032	-1.50007	0.436 941	3.43 312	0.000 597	0.024 006	Tex12	protein_coding
ENSMUSG00000027579.6	22.11 446	-1.49186	0.496 402	3.00 534	0.002 653	0.071 637	Srms	protein_coding
ENSMUSG00000072568.4	641.8 383	-1.41222	0.161 175	8.76 206	1.92 E-18	1.04 E-15	Lratd2	protein_coding
ENSMUSG00000039081.10	138.0 527	-1.40126	0.324 039	4.32 436	1.53 E-05	0.001 213	Zfp503	protein_coding
ENSMUSG00000049804.9	222.4 013	-1.36938	0.309 854	4.41 944	9.9E- 06	0.000 834	Armcx4	protein_coding

ENSMUSG0000 0106164.1	24.59 243	-1.33251	0.437 793	3.04 369	0.002 337	0.065 71	9430085 M18Rik	lncRNA
ENSMUSG0000 0048826.7	343.0 35	-1.28506	0.268 155	4.79 222	1.65 E-06	0.000 176	Dact2	protein_coding
ENSMUSG0000 0104664.1	37.00 942	-1.28029	0.421 415	3.03 807	0.002 381	0.066 096	Gm35570	processed_pseudogene
ENSMUSG0000 0045314.5	236.9 44	-1.27794	0.194 455	6.57 191	4.97 E-11	1.24 E-08	Sowahb	protein_coding
ENSMUSG0000 0035455.12	37.88 364	-1.26172	0.424 482	2.97 237	0.002 955	0.076 768	Figl1	protein_coding
ENSMUSG0000 0078234.6	1047. 945	-1.23451	0.261 895	4.71 375	2.43 E-06	0.000 245	Klhdc7a	protein_coding
ENSMUSG0000 0036306.13	40.09 857	-1.22494	0.411 557	2.97 636	0.002 917	0.076 018	Lzts1	protein_coding
ENSMUSG0000 0042745.9	1050. 72	-1.22277	0.275 253	4.44 234	8.9E- 06	0.000 766	Id1	protein_coding
ENSMUSG0000 0054417.5	8355. 147	-1.15469	0.223 29	5.17 123	2.33 E-07	2.86 E-05	Cyp3a44	protein_coding
ENSMUSG0000 0101211.1	53.20 226	-1.14132	0.373 477	3.05 592	0.002 244	0.064 201	Gm28818	lncRNA
ENSMUSG0000 0030364.6	80.70 52	-1.13182	0.388 83	2.91 083	0.003 605	0.087 491	Clec2h	protein_coding
ENSMUSG0000 0046179.17	215.3 33	-1.11797	0.227 89	4.90 574	9.31 E-07	0.000 102	E2f8	protein_coding
ENSMUSG0000 0028525.16	569.5 701	-1.10695	0.206 555	5.35 912	8.36 E-08	1.15 E-05	Pde4b	protein_coding
ENSMUSG0000 0029414.11	41.93 847	-1.10582	0.364	3.03 797	0.002 382	0.066 096	Kntc1	protein_coding
ENSMUSG0000 0024327.16	29.61 882	-1.09992	0.348 816	3.15 329	0.001 614	0.050 979	Slc39a7	protein_coding
ENSMUSG0000 0021198.16	45.85 669	-1.09729	0.384 947	2.85 049	0.004 365	0.098 455	Unc79	protein_coding
ENSMUSG0000 0024042.7	726.6 621	-1.07489	0.224 004	4.79 854	1.6E- 06	0.000 172	Sik1	protein_coding
ENSMUSG0000 0018983.9	91.64 714	-1.06153	0.322 739	3.28 915	0.001 005	0.035 759	E2f2	protein_coding

ENSMUSG0000 0022651.6	39.59 873	-1.05201	0.342 936	3.06 766	0.002 157	0.062 538	Retnlg	protein_coding
ENSMUSG0000 0047492.5	1988. 615	-1.03139	0.269 207	3.83 12	0.000 128	0.007 078	Inhbe	protein_coding
ENSMUSG0000 0062061.3	33.85 095	-1.02848	0.357 768	2.87 47	0.004 044	0.094 085	Obp2a	protein_coding
ENSMUSG0000 0049907.8	122.2 551	-1.02723	0.306 482	3.35 167	0.000 803	0.030 237	Rasl11b	protein_coding
ENSMUSG0000 0025185.14	47.89 64	-1.00584	0.277 483	3.62 486	0.000 289	0.013 588	Loxl4	protein_coding

APPENDICES VIII: RNA SEQ ARNT IKO DOWNREGULATED GENES

gene_id	base Mean	log2Fold Change	lfcSE	stat	pval ue	padj	gene_na me	gene_type
ENSMUSG00000062170.12	39.26 073		0.78 0604	5.15 283	2.57 E-07	0.00 0251	Fmr1nb	protein_coding
ENSMUSG00000081207.4	162.3 454	-3.90606	0.98 1458	3.97 986	6.9E- 05	0.01 5001	Gm1377 5	lncRNA
ENSMUSG00000106069.1	1923. 402	-2.56152	0.56 3662	4.54 442	5.51 E-06	0.00 2867	Gm6135	lncRNA
ENSMUSG00000106303.1	94.31 898	-2.53371	0.53 6763	4.72 035	2.35 E-06	0.00 1492	Gm7652	unprocessed_pseudo gene
ENSMUSG00000024112.16	37.96 387	-2.53319	0.68 0882	3.72 046	0.00 0199	0.02 988	Cacna1h	protein_coding
ENSMUSG00000084803.8	36.06 719	-2.23618	0.51 0439	4.38 09	1.18 E-05	0.00 4897	5830444 B04Rik	lncRNA
ENSMUSG00000074375.3	524.8 315	-2.13807	0.33 6822	6.34 777	2.18 E-10	1.06 E-06	Sult2a3	protein_coding
ENSMUSG00000027870.8	647.8 058	-2.04331	0.50 2237	4.06 842	4.73 E-05	0.01 1693	Hao2	protein_coding
ENSMUSG00000078675.9	59.55 069	-2.03453	0.58 7925	3.46 052	0.00 0539	0.05 8076	Mup16	protein_coding
ENSMUSG00000106164.1	24.59 243	-1.91292	0.49 9559	3.82 921	0.00 0129	0.02 3717	9430085 M18Rik	lncRNA
ENSMUSG00000049515.12	17.02 004	-1.8597	0.55 3616	3.35 918	0.00 0782	0.07 447	Espnl	protein_coding
ENSMUSG00000000305.12	39.56 243	-1.70335	0.44 6997	3.81 065	0.00 0139	0.02 4636	Cdh4	protein_coding
ENSMUSG00000092008.2	9790. 974	-1.65663	0.29 6284	5.59 135	2.25 E-08	3.28 E-05	Cyp2c69	protein_coding

ENSMUSG0000	4979.		0.27	5.88	3.98	6.44			
0035836.2	116	-1.63906	8505	522	E-09	E-06	Ugt2b1	protein_coding	
ENSMUSG0000	122.8		0.47	3.33	0.00	0.07	Mup-	unprocessed_pseudo	
0095532.1	635	-1.59364	7947	435	0855	6922	ps13	gene	
ENSMUSG0000	8986.		0.26	5.96	2.43	5.05			
0078798.4	295	-1.58415	5515	634	E-09	E-06	Sult2a1	protein_coding	
ENSMUSG0000	118.9		0.47	3.20	0.00	0.09	Gm2418		
0088609.1	78	-1.52079	4822	287	1361	8177	7	miRNA	
ENSMUSG0000	83.01		0.36	3.99	6.36	0.01			
0024411.11	897	-1.47351	8463	908	E-05	4043	Aqp4	protein_coding	
ENSMUSG0000	871.7		0.28	5.09	3.42	0.00			
0040026.8	207	-1.45976	6287	895	E-07	0311	Saa3	protein_coding	
ENSMUSG0000	74.64		0.43	3.29	0.00	0.08	Gm2638		
0076137.1	567	-1.44012	6522	909	097	1721	4	miRNA	
ENSMUSG0000	857.8		0.30	4.66	3.1E-	0.00			
0073940.3	396	-1.41919	4261	439	06	1805	Hbb-bt	protein_coding	
ENSMUSG0000	6638.		0.26	5.15	2.58	0.00			
0003555.8	756	-1.36953	584	17	E-07	0251	Cyp17a1	protein_coding	
ENSMUSG0000	3386.		0.30	4.40	1.05	0.00			
0069919.7	622	-1.35866	834	636	E-05	4643	Hba-a1	protein_coding	
ENSMUSG0000	61.41		0.38	3.47	0.00	0.05	A130071		
0104291.1	424	-1.35201	895	606	0509	7049	D04Rik	TEC	
ENSMUSG0000	189.5		0.22	5.89	3.85	6.44	Gm4966		
0116718.1	045	-1.32257	4524	055	E-09	E-06	8	lncRNA	
ENSMUSG0000	1205.		0.31	4.17	2.95	0.00			
0067656.13	858	-1.30128	153	705	E-05	8784	Slc22a27	protein_coding	
ENSMUSG0000	101.0		0.36	3.50	0.00	0.05			
0014030.15	117	-1.28428	601	885	045	2058	Pax5	protein_coding	
ENSMUSG0000	6175.		0.28	4.35	1.31	0.00			
0052305.6	664	-1.24082	4678	869	E-05	5019	Hbb-bs	protein_coding	
ENSMUSG0000	6146.		0.14	8.25	1.55	2.25			
0025004.16	701	-1.23486	9626	296	E-16	E-12	Cyp2c40	protein_coding	

-									
ENSMUSG0000	71.34		0.36	3.30	0.00	0.08	Gm4744		
0111361.1	885	-1.21185	6783	4	0953	077	5		TEC
-									
ENSMUSG0000	39.27		0.35	3.32	0.00	0.07			
0022838.14	66	-1.19667	9661	722	0877	7698	Eaf2		protein_coding
-									
ENSMUSG0000	102.5		0.33	3.61	0.00	0.04			
0025889.13	343	-1.19399	0697	053	0306	0859	Snca		protein_coding
-									
ENSMUSG0000	76.99		0.30	3.86	0.00	0.02	3110045		
0097503.2	866	-1.17874	4762	774	011	1602	C21Rik		lncRNA
-									
ENSMUSG0000	59.20		0.34	3.35	0.00	0.07	Gm3023		
0103898.1	006	-1.1609	5898	619	079	479	8		TEC
-									
ENSMUSG0000	112.0		0.23	4.91	8.89	0.00	1810053		
0100277.8	823	-1.1472	3422	467	E-07	0682	B23Rik		lncRNA
-									
ENSMUSG0000	394.2		0.18	6.07	1.22	2.97	Gm3158		
0115919.1	995	-1.1404	7659	699	E-09	E-06	3		lncRNA
-									
ENSMUSG0000	254.0		0.26	4.21	2.5E-	0.00	Gm3604		
0112774.1	052	-1.13739	9873	452	05	8107	1		lncRNA
-									
ENSMUSG0000	106.9		0.31	3.52	0.00	0.04			
0055301.8	623	-1.12545	8839	984	0416	9676	Adh7		protein_coding
-									
ENSMUSG0000	306.1		0.28	3.94	8.13	0.01			
0091780.3	56	-1.12386	5196	067	E-05	7163	Sco2		protein_coding
-									
ENSMUSG0000	54.18		0.29	3.73	0.00	0.02	B230217		
0097785.2	459	-1.1033	5159	797	0186	8684	O12Rik		lncRNA
-									
ENSMUSG0000	273.5		0.27	3.87	0.00	0.02	AI46322		
0114828.1	268	-1.08531	998	639	0106	153	9		lncRNA
-									
ENSMUSG0000	230.0		0.32	3.33	0.00	0.07	D130043		
0006711.15	26	-1.0779	2768	956	0839	6922	K22Rik		protein_coding
-									
ENSMUSG0000	42.24		0.30	3.45	0.00	0.05	Gm2223		
0087710.1	57	-1.06848	8887	914	0542	8076	5		snRNA
-									
ENSMUSG0000	72.13		0.30	3.33	0.00	0.07			
0079457.10	046	-1.03371	9578	909	0841	6922	Gm7609		transcribed_unprocessed_pseudogene
-									
ENSMUSG0000	57.11		0.29	3.40	0.00	0.06			
0110165.1	771	-1.01607	8618	258	0668	6639	Gm4032		transcribed_processed_pseudogene

APPENDICES IX: RNA SEQ ARNT FLOX DOWNREGULATED GENES PATHWAY ANALYSIS

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Signaling pathways regulating pluripotency of stem cells	0.002581	0.7821	10.95	65.25
2	Steroid hormone biosynthesis	0.01369	1.000	11.24	48.22
3	TGF-beta signaling pathway	0.01428	1.000	10.99	46.69
4	Primary bile acid biosynthesis	0.03154	1.000	31.25	108.02
5	Maturity onset diabetes of the young	0.05265	1.000	18.52	54.52
6	Kaposi sarcoma-associated herpesvirus infection	0.06930	1.000	4.63	12.36
7	Bladder cancer	0.07888	1.000	12.20	30.97
8	Cholesterol metabolism	0.09355	1.000	10.20	24.18
9	Linoleic acid metabolism	0.09536	1.000	10.00	23.50
10	Non-small cell lung cancer	0.1240	1.000	7.58	15.82
11	Prolactin signaling pathway	0.1345	1.000	6.94	13.93
12	Melanoma	0.1345	1.000	6.94	13.93
13	Bile secretion	0.1345	1.000	6.94	13.93
14	Pancreatic cancer	0.1397	1.000	6.67	13.12
15	Glioma	0.1397	1.000	6.67	13.12
16	Chronic myeloid leukemia	0.1414	1.000	6.58	12.87
17	PPAR signaling pathway	0.1568	1.000	5.88	10.90
18	Retinol metabolism	0.1669	1.000	5.49	9.84
19	Morphine addiction	0.1686	1.000	5.43	9.68
20	Small cell lung cancer	0.1686	1.000	5.43	9.68
21	Chemical carcinogenesis	0.1719	1.000	5.32	9.37
22	Staphylococcus aureus infection	0.1736	1.000	5.26	9.22
23	Prostate cancer	0.1769	1.000	5.15	8.93

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
24	Glucagon signaling pathway	0.1851	1.000	4.90	8.27
25	Parathyroid hormone synthesis, secretion and action	0.1933	1.000	4.67	7.68
26	Cell cycle	0.2189	1.000	4.07	6.18
27	Purine metabolism	0.2391	1.000	3.68	5.26
28	Breast cancer	0.2557	1.000	3.40	4.64
29	Gastric cancer	0.2602	1.000	3.33	4.49
30	Cushing syndrome	0.2735	1.000	3.14	4.08
31	Hippo signaling pathway	0.2735	1.000	3.14	4.08
32	Hepatitis C	0.2750	1.000	3.13	4.03
33	Hepatitis B	0.2794	1.000	3.07	3.91
34	JAK-STAT signaling pathway	0.2808	1.000	3.05	3.87
35	Hepatocellular carcinoma	0.2909	1.000	2.92	3.61
36	Cellular senescence	0.3107	1.000	2.70	3.16
37	Rap1 signaling pathway	0.3434	1.000	2.39	2.56
38	cAMP signaling pathway	0.3460	1.000	2.37	2.51
39	Epstein-Barr virus infection	0.3694	1.000	2.18	2.17
40	Human T-cell leukemia virus 1 infection	0.3895	1.000	2.04	1.92
41	Human cytomegalovirus infection	0.4018	1.000	1.96	1.79
42	MicroRNAs in cancer	0.4325	1.000	1.78	1.49
43	Cytokine-cytokine receptor interaction	0.4451	1.000	1.71	1.39
44	Pathways in cancer	0.6623	1.000	0.93	0.39

APPENDICES X: RNA SEQ ARNT iKO DOWNREGULATED GENES PATHWAYS

Name	P-value	Adjusted p-value	Odds Ratio	Combined score	
1	Chemical carcinogenesis	0.000002169	0.0006572	23.64	308.31
2	Metabolism of xenobiotics by cytochrome P450	0.00001455	0.002205	26.94	300.00
3	African trypanosomiasis	0.00009191	0.009283	34.19	317.77
4	Malaria	0.0001824	0.01382	27.21	234.27
5	Bile secretion	0.0005695	0.03451	18.52	138.35
6	Steroid hormone biosynthesis	0.001056	0.05332	14.98	102.67
7	Retinol metabolism	0.001126	0.04874	14.65	99.47
8	Cortisol synthesis and secretion	0.01055	0.3997	12.88	58.63
9	Drug metabolism	0.02719	0.9153	7.80	28.11
10	Cushing syndrome	0.04977	1.000	5.59	16.77
11	Ascorbate and aldarate metabolism	0.05904	1.000	16.46	46.58
12	Glyoxylate and dicarboxylate metabolism	0.06750	1.000	14.34	38.65
13	Pentose and glucuronate interconversions	0.07379	1.000	13.07	34.07
14	Tyrosine metabolism	0.08624	1.000	11.11	27.23
15	Porphyrin and chlorophyll metabolism	0.08830	1.000	10.84	26.31
16	Vasopressin-regulated water reabsorption	0.09241	1.000	10.34	24.62
17	Linoleic acid metabolism	0.1066	1.000	8.89	19.90
18	Fatty acid degradation	0.1066	1.000	8.89	19.90
19	Ovarian steroidogenesis	0.1207	1.000	7.80	16.49
20	Central carbon metabolism in cancer	0.1344	1.000	6.94	13.93

Name	P-value	Adjusted p-value	Odds Ratio	Combined score	
21	Glycolysis / Gluconeogenesis	0.1403	1.000	6.63	13.03
22	Prolactin signaling pathway	0.1500	1.000	6.17	11.71
23	Peroxisome	0.1727	1.000	5.29	9.29
24	Arachidonic acid metabolism	0.1820	1.000	4.99	8.51
25	Circadian entrainment	0.2003	1.000	4.49	7.22
26	Aldosterone synthesis and secretion	0.2057	1.000	4.36	6.89
27	Inflammatory mediator regulation of TRP channels	0.2495	1.000	3.50	4.86
28	Serotonergic synapse	0.2579	1.000	3.37	4.56
29	Parkinson disease	0.2779	1.000	3.09	3.95
30	Cell adhesion molecules (CAMs)	0.3192	1.000	2.61	2.99
31	Alzheimer disease	0.3269	1.000	2.54	2.84
32	Transcriptional misregulation in cancer	0.3391	1.000	2.43	2.63
33	Calcium signaling pathway	0.3480	1.000	2.35	2.48
34	MAPK signaling pathway	0.4868	1.000	1.51	1.09

VITA

Karen Larissa Pereira de Castro was born and raised in Belo Horizonte, Minas Gerais, Brazil. Before attending the University of Texas Medical Branch, she attended the Pontifícia Universidade Católica de Minas Gerais, in Belo Horizonte, Minas Gerais, Brazil, where she earned a Bachelor in Biology in 2011. From 2012 to 2014, she also attended the Universidade Federal de Minas Gerais, in Belo Horizonte, Minas Gerais, Brazil, where she earned a Master degree in Neuroscience. This work led to the publication of two first author manuscripts and several co-authorships. Karen joined Dr. Cornelis Elferink laboratory in 2016, where she published as first author in a literature review and as a co-author in a collaboration study. An additional first author publication is expected (chapter 2 of dissertation).

Permanent address : Rua Icaraí, 775, apto 01. Bairro Caiçaras. Belo Horizonte, MG, Brazil. Zip code: 30770-160

This dissertation was tipped by: Karen Larissa Pereira de Castro.