Endogenous Signaling and Regulation of the Aryl Hydrocarbon Receptor Pathway

By

Jessica C. Parrott

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Molecular & Environmental Toxicology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2023

Date of final oral examination: 01/20/2023

The dissertation is approved by the following members of the Final Oral Committee:

Christopher A. Bradfield, Professor, Oncology, UW-Madison

Sean Ronnekleiv-Kelly, Assistant Professor, Surgery, UW-Madison

Kristen Malecki, Professor, Environmental and Occupational Health Sciences, UI-Chicago

C. Dustin Rubinstein, Client-Based Researcher III, Biotechnology Center, UW-Madison

DEDICATION

This thesis is dedicated to my parents, Tom and Lee Parrott. Thank you and I love you.

ACKNOWLEDGEMENTS

Thank you first and foremost to the senior scientists in lab that made this thesis possible. To Ed Glover, Anna Shen, and Manabu Nukaya: thank you for being resources I could always turn to when I had questions, difficulties, or needed help in any way. It was a comfort to know I always had your expertise to draw on. Also, a special thank you to Ed for training me in Western blotting and trusting me with precious antibodies.

Thank you to Dustin Rubinstein and Brent Lehman at the UW Biotechnology Center Genome Editing Core for their expertise in cell culture and CRISPR editing. Their training and our scientific discussions provided a basis for much of the cell culture and gene editing work of this thesis.

Mark Berres at the UWBC Bioinformatics Resource Center provided expert assistance with our RNA-seq analysis and the laboratory considerations necessary for a successful experiment. Yana Stackpole at Qlucore helped me through the difficulties of learning a new software and a new type of data to analyze. Her assistance gave me the confidence and skills to intimately understand complex RNA-seq data and derive conclusions from there. Thank you both.

A special thank you is reserved for my friends from my lab and my program. To Brenda Rojas, Emmanuel Vázquez Rivera, Alex Veith, Morgan Walcheck, Rachel Wilson, and Jeremiah Yee: thank you for being the best lab mates possible. Whether it was discussing science, venting about the difficulties of research and graduate school, or laughing about nothing, you made this journey a lot more fun. To Kenneth Sandro Rivera González, Glorimar Guzmán Pérez, and Dominic

Sanchez: I love you guys. Meeting you all our first day of graduate school, I didn't realize the journey we were going on together. Thanks for being there.

Thank you to Mark Marohl for being the best graduate program coordinator ever. Your presence and overwhelming competence made the Molecular and Environmental Toxicology program a more welcoming place. Your support and genuine kindness will never be forgotten.

To Chris Bradfield: thank you. Your guidance has helped make me a better scientist, and your attitude and mindset have helped me learn and grow as a person. I consistently feel lucky I joined your lab as a first-year student.

Last but not least, a huge thank you goes to my family. Words cannot express how blessed I feel to have you in my life. You've been there for me from the very beginning, and I look forward to our next adventures together. To my parents, to whom this thesis is dedicated: thank you for encouraging my curiosity and loving me completely. To my brother Jake: thank you for always being there for me. You're the best partner in crime. And to my sister-in-law Maura: thank you for adding a twist to our particular brand of madness. We're all grateful to have you in the family.

And thank you, dear reader, for picking up this thesis. I hope you enjoy it.

TABLE OF CONTENTS

ABSTRACT

This thesis is an exploration into the workings of the aryl hydrocarbon receptor (AHR) pathway. Topics investigated include endogenous mechanisms of pathway induction and cessation, the former being through tryptophan metabolism and the latter through the workings of the AHR repressor (AHRR). Also explored was the creation and utility of gene-edited cell models. In the first chapter, we review considerations for creating edited cell lines with various methods. Additionally, commonly utilized cell lines specifically for AHR pathway research, along with their feasibility for gene editing, are reviewed. The second chapter describes a fellowship undertaken at the University of Wisconsin Biotechnology Center and its impact on marketable skills for graduate students interested in non-academic careers. The third chapter characterizes several endogenous pathways and their common production of indole-3-pyruvate as the AHR pathway proligand of interest. Several putative AHR pathway genes are also identified in this chapter. The establishment of an auxin-inducible degron (AID) system in HCT 116 cells comprises the fourth chapter. Included there are also the method for cloning and its efficiency, as well as pitfalls of the AID system within this cell line. The fifth chapter is an investigation into the AHRR both as a repressor and a possible secondary transcription factor in the AHR pathway. This chapter suggests AHRR may be able to repress the pathway in a DNA-independent manner; additionally, the AHRR can bind DNA and affect gene transcription at loci beyond those with which the AHR interacts. Taken together, the experiments in this thesis shed light on the endogenous activation and repression of a pathway classically involved in toxicology and xenobiotic metabolism.

Chapter 1: Considerations and Challenges in Creating Edited Cell Lines for Aryl Hydrocarbon Receptor Pathway Research

ABSTRACT

The aryl hydrocarbon receptor (AHR) pathway plays important roles in human health and development. It has been studied for decades using both whole organisms and cell lines. Whole organisms are valuable for toxicity studies while cell lines more easily capture pathway signaling events. Animal models are attractive because they produce more physiologically relevant results. However, animal models are more difficult to edit, grow, and screen. Immortalized cell lines have many advantages over whole organisms including speed, cost, and ease of use. Cells can be edited easily using CRISPR, random integration, or homologous recombination, and many methods exist to deliver DNA or RNA to cells. Protein knockdown methods and reporter assay designs are also numerous. The multiplicity of options in these areas give a flexibility to cell work that allows workarounds to problems within a specific cell line, and it also supports creativity in experimental design. Many human and non-human cell lines have been used over the years, but this review describes the most popular. This review is meant as a practical guide to designing an edited immortalized cell line for AHR pathway research.

INTRODUCTION

The aryl hydrocarbon receptor (AHR) is a ligand-activated transcription factor that makes up part of the Per-Arnt-Sim (PAS) protein family. A large body of research suggests the AHR mediates xenobiotic toxicity and metabolism, modulates physiological processes, and affects cancer development. The aryl hydrocarbon receptor nuclear translocator (ARNT) is the dimerization partner for AHR. The ARNT protein when globally knocked out is embryonic lethal, suggesting it is a vital PAS signaling partner in early development[1,2]. The ARNT dimerization with AHR allows the complex to bind DNA, precipitating downstream effects. Another important player in the AHR pathway is the aryl hydrocarbon receptor repressor (AHRR). The AHRR was first discovered in a screen to identify more PAS family proteins[3]. While the AHRR shares genetic similarity with AHR, their functions differ[4]. The AHRR represses the AHR pathway via an undetermined mechanism[5–8]. The AHRR's physiological relevance and impact on human health is still being investigated. However, the AHRR is useful as a biomarker for smoking and lung cancer, and the transcription factor is a putative tumor suppressor as well as potentially playing a role in immune function[9–13].

In an uninduced state, the AHR exists in the cytosol as a complex with chaperones p23, the AHR-associated protein 3 (ARA3), the AHR interacting protein (AIP, aka ARA9), and heat-shock protein 90 (Hsp90)[14–16]. Upon exposure to ligand, the AHR pathway is thought to signal as follows. First, ligand binds the AHR in the cytosol, causing a conformation shift in the protein that reveals a nuclear localization sequence (NLS). Second, the liganded AHR complex translocates to the nucleus, where the AHR chaperone proteins dissociate. Third, ARNT dimerizes with AHR using the basic helix-loop-helix (bHLH) and PAS domains of each protein.

Fourth, the AHR-ARNT complex binds the dioxin responsive elements (DREs) upstream of the following canonical genes: AHRR and the cytochromes P450 CYP1A1, CYP1A2, and CYP1B1. Fifth, the AHRR represses AHR while the upregulated CYPs metabolize the inducing ligand(s). Lastly, the pathway is attenuated due to metabolism and clearance of the ligand as well as repression by AHRR (Figure 1).

The AHR was first characterized in the mid-20th century through toxicological studies on dioxin and related chlorinated pollutants, contaminants in agricultural feeds and the defoliant Agent Orange. Interestingly, the AHRR was not discovered until the 1990s when studies were designed to expand the members of the PAS protein family[3]. Many reviews exist for those seeking more information on the history and evolution of the AHR pathway[14,17–21]. The purpose of this review is to characterize cell lines used for AHR research and identify each line's utility in studying the AHR pathway's impacts on human health. This review is meant to consolidate information on common, previously immortalized cell lines used to study AHR. This information can then be used for the generation of new edited cell line models. It is the authors' hope this review saves the reader time and effort in designing cell line experiments and models.

BACKGROUND

The AHR pathway has been studied for decades using both whole-organism models (often mice) and cell lines. Each approach has strengths and disadvantages. Together, results from *in vitro* and *in vivo* experiments provide a holistic view of the AHR pathway. Cells have generally been used to study the molecular signaling pathway, while whole organisms are often a better model in which to study toxicity and physiological events in complex systems. In this regard, a survey of nearly two dozen cell lines exposed to the most potent xenobiotic ligand of the AHR, 2,3,7,8-tetrachlorordibenzo-p-dioxin (TCDD), failed to induce overt toxicity[22]. Yeast cells have more recently been used to study the AHR pathway to great benefit, but due to the unique nature of yeast as a model, it is out of the scope of this review.

Advantages of cell line experiments are multifold (Table 1). First, experiments can be tailored to an area of interest based on choice of cell line. For example, a group interested in breast cancer may choose MCF7 cells, while another enamored with liver may opt for HepG2. The AHR is a pathway with tissue-specific expression, so differences in the organ of origin of a cell line may make a difference. Second, the simplicity of a monoculture allows conclusions to be drawn without the confounding effects of multiple cell types interacting within a tissue. Third, cells allow more chronic administration of chemicals with decreased compound clearance. In so doing, cell lines often allow a researcher to tease out the workings of the pathway in its simplest and clearest form. Fourth, cell lines in comparison to animal models are lower cost and often provide more rapid results. Use of cells eliminates the need for animals to be bred, grown, genotyped, fed, housed, et cetera. Cells grow quickly, are easy to edit, have low-cost inputs, and

are simple to work with. Depending on the type of experiments being run and the research questions asked, cells may be preferable to the complexity of working with whole organisms.

Though cell lines are an advantageous system in which to model the AHR pathway, they also have drawbacks (Table 1). First and foremost, cells are often not the best environment to recapitulate the complex phenomena of the human body. Immortalized cell genomes are strange, often containing chromosome duplications, rearrangements, losses, or fusions. Gene expression can differ in these cells versus what might be seen in a human. The abnormal immortalized cell genome can make editing difficult, since achieving a gene editing event can be difficult if there are multiples of the target locus. For example, knocking in a construct at two copies of a gene is much easier and more attainable than having to knock the construct in four times to a tetraploid locus. Second, cells in culture are not always an accurate representation of the cell type or organ they come from. Cells in culture tend to de-differentiate in comparison to their tissue of origin. As such, cells in culture may respond uniquely from the original organ. Third, depending on the desired application a cell line may be unsuitable for the physical manipulations that must happen in lab. Cell lines with contact-dependent growth tend to have low cloning efficiency. Other cell lines may not handle transfection or genetic manipulation well. Each cell line is different, and often the realities and limitations of a line are not apparent until one works with them in lab. Finally, cell lines are prone to contamination and are not always what a researcher thinks they are[23–25]. That is, a lab uses “HepG2” that is actually contaminated with HeLa. Despite the drawbacks existent with using cell lines for research, their use for *in vitro* experiments is an important part of science due to the advantages above, and cells have been used extensively to study the AHR pathway.

For the purposes of this review, a master list of cell lines was created through an extensive literature and web search, pulling lines from papers, company catalogs, and resources like the Cancer Cell Line Encyclopedia (CCLE) and NCI-60[26–28]. Once collected, each cell line was run through PubMed searching for "((aryl hydrocarbon hydroxylases[MeSH Terms]) OR (aryl hydrocarbon receptor[MeSH Terms])) AND (x)" where x is the cell line name in all lower-case searched in the "All Fields" field (e.g. "hepg2" for HepG2 cells) (Table 2). The lines were then ranked by number of items returned. Many other cell lines exist and may be suitable for AHR pathway research. The following discussion contains those most-published and best-characterized for the AHR pathway.

METHODS OVERVIEW

MOLECULAR BIOLOGY

BASICS OF CELL EDITING

Recent advances have allowed gene-editing through the use of clustered regularly interspaced short palindromic repeats (CRISPR) along with CRISPR-associated (Cas) proteins to cut DNA at a targeted sequence. Currently, there are many types of Cas proteins, but Cas9 is the best characterized and most widely used[29,30]. Many extensive reviews exist on CRISPR gene editing, so an in-depth discussion will not be reiterated here[31–36]. In brief, CRISPR is a novel gene-editing technique derived from bacterial viral-defense mechanisms (Figure 2). For gene editing, a guide RNA (gRNA) is designed to the desired portion of the genome with a protospacer adjacent motif (PAM) sequence at its end. The PAM is where the bacterial Cas9 protein recognizes a specific nucleotide sequence. At this sequence, Cas9 makes a double-strand break (DSB) in the DNA three nucleotides upstream of the beginning of the PAM. Flanking this cut site are homology arms which can be used to target the introduction of a new sequence between them. This "knock-in" sequence is generally introduced via a plasmid. The plasmid also contains homology arms complimentary to those of the locus of interest so that the insert between them can be used as a template for homology directed repair (HDR). Minimizing off-target effects is an important consideration when designing gRNAs to edit a cell line[37,38]. Other techniques, which have been used for decades before the advent of CRISPR, include zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and homologous recombination[39–43]. Each system has strengths and weaknesses.

METHODS TO DELIVER EDITING SYSTEMS

Regardless of the gene-editing system used for a project, the editing materials (recombinases, gRNAs, plasmids, Cas9) must be delivered to the cell. Common ways to introduce these materials into a cell are through transient and stable transfection. Transfection refers to the introduction of DNA, RNA, or protein into a cell and can be done through many methods including viral transduction (often with lentivirus), electroporation, or lipofection (Figure 3). Transient versus stable refers to the time of expression of the inserted DNA, RNA, or protein. Transient transfections are not stably integrated into the genome, so expression lasts for a short period of time before the cells eliminate the transfected DNA. In a stable transfection, the DNA has been integrated into the genome of the transfected cell line, and expression may continue indefinitely and be passed from generation to generation. Whether material is introduced via a transient or stable transfection, the goal is to produce an editing event that is stable and integrated in the genome. Cells must be screened to determine whether a stable event has transpired, and screening is often done with antibiotics or fluorescence. It is possible for introduced DNA to integrate randomly. However, precise editing is possible with the use of CRISPR/Cas, ZFNs, or TALENs, and one of these techniques is necessary for tagging an endogenous gene locus.

ACHIEVING MONOCLONAL POPULATIONS

Once cells have been edited, ideally they should be reduced to a single cell state and grown in isolation to achieve a monoclonal population. This ensures all cells have the same genome, so resulting data is more attributable to the introduced edit. Fluorescence-activated Cell Sorting (FACS) is one common way to obtain clonal populations after editing. FACS is especially useful when screening for fluorescently tagged cell lines, such as ones with a reporter gene. If FACS is

inaccessible or undesirable, a classic, simple, inexpensive technique to achieve monoclonality is limiting dilution. This method dilutes a population of cells to a concentration where a single cell may be deposited to each well of a 96-well plate. If the dilution is kept in a larger container such as a 10cm dish, single cells will grow into clusters at many areas of the dish and can be isolated with cloning cylinders. In the case of a stable transfection, cells are also commonly cultured with a selection antibiotic (i.e. puromycin, neomycin, geneticin). Monoclonality is an important requirement for generating dependable edited cell lines, and the use of different clones as biological replicates strengthens conclusions drawn from cell line data.

APPLICATIONS

PROTEIN KNOCKDOWN/KNOCKOUT

Editing-based systems: Protein "knockdown" or "knockout" can be achieved through various methods to study the effects a certain protein has on a pathway in question (Table 3). One of the most common methods, using CRISPR, is to introduce a DSB using a gRNA targeted to the beginning of a gene. Non-homologous end-joining (NHEJ) then repairs the DSB. Since NHEJ is an error-prone process, it can often introduce a frameshift mutation. Frameshifts may cause early termination of translation (nonsense mutation) or a drastic change in the amino acid sequence (missense mutation). Both of these outcomes would affect protein function and, likely, half-life. This in turn can allow observations on pathway function without the wild-type protein of interest. Each step in the editing workflow—transfection, gRNA cut, screening, outgrowth—has an inherent efficiency that depends on the approaches and cell line used. However, to obtain one successfully edited and viable clone can take two hundred cells that are negative for the desired

edit. Since the overall editing efficiency is around 0.5% to 1%, starting with large inputs of cells can be beneficial.

A newer technique to control protein levels, known as the auxin-inducible degron (AID) system, allows for reversible protein destruction through ubiquitination[44]. In this system, a plant-derived F-box protein is inserted into the genome. This F-box protein, commonly *Oryza sativa* TIR1 (OsTIR1), forms a Skp1-Cul1-F-box (SCF) ubiquitinating complex[44]. A second edit introduces another plant-derived motif, termed "degron", which can be attached as a tag on the protein of interest. The SCF complex is inactive until exposure to auxin, a plant hormone. The auxin-activated ubiquitinating complex then targets the degron-tagged protein and destroys it. When the auxin is removed, the protein expression returns in a few hours. Because this system requires two separate gene-editing events (knock-ins), generating an AID cell line takes time and resources. Additionally, the AID system has issues with basal degradation and can require large doses of auxin to work in some cell lines[45]. However, recently several systems have been developed to solve these problems[44,46].

Non-editing-based systems: Another common technique for protein knockdown is RNA interference (RNAi). In this technique, a synthetic small interfering RNA (siRNA) is designed to a gene of interest. This approach takes advantage of the RNA-induced silencing complex (RISC) for post-transcriptional gene silencing[47]. The RNAi technique is target-specific but often does not result in a complete knockdown. Consequently, often multiple siRNAs need to be designed to the same gene and used in parallel. Also, as with any nucleic acid-mediated technology, off-target effects are always a concern. Despite these challenges, RNAi is being pursued as a

promising biomedical therapeutic avenue[48,49]. Lastly, the Tet system can be used to turn a gene of interest on or off based on administration of tetracycline or a derivative like doxycycline[50]. The Tet system has been noted to have leaky expression, so characterization of the system is important to define the control of the gene of interest[51]. Additionally, this method is not suitable for all mammalian cell lines[52]. Many popular methods exist for protein knockdown or knockout. The method chosen depends upon the cell line and system of interest, but each technique will always have challenges to overcome.

REPORTER ASSAYS

Reporter assays are used in many areas of research, including toxicology. They are especially useful for the AHR pathway since so many structurally disparate ligands activate this receptor. In this regard, reporter genes can be used to assay whether a particular compound activates a target gene such as CYP1A1. Upregulation of CYP1A1 is a hallmark of AHR activation, so it often is used as a proxy for AHR signaling[53]. The reporter molecule that is used as a pathway readout can differ. There are many reporters available depending on the pathway of interest and technology available to a lab. Luciferase commonly is used because it is easy and lower cost. A disadvantage is that luciferase is most commonly an endpoint assay; cells must be killed to measure the relative light units (RLUs) of luciferase activity. Fluorescence is another common reporter approach that is easy and low-cost, and it has the added advantage of cell survival past the time of measurement. Additionally, many fluorophores exist, so multiple reporters within the same cell could be assayed if desired. A disadvantage of fluorescence is that its effectiveness can depend upon the available, often expensive, microscope and hardware such as filter cubes and light sources. Historically, enzymes such as chloramphenicol acetyl transferase (CAT) have been

used as reporters, but they have fallen out of favor with the increased ease and simplicity of luciferase and fluorescence[54]. Luciferase dominates as the reporter system for the AHR pathway, but fluorescence is increasing in popularity[55–62].

CELL CULTURE

Though few cell lines accurately model AHR pathway toxicity, immortalized cancer cell lines have a major utility in elucidating molecular signaling mechanisms[22]. Cell lines have been isolated from human cancers as well as many other species. Human cell lines more accurately model AHR signaling in human biology, and cell lines have been isolated from many different organ systems. However, many immortalized cancer cell lines are genomically aberrant and don't reflect a realistic portrait of the human genome (Table 4). These abnormal genomes can present challenges when editing cell lines at specific loci. Fortunately, the AHR pathway has long been studied using transient transfection, which eliminates most concerns over the ploidy and any mutations at a locus of interest. Non-human cell lines can be used in the subdiscipline of ecotoxicology to assay the effects environmental contaminants may have on wildlife and ecosystems. Additionally, the AHR pathway has long been studied in model organisms like mice and rats, and some of the earliest cell lines were derived from mouse and rat liver cells[63–68]. Non-human cell lines can produce prodigious amounts of protein, and many come from systems that are suitable models for human biology as well. Specifically to the AHR pathway, many strains of mice harbor the high-responding AHR-b allele, which makes them more sensitive to ligand and the effects of AHR induction[69,70]. The following discussion focuses on the top five cell lines from Table 2.

CELL LINES

HepG2: HepG2 cells were derived from a human hepatoblastoma from a 15-year-old Caucasian male, and they are an exceedingly popular cell line for AHR biology, as well as other sytems. Their genome is well-studied and characterized due to this popularity[71,72]. HepG2 cells grow in a contact-dependent manner. Therefore, isolating clones from a single cell can be difficult, but culturing cells on a matrix like gelatin or matrigel improves outcomes. Their gene expression profile for AHR biology is ideal (i.e. CYP1A1, CYP1A2, CYP1B1, and AHRR are all inducible)[73]. Additionally, at most loci HepG2 cells are diploid[27].

MCF7: The MCF7 cell line was derived from the breast cancer of a 69-year-old Caucasian female, and this cell line is commonly used for AHR research. Since they are estrogen-sensitive, MCF7 cells can also be used to study the estrogen receptor pathway[74,75]. In regards to the AHR pathway, MCF7 cells are tetraploid at the locus for AHRR, though the AHR pathway is still inducible in these cells[27,73]. They do not have the same issues with contact-dependent growth as HepG2.

H4IIE: The H4IIE cell line was derived from a rat hepatocellular carcinoma and is one of the most-used non-human cell lines for AHR biology (Table 2). The H4IIE cell line has been used to study the AHR pathway since the early 1980s. A popular version of this cell line contains a luciferase reporter downstream of a DRE; it has been used to assay the AHR-activating capabilities of many compounds as a surrogate for toxicity[56,76,77]. As such, these reporter cells have been used to study ligand potency and strength of activation. The H4IIE cell line has also

been used to elucidate AHR pathway characteristics and is often used as a comparator of HepG2 cells[56,77].

Hepa1c1c7: Hepa1c1c7 (Hepa1) cells were derived from a mouse hepatocellular carcinoma and have also been a popular system to study AHR (Table 2). Hepa1 cells have been edited to form various reporter systems as well[62,78,79]. These cells are similar to H4IIE cells in that they have been used since the early 1980s for similar purposes, namely studies into compound toxicity and AHR pathway mechanisms. They can be used in comparison with HepG2 and H4IIE cells[80], and they are optimal for AHR studies since they contain the more-sensitive AHR-b allele[70].

HeLa: HeLa cells were derived from a cervical tumor from Henrietta Lacks, a 30-year-old African American woman, in the early 1950s. They are notable as the first immortalized human cell line, and they have been used as a scientific workhorse since then. Due to their popularity and early availability, the HeLa line genome is well-studied[81]. HeLa cells are easy to work with, grow prodigiously, and accept edits readily; unfortunately, they also contaminate many other cell lines[23,25]. They have a naturally high amount of AHRR produced due to having six copies of chromosome five; additionally, the AHR pathway genes CYP1A1, CYP1A2, and CYP1B1 are relatively uninducible[73]. Despite this genetic profile, HeLa cells have been used for AHR research for many years and may act as an interesting comparator to other, more-inducible cell lines.

SUMMARY

Immortalized cell lines provide a quick, low-cost way to study the details of AHR pathway signaling mechanisms in a monoculture. This allows researchers to interpret experimental results without the added complexity of a whole organism. Cell lines are generally not ideal for studying AHR pathway toxicity, but cell lines constitute a valuable tool for elucidating AHR molecular biology. Knowledge gained in cell culture can then be applied to whole-organism models to collect relevant physiological data.

Creating edited cell lines is a much faster process than establishing an edited whole-organism model, and many techniques exist to manipulate cell line genomes. Cells may be edited with CRISPR, transient transfection, or stable transfection, and FACS and limiting dilution are two ways to achieve a monoclonal cell population. Creating a cell line model takes on the scale of weeks, and running experiments once the tools have been created usually takes only days to get data.

The AHR pathway has many cell lines derived from different tissues that may be edited to study signaling. Though the most popular are discussed in this review, many other ligand-inducible cell lines exist and may be used to interrogate the AHR signaling pathway.

TABLES

Table 1. Advantages and disadvantages of in vitro and in vivo model systems.

Method	Overall Benefit	Advantages	Disadvantages
Cell line (in vitro)	Exploration of basic signaling mechanisms and pathway characterization	Models pathway mechanics	Doesn't recapitulate complex biology
		Fast, easy, low-cost	Each cell line behaves in a unique manner
		Simplifies biology to a monolayer	
		Can tailor to tissue type	
Whole organism (in vivo)	Models the interplay between tissues/biological systems in a whole organism	Models toxicity	Expensive
		Results are physiologically relevant	Models take a long time to generate and/or breed
		Captures biological complexity	

Table 2A. All human cell lines identified for this review ranked most to least publications.

	Cell Line	Species of origin	Tissue of origin	Search term	Total
1	HepG2	human	liver	hepg2	725
2	MCF7	human	breast	mcf7	487
3	HeLa	human	cervix	hela	174
4	HEK-293	human	kidney	hek293	156
5	Caco2	human	colon	caco2	143
6	general	human	blood	hematopoeitic stem cells	125
7	general	human	skin	keratinocytes AND cell lines	113
8	MDA-MB-231	human	breast	mda-mb-231	104
9	A549	human	lung	a549	102
10	HaCaT	human	skin	hacat	78
11	Huh7	human	liver	huh7	67
12	HL 60	human	blood	hl 60	60
13	general	human	epithelium	squamous cell carcinoma AND cell lines	59
14	LNCaP	human	prostate	lncap	39
15	THP-1	human	blood	thp-1	36
16	LS-180	human	colon	ls180	34
17	HCT 116	human	colon	hct116	33
18	Hep3B	human	liver	hep3b	30
19	Jurkat	human	blood	jurkat	22
20	K562	human	bone marrow	k562	15
21	U87	human	brain	u87	11
22	ARPE-19	human	retina	arpe-19	7
23	SW-480	human	colon	sw480	7
24	ACHN	human	kidney	achn	6
25	DLD-1	human	colon	dld1	6
26	OVCAR-3	human	ovary	ovcar3	5
27	RT4	human	bladder	rt4	4
28	T24	human	bladder	t24	4
29	general	human	skin	sk-mel	4
30	OMC-3	human	ovary	omc-3	3
31	HT-1197	human	bladder	ht-1197	2
32	NEC14	human	testes	nec14	2
33	hFOB	human	bone	hfob	1
34	HME1	human	breast	hme1	0

Table 2B. All animal cell lines identified for this review ranked most to least publications.

	Cell Line	Species of origin	Tissue of origin	Search term	Total
1	H4IIE	rat	liver	h4iie	221
2	Hepa1	mouse	liver	hepa1	161
3	3T3	mouse	fibroblast	3t3	80
4	COS7	african green monkey	kidney	cos7	63
5	COS1	african green monkey	kidney	cos1	54
6	CHO	hamster	ovary	chinese hamster ovary	42
7	GH3	rat	pituitary gland	gh3	23
8	CV1	african green monkey	kidney	cv1	20
9	MDCK	dog	kidney	mdck	12
10	LMH	chicken	liver	lmh	10
11	LLC-PK1	pig	kidney	llc-pk1	8
12	Vero	african green monkey	kidney	vero	7
13	YAMC	mouse	colon	yamc	6
14	BHK21	hamster	kidney	bhk21	3
15	AML12	mouse	liver	aml12	1
16	DT40	chicken	b cell	dt40	1
17	ND-E	mouse/rat hybrid	brain	nd-e	0
18	ZF4	zebra fish	unknown	zf4	0
19	AB9	mouse	hybridoma	ab9	0

Table 3. Advantages and disadvantages of protein knockdown methods.

Protein Knockdown Method	Advantages	Disadvantages
CRISPR KO	Easy-to-make edit that results in reliable protein knockout	Non-reversible
	Cells require no treatment to get rid of protein	Another independently derived clone should be used for drawing conclusions compared to WT cells
AID	Reversible, removal of auxin rescues protein expression	Getting the cells edited and a line established can be tricky due to the number of editing events required
	Protein degradation is fast and near complete	Expression may be leaky depending on AID proteins used
siRNA	Fast, easy, cheap, can be used for any gene	Likely need more than 1 siRNA designed per gene
	No additional treatment to knock down protein	Incomplete gene silencing means protein may still be made
Tet ON/OFF	Reversible: removal of tetracycline rescues cells to beginning phenotype	Leaky expression results in incomplete silencing/activation
	Only one piece to edit in	Must treat cells with antibiotics to get desired effect

Table 4. Genomic information and special considerations for most popular human cell lines used in AHR pathway research. Numbers denote ploidy at specified gene.

Cell Line Name	Tissue of origin	Cellosaurus ID	AHR	CYP 1A1	CYP 1A2	CYP 1B1	AHRR	Advantage/Disadvantage
HepG2	Liver	CVCL_0027	2	2	2	3	2	Contact-dependent growth, recommended to grow on a matrix
HuH7	Liver	CVCL_0336	2	3	3	2	2	Alternative liver cell line for HepG2
MCF7	Breast	CVCL_0031	2	3	3	3	4	Genetically abnormal at most AHR pathway loci; estrogen-responsive breast cancer line
MDA-MB-231	Breast	CVCL_0062	3	2	2	3	4	Estrogen, progesterone, and androgen receptor negative breast cancer line
HeLa	Cervix	CVCL_0030	2	2	2	2	6	Naturally high AHRR levels; AHR pathway is barely inducible
HCT 116	Colon	CVCL_0291	2	2	2	2	2	Less inducible than HepG2; grow robustly
A549	Lung	CVCL_0023	3	2	2	4	3	Naturally low AHRR levels; AHR pathway is barely inducible
LNCaP	Prostate	CVCL_0395	2	2	2	2	2	Mutated androgen receptor, responsive to broad hormones
HaCaT	Skin	CVCL_0038	4	4	3	3	4	Models dioxin toxicity; spontaneously immortalized, not from a cancer

FIGURES

Figure 1. Canonical AHR pathway signaling cascade. (1) Lipophilic ligand diffuses through cell membrane. (2) Ligand binds AHR PAS B domain and shifts AHR conformation to reveal nuclear localization signal (NLS). (3) AHR-ligand complex translocates to nucleus due to NLS. (4) Chaperone proteins fall off AHR once in nucleus. (5) AHR dimerizes with partner ARNT. (6) AHR-ARNT dimer binds DNA at dioxin-responsive element (DRE). (7) Canonical AHR pathway genes CYP1A1, 1A2, 1B1, and AHRR are upregulated. (8) A. AHRR, localized in nucleus, dimerizes with ARNT. B. CYPs, anchored in endoplasmic reticulum, bind available ligand. (9) A. AHRR-ARNT complex represses pathway signaling via unknown mechanism. B. CYPs, especially CYP1A1, metabolize ligand, thereby removing impetus for signaling.

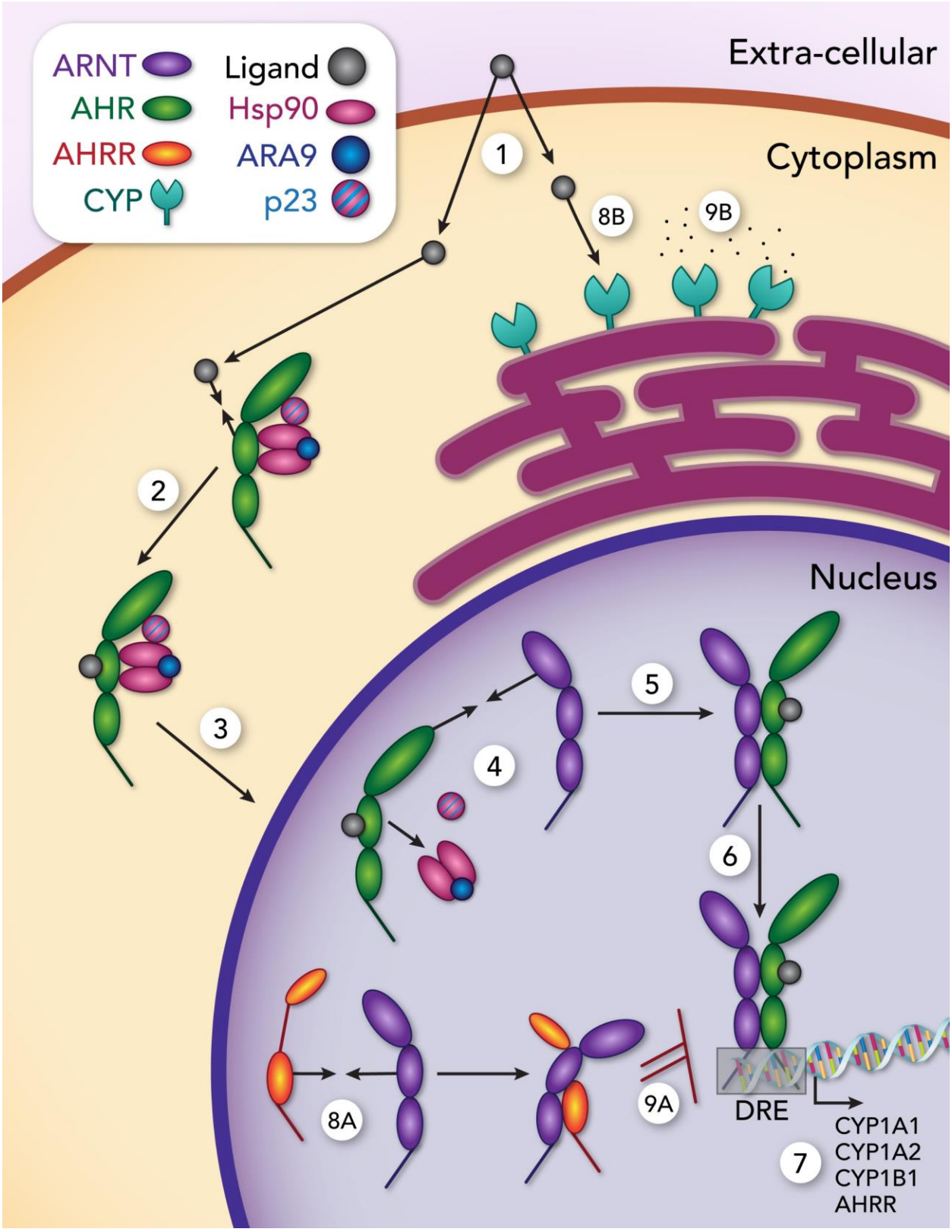

ARNT
AHR
AHRR
CYP
Ligand
Hsp90
ARA9
p23
Extra-cellular
Cytoplasm
Nucleus
1
2
3
4
5
6
7
8A
8B
9A
9B
DRE
CYP1A1
CYP1A2
CYP1B1
AHRR

Figure 2. General mechanism of CRISPR/Cas gene editing. (1) gRNA and Cas9 form complex. (2) gRNA-Cas9 and knock-in (KI) plasmid are delivered to cell nucleus. (3) gRNA targets Cas9 to locus of interest. (4) Cas9 cuts 3 bp upstream of beginning of PAM. (5) Plasmid homology arms target the KI sequence to location of cut. (6) DNA polymerase uses plasmid sequence as template. (7) KI sequence is edited into genome.

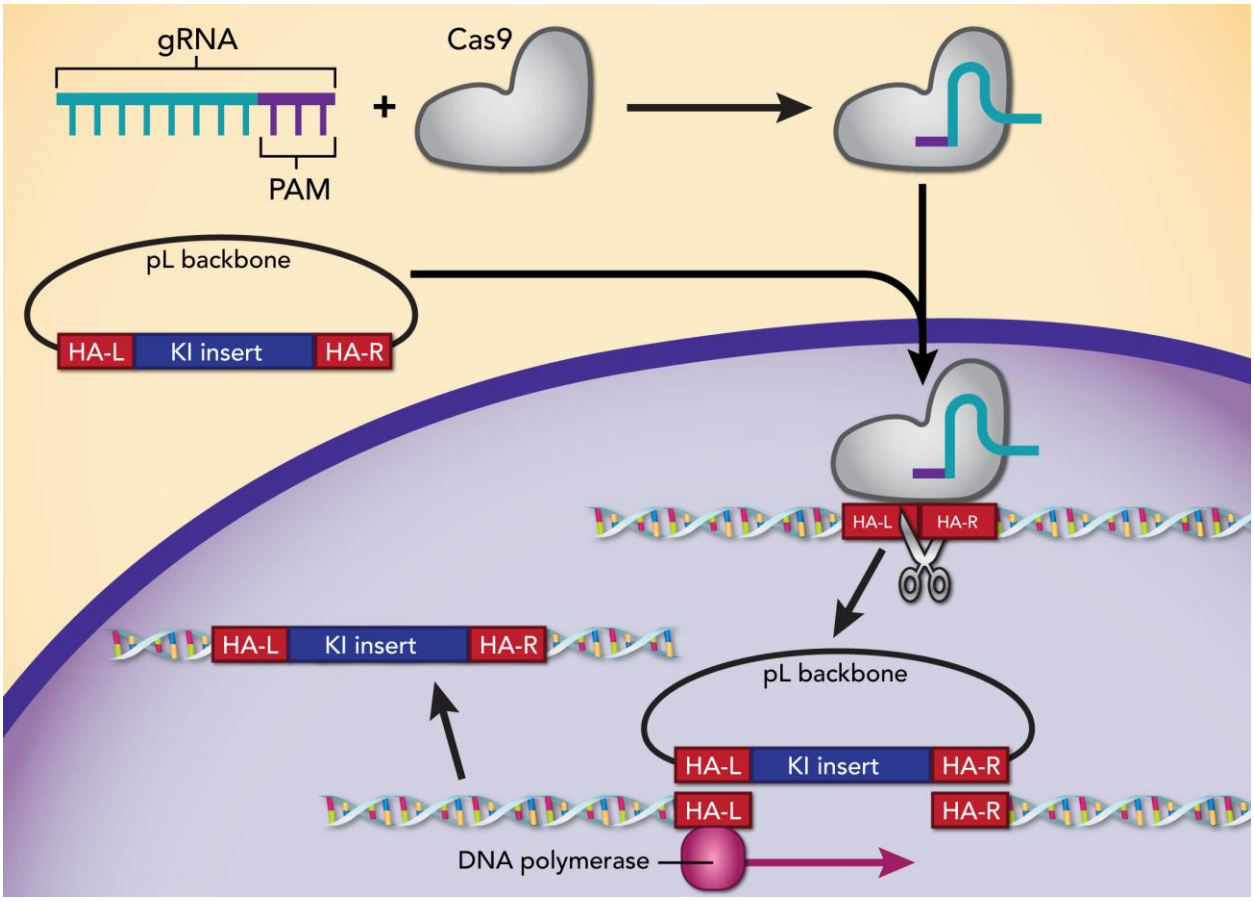
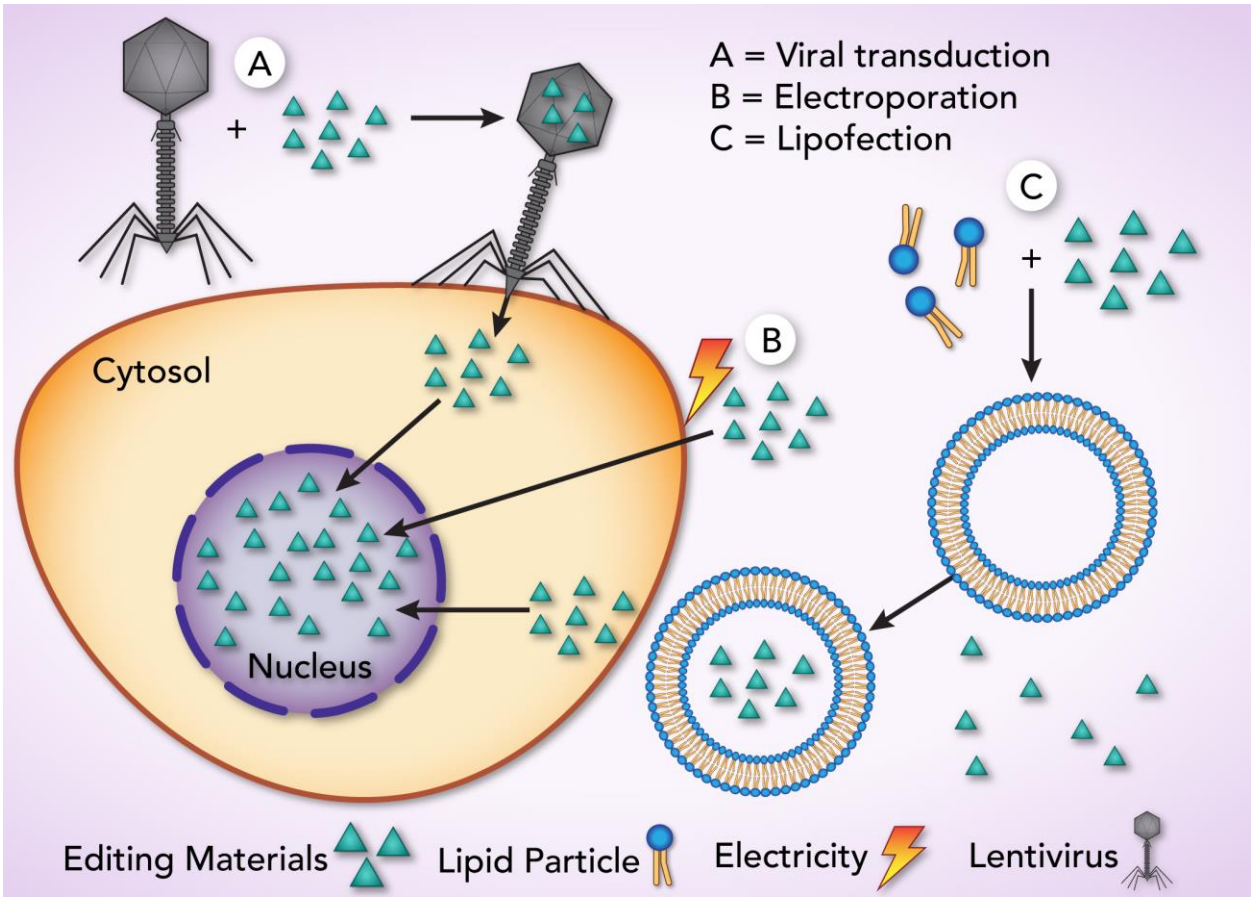

gRNA
Cas9
PAM
pL backbone
HA-L
KI insert
HA-R
HA-L
HA-R
HA-L
KI insert
HA-R
pL backbone
HA-L
KI insert
HA-R
HA-L
HA-R
DNA polymerase

Figure 3. Methods of editing material introduction to a cell. A. (1) Editing materials are packaged in lentiviral shell. (2) Cells are infected with lentivirus. (3) Lentivirus delivers materials into cell. B. (1) Cells are zapped with electric current, making the cell membrane porous. (2) Editing materials enter cell through membrane gaps. C. (1) Editing materials and lipid particles are mixed. (2) Lipid particles form liposomes and coat editing materials. (3) Materials enter cell through endocytosis.


A = Viral transduction
B = Electroporation
C = Lipofection
A
B
C
Cytosol
Nucleus
Editing Materials
Lipid Particle
Electricity
Lentivirus

REFERENCES

1. 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. Off. J. Soc. Toxicol.* **118**, 554–563 (2010).

2. Nguyen, T. *et al.* Endothelial Aryl Hydrocarbon Receptor Nuclear Translocator Mediates the Angiogenic Response to Peripheral Ischemia in Mice With Type 2 Diabetes Mellitus. *Front. Cell Dev. Biol.* **9**, 691801 (2021).

3. Mimura, J., Ema, M., Sogawa, K. & Fujii-Kuriyama, Y. Identification of a novel mechanism of regulation of Ah (dioxin) receptor function. *Genes Dev.* **13**, 20–25 (1999).

4. Baba, T. *et al.* Structure and Expression of the Ah Receptor Repressor Gene. *J. Biol. Chem.* **276**, 33101–33110 (2001).

5. Evans, B. R. *et al.* Repression of Aryl Hydrocarbon Receptor (AHR) Signaling by AHR Repressor: Role of DNA Binding and Competition for AHR Nuclear Translocator. *Mol. Pharmacol.* **73**, 387–398 (2008).

6. Oshima, M., Mimura, J., Yamamoto, M. & Fujii-Kuriyama, Y. Molecular mechanism of transcriptional repression of AhR repressor involving ANKRA2, HDAC4, and HDAC5. *Biochem. Biophys. Res. Commun.* **364**, 276–282 (2007).

7. Oshima, M., Mimura, J., Sekine, H., Okawa, H. & Fujii-Kuriyama, Y. SUMO Modification Regulates the Transcriptional Repressor Function of Aryl Hydrocarbon Receptor Repressor. *J. Biol. Chem.* **284**, 11017–11026 (2009).

8. Sakurai, S., Shimizu, T. & Ohto, U. The crystal structure of the AhRR-ARNT heterodimer reveals the structural basis of the repression of AhR-mediated transcription. *J. Biol. Chem.* **292**, 17609–17616 (2017).

9. Kemp Jacobsen, K., Johansen, J. S., Mellemgaard, A. & Bojesen, S. E. AHRR (cg05575921) methylation extent of leukocyte DNA and lung cancer survival. *PloS One* **14**, e0211745 (2019).

10. Zudaire, E. *et al.* The aryl hydrocarbon receptor repressor is a putative tumor suppressor gene in multiple human cancers. *J. Clin. Invest.* **118**, 640–650 (2008).

11. Li, Y. *et al.* Poor Prognosis of Gastric Adenocarcinoma with Decreased Expression of AHRR. *PLOS ONE* **7**, e43555 (2012).

12. Yang, X. *et al.* The aryl hydrocarbon receptor constitutively represses c-myc transcription in human mammary tumor cells. *Oncogene* **24**, 7869–7881 (2005).

13. Vogel, C. F. A. *et al.* Transgenic Overexpression of Aryl Hydrocarbon Receptor Repressor (AhRR) and AhR-Mediated Induction of CYP1A1, Cytokines, and Acute Toxicity. *Environ. Health Perspect.* **124**, 1071–1083 (2016).

14. Stockinger, B., Shah, K. & Wincent, E. AHR in the intestinal microenvironment: safeguarding barrier function. *Nat. Rev. Gastroenterol. Hepatol.* **18**, 559–570 (2021).

15. LaPres, J. J., Glover, E., Dunham, E. E., Bunger, M. K. & Bradfield, C. A. ARA9 modifies agonist signaling through an increase in cytosolic aryl hydrocarbon receptor. *J. Biol. Chem.* **275**, 6153–6159 (2000).

16. Nukaya, M. *et al.* The Aryl Hydrocarbon Receptor-interacting Protein (AIP) Is Required for Dioxin-induced Hepatotoxicity but Not for the Induction of the Cyp1a1 and Cyp1a2 Genes. *J. Biol. Chem.* **285**, 35599–35605 (2010).

17. Avilla, M. N., Malecki, K. M. C., Hahn, M. E., Wilson, R. H. & Bradfield, C. A. The Ah Receptor: Adaptive Metabolism, Ligand Diversity, and the Xenokine Model. *Chem. Res. Toxicol.* **33**, 860–879 (2020).

18. Gu, Y. Z., Hogenesch, J. B. & Bradfield, C. A. The PAS superfamily: Sensors of environmental and developmental signals. *Annu. Rev. Pharmacol. Toxicol.* **40**, 519–561 (2000).

19. Hahn, M. E. Aryl hydrocarbon receptors: diversity and evolution. *Chem. Biol. Interact.* **141**, 131–160 (2002).

20. Nebert, D. W., Dalton, T. P., Okey, A. B. & Gonzalez, F. J. Role of Aryl Hydrocarbon Receptor-mediated Induction of the CYP1 Enzymes in Environmental Toxicity and Cancer*. *J. Biol. Chem.* **279**, 23847–23850 (2004).

21. Vazquez-Rivera, E. *et al.* The aryl hydrocarbon receptor as a model PAS sensor. *Toxicol. Rep.* **9**, 1–11 (2022).

22. Knutson, J. C. & Poland, A. 2,3,7,8-Tetrachlorodibenzo-p-dioxin: Failure to demonstrate toxicity in twenty-three cultured cell types. *Toxicol. Appl. Pharmacol.* **54**, 377–383 (1980).

23. Lucey, B. P., Nelson-Rees, W. A. & Hutchins, G. M. Henrietta Lacks, HeLa cells, and cell culture contamination. *Arch. Pathol. Lab. Med.* **133**, 1463–1467 (2009).

24. Ye, F., Chen, C., Qin, J., Liu, J. & Zheng, C. Genetic profiling reveals an alarming rate of cross-contamination among human cell lines used in China. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **29**, 4268–4272 (2015).

25. Capes-Davis, A. *et al.* Check your cultures! A list of cross-contaminated or misidentified cell lines. *Int. J. Cancer* **127**, 1–8 (2010).

26. Verma, A., Verma, M. & Singh, A. Animal tissue culture principles and applications. *Anim. Biotechnol.* 269–293 (2020) doi:10.1016/B978-0-12-811710-1.00012-4.

27. Broad Institute Cancer Cell Line Encyclopedia (CCLE). https://portals.broadinstitute.org/ccle.

28. Cell Lines in the In Vitro Screen | NCI-60 Human Tumor Cell Lines Screen | Discovery & Development Services | Developmental Therapeutics Program (DTP). https://dtp.cancer.gov/discovery_development/nci-60/cell_list.htm.

29. Makarova, K. S. *et al.* Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477 (2011).

30. Swarts, D. C. & Jinek, M. Cas9 versus Cas12a/Cpf1: Structure-function comparisons and implications for genome editing. *Wiley Interdiscip. Rev. RNA* **9**, e1481 (2018).

31. Lino, C. A., Harper, J. C., Carney, J. P. & Timlin, J. A. Delivering CRISPR: a review of the challenges and approaches. *Drug Deliv.* **25**, 1234–1257 (2018).

32. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).

33. Zhang, F., Wen, Y. & Guo, X. CRISPR/Cas9 for genome editing: progress, implications and challenges. *Hum. Mol. Genet.* **23**, R40-46 (2014).

34. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).

35. Jiang, F. & Doudna, J. A. CRISPR-Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **46**, 505–529 (2017).

36. Torres-Ruiz, R. & Rodriguez-Perales, S. CRISPR-Cas9 technology: applications and human disease modelling. *Brief. Funct. Genomics* **16**, 4–12 (2017).

37. Naeem, M., Majeed, S., Hoque, M. Z. & Ahmad, I. Latest Developed Strategies to Minimize the Off-Target Effects in CRISPR-Cas-Mediated Genome Editing. *Cells* **9**, E1608 (2020).

38. Ma, Y., Zhang, L. & Huang, X. Genome modification by CRISPR/Cas9. *FEBS J.* **281**, 5186–5193 (2014).

39. Carroll, D. Genome engineering with zinc-finger nucleases. *Genetics* **188**, 773–782 (2011).

40. Bhardwaj, A. & Nain, V. TALENs-an indispensable tool in the era of CRISPR: a mini review. *J. Genet. Eng. Biotechnol.* **19**, 125 (2021).

41. Wright, W. D., Shah, S. S. & Heyer, W.-D. Homologous recombination and the repair of DNA double-strand breaks. *J. Biol. Chem.* **293**, 10524–10535 (2018).

42. Bak, R. O., Gomez-Ospina, N. & Porteus, M. H. Gene Editing on Center Stage. *Trends Genet. TIG* **34**, 600–611 (2018).

43. 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).

44. Yesbolatova, A. *et al.* The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nat. Commun.* **11**, 5701 (2020).

45. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* **6**, 917–922 (2009).

46. Li, S., Prasanna, X., Salo, V. T., Vattulainen, I. & Ikonen, E. An efficient auxin-inducible degron system with low basal degradation in human cells. *Nat. Methods* **16**, 866–869 (2019).

47. Han, H. RNA Interference to Knock Down Gene Expression. *Methods Mol. Biol. Clifton NJ* **1706**, 293–302 (2018).

48. Gavrilov, K. & Saltzman, W. M. Therapeutic siRNA: Principles, Challenges, and Strategies. *Yale J. Biol. Med.* **85**, 187–200 (2012).

49. Dana, H. *et al.* Molecular Mechanisms and Biological Functions of siRNA. *Int. J. Biomed. Sci. IJBS* **13**, 48–57 (2017).

50. Das, A. T., Tenenbaum, L. & Berkhout, B. Tet-On Systems For Doxycycline-inducible Gene Expression. *Curr. Gene Ther.* **16**, 156–167 (2016).

51. Zhou, Y., Lei, C. & Zhu, Z. A low-background Tet-On system based on post-transcriptional regulation using Csy4. *PloS One* **15**, e0244732 (2020).

52. Nishijima, H., Yasunari, T., Nakayama, T., Adachi, N. & Shibahara, K. Improved applications of the tetracycline-regulated gene depletion system. *Biosci. Trends* **3**, 161–167 (2009).

53. Whitlock, J. P. Induction of Cytochrome P4501a1. *Annu. Rev. Pharmacol. Toxicol.* **39**, 103–125 (1999).

54. Fujisawa-Sehara, A., Sogawa, K., Yamane, M. & Fujii-Kuriyama, Y. Characterization of xenobiotic responsive elements upstream from the drug-metabolizing cytochrome P-450c gene: a similarity to glucocorticoid regulatory elements. *Nucleic Acids Res.* **15**, 4179–4191 (1987).

55. Ziccardi, M. H., Gardner, I. A. & Denison, M. S. Development and Modification of a Recombinant Cell Bioassay to Directly Detect Halogenated and Polycyclic Aromatic Hydrocarbons in Serum. *Toxicol. Sci.* **54**, 183–193 (2000).

56. Doan, T. Q., Connolly, L., Igout, A., Muller, M. & Scippo, M. L. In vitro differential responses of rat and human aryl hydrocarbon receptor to two distinct ligands and to different polyphenols. *Environ. Pollut. Barking Essex 1987* **265**, 114966 (2020).

57. Chang, H. J. *et al.* Ascorbic acid suppresses the 2,3,7,8-tetrachloridibenxo-p-dioxin (TCDD)-induced CYP1A1 expression in human HepG2 cells. *Toxicol. In Vitro* **23**, 622–626 (2009).

58. Haarmann-Stemmann, T. *et al.* Analysis of the transcriptional regulation and molecular function of the aryl hydrocarbon receptor repressor in human cell lines. *Drug Metab. Dispos. Biol. Fate Chem.* **35**, 2262–2269 (2007).

59. Megna, B. W. *et al.* The aryl hydrocarbon receptor as an antitumor target of synthetic curcuminoids in colorectal cancer. *J. Surg. Res.* **213**, 16–24 (2017).

60. Hayashi, A. & Denison, M. S. Development of a novel recombinant cell line for detection and characterization of Ah receptor nuclear translocation in intact cells. *Toxicol. Vitro Int. J. Publ. Assoc. BIBRA* **66**, 104873 (2020).

61. Peters, A. K. *et al.* Interactions of polybrominated diphenyl ethers with the aryl hydrocarbon receptor pathway. *Toxicol. Sci. Off. J. Soc. Toxicol.* **92**, 133–142 (2006).

62. Nagy, S. R., Sanborn, J. R., Hammock, B. D. & Denison, M. S. Development of a Green Fluorescent Protein-Based Cell Bioassay for the Rapid and Inexpensive Detection and Characterization of Ah Receptor Agonists. *Toxicol. Sci.* **65**, 200–210 (2002).

63. Whitlock, J. P. & Gelboin, H. V. Aryl hydrocarbon (benzo(a)pyrene) hydroxylase induction in rat liver cells in culture. *J. Biol. Chem.* **249**, 2616–2623 (1974).

64. Legraverend, C. *et al.* Aryl hydrocarbon hydroxylase induction by benzo[a]anthracene: regulatory gene localized to the distal portion of mouse chromosome 17. *Genetics* **107**, 447–461 (1984).

65. Greiner, J. W., Malan-Shibley, L. B. & Janss, D. H. Characteristics of aryl hydrocarbon hydroxylase activity in rat mammary epithlial cells grown in primary culture. *Chem. Biol. Interact.* **27**, 323–334 (1979).

66. Bradlaw, J. A. & Casterline, J. L. Induction of enzyme activity in cell culture: a rapid screen for detection of planar polychlorinated organic compounds. *J. - Assoc. Off. Anal. Chem.* **62**, 904–916 (1979).

67. Bigelow, S. W. & Nebert, D. W. The Ah regulatory gene product. Survey of nineteen polycyclic aromatic compounds' and fifteen benzo[a]pyrene metabolites' capacity to bind to the cytosolic receptor. *Toxicol. Lett.* **10**, 109–118 (1982).

68. Safe, S. *et al.* Effects of structure on binding to the 2,3,7,8-TCDD receptor protein and AHH induction--halogenated biphenyls. *Environ. Health Perspect.* **61**, 21–33 (1985).

69. Poland, A. & Glover, E. 2,3,7,8-Tetrachlorodibenzo-p-dioxin: Segregation of Toxicity with the Ah Locus. *Mol. Pharmacol.* **17**, 86–94 (1980).

70. Poland, A. & Glover, E. Characterization and strain distribution pattern of the murine Ah receptor specified by the Ahd and Ahb-3 alleles. *Mol. Pharmacol.* **38**, 306–312 (1990).

71. Zhou, B. *et al.* Haplotype-resolved and integrated genome analysis of the cancer cell line HepG2. *Nucleic Acids Res.* **47**, 3846–3861 (2019).

72. Ruoß, M. *et al.* Epigenetic Modifications of the Liver Tumor Cell Line HepG2 Increase Their Drug Metabolic Capacity. *Int. J. Mol. Sci.* **20**, 347 (2019).

73. Tsuchiya, Y., Nakajima, M., Itoh, S., Iwanari, M. & Yokoi, T. Expression of aryl hydrocarbon receptor repressor in normal human tissues and inducibility by polycyclic aromatic hydrocarbons in human tumor-derived cell lines. *Toxicol. Sci. Off. J. Soc. Toxicol.* **72**, 253–259 (2003).

74. Kanno, Y., Takane, Y., Takizawa, Y. & Inouye, Y. Suppressive effect of aryl hydrocarbon receptor repressor on transcriptional activity of estrogen receptor alpha by protein-protein interaction in stably and transiently expressing cell lines. *Mol. Cell. Endocrinol.* **291**, 87–94 (2008).

75. 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**, e29545 (2012).

76. Larsson, M., Orbe, D. & Engwall, M. Exposure time-dependent effects on the relative potencies and additivity of PAHs in the Ah receptor-based H4IIE-luc bioassay. *Environ. Toxicol. Chem.* **31**, 1149–1157 (2012).

77. Vrzal, R. *et al.* Activation of the aryl hydrocarbon receptor by berberine in HepG2 and H4IIE cells: Biphasic effect on CYP1A1. *Biochem. Pharmacol.* **70**, 925–936 (2005).

78. Hoffman, T. E. *et al.* Ultrasensitivity dynamics of diverse aryl hydrocarbon receptor modulators in a hepatoma cell line. *Arch. Toxicol.* **93**, 635–647 (2019).

79. Shi, H. *et al.* Concentration dependence of human and mouse aryl hydrocarbon receptor responsiveness to polychlorinated biphenyl exposures: Implications for aroclor mixtures. *Xenobiotica Fate Foreign Compd. Biol. Syst.* **49**, 1414–1422 (2019).

80. Dere, E., Lee, A. W., Burgoon, L. D. & Zacharewski, T. R. Differences in TCDD-elicited gene expression profiles in human HepG2, mouse Hepa1c1c7 and rat H4IIE hepatoma cells. *BMC Genomics* **12**, 193 (2011).

81. Adey, A. *et al.* The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. *Nature* **500**, 207–211 (2013).

Chapter 2: A Biotechnology Fellowship Prepares Graduate Students for Non-Academic Careers

ABSTRACT

The University of Wisconsin Biotechnology Center (UWBC) Genome Editing and Animal Models (GEAM) core provides cell and animal model generation services to campus investigators. The GEAM core utilizes cutting-edge technology to carry out their mission. Importantly, their available resources are focused on generating revenue as a fee-for-service facility. This focus leaves little time or capital for innovation. Core facilities like GEAM must find ways to use constrained resources to continue innovating and providing advanced services to their clients.

Graduate student fellows provide a source of untapped potential that would allow GEAM and other cores to innovate while meeting client demands. Graduate students are primed to explore new avenues of research, and they bring experience and expertise in learning new techniques or systems quickly and optimizing them for most efficient use.

Implementation of graduate student fellows would be a mutually beneficial arrangement for the student and the core in which they work. The core gains an extra worker who can take on projects permanent personnel might not have time for. The fellow, once trained, can contribute to customer projects as well. The fellow learns the workings of a non-academic lab, and she gains marketable skills that contribute to her thesis work. Additionally, these abilities and experiences increase student desirability to future employers.

This white paper details the author's contributions to GEAM as a pilot graduate student fellow. It details the benefits both to the core facility and the student, and it establishes the viability and usefulness of a fellow program.

INTRODUCTION

The University of Wisconsin Biotechnology Center (UWBC) Genome Editing and Animal Models (GEAM) core provides cell and animal editing services to the UW campus. They provide years of expertise in their area of specialization, and they are a cutting-edge resource for researchers on campus. However, as a fee-for-service entity, grants and revenue from services are not always reliable sources of income. The GEAM core must find creative ways to test new technologies while maintaining a steady flow of customer projects.

Graduate students present an untapped resource on campus for facilities like GEAM. Establishing a graduate student fellow program would provide skilled, intelligent workers focused on implementation and optimization of new techniques. Graduate students have an appetite for exploration, knowing their efforts may result in an exciting discovery at best or elimination of a research project path at worst. They also have an ability to learn new and complex information quickly and utilize it practically. Graduate students are ideal candidates to handle the non-standard tasks required for establishing a new technology or workflow while normal core projects proceed to meet required deadlines. Once the graduate student has optimized a new procedure, she can then translate the new technology to the required personnel. This model would allow for easier development and implementation of new technologies into the core without personnel wasting precious time and resources on projects that may go nowhere.

Areas in which the student would strengthen the core can be broken down into three categories: Standard Projects, Maintenance, and Special Projects (Figure 1). “Standard Projects” refers to the work customers request from the core, all of which flows into a set pipeline and is expected

to be completed in a timely manner. “Maintenance” refers to the day-to-day lab maintenance work vital to a smooth operation. “Special Projects” refers to the exploration and optimization of new techniques that clients never see but which benefits the projects they submit; this is the area where graduate student fellows can especially benefit core facilities.

The purpose of this white paper is to summarize the author’s work in GEAM and to establish the usefulness and value of integrating a graduate student fellow into the project-focused, industry-like setting of a core facility. Additionally, the white paper will detail the benefits to the participating fellow.

METHODS

The Methods section details the fellow's work in the GEAM core under Standard Projects, Maintenance, and Special Projects. The fellow spent a year at GEAM developing her skills and contributing to the core's function.

STANDARD PROJECTS

Cell Lines: Cell line editing projects form the bulk of GEAM's project load. Moving projects forward in a timely fashion both increases customer satisfaction and allows the core to take on more projects, thereby generating optimal revenue flows to fund core improvements and innovation. A graduate student fellow can be a great help in this area. Fellows provide a skilled, intelligent laborer that, once trained, can advance standard cell line editing projects. This additional specialist creates a mutually beneficial working relationship; the GEAM core increases its project output while the graduate student learns the foundations of cell culture technique and begins to build her expertise. As the graduate student gains facility with cell culture, she can become a trusted resource to autonomously advance customer projects.

In her time at the UWBC GEAM core, this graduate student fellow contributed to many standard cell line projects. This work included cell maintenance to keep deliverable lines healthy or prepare unedited lines for CRISPR; clone expansion to grow large quantities of deliverable lines; and the hands-on work of editing cells with CRISPR and sorting for single cells via fluorescence-activated cell sorting (FACS). She participated in every area of the GEAM cell editing pipeline, including cell intake, editing, sorting, expansion, and product delivery. This work advanced core projects to completion, and it gave this fellow a comfort and familiarity with

genome editing and cell culture. She then used these skills to derive her thesis cell lines reported later in this document.

Miscellaneous Projects: As a core facility, GEAM must keep projects passing through at a steady rate to produce consistent income. The daily work, though not glamorous, provides important materials and, eventually, products that are delivered to the customer and further that lab's research. This is another area in which a graduate student fellow can and should contribute to the core's workflow.

The graduate student fellow participating in this pilot project contributed to the core by taking up backlogged or difficult projects languishing because the permanent staff were needed elsewhere. In many cases, this fellow successfully moved these projects to completion and learned techniques such as Sanger sequencing and PCR screening. She also gained valuable experience with troubleshooting said techniques. The fellow gained insight into the standard, day-to-day workings of a core facility and learned to think in terms of project delivery, timelines, and customer satisfaction, all aims not overtly necessary in an academic lab setting. This exposure allowed the fellow to get a feel for the mindset outside her usual setting and helped her prepare for a career in industry.

MAINTENANCE

In addition to hands-on project work, the fellow took care of maintenance duties essential in any lab setting. These duties included autoclaving biohazardous waste, ensuring incubator humidity and sterility, and assisting staff with lab upkeep. Additionally, in the fellow's time at GEAM she

aided the core in moving to and setting up a larger lab space. These endeavors helped the core run smoothly and freed up time for the permanent staff to focus on required work. The fellow learned to appreciate the work necessary to keep the core running, and she gained a foundation in lab maintenance critical for any level of scientist.

SPECIAL PROJECTS

Cell Lines: The GEAM core is responsible for cell editing projects brought to them by researchers on campus. GEAM relies on a familiar set of cell lines that are easily edited, grow quickly, and which the core has experience with. These cell lines include HEK293 cells and HeLa cells. Additions to this list of standard cell lines are valuable because they increase the core's offerings to potential customers. If a customer's desired cell line does not meet growth, expansion, or survival thresholds, one of these standard cell lines will be chosen. HepG2 cells are an example of a non-ideal editing candidate.

HepG2 cells are widely used for research in many areas but are particularly popular for toxicology studies. Unfortunately, HepG2 cells grow in a contact-dependent manner (Figure 2B) and do not easily survive dilution to a single-cell state, something required for establishing a monoclonal, CRISPR-edited population. The author tried several iterations of CRISPR editing and FACS sorting. Conditions to improve clonal outgrowth included conditioned media, treatment with EDTA to improve single cell suspension pre-FACS, and increased FBS percentage in the growth media to 20% up from 10% (volume). The above conditions did nothing to allow clonal outgrowth post-FACS (Figure 2C). The GEAM core prefers not to work with HepG2 knowing the above challenges.

Additionally, since HepG2 cells proved difficult in gene editing protocols, the author switched to HCT 116 cells for her thesis. HCT 116 cells were easy to edit and work with, and most importantly they exhibited high clonal outgrowth rates (~50%) and grew to confluence quickly from a single-cell state (approximately 10 days for the fastest growers). Based on the author's results, GEAM has now incorporated HCT 116 cells on their list of cell lines to offer clients.

A graduate student fellow working in the GEAM core benefits greatly from learning cell culture techniques and working with multiple cell lines. The fellow also may make progress on her thesis projects while contributing to the core; the author has been able to make three edited cell lines during her time at GEAM while handling other projects.

Automation: Automation is quickly becoming a staple in laboratory settings. Automation reduces hands-on time for personnel, allowing them to advance other tasks while running something in the background. Additionally, automation can reduce the time and resources spent on a repetitive workflow. Liquid handlers are one way of automating protocols that are run consistently and require many simple and repetitive pipetting steps. The GEAM core during this fellowship bought an Opentrons liquid handling robot, which uses Python scripting to automate workflows from nucleic acid bead cleanups to clone picking. Many scripts are pre-made and available on the Opentrons website, making the technology accessible to biologists with very little coding experience. On-boarding the Opentrons robot was one project of the author in her time at GEAM. The standard script from Opentrons was edited to reduce reagent and tip waste. The machine was optimized for the core's standard nucleic acid bead cleanup and library preparation

for Next-Gen Sequencing (NGS). Onboarding the Opentrons freed GEAM personnel from the physically wearing process of repetitive pipetting and allowed them to handle more complex tasks.

The GEAM core during this fellowship also added a CloneSelect Imager (CSI) from Molecular Devices to increase their throughput for cell cloning projects. The CSI machine allows clonal monitoring of 96-well plates from the single-cell state to an expanded clone without a scientist having to physically check growth in a microscope. Additionally, the CSI can image in fluorescent channels, making fluorescent screening faster and easier. The machine also keeps a photographic record that can be compiled over days. This allows one to track growth over time. Another automation project of the author was on-boarding the CSI machine, its robot arm, and a Liconic incubator. These three machines were developed to form a fully automated unit which is remote-controllable. The author set up this system for daily use in the core, troubleshooting various software and user issues, which took several months.

Due to her presence, the GEAM core had four separate machines in two systems on-boarded and optimized for use, and the fellow could then teach core personnel how to use each machine. This saved the core time and effort in onboarding, and the automation systems will save resources moving forward. The fellow gained a familiarity with automation and the various problems associated with it. This awareness will be a valuable and marketable skill for her future career. She also improved her teaching abilities by transmitting what she had learned to the core personnel who will use it after she finishes her fellowship.

Fluorescence Microscopy: Fluorescence microscopy is a widely used technique which uses fluorophores to image cellular events and structures. Within the GEAM core, fluorescence microscopy is used as a screening technique to ensure gene edits have been successful. A working and efficient microscopy system has hardware (filter cubes, light source) that is optimized for the excitation and emission spectra of the fluorophore(s) of choice. A mismatch in the wavelength ranges of these components results in decreased or absent fluor visibility, since the fluorescent protein (e.g. EGFP, mCherry) isn't excited properly or the filter cube is not allowing the excitation light to be transmitted. Several factors, including cell media, can affect background noise and, therefore, image clarity. An optimized fluorescence microscopy system allows an easy check for edited cell line success and guides subsequent clone picking and outgrowth.

One of the projects of this author was to troubleshoot the current core fluorescence microscope and set up a standard operating procedure (SOP) for core employees. The microscope in-use at the GEAM core had issues with imaging in the green channel with mNeonGreen as the fluor, and there were also issues with high background noise especially in the red channel.

To reduce background, several different types of media were assayed. This fellow learned Gibco FluoroBrite DMEM (cat. no. A1896701) is ideal for faint fluorophore expression (one to two copies knocked into a gene) because it has very little background. For cells that expressed a lot of the fluorophore (e.g. transient transfection of EGFP under the CMV promoter), regular DMEM containing phenol red was sufficient to achieve a clear image. Optimized imaging media allowed the core greater clarity and confidence when screening cells.

This fellow discovered the issues with mNeonGreen were due to a mismatch between the filter cube and the excitation and emission spectra of the fluorophore. A comparison between mNeonGreen and EGFP clearly demonstrated the mismatch, as EGFP shone bright and clear while mNeonGreen barely emitted over the background fluorescence. The GEAM core made a pivot to using EGFP over mNeonGreen where possible in cell editing projects. For existing projects the scope was optimized for maximum emission. Though these measures helped, mNeonGreen is still a suboptimal fluor for use with the hardware setup of GEAM's fluorescence scope.

The author's efforts optimized the fluorescence microscope for daily use in the core. Knowing the limitations of the system also allowed for a fluorophore pivot, and these limitations explained the issues with mNeonGreen. The fellow undertaking these projects gained a familiarity with fluorescence microscopy and considerations necessary before setting up a new system or choosing fluorophores for projects. Knowing fluorescence microscopy basics also gives flexibility in her thesis projects since the fellow may choose this as a form of imaging or gene expression quantification (if a reporter system is made).

CONCLUSION

Core facilities are an integral part of a vibrant and cutting-edge research community. Specifically, the GEAM core at UW-Madison provides state-of-the-art cell line editing services to campus, furthering research for many groups in many areas. Graduate student fellows not only help the core deliver standard projects to the labs they work with; they also are a source of experimentation and onboarding of new techniques and instrumentation to keep the core current and efficient (Table 1).

In the year she spent learning and contributing to GEAM, the graduate student fellow was able to add a cell line to GEAM's preferred list, onboard two separate automation systems, optimize GEAM's fluorescence microscopy system, and advance difficult miscellaneous projects to completion. This fellow gained priceless insight into the workings of a non-academic facility in addition to the technical skills she built working on standard and special projects. She can now add cell culture, fluorescence microscopy, Opentrons scripting and troubleshooting, CSI automation, NGS and Sanger sequencing, and technical communication to her list of skills. Her contributions strengthened the core but also advanced her thesis, and her graduate work is stronger because of it.

A post-internship evaluation reveals certain areas of strength and weakness that could improve the experience of future graduate students and cores entering into this relationship. A key to success for this internship was the substantial separation of the student from her thesis advisor's lab. It was understood that she would devote 100% of her time and effort to the core facility, and she was not forced to split herself between locations and supervisors. Without this understanding

in place, it is doubtful the internship would have been as successful as described, and it may have negatively impacted the student's experience through undue stress. An additional strength that contributed to the core's and student's success was the flexibility exhibited by both parties. The student was allowed to pursue projects that furthered her expertise and added value to her career prospects and graduate work; she also was willing to complete work that had no bearing on her graduate thesis but was important for the core. A balance between benefit to the core and benefit to the student must be found, but striking the right equilibrium can result in mutually beneficial relationships that further both parties' abilities beyond what they could accomplish separately.

An area for improvement exists within the realm of formalizing how the internship will be structured. A formalized understanding of what is expected of the student at what times would help reduce stress and provide a starting point for any communication about the nature of the relationship between core and student. As a pilot program, necessarily the experience described in this paper did not have a set format. No blueprint existed for how a core could integrate a graduate student fellow into its operation. After completing her fellowship at the UWBC GEAM core, this graduate student proposes the following model: the first four to six months of the fellowship will be devoted to training and onboarding the student to the core (Phase One). The graduate student will in effect be a new employee and will be taught every lab skill to the core's specifications. She will complete necessary, routine tasks like autoclaving biohazardous waste and miscellaneous upkeep integral to the efficient function of any lab. While not exciting, this first portion teaches the student important and marketable skills for her future career, and it provides familiarization with laboratory protocols and a solid foundation from which to build. In the next six to eight months of the fellowship, the student will use her expertise from Phase One

to complete special projects. She will pursue one or two projects outside the normal function of the core but that will add value to its future workings. For this graduate student, those projects included onboarding the CSI and Opentrons systems. In the future, these projects should be mutually agreed upon between student and supervisor. Additionally, in this second phase of the internship, the student will pursue work that directly relates to her thesis project. This is essential; the student must have a direct benefit to her thesis work in order to gain value from the internship. This work also provides a value to the core in that, while she is using their resources and space, no permanent core staff is responsible for the work done, so they may pursue their own support objectives for the campus. Her thesis project may also benefit the core in unexpected ways.

TABLES

Table 1. Benefits of fellow work to the core and the student by area of specialization. Impact of fellowship on thesis/career.

Specialization Area	Core Benefits	Student Benefits	Thesis/Career Progress
Cell Lines	New cell lines ID and screening; new cell lines for GE standard list	Facility with cell culture techniques for different cell lines; progress on thesis cell line work	Generated several cell lines used for thesis experiments; learned how to work with cells
Automation	Reduction of time and resources used to on-board automation; automation frees personnel from repetitive tasks, reduces time and reagents spent	Familiarization with automation use, issues, troubleshooting; exposure to high-throughput thinking; experience teaching others	Marketable skill for future career endeavors; facility with scripting and high-throughput technology bonus to biological skills
Fluorescence Microscopy	Hardware, fluorophores, workflow for easy screening of fluorescently labeled cells established and optimized	Facility with fluorescence microscopy, hardware and fluorophore considerations, and troubleshooting	Opened a new avenue for experiments to take, increased experiment possibilities
Miscellaneous	Advancement of backlogged or time-consuming projects	Introduction to standard core workflows, project delivery, non-academic lab considerations	Exposure to non-academic lab; ability to do basic lab maintenance
Overall	Addition and on-boarding of new technologies; optimization of existing workflows; advancement of difficult projects; all without disrupting existing personnel rhythms	Progress in thesis work, especially related to cell lines; improve skills with marketable techniques like automation and cell editing; familiarity with non-academic lab workings and goals	Accelerated thesis progress; experience with many in-demand skills

FIGURES

Figure 1) Fellow program hierarchy. Arrow size indicates allocation of resources; larger arrows equate to more time spent or greater output. Core Director oversees all staff, including fellow, and projects/lab operation. Principal Investigator is independent of core; fellow still reports to PI and is responsible for thesis progress. Core Staff advances daily work, focusing on standard projects, then maintenance, then special projects. Fellow divides time between maintenance, standard, or special projects. Thesis advances as a result of special project work. Customers only see work from standard core projects, but maintenance and special projects improve standard project quality and efficiency. Special projects strengthen the core while also feeding back into student thesis work.

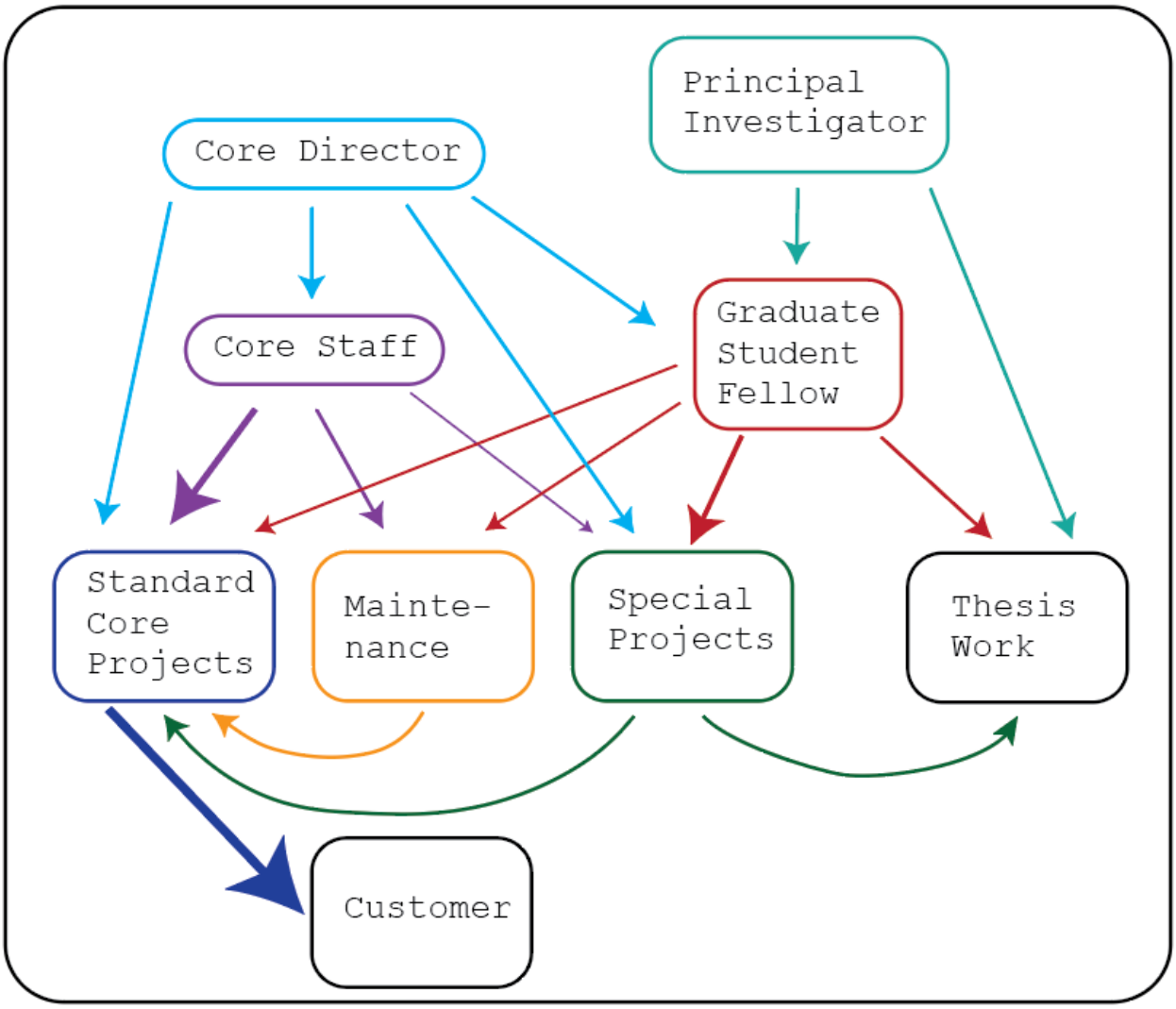
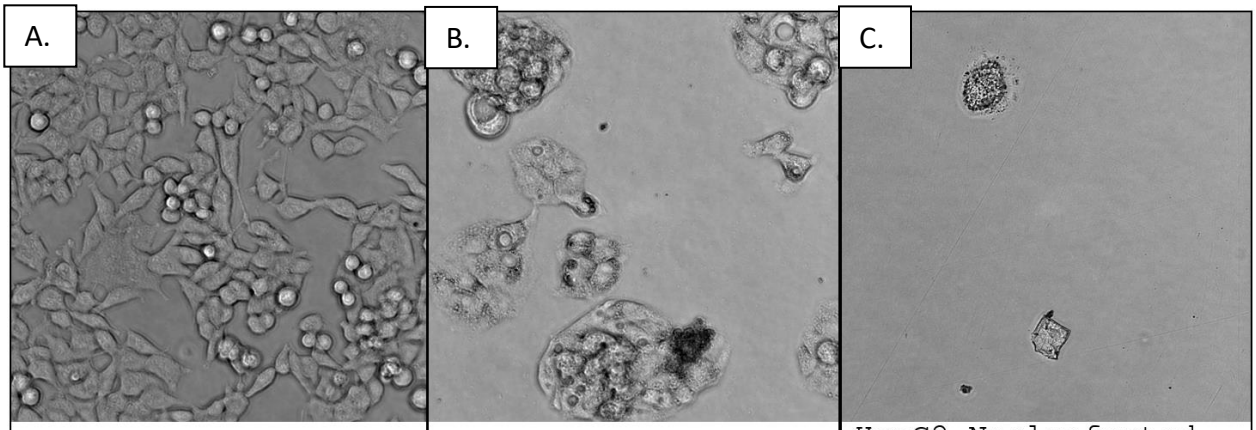

Principal Investigator
Core Director
Core Staff
Graduate Student Fellow
Standard Core Projects
Mainte-nance
Special Projects
Thesis Work
Customer

Figure 2) Comparison of HCT 116 and HepG2 cell morphology. A) Wild-type HCT cells; cells grow as individuals and are easily separated. **B)** Wild-type HepG2 cells; cells grow as compact nodes which are difficult to separate. **C)** Representative picture of HepG2 cells after cell sorting and mentioned interventions; cells quickly turn into apoptotic debris.



HCT 116, 40x	HepG2, 40x	HepG2-Nucleofected and Bulk Sorted, 40x

Chapter 3: Development of a Cell Culture Model for Endogenous Aryl Hydrocarbon Receptor Ligand Signaling

ABSTRACT

The aryl hydrocarbon receptor (AHR) pathway is responsible for regulating the metabolism of certain xenobiotics, sensing external chemical stimuli with planar aromatic structures, and additional unknown endogenous physiological roles. Tryptophan (TRP) and its metabolites have been proposed as a major endogenous source of AHR ligands. Here, we compare multiple TRP metabolizing paths that induce AHR signaling; i.e., alanine transaminase (ALT), aspartate aminotransferase (AST), D-amino acid oxidase (DAO), and ultraviolet (UV) light. In this chapter, we develop and optimize a cell culture model system for the study of this putative endogenous pathway. Our results support TRP metabolism as a major source of endogenous AHR ligands and support indole-3-pyruvate (I3P) as the common proligand each metabolic pathway produces. RNA-seq comparisons between DAO, I3P, and UV stimulation was used to identify a list of core AHR pathway genes upregulated between all pathways. This list was compared to published literature to further develop AHR pathway candidates, which may shed light on AHR biology.

INTRODUCTION

The aryl hydrocarbon receptor (AHR) has been known since the mid-1900s as the mediator of the toxicity of "Dioxin-Like-Compounds" (DLCs) and as a regulator of the metabolic activation and deactivation of certain polycyclic aromatic hydrocarbons (PAHs). The AHR is a ligand-activated transcription factor in the Per-ARNT-Sim (PAS) protein family[1–3]. The genes it upregulates include the AHR repressor (AHRR), and the cytochromes P450 (CYP) 1A1, 1A2, and 1B1[4–6]. In aggregate, these inducible gene products are thought to act in a feedback inhibition mode to attenuate or block the AHR signaling pathway by eliminating ligands and repressing signaling. More recent evidence supports the idea that the AHR is an integral part of normal development and immunity[7–11]. This idea is also consistent with the possibility that the AHR is activated in abnormal physiological processes via the binding of unknown endogenously generated ligands.

We have argued that tryptophan (TRP) metabolites and related aromatic amino acids (AAAs) are a major source of endogenous ligands of the AHR pathway[12,13]. Most important of these is L-tryptophan (L-TRP), known to be essential for eukaryotic protein synthesis and organism survival. Not only is L-TRP an essential component of cellular protein, this AAA is a precursor to many compounds such important signaling molecules such as serotonin and kynurenine[14]. The D enantiomer of tryptophan (D-TRP) is less utilized in eukaryotic systems but can be produced in the gut microbiome as a component of the bacterial cell wall. Moreover, D-TRP is also a contaminant formed in processed foods when L-TRP spontaneously racemizes due to high heat or pH[14].

An emerging idea is that AAAs like TRP, phenylalanine and tyrosine serve as metabolic precursors to AHR proligands in vivo. This TRP-proligand concept arose from the study of the plant metabolite indole-3-carbinol (I3C), a 3-substituted indole (like TRP) which was shown to be converted to ICZ and a number of related but uncharacterized AHR ligands in the low-pH environment of the stomach[15,16] (Figure 1A). This observation led to the model where certain biologically ubiquitous indoles, by virtue of a reactive leaving groups at the 3-position, might generally serve as AHR "proligands." We coined the term AHR proligand from the chemical carcinogenesis literature (procarcinogen) and is meant to denote the chemical's structural proximity to the ultimate AHR ligands generated by chemical condensation into multiple ring systems such as carbazoles (either spontaneous, pH induced, or UV induced).

We propose that the TRP metabolite indole-3pyruvate (I3P) is a biologically central AHR proligand in humans that can be derived enzymatically or through UV irradiation of TRP. For example, the aspartate aminotransferase (AST) enzyme, also known as the glutamic oxaloacetic transaminase (GOT1, ENSG00000120053, or GOT2, ENSG00000125166, for cytoplasmic or mitochondrial version, respectively), was the first enzyme known to produce AHR activation (via I3P) and was shown to be dependent upon L-TRP as a substrate (Figure 1B)[12]. This was followed by the later report that D-amino acid oxidase (DAO, ENGS00000110887) served as another enzyme that metabolizes D-TRP to the proligand I3P and activates the AHR pathway (Figure 1C)[17]. Other groups have demonstrated that ultraviolet radiation (UV) in the B spectrum (UVB) can activate the AHR pathway through an I3P intermediate step (Figure 1D). All of the pathways depicted in Figure 1 generate reactive 3-substituted indoles that condense to form

higher order structures, like carbazoles such as 6-formylindolo[3,2-b]carbazole (FICZ), that are potent AHR agonists[18–21].

Our model of TRP in AHR barrier physiology is as follows: Barriers tissues such as the skin, intestine and liver are under constant assault from pathogens, chemicals, and UV stress. Barrier maintenance requires physiological responses by local epithelia as well as immunological surveillance from resident and infiltrating lymphocytes and macrophages in times of stress. A large body of evidence leads to the suggestion that TRP metabolites regulate this physiology through AHR. This may occur through cellular release of these enzymes upon cell death or through their secretion which can generate local increases in I3P, leading to its condensation into carbazoles. In turn, these carbazoles serve as signals that influence local barrier gene expression as well as gene expression in and recruitment of infiltrating immune cells. What is not yet clear is if these carbazoles modulate barrier physiology through an unknown AHR-mediated transcriptional output or through stimulation of a unique set of genes that is AHR independent. A final point to state explicitly is that the most robustly upregulated AHR targets CYP1A1, CYP1A2, CYP1B1, and AHRR, broadly known to attenuate responses to environmental PAHs, also may serve to reduce endogenous ligand concentrations locally and repress AHR signaling by a similar manner. This observation is consistent with the idea that signaling requires rapid downregulation to avoid deleterious cellular consequences (e.g., inflammation).

In an effort to generate support for the above model, we set out to build a cell culture model system of this biology and initially use it to test three hypotheses. First, is that if AST, DAO, and UV all work through I3P as a proligand, then each of these activators should induce the same

AHR mediated pleiotropic response. To address this idea, we employ gene expression studies in a relevant human biological target, intestinal epithelial cells (here we use HCT 116). Second, if I3P products activate cellular physiology through some secondary set of AHR regulated genes or through a separate signaling pathway, then this pathway should be regulated similarly upon activation of I3P by multiple mechanisms. The purpose of this paper was to compare the various TRP-metabolizing pathways and verify I3P as a proligand is contributing to AHR signaling via all of these pathways. Third, we hypothesized that signaling from AHR ligands is a paracrine event that occurs outside of the cell, perhaps after secretion or cell lysis.

MATERIALS AND METHODS

Biochemicals: Cytosolic aspartate aminotransferase (ASTcyt, cat. no. G2751), D-amino acid oxidase (DAO, cat. no. A5222), alanine aminotransferase (ALT, cat. no. G8255), D-tryptophan (cat. no. T9753), L-tryptophan (cat. no. T8941), L-tyrosine (cat. no. T8566), L-phenylalanine (cat. no. P5482), and indole-3-pyruvic acid (I3P, cat. no. I7017) were purchased from Sigma Aldrich (St. Louis, MO). Lipofectamine 3000 (cat. no. L3000015) and 4',6-diamidino-2-phenylindole (DAPI, cat. no. D1306) were purchased from ThermoFisher (Waltham, MA). Dulbecco's Modified Eagle Medium (DMEM) cell culture media (cat. no. 11960044) was purchased from Gibco (Waltham, MA), along with fetal bovine serum (FBS), L-glutamine (cat. no. 25030081), and 0.25% trypsin-EDTA (cat. no. 25200056). Aromatic amino acid dropout media was purchased from Boca Scientific (Dedham, MA) as a custom order. The media was formulated without L-TRP, L-tyrosine, L-phenylalanine, L-glutamine, and phenol red, then the custom media was formulated to match the Gibco DMEM above.

Ultraviolet Radiation: UV bulbs were obtained from various companies. The Germicidal Puritec HNS 15W G13 bulb (cat. no. 4008321398826) came from Osram (Premstätten, Austria). The wideband UVB bulb (cat. no. 3000318) was the Ushio G8T5E (Tokyo, Japan). The narrowband UVB bulb (cat. no. 871150086891680) was the Philips PL-S 9W/01/2P (Amsterdam, Netherlands). Power was measured with a General Tools (Secaucus, NJ) Digital UVA/UVB Meter (cat. no. UV513AB). Calculations were then made using the power in uW/cm^2 multiplied by time of exposure.

Cell culture and induction conditions: The HCT 116 (cat. no. CCL-247) and HepG2 (cat. no. HB-8065) cells were obtained from the American Type Culture Collection (ATCC, Manassas, VA) (Table 1 and 2). Both cell lines were cultured in DMEM with 10% FBS and 2mM L-glutamine (complete DMEM). Cells were kept at 5% CO_2 and 37ºC in a humidified incubator. For proligand treatments, cells were plated at an initial density of 25,000 cells per well in a 96-well plate or 500,000 cells per well in a 6-well plate and given 24 hours to adhere. Stock solutions of I3P, AST, ALT, DAO, and D-TRP were diluted to proper working concentrations using complete DMEM (Table 3). To expose cells, culture media was removed from cells after 24 hours, and the induction mixture/media was pipetted onto cells for the specified periods of time (see figure legends). For induction with UV, media solutions were plated in either a 96-well plate at a volume of 100 uL or in a 6-well plate at a volume of 2 mL. The media was then placed under an appropriate UV bulb and exposed for the specified time. After exposure, the media was pipetted onto cells as described above. For all conditions, cells were exposed to UV treated media for 24 hours unless otherwise indicated.

UV exposure time and calculations: The power of UV output in the A and B spectrum was measured with the previously described light meter. To obtain energy of exposure, measurements in power (in uW/cm^2) was multiplied by time of exposure (in seconds) to find energy of exposure in uJ/cm^2. Assuming 1 standard erythema dose (SED) is equivalent to 100 J/m^2, exposures were then converted to SEDs (Table 4). For comparisons to "natural" sun exposure, it was assumed the sun emits UVA and B at 250 mW/m^2, which translates to summer sun emissions in northern climes[22].

Plasmids: Plasmids for expression of human DAO, ASTcyt, and ASTmt were synthesized by VectorBuilder (Chicago, IL). All plasmids were constructed with the gene of interest under the cytomegalovirus (CMV) promoter with mCherry in the same plasmid under a separate CMV promoter. All genes were codon optimized for human expression, and the Ensembl accession numbers are as follows: DAO, CCDS_9122.1; AST cyt, CCDS_7479.1 (aka GOT1); ASTmt, CCDS_10801.1 (aka GOT2). Plasmid maps are included in appendix.

Lipofection and cell sorting: Cells were plated in a 96-well plate at 25,000 cells per well or in a 6-well plate at 500,000 cells per well if destined for cell sorting. After 24 hours, lipofection reagents were mixed according to Lipofectamine 3000 protocol and added to wells. Lipofected cells were treated for 48 hours. If cells were used in assays unsorted, RNA or protein was obtained and used for RT-qPCR or enzyme activity assay, respectively (methods in following sections). For cell sorting, cells at 48 hours post-lipofection were collected from plates via trypsinization and washed with wash buffer (WB, 2% FBS in PBS). Then, cells were resuspended in WB with 1ug/uL DAPI and filtered to select for single cells. Filtered cells were subjected to flow cytometry using a BD FACSAria Cell Sorter. Cells were collected in bulk populations. Cells positive for fluorescent markers by sorting subjected to immediately RNA preparation for use in downstream protocols.

Enzyme activity assays: Enzyme activity assay kits for DAO (cat. no. ab273325) and AST (cat. no. ab105135) were purchased from Abcam (Boston, MA). Plasmids containing the ORF of DAO, ASTcyt, and ASTmt were lipofected into HCT 116 cells, plated 24 hours previously at a density of 500,000 cells per well in a 6-well plate. Lipofected cells were given 48 hours to grow

and express plasmid. Then, cells were collected, washed, resuspended in endpoint assay buffer, and lysed using two freeze-thaw cycles on dry ice. Cell lysate then was assayed per kit instructions by the manufacturer (DAO and AST enzyme activity kits). Results were read on a ClarioStar plate reader (BMG LabTech, Ortenberg, Germany).

RNA prep, RT-qPCR, and RNA-seq: The Promega (Madison, WI) "ReliaPrep RNA Miniprep System" (cat. no. Z6012) was used for all RNA preparations. The RT-qPCR analyses were performed with the Promega GoTaq Probe 1-Step RT-qPCR System (cat. no. A6121). Predesigned primer and probe assays were purchased from Integrated DNA Technologies (IDT, Coralville, IA). The probe dye for CYP1A1 (Hs.PT.58.219047), CYP1A2 (Hs.PT.58.45671878), CYP1B1 (Hs.PT.58.25328727.g) and AHRR (Hs.PT.58.23243692) was FAM. The probe dye for HPRT1 (Hs.PT.58v.45621572) was HEX. The RT-qPCR reactions were analyzed on an Applied Biosystems (Bedford, MA) QuantStudio Flex 7 instrument.

Samples for RNA-seq were prepared as above and submitted to the University of Wisconsin Biotechnology Center (UWBC) Gene Expression Core (GEC) for quality control and library preparation. Prepared libraries were then submitted to the UWBC Next-Gen Sequencing Core and sequenced on the Illumina (San Diego, CA) NovaSeq6000 at a depth of 30 million reads per sample. Differential gene expression (DGE) analysis was done by the UWBC Bioinformatics Resource Core (BRC), as well as with Qlucore (New York, NY) Omics Explorer.

Statistical analysis and calculations: Results of RT-qPCR were analyzed using the ddCt method[23]. In situations where genes had high initial Cts (~35 or above), the percent increase over

the minimum response was used to adjust for the low baseline of these genes. The calculation used was as follows: $100 - ((dCt/dCt_{max})*100)$ where dCt is the difference between the gene of interest (GOI) and the housekeeping gene (HK) and dCt_{max} is the maximum dCt for a given experimental control group. Statistical analyses and graphs were made using GraphPad Prism (San Diego, CA). Data was analyzed using 2-way ANOVA and multiple unpaired t-tests or 1-way ANOVA and student's t-test.

RESULTS

Choice of a cell culture model: In our effort to develop an easily studied human cell culture model that would be amenable to genetic manipulation through gene editing and recombination, we initially focused on the potential of two cell culture model systems. The hepatocellular system was chosen to reflect the role of the AHR in the biology of liver as a barrier to the external environment, and the intestinal epithelial system to reflect the receptor's influence on that barrier and it role in inflammatory bowel (IBD) disease[24,25]. An initial survey of the literature revealed two immortalized cell lines of potential utility, the HepG2 hepatocellular carcinoma line and the HCT 116 colorectal carcinoma line. An initial focus on these lines arose from an analysis of their known genomic integrity that indicated diploidy at important AHR target genes, thus simplifying gene editing success in future experiments (Table 1). Both lines were diploid at AHRR, CYP1A1 and CYP1A2, but CYP1B1, was found to harbor three loci in HepG2 (triploid). This was an initial indication that HCT 116 may be preferable in future experiments.

In our second level of analysis, we asked which cell line was more sensitive in its transcriptional response to our set of test stimuli related to TRP metabolism (Figure 1). To answer this question, we performed qPCR analysis of the response of CYP1A1 as in indicator. Cell lines were treated with UV exposed media (unexposed media as control); I3P (0.7% ethanol as control); or one of AST, DAO, D-tryptophan, or DAO and D-tryptophan (phosphate buffered saline, PBS, was also included as a control). Both HepG2 and HCT 116 cells responded to all induction conditions to varying degrees (Figure 2). HepG2 cells responded 5- to 10-times higher fold induction than

HCT 116 cells in most conditions, with 500 uM I3P showing the largest difference between cell types.

The final criteria for choice of a cell culture model were characteristics that would affect gene editing: i.e. transfection efficiency and clonal outgrowth success. Given our interest in generating null and edited alleles at target loci in future experiments, efficiency of editing and cloning was of importance for future success (Table 2). Results from multiple transfection and cloning experiments reveal that HCT 116 was the preferred cell line based on these properties. While sensitivity of response (Figure 2) was an important criterion, the difficulty of cloning HepG2 cells, along with its triploidy at the CYP1B1 locus, led us to the choice of HCT 116 cells for further study. This decision was supported by additional experiments to both demonstrate the sensitivity of the HCT 116 cell line to our proagonist stimuli as well as optimize this response for further studies.

HCT 116 cells are a suitable model for multiple pathways of AHR activation. To document HCT 116 cell sensitivity and to determine their suitability to respond to an intestinal enzyme's (i.e., DAO's) production of the proligand I3P, we performed an induction experiment measuring all four classical AHR mediated responses (i.e., AHRR, CYP1A1, CYP1A2, and CYP1B1). To this end, HCT 116 cells were dosed with media containing DAO (2U/mL) and D-tryptophan (21uM) to make the AHR ligand I3P (Figure 1). Cells were incubated in the mixture for 24 hours, at which time, cells were harvested, and RNA was prepared. The RT-qPCR assays was performed using primers and probes for CYP1A1, CYP1A2, CYP1B1, and AHRR, revealing induction of all targets except CYP1A2 (Figure 3). The lack of a CYP1A2 response was not

surprising given that this is commonly known as a hepatocyte-restricted gene product. Given the highly sensitive nature of the CYP1A1, CYP1B1, and AHRR response, these conditions were chosen as the common workflow for induction throughout (i.e., plate 25,000 cells per well in a 96-well plate, wait 24 hours, induce, wait 24 hours, make RNA, perform RT-qPCR).

To further establish this HCT 116 cell culture model, we examined our standard condition's influence on the I3P response, the UV response, and the influence of additional pro-oxidants to modify production of agonists from I3P (Figure 4). We observed a sensitive response in HCT 116 cells to I3P, AST, DAO, and UV when using CYP1A1 as a standard endpoint. The HCT 116 cells were also induced with AST and $FeCl_3$, and RNA was collected at various timepoints post-induction. While we did observe some increase in response when media was pretreated with $FeCl_3$ using AST as an enzyme, this response increase was not as robust when using DAO. Thus $FeCl_3$ was not used in the remaining experiments. HCT 116 cells responded maximally at 24 hours (Figure 4E).

Transient transfection of DAO and AST plasmids fail to induce the AHR pathway despite production of active enzyme. To determine whether the enzymes DAO or AST were producing I3P intracellularly, we increased expression of intracellular levels of these enzymes using transient transfection. The HCT 116 cells were lipofected with DAO or AST expression plasmids and were prepared for the matching Abcam enzyme activity assay. All plasmids produced active enzyme as determined using commercial assays for each activity (Figure 5A and B). For AST, the cytosolic version produced roughly three times as much active enzyme as the mitochondrial version, while the DAO plasmid produced over 300 mU/mL of active enzyme. All results were

significantly increased over the mCherry control. Interestingly, in a parallel experiment, cells that that were transfected with either DAO or AST expression plasmids did not demonstrate increases in the AHR response, CYP1A1 (Figure 5C). Included in this experiment was addition of D-TRP which is not typically found in standard cell culture media.

UV exposure of AAA-deficient cell culture media induces the AHR pathway via tryptophan, tyrosine, and phenylalanine photodegradation. Custom cell culture media lacking all three AAAs, was prepared commercially and FBS, phenol red, L-TRP, L-tyrosine, and L-phenylalanine were added back as indicated in methods and Figure Legends (Figure 6). The experimental groups were plated in a 96-well plate at a volume of 100uL for UV exposure from the narrowband source. The exposed media then replaced regular growth DMEM on cells plated 24 hours earlier at 25,000 cells per well in a 96-well plate. Cells were incubated 24 hours in the UV exposed media then extracted for RNA to be used in RT-qPCR. We observed that while response displayed highly variable results correlated with well location, the overall response suggested L-TRP, L-tyrosine, and FBS may play a role in ligand production from UV exposure (Figure 6).

Subsequent experiments using the same workflow detailed above except media were exposed in 6-well plates at a volume of 2mL. This adjustment provided tighter results presumably by removing the well-by-well variability. Using this improved protocol, the cultures were also exposed to different UV light sources as well as the addition of AAAs. The bulbs we employed included: in the germicidal/UVC range, in the wideband UVB range, and in the narrowband UVB range (for exact spectra, see Figure 7). Again, we observed that L-TRP and FBS were the

main sources of proligand. When exposed to UVC, both L-TRP and FBS caused a roughly four-fold increase in CYP1A1 expression (Figure 8A). The wide- and narrowband UVB sources showed results similar to each other. Both UVB sources were matched for power output in uW/cm^2 (UVB wide = 1,800 uW/cm^2; UVB narrow = 2,000 uW/cm^2), and media was exposed in both for 1 hour with matched controls covered by aluminum foil. For the wideband UVB source, L-TRP caused a roughly 17-fold increase in CYP1A1 expression, L-tyrosine and L-phenylalanine each caused a 3-fold increase, FBS caused a 40-fold increase, and all four compounds together induced CYP1A1 by approximately 250-fold (Figure 8B). Finally, the narrowband UVB source caused a 4-fold increase in the custom media alone, which was unique to this light source. Additionally, L-TRP caused an 11-fold increase, L-tyrosine and L-phenylalanine again each caused a 3-fold increase, FBS caused an approximately 14-fold increase, and all compounds together upregulated CYP1A1 by about 40-fold (Figure 8C).

The DAO enzyme, I3P, and UV upregulate a similar battery of AHR genes. RNA samples were prepared from cells treated with 2U/mL DAO and 21uM D-TRP, 500uM I3P, 1% PBS, UV exposed media, or unexposed UV control media. The RNA was then sent to the UWBC for RNA-seq analysis. Resulting data was analyzed in Qlucore Omics Software. Differential gene expression (DGE) analysis was performed using a q-value of 0.1; the PBS and UV controls were pooled to provide a sample size of six (Figure 9A). The DAO, I3P, and UV exposed samples were then compared to the pooled controls. The intersecting differentially expressed genes (DEGs) of each condition were then taken to provide a master list of overlapping DEGs in all conditions (i.e. the genes DAO, I3P, and UV all shared). Genes were visualized in Qlucore via a heatmap, which revealed 15 of 16 genes were similarly upregulated across induction methods;

MAGI1 was the only gene upregulated in one condition (UV) and downregulated in the others (I3P and DAO) (Figure 9B). RT-qPCR was done to verify the RNA-seq results for CYP1A1 (Figure 9C). All induction mechanisms upregulated CYP1A1, and other pathways hypothesized to make I3P as a proligand, namely ALT and AST, also increased CYP1A1 expression. The list of DEGs shared between DAO, I3P, and UV induction were compared to published lists of AHR-regulated genes (Figure 10)[26,27]. The list elucidated in this experiment had major overlap with the AHR genes identified in Yang et al., with 10 out of 16 genes shared between lists[26]. Of those genes, three were also shared with a list of AHR-regulated genes from ChIPBase[27]. These three genes were CYP1A1, GDF15, and NQO1. Notably, the ChIPBase list of AHR-driven genes did not contain known AHR-regulated genes such as AHRR or TIPARP. Genes uniquely identified in the current study were AC015712.2, ADGRF1, FTH1, GDA, MAGI1, and SLC16A6.

DISCUSSION

HCT 116 cells are a suitable model for AHR pathway study and have several advantages over HepG2 cells. Historically, HepG2 cells are a common model for the human AHR pathway. This use is primarily driven by their physiology being related to hepatocyte function, and their sensitivity to a broad spectrum of agonists including AHR proligands (Figure 1)[6]. However, this cell line seemed more physiologically distant from intestinal biology and disease states such as IBD. Therefore, we investigated the use of an alternative model, such as HCT 116 (Figure 2). Interestingly, the HCT 116 cells have been infrequently used to study the AHR pathway. Our results suggests that this may change as we have shown that HCT 116 cells are inducible at CYP1A1, CYP1B1, and AHRR loci (Figure 3) and respond to a diverse assortment of proligand producing pathways. These pathways include, I3P, DAO plus D-TRP, AST, and UV (Figure 1). Additionally, we performed preliminary assays with ALT and also observed a modest response in the presence of FBS and L-TRP (Figure 9C and data not shown). Both HepG2 and HCT 116 exhibit a diploid genome at many AHR target genes of interest, making them promising candidates for CRISPR editing, CYP1B1 is triploid in HepG2 making targeting of that locus more problematic compared to HCT 116 (Table 1). While HCT 116 cells may be less responsive than HepG2 cells, HCT 116 cells had higher transfection, editing, and single-cell sorting survival rates than HepG2 cells (Table 2). Additionally, we observed some doses of AST (for example, 15 U/mL) that killed HepG2 cells but did not adversely affect HCT 116 cells (data not shown). In addition, HCT 116 cells also are physiologically related to the AHR pathway that is related to IBD. In this regard, D-TRP is produced by microbes populating the gut, and HCT 116 is a colon-derived epithelial cell line. Therefore, this may be a more suitable model for AHR-related enteric diseases like IBD.

Several I3P-producing pathways activate the AHR pathway. The methods used to activate the AHR pathway in this paper are all derived from endogenous, physiologically relevant biochemical pathways. All are thought to work through TRP degradation, specifically through I3P as a common proligand for each pathway (Figure 1). In this report, each of the employed pathways (DAO, AST, and UV) produced obvious, measurable responses at doses reflecting abnormal physiological states (Figure 4 and Table 3). When plasmids were expressed that made endogenous-like levels of AST and DAO intracellularly, no induction of CYP1A1 was seen (Figure 5). For example, the inducing dose of AST was 10,000 mU/mL, and the normal endogenous level of AST is 40 mU/mL; 10,000 mU/mL would be indicative of injury from acetaminophen toxicity or cardiac arrest[28]. The doses used for UVB exposure would be roughly equivalent to multiple days of unprotected summer sun exposure in northern climes; at these doses, a person would have burns and skin damage which might also release AST locally (Table 4). Since the standard physiological levels of DAO, in gut and elsewhere, were not readily apparent in our literature search, the relationship of DAO amounts used in this paper to regular endogenous amounts was based on speculation and similar to a dose used previously by our lab[17]. The results from these experiments support AHR's role in maintaining homeostasis at barrier organs when the body faces pathogen invasion, extreme perturbation, or organ injury.

Cell culture media with AAA dropouts exposed to UV activates the AHR pathway through various AAAs. The previous hypothesis for AHR pathway activation via UVB exposure singled out TRP and its metabolites as the primary ligands responsible for induction[18,19]. This work supports the concept of TRP photodegradation as a key contributor to AHR pathway induction at

skin since tryptophan and tryptophan-containing solutions like FBS had the highest levels of CYP1A1 expression (Figure 6). However, groups containing tyrosine and phenylalanine also induced the pathway, albeit at a lower level. Tyrosine and phenylalanine each induced CYP1A1 three-fold over a control containing no AAAs or FBS. This suggests UV as a mechanism for AHR pathway induction is more complicated than previously thought, with tyrosine and phenylalanine contributing to the milieu of ligands created from photodegradation products.

UV source spectra can be broader than initially thought. In the literature, the term UVB has traditionally been identified as the spectrum necessary for AHR activation due to its ability to produce TRP photodegradation products. These results support the idea that a UV light source with a narrow band at 310nm is convenient and reproducible for AHR UV research. Our results indicate the wavelengths of light able to produce AHR agonists can arise from a broader spectrum than previously reported (Figure 7). This observation may have experimental implications when such wavelengths are used for sterilization of reagents and instruments. In our experiments, UV light in the germicidal UVC range induced CYP1A1 at a level equivalent to much of the published literature (Figure 8A). Additionally, wideband and narrowband UVB sources both induced CYP1A1 at high levels through all AAAs (Figure 8B and C). The increase in induction from UVC to UVB can likely be contributed to an increase in source power rather than an innate quality of UVB which UVC lacks. Interestingly, the wideband UVB source was able to induce to a greater magnitude than the narrowband source even though its output energy was slightly less (wideband = 64.8 kJ/m^2 and narrowband = 72 kJ/m^2). This result suggests UV sources can be of a broader spectrum of UVB or even UVC may be even more appropriate for AHR pathway research.

Disparate induction mechanisms activate a core battery of target AHR genes. A core list of AHR-regulated genes was generated by taking the intersection of DEGs between DAO, I3P, and UV treatment each compared to pooled controls (Figure 9). Visualization via a heatmap reveals 15 of the 16 identified genes share the same pattern of upregulation across all three induction mechanisms (Figure 9B). Only MAGI1 was upregulated when inducing with UV but downregulated compared to the pooled controls when inducing with DAO and I3P. The conservation in pattern across all induction mechanisms for the majority of genes provides strong evidence these genes comprise a core battery in the AHR response to ligand. Further, it suggests I3P is the common proligand produced by DAO and UV. The proligand I3P is known to spontaneously degrade into a milieu of compounds that induce AHR[12]. To confirm these results for a well-known AHR pathway gene included in the identified list (CYP1A1), RT-qPCR was done on RNA samples induced with multiple pathways (Figure 9C). All pathways are known to make I3P, and all induced CYP1A1 expression, supporting the RNA-seq results.

RNA-seq identifies novel AHR pathway genes suitable for further study. The list of AHR pathway genes identified in this chapter was compared to published AHR pathway gene lists (Figure 10). The published lists were derived from ChIP-seq experiments rather than RNA-seq as in this chapter[26,27]. Of the 16 identified genes, 10 were also shared in Yang et al., and 3 of those 10 overlapped with the ChIPBase list. However, the ChIPBase list of AHR genes omitted known AHR targets such as AHRR and TIPARP. This suggests an incomplete sampling likely due to experimental or biological constraints, and it presents the possibility that more AHR genes are shared between all compared lists than reported. Concurrence with published AHR pathway

genes suggests the list identified in this experiment accurately reflects AHR biology. Additionally, six genes identified in the current experiment did not overlap with published lists. Due to their conserved response across multiple induction mechanisms, these novel genes are core-AHR-battery candidates. Further investigation into their ligand responsiveness and biological role may shed further light on AHR signaling in vivo.

TABLES

Table 1. Ploidy at AHR genes of interest in HepG2 and HCT 116 cell lines.

Cell Line	Cellosaurus ID	AHR	CYP1A1	CYP1A2	CYP1B1	AHRR
HepG2	CVCL_0027	2	2	2	3	2
HCT 116	CVCL_0291	2	2	2	2	2

Table 2. Comparison of cloning and transfection efficiency between HepG2 and HCT 116. Observations from at least four independent experiments reveal the following estimates for transfection efficiency (visible expression of an mCherry plasmid under fluorescence). Clonal outgrowth as number of proliferating wells as visualized.

Cell Line	**HepG2**	**HCT 116**
Transfection Efficiency	15-20%	50-75%
Clonal Outgrowth Efficiency	0%	50%

Table 3. Working concentrations of inducing compounds and amino acids. * is in patients with acute coronary symptoms.

Compound	Working concentration (unless otherwise specified)	Solvent	Normal endogenous level
Indole-3-pyruvate (I3P)	500 uM	70% Ethanol	Unknown
Aspartate aminotransferase (AST)	10 U/mL	Ammonium sulfate	0.040 U/mL[28]
Alanine aminotransferase (ALT)	25 U/mL	PBS	0.020 U/mL[29]
D-Amino acid oxidase (DAO)	2 U/mL	PBS	Unknown
D-Tryptophan	21 uM	PBS	Unknown
L-Tryptophan	23 uM	PBS	63.7 uM[30]
L-Tyrosine	23 uM	PBS	35.5 uM[31] *
L-Phenylalanine	23 uM	PBS	31.2 uM[31] *

Table 4. UV source power, exposure, output energy, SEDs.

UV Source	UVA/UVB Power (mW/m^2)	Exposure Time (seconds)	Energy (kJ/m^2)	SEDs
Sun (northern summer)[22]	250	3,600	0.9	9
Germicidal (UVC)	0	3,600	0	0
Wideband UVB	18,000	3,600	64.8	648
Narrowband UVB	20,000	3,600	72	720

FIGURES

Figure 1) Biological pathways that produce 3-substituted indole proligands in vivo. A) Production of Indole 3-carbinol from enzymatic hydrolysis of glucobrassicin in plant homogenates to the generation of ICZ from indole 3-carbinol in the low pH environment of the stomach. **B)** Classically, AST functions as a glutamic oxaloacetic transaminase, converting between aspartic acid and glutamic acid in cells, when TRP is the substrate AST converts L-TRP to I3P. **C)** The enzyme DAO catalyzes the oxidation of D-TRP to I3P. **D)** Wavelengths of UVB light lead to conversion of TRP to I3P. Condensation of I3P creates AHR-activating compounds like FICZ, a compound with high affinity for the AHR.

Figure 1A)

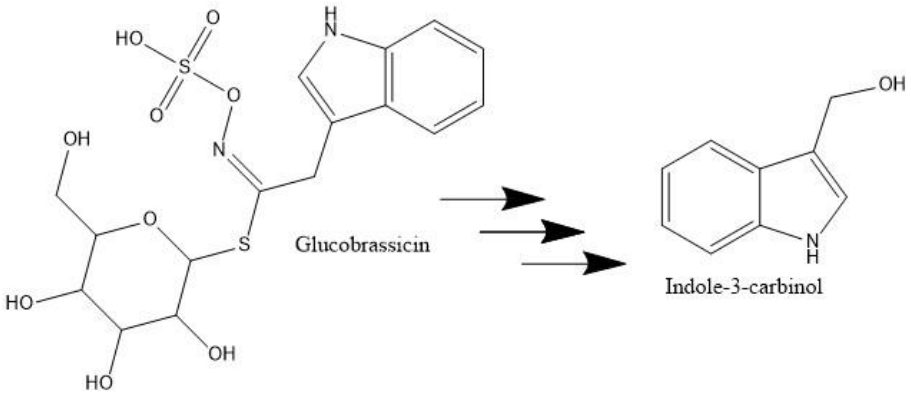

HO
O
S
O
O
N
H
N
OH
O
S
HO
OH
HO
Glucobrassicin
OH
N
H
Indole-3-carbinol

Figure 1B)

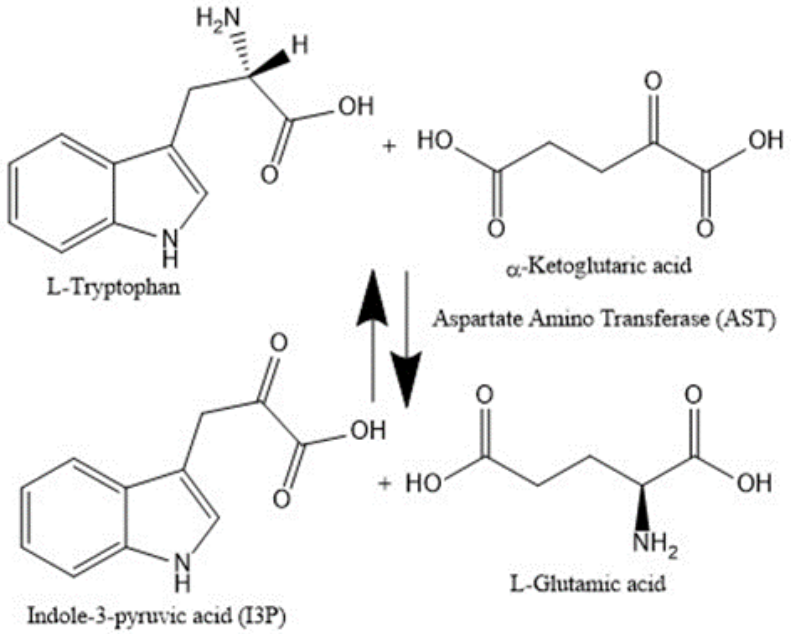

H2N
H
OH
O
N
H
L-Tryptophan
+ HO
O
O
OH
O
α-Ketoglutaric acid
Aspartate Amino Transferase (AST)
O
OH
O
N
H
Indole-3-pyruvic acid (I3P)
+ HO
O
O
OH
NH2
L-Glutamic acid

Figure 1C)

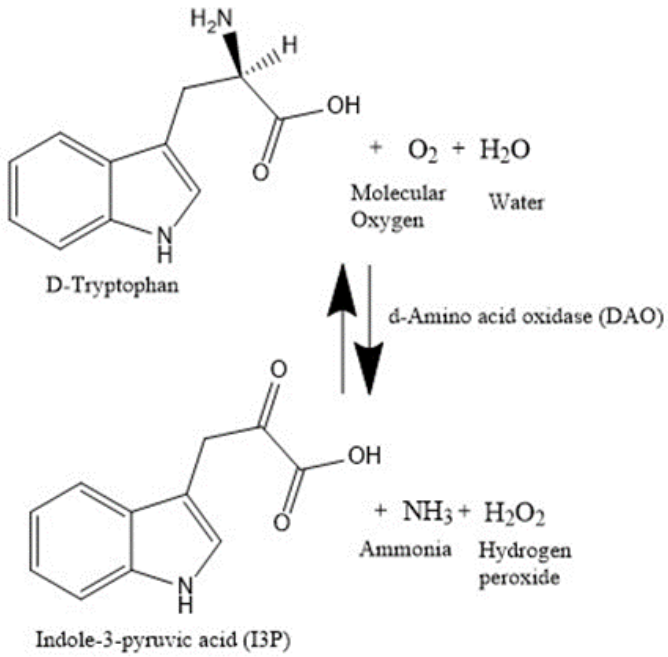

H2N
H
OH
O
N
H
D-Tryptophan
+ O2 + H2O
Molecular Oxygen
Water
d-Amino acid oxidase (DAO)
O
OH
O
N
H
Indole-3-pyruvic acid (I3P)
+ NH3 + H2O2
Ammonia
Hydrogen peroxide

Figure 1D)

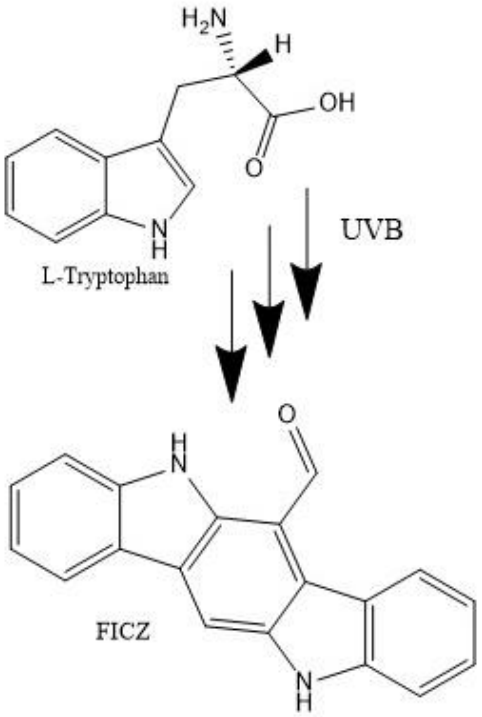


Figure 2) The cell lines HCT 116 and HepG2 cells respond at different magnitudes to AHR ligands. HepG2 cells are roughly five to ten times more inducible than HCT 116 cells when exposed to UV, I3P, DAO+D-tryptophan, or AST for 24 hours. Corrected p-values are as indicated. See text for details.

Figure 2)

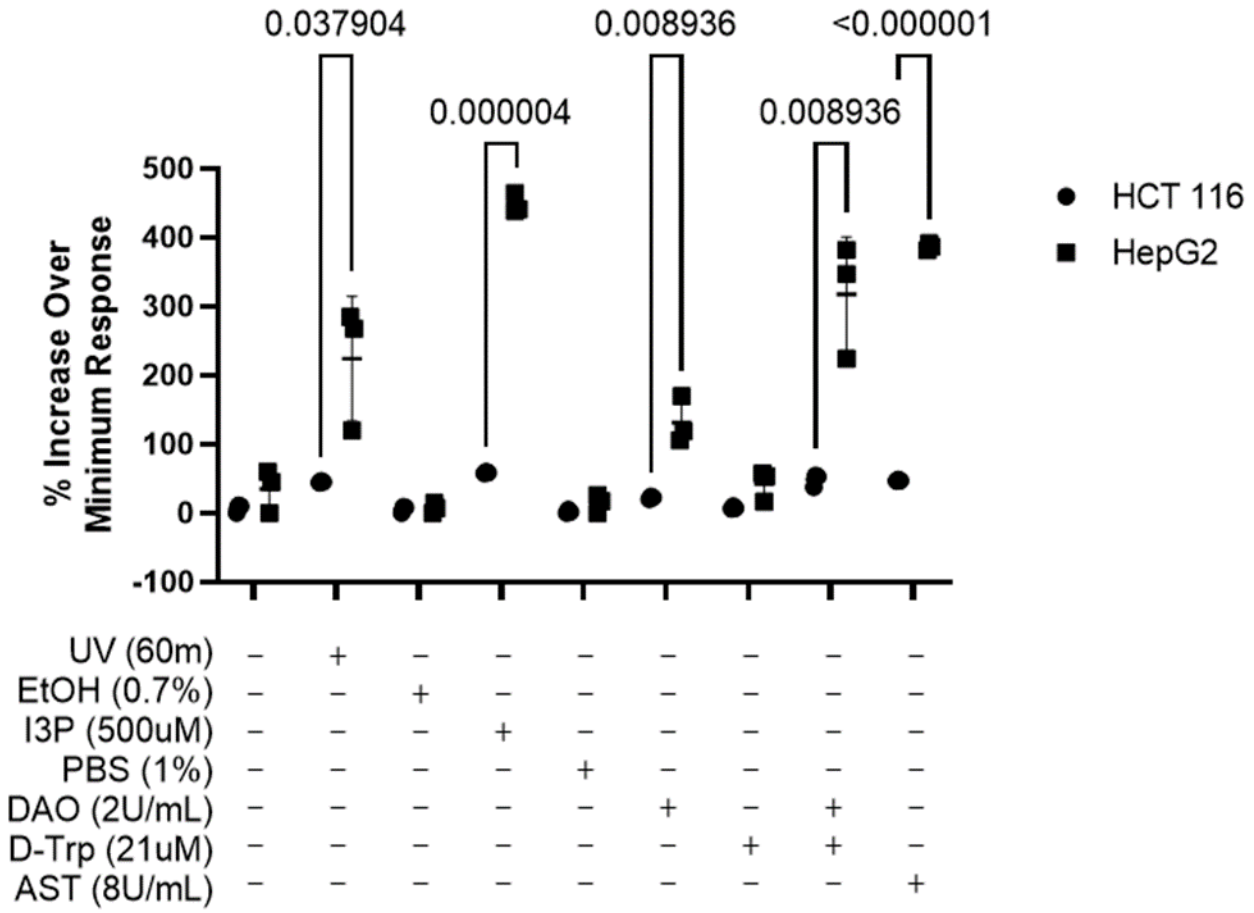

0.037904
0.000004
0.008936
0.008936
<0.000001
% Increase Over Minimum Response
500
400
300
200
100
0
-100
HCT 116
HepG2
UV (60m) – + – – – – – – –
EtOH (0.7%) – – + – – – – – –
I3P (500uM) – – – + – – – – –
PBS (1%) – – – – + – – – –
DAO (2U/mL) – – – – – + – + –
D-Trp (21uM) – – – – – – + + –
AST (8U/mL) – – – – – – – – +

Figure 3) DAO induces four canonical AHR pathway genes. Treatment of HCT 116 cells with 2U/mL DAO and 21uM D-tryptophan for 24 hours upregulates CYP1A1, CYP1B1, and AHRR approximately 50% over baseline. Upregulation for all genes is statistically significant at $p < 0.01$, including for CYP1A2 ($p = 0.009$). CYP1A1 is used as a reporter gene in further experiments.

Figure 3)

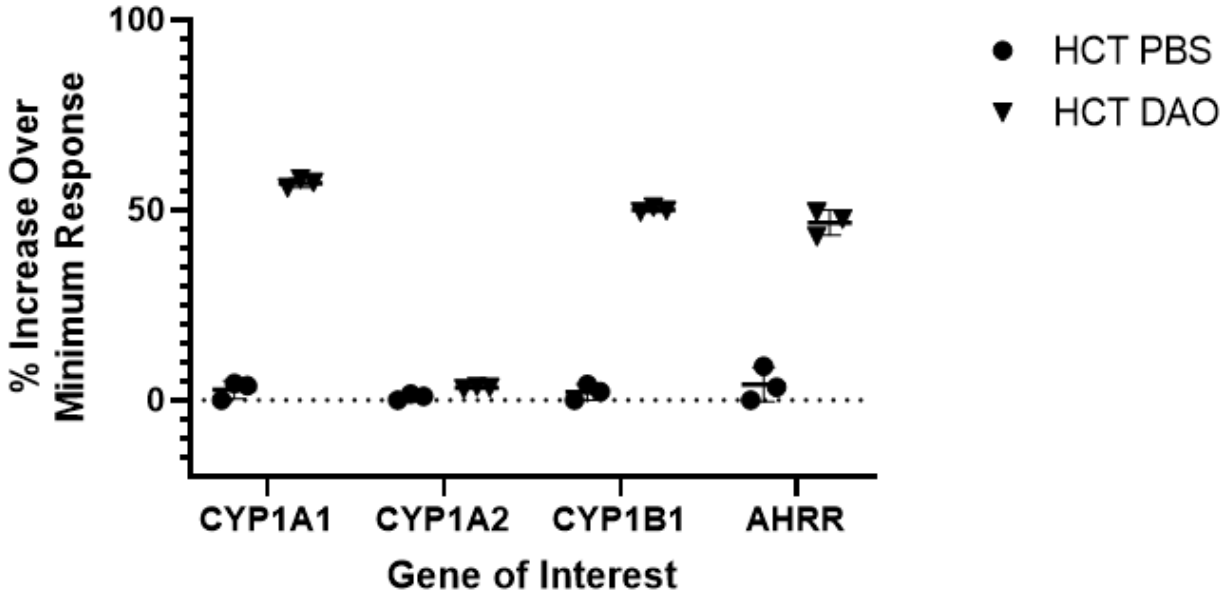

100
50
0
% Increase Over Minimum Response
CYP1A1
CYP1A2
CYP1B1
AHRR
Gene of Interest
HCT PBS
HCT DAO

Figure 4) Optimization of the HCT 116 cell system: A) Time course of I3P response: The AHR dose-response for CYP1A1 is measured and found to be maximal between 20-500 uM I3P. Asterisks indicate statistical significance according to a one-sample t-test with the theoretical mean set to 1. **B) Treatment with UV upregulates the AHR pathway in HCT 116 cells.** The HCT 116 cells were exposed to media treated with UV (DMEM only group), cells were exposed through PBS, or cells were exposed through media. Complete media exposed to UV caused AHR upregulation when cells were and were not exposed to UV directly. Numbers indicate adjusted p-values according to multiple unpaired t-tests for each group, +UV vs -UV. **C) DAO upregulates CYP1A1 in HCT 116 cells upon addition of D-TRP.** Cells were treated with indicated compounds for 24 hours before RT-qPCR. Excess D-tryptophan does not increase cell induction magnitude. Asterisks indicate statistical significance according to a one-sample t-test with the theoretical mean set to 1. **D) AST upregulates the AHR pathway using no additional TRP addition.** The HCT 116 cells were treated as indicated for 24 hours before RT-qPCR. The addition of 10U/mL AST exhibited greatest upregulation with least toxic effects on cells. Asterisks indicate statistical significance according to a one-sample t-test with the theoretical mean set to 1. **E) HCT 116 induction time course with AST and $FeCl_3$ induction**. HCT 116 cells were induced with 10U/mL AST and 50mM $FeCl_3$ for 24 hours before RT-qPCR. The AHR pathway begins to upregulate at 8 hours, but appreciable upregulation did not occur before that. Only the 24-hour timepoint is statistically significant ($p = 0.045$).

Figure 4A)

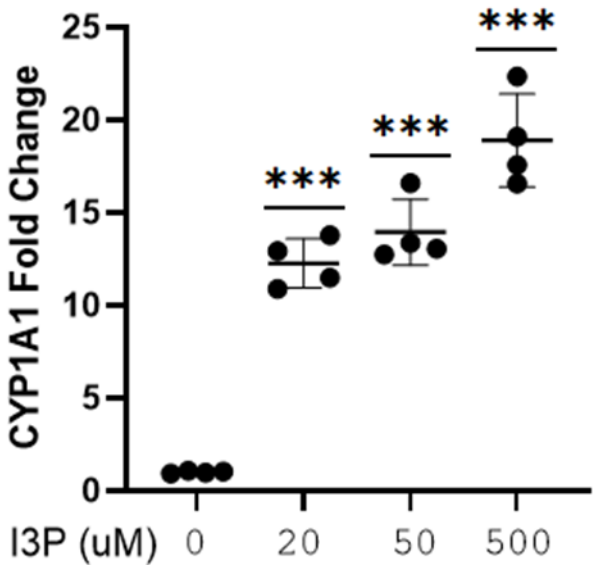


Figure 4B)

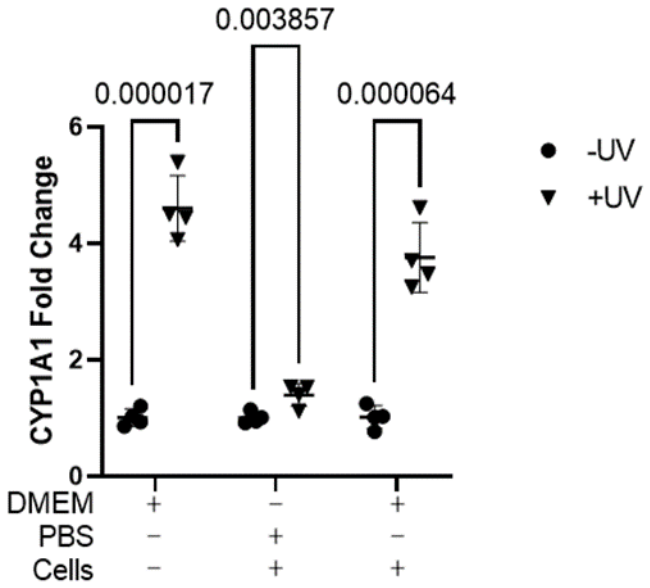

0.003857
0.000017
0.000064
CYP1A1 Fold Change
6
4
2
0
-UV
+UV
DMEM + − +
PBS − + −
Cells − + +

Figure 4C)

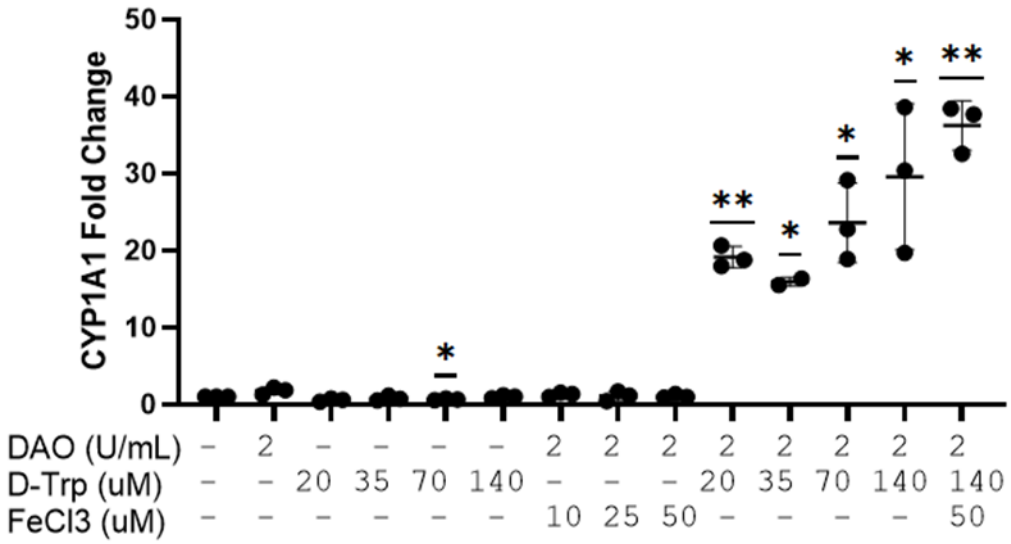

CYP1A1 Fold Change
50
40
30
20
10
0
*
**
*
*
*
**
DAO (U/mL) – 2 – – – – 2 2 2 2 2 2 2 2
D-Trp (uM) – – 20 35 70 140 – – – 20 35 70 140 140
FeCl3 (uM) – – – – – – 10 25 50 – – – – 50

Figure 4D)

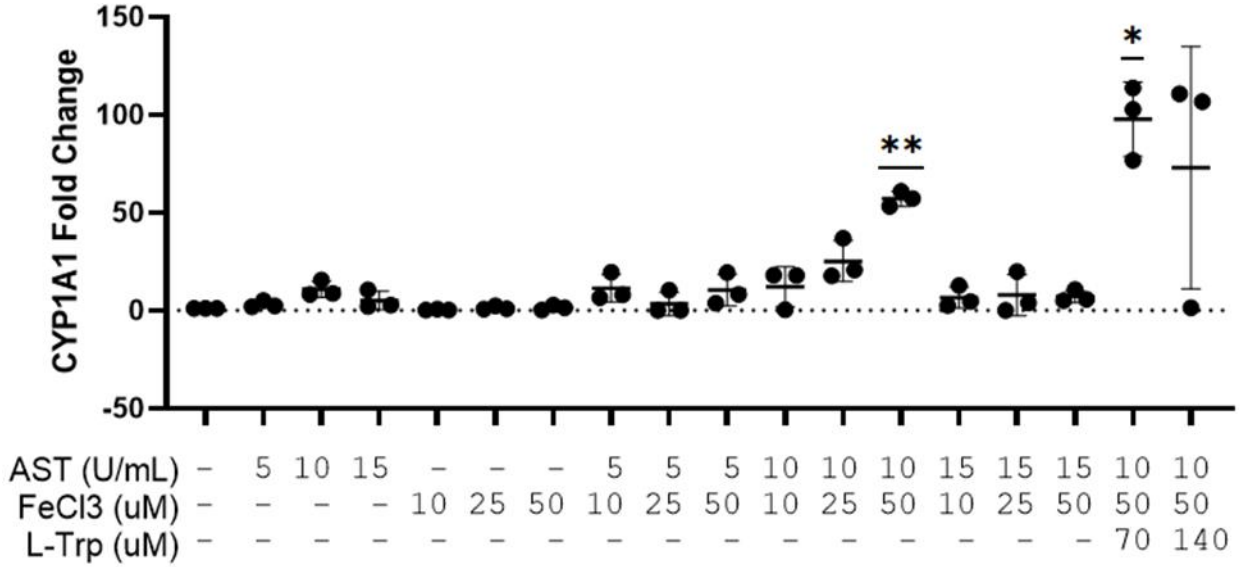

CYP1A1 Fold Change
150
100
50
0
-50
**
*
AST (U/mL) - 5 10 15 - - - 5 5 5 10 10 10 15 15 15 10 10
FeCl3 (uM) - - - - 10 25 50 10 25 50 10 25 50 10 25 50 50 50
L-Trp (uM) - - - - - - - - - - - - - - - - - 70 140

Figure 4E)

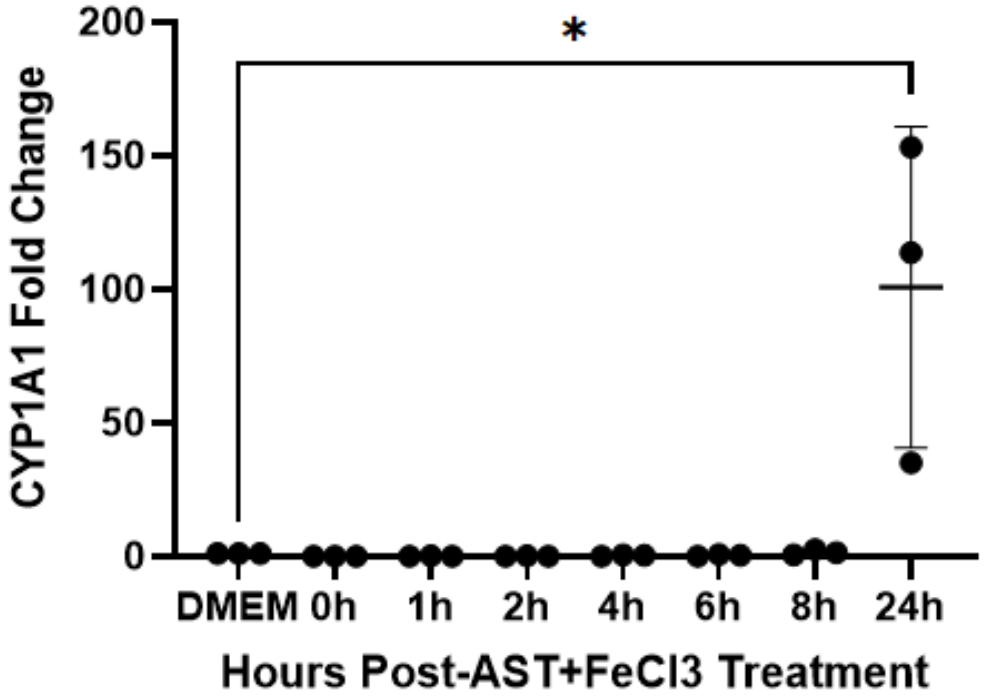

200
150
100
50
0
CYP1A1 Fold Change
*
DMEM 0h
1h
2h
4h
6h
8h
24h
Hours Post-AST+FeCl3 Treatment

Figure 5) The overexpression of DAO and AST via plasmid transfection does not lead to activation of the AHR pathway. A) AST plasmid overexpression produces active enzyme. The lesser activity from the ASTmt plasmid may be from lack of purifying mitochondrial fraction rather than a true biological difference. Asterisks indicate statistical significance according to one-sample t-test with theoretical mean set to the mCherry control sample mean (147.97). **B)** DAO plasmid overexpression produces active enzyme. Asterisks indicate statistical significance according to two-tailed unpaired t-test. **C**) Transient transfection of DAO and AST does not result in pronounced CYP1A1 induction. The ASTmt plasmid group was the only to show statistical significance when analyzed with a one-sample t-test with theoretical mean set to 1. The ASTmt plasmid caused a 1.2-fold increase in CYP1A1 expression which was significant by a t-test.

Figure 5A)

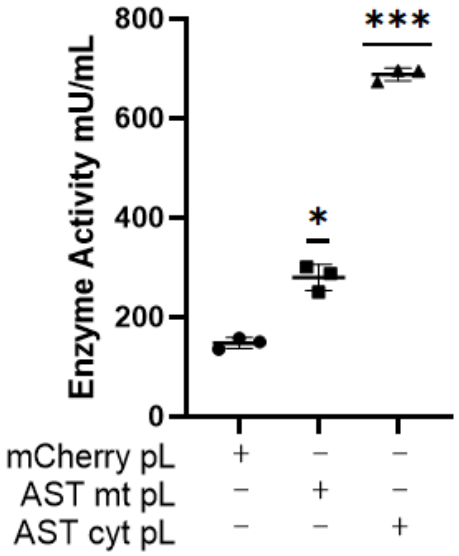

Enzyme Activity mU/mL
800
600
400
200
0
***
*
mCherry pL + − −
AST mt pL − + −
AST cyt pL − − +

Figure 5B)

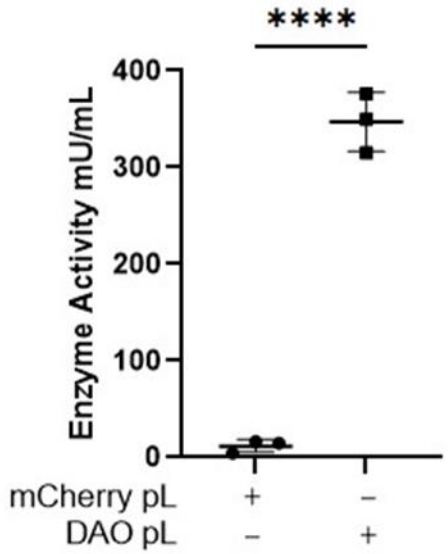


Figure 5C)

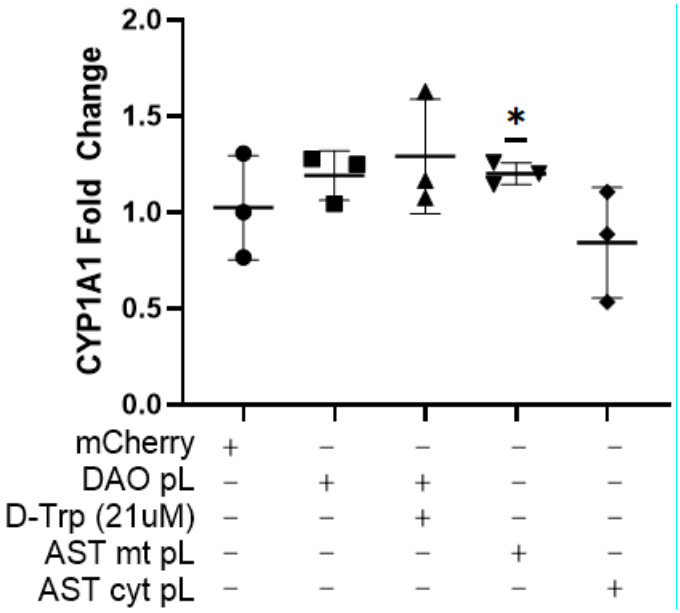

CYP1A1 Fold Change
2.0
1.5
1.0
0.5
0.0
*
mCherry + − − − −
DAO pL − + + − −
D-Trp (21uM) − − + − −
AST mt pL − − − + −
AST cyt pL − − − − +

Figure 6) Initial amino acid dropout with narrowband UV exposure experiment suggests tryptophan photodegradation is the main pathway of AHR ligand production, but tyrosine and phenylalanine make minor contributions to AHR pathway activation. Number indicates corrected p-value from multiple unpaired t-tests. Variation in UV-exposed dropout groups may be due to inconsistency in UV exposure (the bulb used was the narrowband UVB). This was corrected in later experiments.

Figure 6)

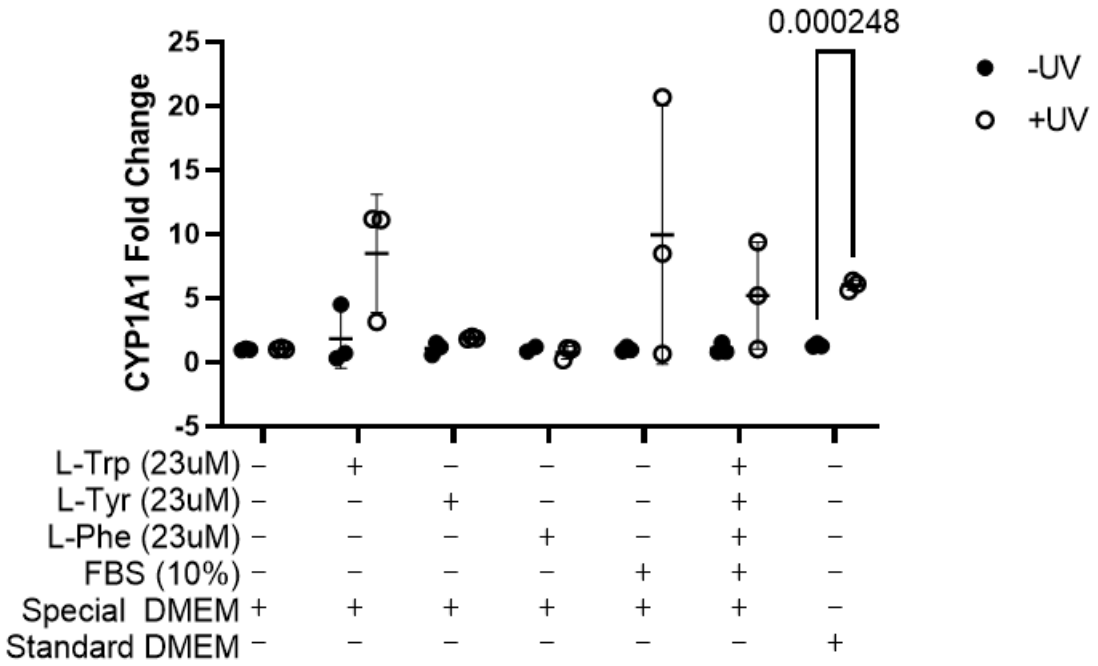

0.000248
-UV
+UV
CYP1A1 Fold Change
25
20
15
10
5
0
-5
L-Trp (23uM) − + − − − + −
L-Tyr (23uM) − − + − − + −
L-Phe (23uM) − − − + − + −
FBS (10%) − − − − + + −
Special DMEM + + + + + + −
Standard DMEM − − − − − − +

Figure 7) Representative UV spectra of three different light sources. The germicidal source has its main peak at approximately 250nm with negligible peaks at higher wavelengths. The wideband UVB source has a broad peak from approximately 250nm to 400nm with its maxima at approximately 320nm. Other peaks occur at 405nm, 440nm, 550nm, and 580nm. The narrowband UVB source has a minor peak at 300nm and its maximal peak at 310nm. Tryptophan maximally absorbs UV at 280nm.

Figure 7)

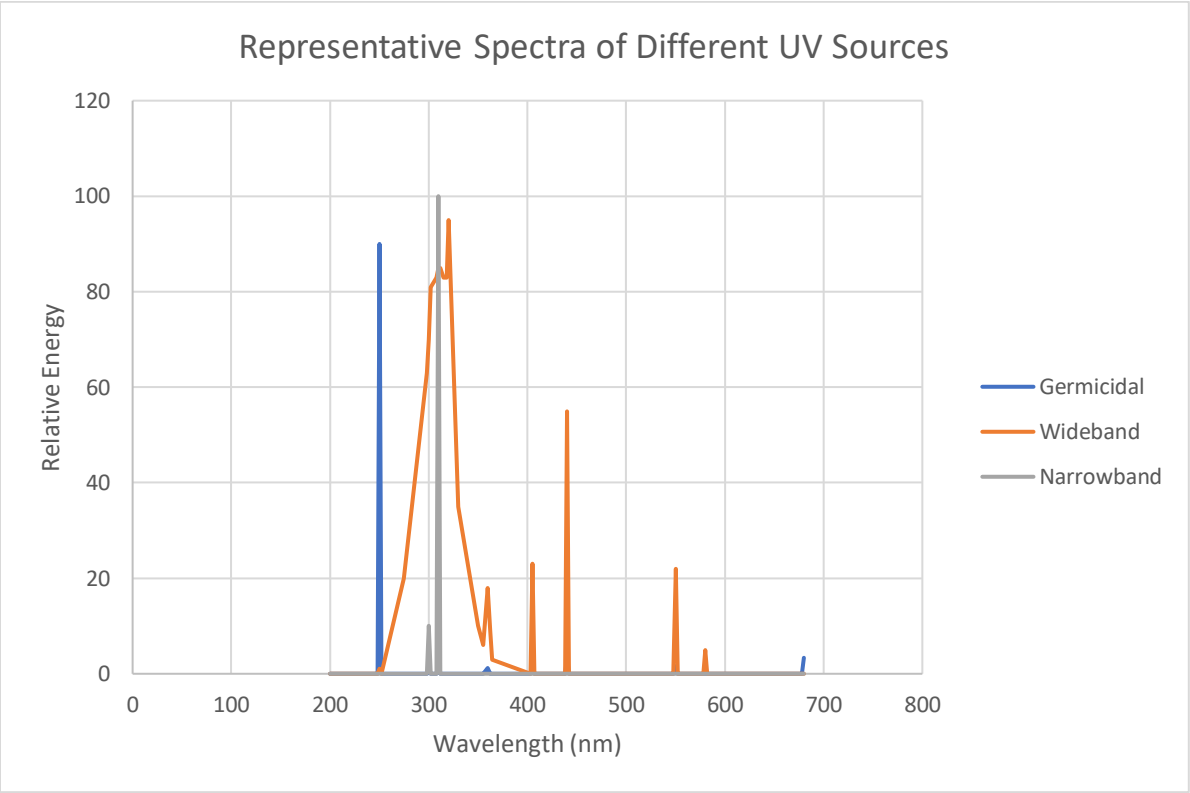

Representative Spectra of Different UV Sources
Relative Energy
120
100
80
60
40
20
0
0
100
200
300
400
500
600
700
800
Wavelength (nm)
Germicidal
Wideband
Narrowband

Figure 8) Comparison of three different UV light sources indicates many light sources commonly employed in the laboratory setting can activate AHR through irradiation of AAAs. Numbers indicate corrected p-values from multiple unpaired t-tests. **A) Germicidal UVC light source.** Groups with L-TRP (L-TRP, FBS, and complete media) had the highest CYP1A1 induction. Special-order DMEM alone had statistically significant induction ability when exposed to UV, indicating integral components of cell culture media may degrade to AHR ligands; alternatively, components of the cell culture plates media was exposed in may release AHR ligands when exposed to UVC. Aromatic amino acids besides tryptophan did not induce CYP1A1 when UV exposed. **B) Wideband UVB light source**. All groups besides the special-order DMEM only were significantly upregulated. The highest upregulation occurred in the complete media group with all AAAs plus FBS; this response appears to be synergistic in nature. L-tyrosine and L-phenylalanine also increased CYP1A1 expression roughly 2.5-fold. **C) Narrowband UVB light source.** All groups had significant CYP1A1 upregulation when exposed to UVB. See A. above for hypotheses as to why special-order DMEM alone may upregulate CYP1A1. Results for this light source agreed with those for a wideband UVB source (B. above). However, the magnitude of CYP1A1 expression was reduced for this light source even though the UV exposure was matched with the wideband source.

Figure 8A) Germicidal Light Source

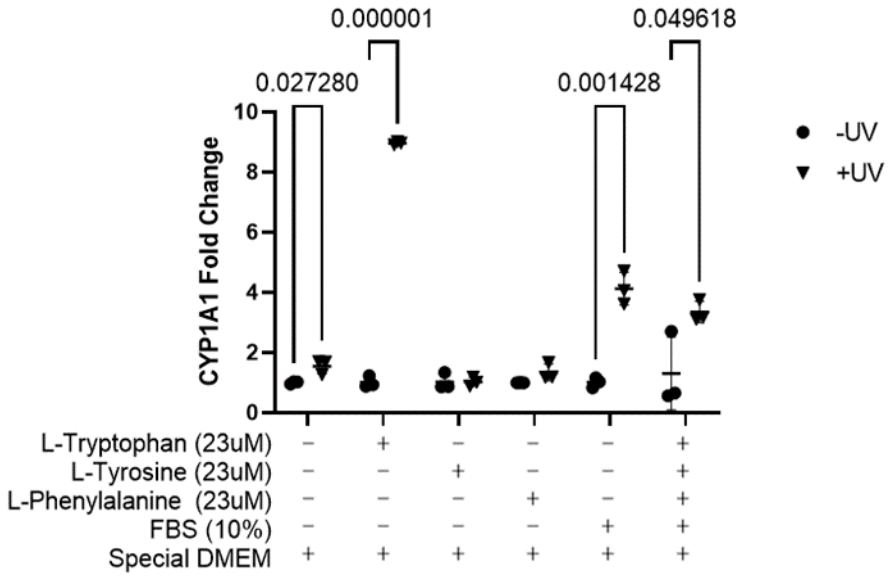


Figure 8B) Wideband UVB light source

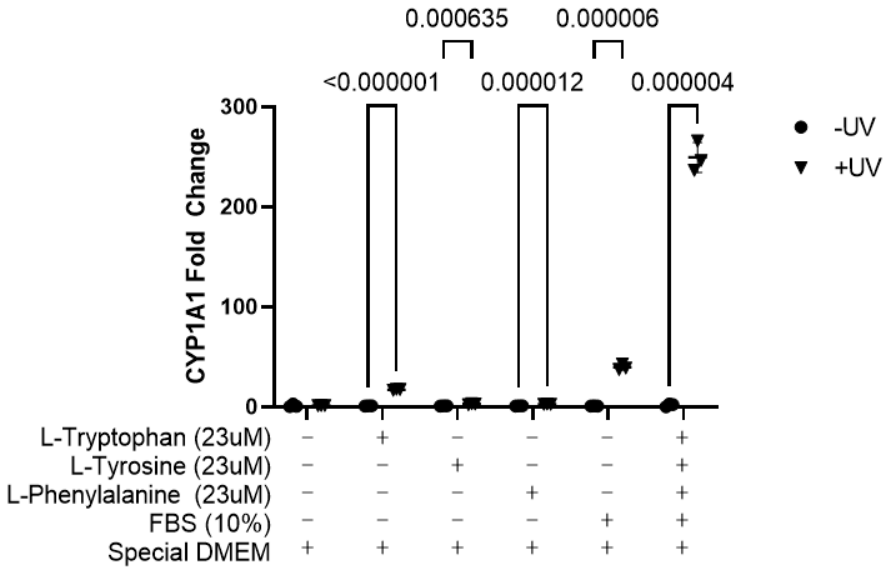


Figure 8C) Narrowband Light Source

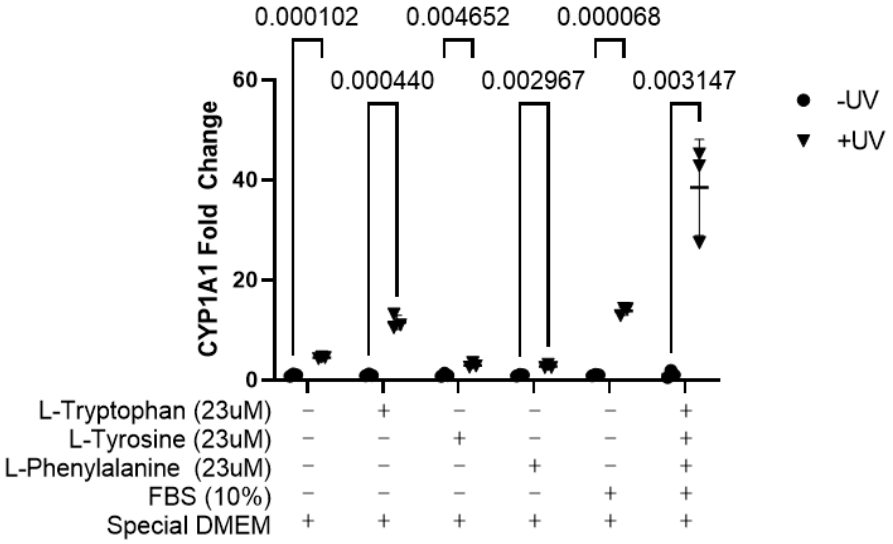


Figure 9) Different induction mechanisms upregulate a shared set of genes through the proligand I3P. A) Pooled RNA-seq controls cluster together. PCA plot analysis reveals the PBS control for I3P and DAO clusters with the unexposed UV control. Pooling controls is therefore appropriate for subsequent RNA-seq analysis. **B) DAO, I3P, and UV upregulate a shared set of genes by RNA-seq.** Heatmap of 16 genes identified with $q < 0.1$ that are differentially expressed by DAO, I3P, and UV treatment when compared to uninduced controls. Red indicates upregulation while blue indicates downregulation; color darkness reflects magnitude of change. All genes besides MAGI1 share a similar expression pattern across all three induction mechanisms. I3P shows the greatest upregulation in the differentially expressed genes. **C) RT-qPCR confirms different induction mechanisms that make I3P as a proligand upregulate CYP1A1**. Induction mechanisms were not controlled for potency, so dosages may not be directly comparable to each other. Asterisks indicate statistical significance according to one-sample t-tests with theoretical mean set to 1.

Figure 9A)

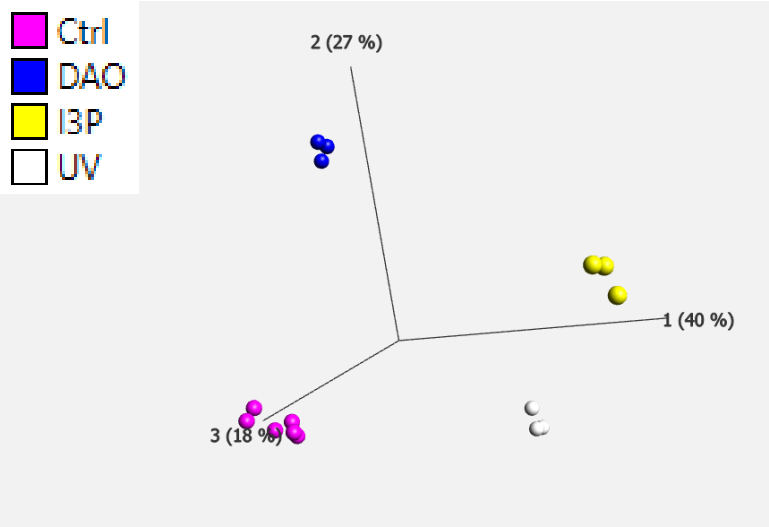


Figure 9B)

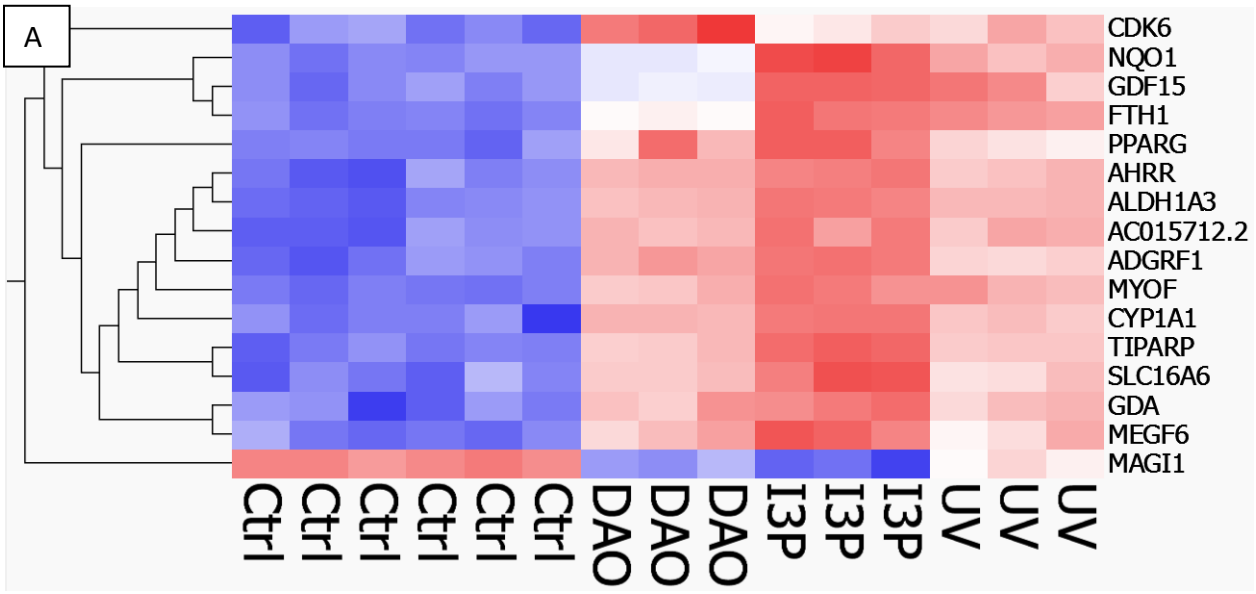

A
CDK6
NQO1
GDF15
FTH1
PPARG
AHRR
ALDH1A3
AC015712.2
ADGRF1
MYOF
CYP1A1
TIPARP
SLC16A6
GDA
MEGF6
MAGI1
Ctrl
Ctrl
Ctrl
Ctrl
Ctrl
Ctrl
DAO
DAO
DAO
I3P
I3P
I3P
UV
UV
UV

Figure 9C)

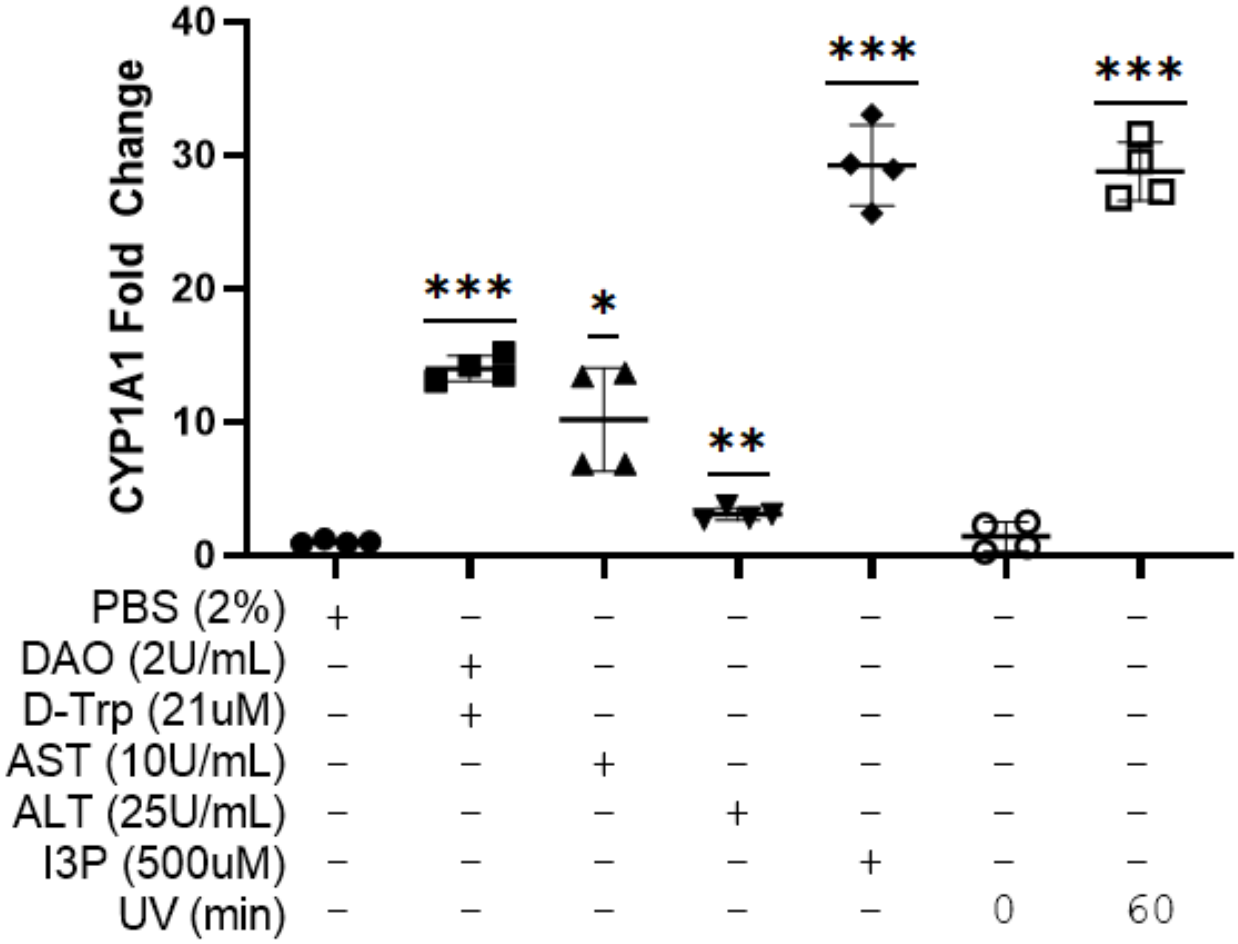
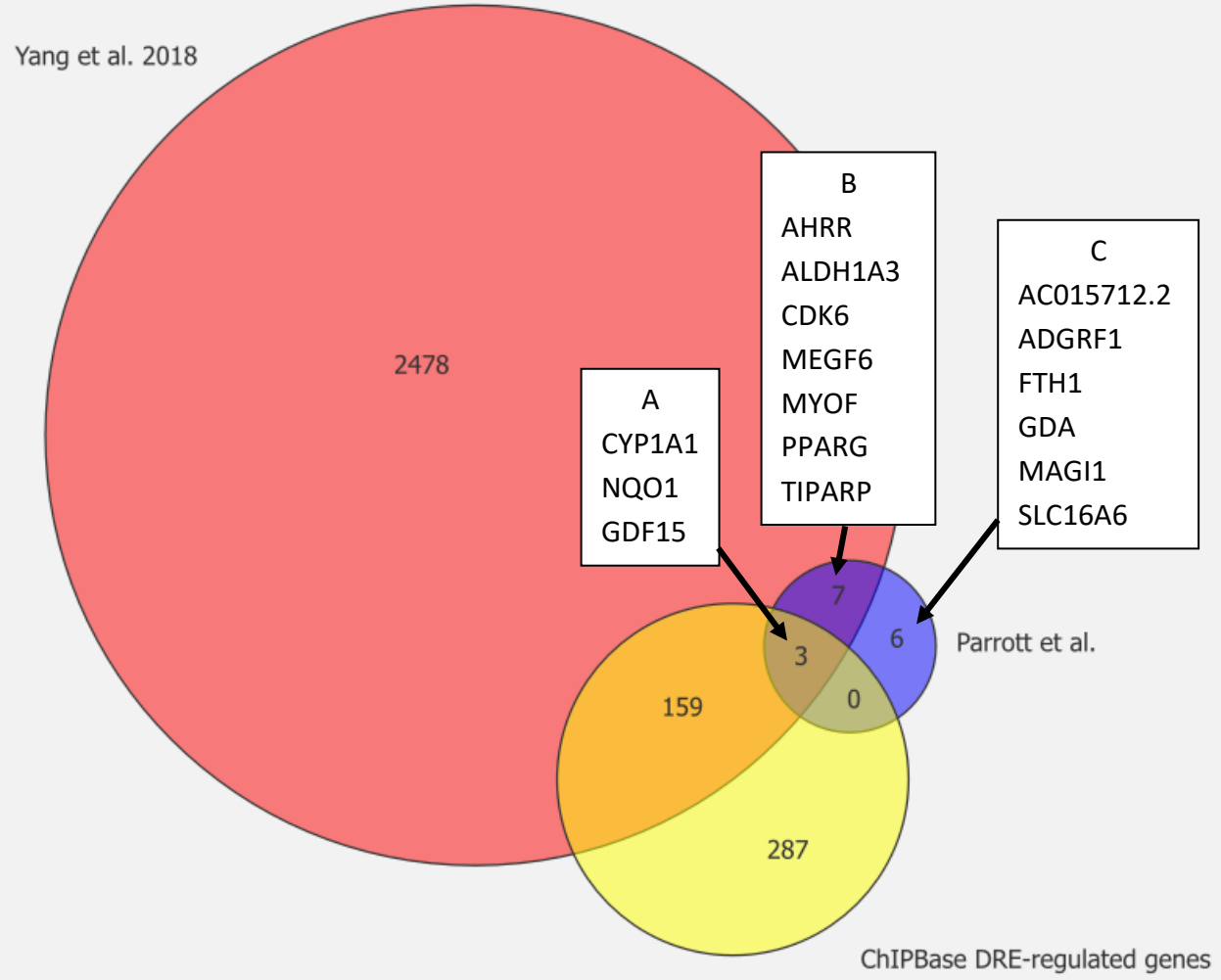

CYP1A1 Fold Change
40
30
20
10
0
***
*
**
***
***
PBS (2%) + – – – – – –
DAO (2U/mL) – + – – – – –
D-Trp (21uM) – + – – – – –
AST (10U/mL) – – + – – – –
ALT (25U/mL) – – – + – – –
I3P (500uM) – – – – + – –
UV (min) – – – – – 0 60

Figure 10) Overlap of current chapter's list of AHR genes with published lists. CYP1A1, NQO1, and GDF15 overlap between all three lists, while other historical core AHR genes like AHRR and TIPARP are shared between this paper and Yang et al.'s list. Novel genes identified in this paper are in box C and are candidates for further research. The ChIPBase list of DRE-regulated genes omits all genes of box B despite some of them having known DREs.

Figure 10)


Yang et al. 2018
2478
A
CYP1A1
NQO1
GDF15
B
AHRR
ALDH1A3
CDK6
MEGF6
MYOF
PPARG
TIPARP
C
AC015712.2
ADGRF1
FTH1
GDA
MAGI1
SLC16A6
7
3
6
0
159
287
Parrott et al.
ChIPBase DRE-regulated genes

REFERENCES

1. Gu, Y. Z., Hogenesch, J. B. & Bradfield, C. A. The PAS superfamily: Sensors of environmental and developmental signals. *Annual Review of Pharmacology and Toxicology* **40**, 519–561 (2000).

2. Nebert, D. W. Aryl hydrocarbon receptor (AHR): 'pioneer member' of the basic-helix/loop/helix per-Arnt-sim (bHLH/PAS) family of 'sensors' of foreign and endogenous signals. *Progress in Lipid Research* **67**, 38–57 (2017).

3. Vazquez-Rivera, E. *et al.* The aryl hydrocarbon receptor as a model PAS sensor. *Toxicol Rep* **9**, 1–11 (2022).

4. Hahn, M. E. Aryl hydrocarbon receptors: diversity and evolution. *Chem Biol Interact* **141**, 131–160 (2002).

5. Nebert, D. W., Dalton, T. P., Okey, A. B. & Gonzalez, F. J. Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. *Journal of Biological Chemistry* **279**, 23847–23850 (2004).

6. Tsuchiya, Y., Nakajima, M., Itoh, S., Iwanari, M. & Yokoi, T. Expression of aryl hydrocarbon receptor repressor in normal human tissues and inducibility by polycyclic aromatic hydrocarbons in human tumor-derived cell lines. *Toxicol. Sci.* **72**, 253–259 (2003).

7. Diaz-Diaz, C. J. *et al.* The Aryl Hydrocarbon Receptor is a Repressor of Inflammation-associated Colorectal Tumorigenesis in Mouse. *Annals of Surgery* **264**, 429–436 (2016).

8. Walisser, J. A., Bunger, M. K., Glover, E. & Bradfield, C. A. Gestational exposure of Ahr and Arnt hypomorphs to dioxin rescues vascular development. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 16677–16682 (2004).

9. Mezrich, J. D. *et al.* An Interaction between Kynurenine and the Aryl Hydrocarbon Receptor Can Generate Regulatory T Cells. *The Journal of Immunology* **185**, 3190 (2010).

10. Stevens, E. A., Mezrich, J. D. & Bradfield, C. A. The aryl hydrocarbon receptor: a perspective on potential roles in the immune system. *Immunology* **127**, 299–311 (2009).

11. 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).

12. Bittinger, M. A., Nguyen, L. P. & Bradfield, C. A. Aspartate Aminotransferase Generates Proagonists of the Aryl Hydrocarbon Receptor. *Molecular Pharmacology* **64**, 550 (2003).

13. Avilla, M. N., Malecki, K. M. C., Hahn, M. E., Wilson, R. H. & Bradfield, C. A. The Ah Receptor: Adaptive Metabolism, Ligand Diversity, and the Xenokine Model. *Chem Res Toxicol* **33**, 860–879 (2020).

14. Friedman, M. Analysis, Nutrition, and Health Benefits of Tryptophan. *Int J Tryptophan Res* **11**, 1178646918802282 (2018).

15. Bjeldanes, L. F., Kim, J. Y., Grose, K. R., Bartholomew, J. C. & Bradfield, C. A. Aromatic hydrocarbon responsiveness-receptor agonists generated from indole-3-carbinol in vitro and in vivo: comparisons with 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Proc Natl Acad Sci U S A* **88**, 9543–9547 (1991).

16. Bradfield, C. A. & Bjeldanes, L. F. Modification of carcinogen metabolism by indolylic autolysis products of Brassica oleraceae. *Adv Exp Med Biol* **289**, 153–163 (1991).

17. Nguyen, L. P. *et al.* D-Amino Acid Oxidase Generates Agonists of the Aryl Hydrocarbon Receptor from D-Tryptophan. *Chem Res Toxicol* **22**, 1897–1904 (2009).

18. Rannug, A. *et al.* Certain photooxidized derivatives of tryptophan bind with very high affinity to the Ah receptor and are likely to be endogenous signal substances. *Journal of Biological Chemistry* **262**, 15422–15427 (1987).

19. Rannug, A. & Fritsche, E. The aryl hydrocarbon receptor and light. *Biol Chem* **387**, 1149–1157 (2006).

20. Bergander, L. *et al.* Metabolic fate of the Ah receptor ligand 6-formylindolo[3,2-b]carbazole. *Chemico-Biological Interactions* **149**, 151–164 (2004).

21. Smirnova, A. *et al.* Evidence for New Light-Independent Pathways for Generation of the Endogenous Aryl Hydrocarbon Receptor Agonist FICZ. *Chemical Research in Toxicology* **29**, 75–86 (2016).

22. UV index, SED, MED, Erythema. *Bentham Instruments Limited* https://support.bentham.co.uk/support/solutions/articles/5000619299-erythemal-radiant-exposure-and-uv-index.

23. Livak, K. J. & Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* **25**, 402–408 (2001).

24. Pande, K., Moran, S. M. & Bradfield, C. A. Aspects of dioxin toxicity are mediated by interleukin 1-like cytokines. *Mol Pharmacol* **67**, 1393–1398 (2005).

25. Esser, C., Rannug, A. & Stockinger, B. The aryl hydrocarbon receptor in immunity. *Trends in immunology* **30**, 447–454 (2009).

26. 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. *Archives of Toxicology* **92**, 225–240 (2018).

27. Zhou, K.-R. *et al.* ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. *Nucleic Acids Res* **45**, D43–D50 (2017).

28. Aspartate Aminotransferase - an overview | ScienceDirect Topics. https://www.sciencedirect.com/topics/neuroscience/aspartate-aminotransferase.

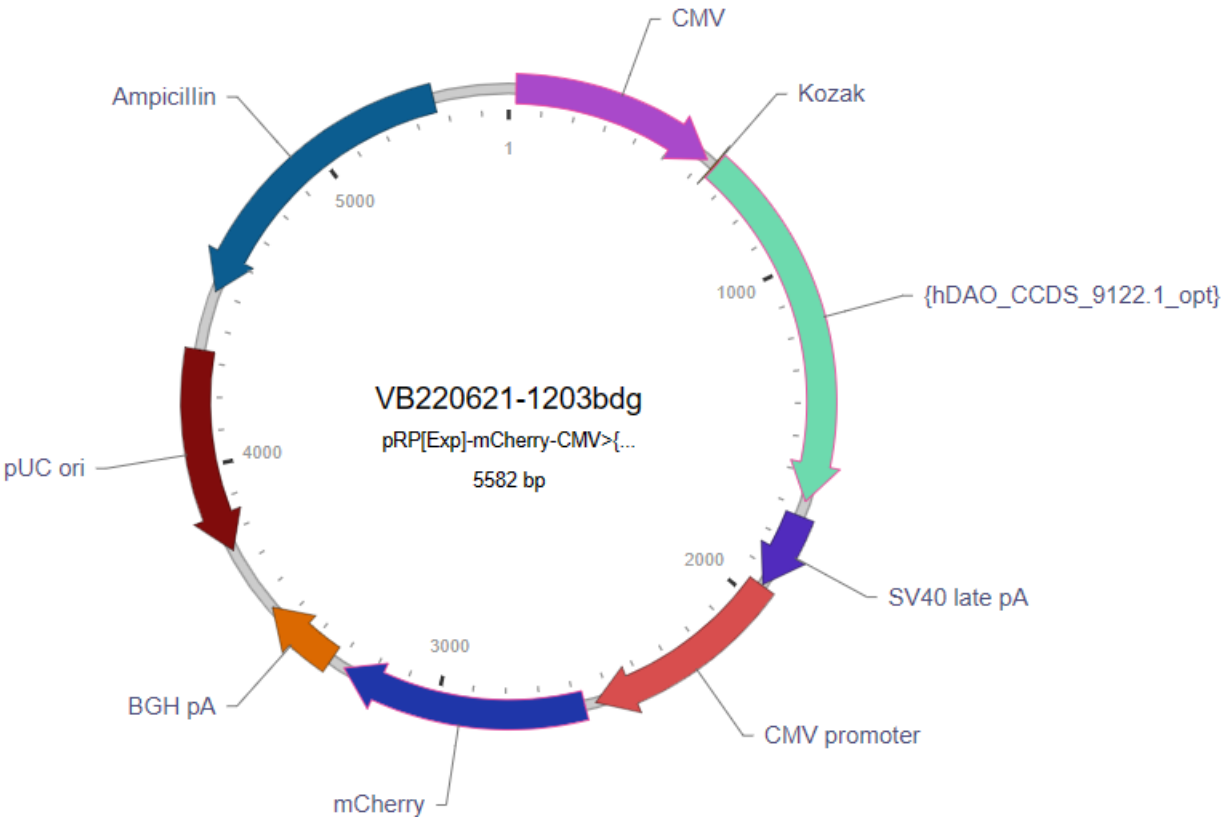
29. Yip, T. C.-F., Wong, V. W.-S. & Wong, G. L.-H. Alanine Aminotransferase Level: The Road to Normal in 2021. *Hepatol Commun* **5**, 1807–1809 (2021).

30. Comai, S. *et al.* Serum Levels of Tryptophan, 5-Hydroxytryptophan and Serotonin in Patients Affected with Different Forms of Amenorrhea. *Int J Tryptophan Res* **3**, 69–75 (2010).

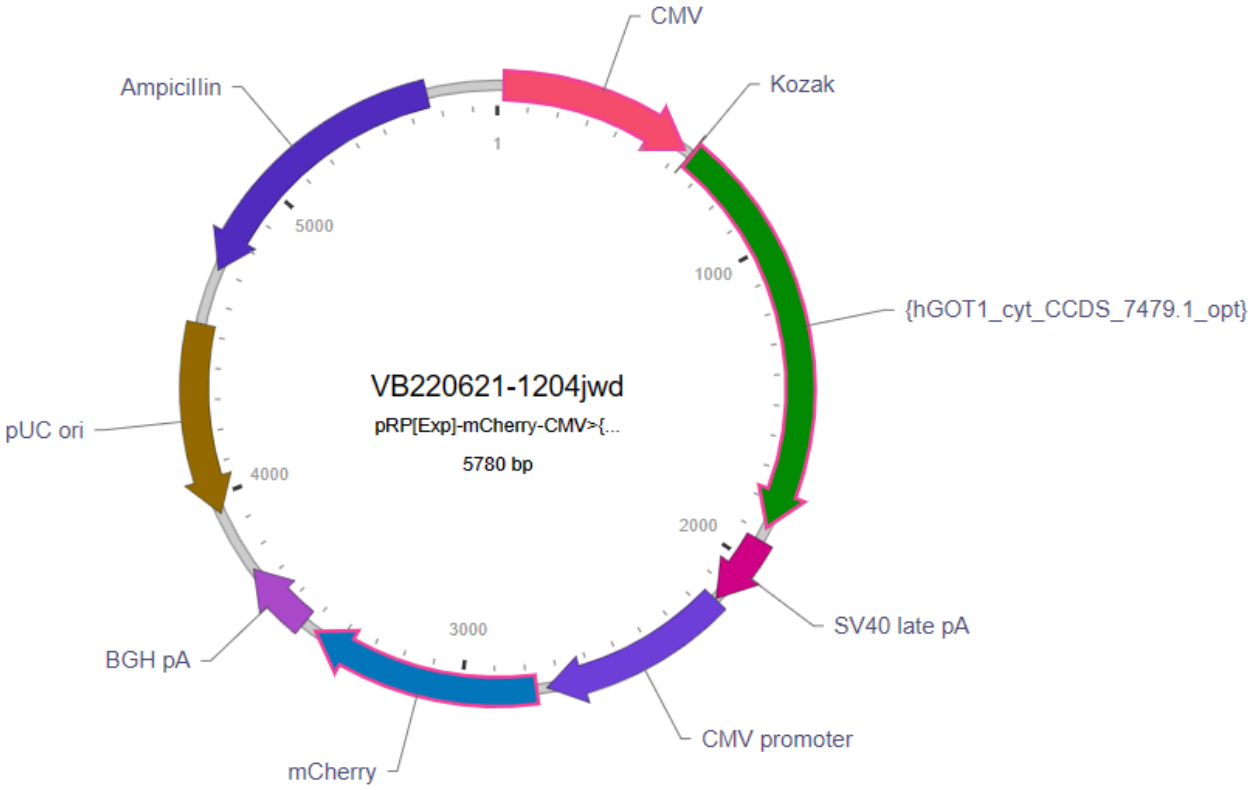
31. Al-Sadoon, I. *et al.* Assessment of serum phenylalanine and tyrosine isomers in patients with ST-segment elevation vs non-ST-segment elevation myocardial infarction. *J Clin Lab Anal* **35**, e23613 (2021).

APPENDIX

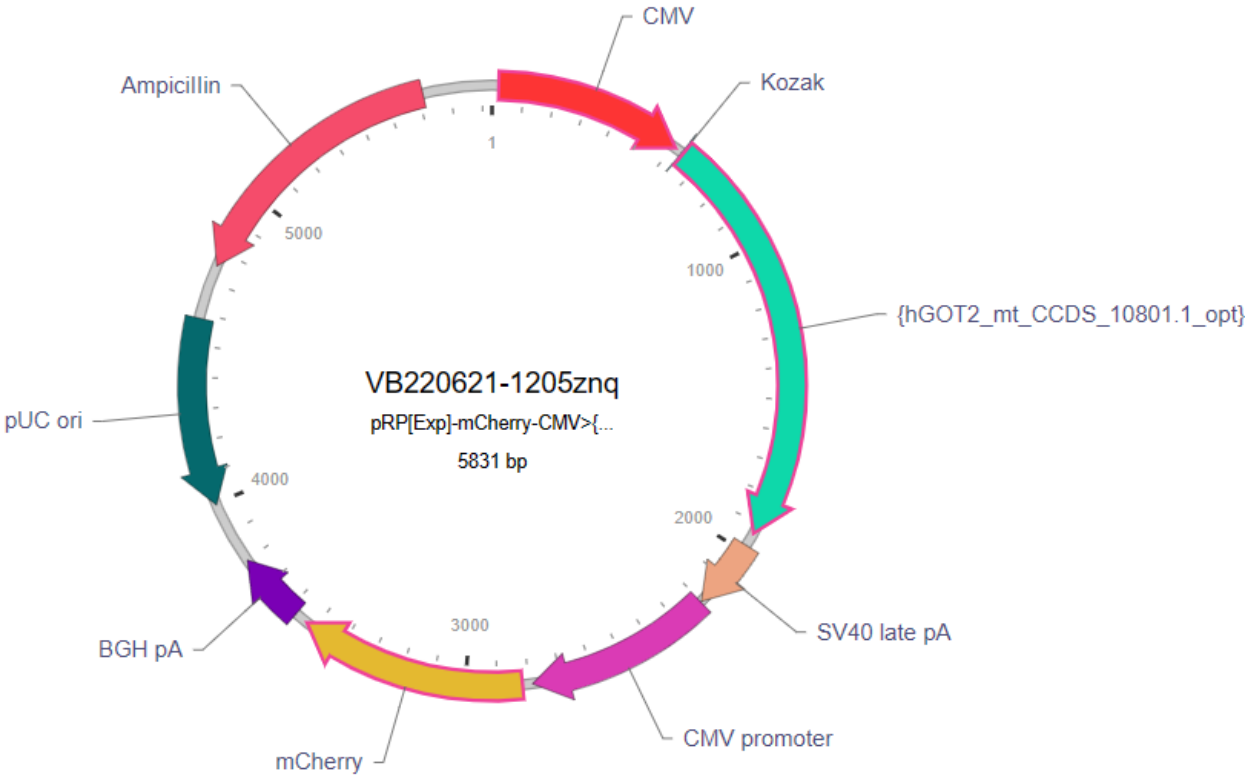
Appendix 1. Human DAO coding sequence plasmid map.



Appendix 2. Human ASTcyt plasmid map.



Appendix 3. Human ASTmt plasmid map.



Chapter 4: Studies to Establish an Auxin-Inducible Method of Protein Knockdown in HCT 116 Cells

ABSTRACT

The auxin-inducible degron (AID) system showed early promise in the reversible degradation of protein products. Because it relies on a plant ubiquitinating protein (OsTIR1), the AID system is proposed to degrade proteins in mammalian systems upon exposure with the plant hormone or auxin known as indole-3-acetic acid (IAA). Published literature suggests the AID system is highly responsive, reversible, and auxin-dependent without basal degradation in a wide variety of mammalian systems[1]. The experiments in provide a stark contrast to this premise, indicating limited influence on protein expression, although reversibility of the AID system could not be tested as no significant degradation was seen in multiple scenarios. Transiently transfected OsTIR1 was able to degrade a tagged reporter in the absence of auxin, and the maximal decrease in reporter expression was less than 10%. While the AID system shows promise in published literature, the results of these experiments suggest the spectrum of its utility may be highly selective and cannot be relied upon to control pathways in many commonly used cell systems commonly used to study Ah receptor signaling.

INTRODUCTION

Regulation of protein expression through “knockdown” or “knockout” techniques have long been used to study diverse biological systems. Many genetic methods exist to totally remove protein products, including the popular clustered regularly interspaced short palindromic repeats (CRISPR) editing[2–4]. These approaches generally delete or mutate a gene into a non-functional form. Additional approaches to control the expression of an mRNA such as RNA interference (RNAi) and the Tet on/off system are also popular[5–7]. Currently, CRISPR and RNAi are perhaps the simplest methods to study the effect a particular gene product has on a specific biological pathway in cell culture models. Of note, neither of these methods are regulatable and typically rely on the comparison of multiple clones or transfected populations. Moreover, neither of these methods control expression of the gene product at the protein level, thus, they are temporally more distant from the pathway under study.

A more recently proposed alternative to the above methods is the auxin inducible degron (AID) system first published in 2011 and improved in subsequent years[1,8–10]. The AID system uses a plant ubiquitinating protein (OsTIR1, part of a Skp1-Cul1-F-box protein complex) to target a protein of interest for degradation. In this novel regulatory method, the protein of interest is tagged with a plant-derived epitope, termed a “degron”, that serves as a ubiquitination site for the OsTIR1, leading to its rapid degradation by the cellular proteosome. Crucially, after generation of a cell line expressing OsTIR1, ubiquitination of the protein of interest happens only when the plant auxin hormone, indole-3-acetic acid (IAA) is added to the system. Of equal importance to this system, removal of the auxin inactivates OsTIR1 and reverses the protein degradation, leading to normal target protein stability and expression in the system. Early problems with the

system, such as auxin-independent basal degradation and large required doses of auxin, have been rectified in more recent years[1,10]. Briefly, OsTIR1 was mutated at the 74th amino acid from phenylalanine to glycine (termed OsTIR1-F74G). This formed a “bump” in the auxin-binding domain where the phenyl group of an alternative auxin, 5-phenyl-indole-3-acetic acid (5-Ph-IAA), fits with high specificity. This mutation resolved the basal degradation and high auxin dosage issues[1]. The ligand-inducibility and reversibility of the AID system make it a potentially powerful tool to study molecular biology. In particular, systems such as Ah receptor (AHR) signaling where repressors such as the Ah receptor repressor (AHRR) are rapidly upregulated and are difficult to study through the control of mRNA levels.

The goal of the authors for this paper was to investigate the power of the AID system to produce a tool useful for the study of the AHR pathway in cell lines such as HCT116. For system evaluation, a successful level of oblation was considered anything at or above 80%. Knockdown of 95% or greater would have been ideal. In our initial attempt, we stably expressed the improved OsTIR1-F74G in HCT 116 cells. Once the universal stable lines were generated, proteins of interest could be tagged with a degron, thereby making the complete AID system. In these initial studies, we planned to make two separate models. One would have the AHRR with the degron tag. The other would have cytochrome P450 1A1 (CYP1A1) tagged with the degron. These two models would allow further investigation of the roles of feedback inhibition on the AHR pathway, which could produce insights into cancer, development, and role of xenobiotic metabolism in normal physiology.

MATERIALS AND METHODS

Chemicals, solutions, and antibodies: Five-phenyl-indole-3-acetic acid (5-Ph-IAA, cat. no. HY-134653) was purchased from MedChemExpress (Monmouth Junction, NJ). Lipofectamine 3000 (cat. no. L3000015) and 4',6-diamidino-2-phenylindole (DAPI, cat. no. D1306), along with the Pierce bicinchoninic acid (BCA) protein assay kit (cat. no. 23227), were purchased from ThermoFisher (Waltham, MA). Dulbecco's Modified Eagle Medium (DMEM) cell culture media (cat. no. 11960044) was purchased from Gibco (Waltham, MA), along with fetal bovine serum (FBS), penicillin/streptomycin (cat. no. 15140122), L-glutamine (cat. no. 25030081), puromycin dihydrochloride (cat. no. A1113803), and 0.05% trypsin-EDTA (cat. no. 25200056). CelLytic M (cat. no. C2978) for cell lysis was purchased from Sigma-Aldrich (St. Louis, MO). Western blot reagents were as follows: PVDF membrane sandwiches (cat. no. 1620219), extra thick blot paper (cat. no. 1703966), Precision Plus Protein All Blue Prestained Protein Standards (cat. no. 1610373), AP Conjugate Substrate Kit (cat. no. 1706432), and 10X PBS (cat. no. 1610780) from Bio-Rad (Hercules, CA); Bolt 8% Bis-Tris Plus gels (cat. no. NW00082BOX), Bolt LDS Sample Buffer (cat. no. B0007), and Bolt Sample Reducing Agent (cat. no. B0009) from Invitrogen (Carlsbad, CA); Bolt MOPS SDS Running Buffer (cat. no. B0001) and Bolt Transfer Buffer (cat. no. BT00061) from Novex (Carlsbad, CA); blotting power source (cat. no. FB 600) from Fisher Scientific (Pittsburgh, PA); SEMI-PHOR semi-dry transfer apparatus (cat. no. TE77) from Hoefer Scientific Instruments (San Francisco, CA); anti-OsTIR1 polyclonal antibody (cat. no. PD048) from MBL International (Woburn, MA); Goat anti-Rabbit IgG (H+L) Secondary Antibody [Alkaline Phosphatase] (Pre-adsorbed) (cat. no. NB120-6024) from Novus Biologicals (Centennial, CO); methanol (cat. no. A452-4) from Fisher Chemical (Pittsburgh, PA); Flavorite non-fat dry milk (NFDM) from Preferred Products, Inc. (Eden Prairie, MN). Tris HCl (cat. no.

T3253), Tris Base (cat. no. T1503), NaCl (cat. no. S9888), and Tween20 (cat. no. P1379) were from Sigma-Aldrich. The TTBS+ buffer concentrations of the above chemicals are as follows: Tris HCl 40.5mM, Tris Base 9.5mM, NaCl 300mM, 0.5% Tween.

Cell culture and puromycin kill curve: HCT 116 (cat. no. CCL-247) cells were obtained from the American Type Culture Collection (ATCC, Manassas, VA). HCT 116 cells were cultured in DMEM with 10% FBS, 100 U/mL penicillin/streptomycin, and 2mM L-glutamine. Cells were kept at 5% CO_2 and 37ºC in a humidified incubator. For puromycin (PURO) kill curve, cells were seeded at 5,000 cells per well in a 96-well plate and given 24 hours to adhere. Then, cells were treated with 0, 0.1, 0.3, 0.5, 0.7, 1, or 2ug/mL PURO in cell culture media. Samples were run in triplicate. Cells were incubated with PURO and viability was measured at days 4, 5, and 6 with the Promega (Madison, WI) CellTiter 96 Aqueous One Solution Cell Proliferation Assay (cat. no. G3582). Graphed results suggested 1ug/mL PURO was the optimal concentration for HCT 116 cells.

PCR primers, guide RNAs (gRNAs), and plasmids: Primers for PCR screening and gRNAs were all ordered from Integrated DNA Technologies (IDT, Coralville, IA) (Table 1). The OsTIR1-F74G (pMK381, cat. no. 140536) and mAID-EGFP (pAY5, cat. no. 140532) plasmids were obtained from AddGene (Watertown, MA). Both were deposited by the Kanemaki lab[1]. The mAID-mCherry plasmid was synthesized through Twist Bioscience (South San Francisco, CA). The CMV control plasmid has no ORF in front of a CMV promoter; it was a construct from the Bradfield lab (pL1966).

Nucleofection and recovery for stable cell line generation: HCT116 cells were edited using plasmid "nucleofection" via the Lonza (Basel, Switzerland) 4D Nucleofector system. The nucleofection mixture was as follows: gRNA, plasmid harboring an OsTIR1-F74G open reading frame (ORF), Cas9 enzyme, and commercial "electroporation enhancer" mixed to 20uM, 10uM, 19.5uM, and 20uM, respectively. The HCT 116 cells of a low passage number (below 10 passages) were counted, washed, resuspended in Lonza SE nucleofection buffer (cat. no. V4XC-1032). The mixture was added to an electroporation cassette, and the cells were then nucleofected in the Lonza 4D system using the EN-113 pulse code. Nucleofected cells were then plated in a 96-well plate for recovery and growth. After approximately two days, the cells were expanded to a 6-well plate when they reached confluency. The progression from 96-well to 6-well plates was used for optimal cell growth conditions, since cell survival directly after nucleofection was most successful when cells were kept relatively confluent.

PURO selection and cloning for stable cell line generation: Three days post-nucleofection, regular cell culture media was replaced with cell culture media plus 1ug/mL PURO. Media containing PURO was replenished at day 7 and day 10 post-nucleofection. This selected for cells likely containing the desired edit, but it did not create single-cell clones. At day 10 after nucleofection, cells were collected from plates via trypsinization, pooled per group (e.g. all no-guide-control (NGC) groups together, all gRNA 1 groups together, all gRNA 2 groups together, etc.), washed with wash buffer (WB, 2% v/v FBS in PBS), and resuspended in WB with 1ug/uL DAPI. Resuspended cells were taken to the UW Carbone Cancer Center Flow Cytometry Laboratory for live/dead sorting on the Becton Dickinson (BD, Franklin Lakes, NJ) FACSAria Cell Sorter to achieve monoclonal populations. Single cells were deposited in each well of a 96-

well plate. Clones were allowed to outgrow for a week before the media was changed to 1ug/uL PURO media for confirmatory selection. Approximately three weeks after single cell sorting, cells surviving PURO selection were expanded for freezing in multiple aliquots of one million cells per Corning (Corning, NY) 1.2mL Internal Threaded Polypropylene Cryogenic Vial (cat. no. 430487).

PCR selection scheme: An aliquot of the expanded clones (approximately 100,000 cells) was examined by PCR amplification as follows. Samples for PCR were boiled at 95ºC for 15 minutes to generate a crude cell lysate which was then used as template for the internal (primers OL2399/OL2400), 5' spanning (primers OL2398/OL2399), and 3' spanning (primers OL2408/OL1741) PCRs (Figure 1). Takara (San Jose, CA) PrimeSTAR GXL DNA Polymerase (cat. no. R050B) kit was used for all PCRs. Clones positive by PCR were grown in 10cm dishes for Western blotting. All primers are described in Table 1.

Western blot analysis of the OsTIR1 protein: Protein extracts for western blot were made by collecting cells via trypsinization, washing, and resuspending in 25 to 50 uL of CelLytic M per 1 million cells. Suspensions were then incubated with rotation at 4ºC for 15 minutes followed by a subsequent 15-minute spin at 4ºC and 12,700 rpm to remove pelleted cell debris. The cleared supernatant was then quantified using the BCA assay per manufacturer's instructions[11]. Protein samples were denatured in Bolt LDS sample buffer and reducing agent (both diluted to 1X) with the following protocol: 10 minutes at 70ºC, 2 minutes at 85ºC, with storage at 4ºC. One hundred micrograms of denatured protein per sample was loaded on the 8% Bis-Tris gel and run at 97 V for approximately 30 minutes for stacking then 204V until the dye front ran out of the bottom of

the gel (approximately 30 minutes). The gel was then equilibrated in Bolt transfer buffer for 30 minutes, along with the methanol-activated PVDF membrane and blotting paper. Separated proteins were transferred to the PVDF membrane via semi-dry transfer at 1mA/cm^2 for 1 hour. The membrane was blocked in 10% NFDM in PBS (Blotto) overnight. Primary anti-OsTIR1 antibody was diluted 1:500 in 5% Blotto, and the membrane was incubated for 1 hour at room temperature (RT) with the antibody mixture. Three washes were done with TTBS+ buffer. Secondary antibody was diluted 1:10,000 in 5% Blotto and incubated with the washed membrane 1 hour at room temperature. Another set of three washes was done with TTBS+ before starting color development with the AP Conjugate Substrate Kit. Briefly, reagents A and B of the kit were diluted to 1X using distilled water and poured over the washed blot. Color development ran approximately 8 minutes before the reaction was stopped with distilled water. The blot was dried before photographing for clearest bands.

Lipofection, auxin treatment, and flow cytometry: Cells were plated at a density of 500,000 cells per well in a 6-well plate and given 24 hours to adhere under standard cell culture conditions. Then, cells were "lipofected" with specified plasmids and noted concentrations using the manufacturer's Lipofectamine3000 protocol[12]. If treatment with auxin was required, cells were given 24 hours post-lipofection and treated with 1uM 5-Ph-IAA in their unchanged cell culture media. Forty-eight hours post-lipofection, cells were collected via trypsinization, washed with wash buffer (WB, 2% FBS in PBS), resuspended in WB with 1 ug/uL DAPI, and taken to the UW Carbone Cancer Center Flow Cytometry Laboratory. Data was recorded on the ThermoFisher (Waltham, MA) Attune NxT machine. For Figure 3 only, cells were treated with 5-Ph-IAA 1 hour before sorting.

Data analysis and gating strategy: Flow cytometry data was analyzed using FCSExpress from De Novo Software (Pasadena, CA). Histogram gate percentages reflect the percent of single, live cells that had EGFP or mCherry expression. Gating was as follows (Y vs X axis): SSC-A vs FSC-A (cells); FSC-H vs FSC-A (single cells); DAPI vs FSC-A (live cells); EGFP-A vs FSC-A (EGFP-positive cells) or mCherry-A vs FSC-A (mCherry-positive cells); Count vs EGFP-A or mCherry-A (histogram of EGFP- or mCherry-positive cells). Percent changes were calculated by dividing the difference between groups by the control level of fluorescence. For example, the difference between an mCherry-IAA group with 30% mCherry positive cells versus an mCherry+IAA group with 25% mCherry positive cells would be 5 / 30 * 100 = 16.67%.

RESULTS

A mismatch in gRNA cut site and plasmid homology arms still results in a functional edited cell line. The first round of clone generation unintentionally used a gRNA which cut 48 base-pairs upstream of where the plasmid homology arms aligned with the genome (Figure 1). The gRNA used was 296D.target.AAVS.4 (Table 1). Identification of the error was made after clones had already been generated, so screening was undertaken. A PCR designed within the OsTIR1 knock-in sequence revealed two of three collected clones had the insert (Figure 1 and 2A). Clones Os.1.A10 and Os.1.B10 were then screened using two additional PCRs. One was designed to span the genomic DNA and plasmid insert on the five-prime end, while the second spanned the genomic DNA and plasmid insert on the three-prime end (Figure 1). These spanning PCRs revealed bands for Os.1.A10 on the five-prime and three-prime ends (Figure 2B). Clone Os.1.B10 had a successful band on the three-prime end only (Figure 2B). However, the three-prime end PCR also generated a band in the NGC lane, suggesting these results were non-specific to the knock-in (Figure 2B). Due to Os.1.A10 and Os.1.B10 generating bands in the internal PCR and spanning PCR(s), both clones were screened using a Western blot (Figure 2C). The Western blot revealed OsTIR1-F74G protein production only in clone Os.1.A10.

Properly designed gRNA cut site and plasmid homology arms generates an additional functional OsTIR1 clone. A secondary round of clone generation was undertaken using a gRNA designed with perfect alignment to the plasmid homology arms (Figure 1). The gRNA was 388B.target.AAVS1.Kanemaki (Table 1). Clones were screened in exactly the same way as described above. Internal and spanning PCRs are not shown due to the volume of clones screened, but results looked similar to those from round one (Figure 2A and B). Clones positive

by PCR were also screened using a Western blot; of the five Os clones from round two, only Os.5.D2 was confirmed to make OsTIR1 protein (Figure 2C). The no-guide-control (NGC) clone, NGC.1.B2, is an unedited wild-type cell line exposed to the same nucleofection mix as the Os clones but lacking a gRNA.

Stable OsTIR1-expressing HCT 116 clones show no significant decrease in degron-reporter expression. Clones positive for OsTIR1-F74G were examined for the capacity to regulate the expression of an mAID-mCherry reporter which had been introduced via lipofection (Appendix 1). Cells were cultured with 5-Ph-IAA in their growth media one hour before flow cytometry (Figure 3). Complete degradation of the reporter in the presence of the auxin was not observed in either clone Os.1.A10 or clone Os.5.D2. This experiment was repeated three times; data not shown.

Transient OsTIR1 expression in HCT 116 cells fails to reduce expression of a previously published EGFP-based degron-reporter, even in the presence of auxin. Wild-type HCT 116 cells were lipofected with the OsTIR1-F74G plasmid and pAY5, an mAID-EGFP reporter plasmid acquired from the Kanemaki lab via AddGene, or pAY5 alone (Figure 4)[1]. For each plasmid, 250ng were lipofected 48 hours before flow cytometry. Cells were treated with 1uM 5-Ph-IAA 24 hours before flow cytometry for maximum exposure. Again, flow cytometry shows limited (if any) decrease in EGFP expression in the OsTIR1 plus 5-Ph-IAA group.

Transiently transfected OsTIR1 plasmid shows strong constitutive degradation activity at high amounts. All groups were transfected with 500ng of EGFP reporter. The OsTIR1-F74G or

CMV-empty control plasmids were transfected at 100ng, 500ng, and 1,000ng. Flow cytometry was done 48 hours post-nucleofection when fluorescence was maximal. The OsTIR1-F74G plasmid showed strong basal degradation at 500ng and up (Figure 5). No groups were treated with auxin in this experiment. The 100ng condition was used in following experiments to mitigate basal degradation.

Addition of auxin to transiently transfected OsTIR1 plus degron-reporter does not cause an increase in degradation over those seen with addition of OsTIR1 alone. Cells were transiently transfected with 500ng reporter-degron constructs and 100ng of either the empty CMV control or OsTIR1 plasmid. Cells transfected with CMV control and reporter-degron showed no significant difference between groups treated with and without 1uM 5-Ph-IAA (Figure 6). In all groups, measurable fluorescent expression was seen.

DISCUSSION

Positive CRISPR events do not always correspond to regulatable protein expression. Two rounds of clone generation and single-cell sorting yielded 99 unscreened monoclonal cell populations that potentially had integrated the OsTIR1-F74G minigene into the HCT 116 genome. Subsequent PCR analysis was used to verify the proper integration of the minigene at its five-prime and three-prime ends (Figure 1). Twenty clones out of the 99 screened were positive by PCR for the edit (Figure 2A and B). Seven PCR-positive clones were chosen for protein extraction based on speed of growth and consistent morphology. These were further examined by Western blot. Only two of these seven clones displayed positive bands for OsTIR1 (Figure 2C). Assuming this proportion of clones positive by Western (28.57%) holds true for the population, then approximately 6 of the 20 PCR-positive clones should have been confirmed via Western, if screened. Therefore, 6 out of the 99 initial clones (6.06%) would be positive for OsTIR1 protein despite more having the knock-in by PCR. Only 20.2% of the initial 99 clones were sufficient for further study based on PCR screening. These results illustrate the low rate of appropriate gene editing, reflecting both inefficiencies in proper insertion of the construct at the target site and of the insert protein being expressed at a measurable level. Additionally, these results portray the volume of material and reagents required to generate a successful knock-in clone. Interestingly, the “best” clone based on PCR and Western results, Os.1.A10, had a successful knock-in despite the misalignment of the gRNA and plasmid homology arms detailed in the Results section (Figure 1).

The OsTIR1 protein is fortunate to have a specific, high-quality antibody for use in Western blot analysis. However, many proteins lack any antibody of quality. As such, these data suggest that

confirmation of successful editing at the DNA level cannot always be reliable when immunological reagents do not exist for a high-fidelity protein production assay. The editing rates and results of screenings in this OsTIR1 clone generation suggest "positive" clones confirmed only with PCR should be treated with skepticism until confirmation by protein analysis is completed with confidence.

Stable OsTIR1 clones are not responsive to auxin even through protein is present via Western blot. When OsTIR1-positive clones (Os.1.A10 and Os.5.D2) were transfected with an mAID-mCherry reporter and dosed with 5-Ph-IAA, mCherry expression didn't significantly decrease (Figure 3). Clone Os.5.D2 didn't respond at all, despite making measurable protein (Figure 2C). The maximum reduction in mCherry occurred in clone Os.1.A10 and was roughly 11%. Accepting the statistical significance of this figure, which is doubtful, it begs the question whether an 11% decrease would allow any useful interrogation into a biological pathway; the published system upon which these experiments are based shows a complete or nearly complete oblation of signal[1]. In this experiment, presence of OsTIR1 and 5-Ph-IAA did not decrease protein expression to a level suitable for study of a "knockout" phenotype.

Failure to respond to auxin is not a feature unique to stably generated OsTIR1 lines. Wild-type HCT 116 cells transfected with OsTIR1-F74G and mAID-EGFP reporter plasmids were treated with and without 5-Ph-IAA. This experiment recapitulated the results seen with the stably generated clones (Figure 4). Cells with OsTIR1-F74G plasmid treated with 5-Ph-IAA saw a 19.1% decrease in EGFP expression, but overall EGFP expression was still around 40%. The

oblation of target protein achieved with transient transfection in HCT 116 cells was still insufficient to model a knockout phenotype.

OsTIR1 degrades degron-reporter without the presence of auxin in a dose-responsive manner. Transient transfection into wild-type HCT 116 cells with increasing amounts of OsTIR1 or CMV control plasmid, but constant levels of mAID-EGFP reporter, provide evidence that OsTIR1 is constitutively active at high levels of expression in HCT 116 cells (Figure 5). In cells never treated with auxin, increasing the OsTIR1 plasmid concentration from 100ng to 500ng resulted in a 93.8% decrease in EGFP expression. This represents the largest decrease in any experiment conducted in our hands with OsTIR1-F74G. The CMV control plasmid at the same level of increase only had a 3.51% decrease in EGFP expression. This result suggests auxin addition is unnecessary for OsTIR1 to destroy a degron-tagged protein in HCT 116 cells when OsTIR1 is expressed at high levels. Additionally, this experiment reinforces the importance of the details and controls of a transient transfection experiment; the dose of OsTIR1 must be tailored to prevent basal degradation, but the dose cannot be so low as to not function.

Addition of auxin does not increase the level of degradation in cells transiently transfected with OsTIR1 and degron-reporter. Cells were transfected with 100 ng of OsTIR1 or CMV control and the degron-reporter. These groups were then treated with and without 5-Ph-IAA. Addition of auxin caused a 24.1% (EGFP) and 39.3% (mCherry) decrease in reporter expression (Figure 6). However, reporter expression was still detectable at near-baseline levels. Even with an optimized amount of OsTIR1 transfected, degradation of mAID-EGFP and mAID-mCherry is modest. This keeps with the previous experiments done to characterize this system.

Utility of the OsTIR1 degradation system in the HCT 116 cell line. The experiments in this paper sought to establish an auxin-inducible degron system in HCT 116 cells by inserting the OsTIR1 protein at the AAVS1 safe-harbor locus (Figure 1). The authors defined sufficient knockdown as a percent decrease at or above 80%, but a decrease of 95% or above would have been ideal. Two stably expressing clones, Os.1.A10 and Os.5.D2, were generated and confirmed for genomic integration and protein expression. Despite being positive for protein via Western blot (Figure 2), both clones failed to markedly reduce reporter-degron proteins to a level consistent with published literature (Figure 3). Transient transfection experiments with this system recapitulated the observation that this is a system with a poor ability to degrade tagged proteins (Figures 4-6). The maximum amount of degradation seen was 93.8% in cells untreated with auxin and having OsTIR1 expressed at a much higher level than the reporter construct (Figure 5). Additionally, this high level of degradation was abnormal when considered in the context of the other experiments; all other levels of protein oblation were below 50%, with most being at 25% or less.

Published literature suggests the AID system is highly responsive, reversible, and auxin-dependent without basal degradation. When employed in the HCT 116 cell line, our experiments argue against the utility of this system in many cell culture lines. Our data are not in agreement with the auxin inducibility of this system and suggest that in some cases the activity of OsTIR1 may be constitutive. One aspect we could not examine was reversibility of the AID system. No significant, auxin-dependent degradation was observed, so we could not test reversibility. The AID system may be suitable for the study of reversible protein degradation; however,

establishing a stable edit in the HCT 116 cell line may require significant optimization in addition to the experiments conducted in this chapter. Alternatively, a reevaluation of knock-in strategy, type of AID system introduced, and a new iteration of clone generation may result in a more functional system.

TABLES

Table 1. Sequences of PCR primers and gRNAs.

Name	Sequence	Direction
OL2398	CCTCTCCATCCTCTTGCTTTC	5’ > 3’
OL2399	TGCACACAGAGGGTAGATAAAG	3’ > 5’
OL2400	CAGGAGAATGAAGTGGAGGATAG	5’ > 3’
OL2408	ATTAAGGGCCAGCTCATTCC	5’ > 3’
OL1741	AGAGTGAGTTTGCCAAGCAGTCACC	3’ > 5’
296D.target.AAVS.4	ATTCCCAGGGCCGGTTAATGTGG	5’ > 3’
388B.target.AAVS1.Kanemaki	CACCCCACAGTGGGGCCACTAGG	5’ > 3’

FIGURES

Figure 1) Knock in scheme for stable OsTIR1 clone generation. Black text is intron 1 sequence at site of interest. Blue text is gRNA designed to locus. Letters in bold are PAM sequence. The ^ denotes cut site. Vertical lines in insert mark beginning/end of homology arms (HA), denoted left (L) and right (R). Boxes denote indicated feature sequence, in order: CMV promoter, OsTIR1-F74G insert, poly A site, LoxP site, PGK promoter, puromycin resistance gene, LoxP site. Homology arm sequences match sequences flanking the left and right of the AAVS1 cut site. Primers are denoted by horizontal portion of L-shapes labeled by primer name. Forward primers are depicted above the insert with stem on the left; reverse primers are below the insert with stem on the right. PCR schemes were as follows: 5 prime-spanning pair = OL2398/OL2399; internal pair = OL2400/OL2399; 3 prime-spanning pair = OL2408/OL1741. The black star denotes the cut site of the 296D.target.AAVS.4 gRNA. The knock-in was targeted to the AAVS1 safe harbor locus in the PPP1R12C gene.

AAVS1 Safe Harbor Locus
(PPP1R12C Gene)
Intron 1

TAT CTG TCC CCT CCA CCC CAC AGT GGG GCC ACT **AGG** GAC
CA CCC CAC AGT GGG GCC^ACT **AGG>**

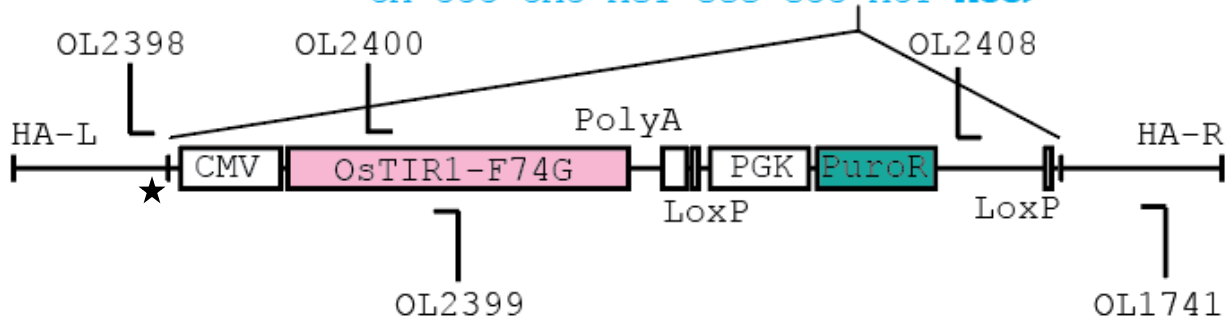


Figure 2) Characterization of selected positive clones. A) Example of internal PCR screen done to over 100 clones. A small portion of the insert was amplified to identify putative knock-ins. **B)** Example of an "in/out" PCR screen done to over 100 clones. Primers sit in the genomic DNA around the insert and within the insert itself. Presence of bands on both the 5 prime and 3 prime ends indicates strong likelihood the insert was integrated in the DNA. **C)** Western confirmation of OsTIR1 protein presence in clones identified by PCR as positive. Of the seven clones identified via PCR as positive, only two showed a protein band at the proper size. NGC cell lines were mock-transfected and serve as negative controls.

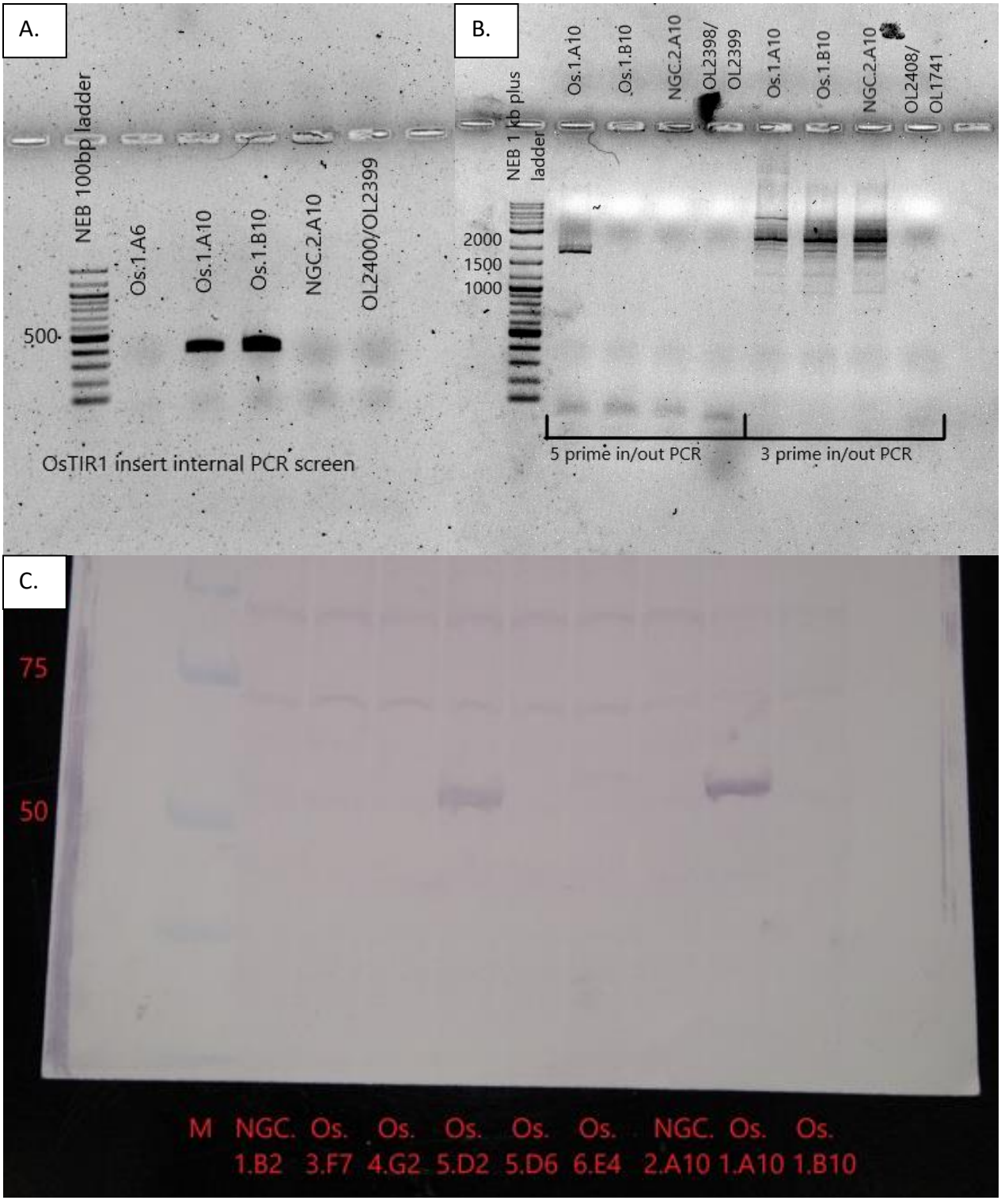

A.
NEB 100bp ladder
Os.1.A6
Os.1.A10
Os.1.B10
NGC.2.A10
OL2400/OL2399
500
OsTIR1 insert internal PCR screen
B.
NEB 1 kb plus ladder
Os.1.A10
Os.1.B10
NGC.2.A10
OL2398/ OL2399
Os.1.A10
Os.1.B10
NGC.2.A10
OL2408/ OL1741
2000
1500
1000
5 prime in/out PCR
3 prime in/out PCR
C.
75
50
M NGC. Os. Os. Os. Os. Os. NGC. Os. Os.
1.B2 3.F7 4.G2 5.D2 5.D6 6.E4 2.A10 1.A10 1.B10

Figure 3) Stable knock-in clones fail to oblate degron-reporter expression in the presence of auxin. Clones were transfected with the mAID-mCherry reporter plasmid, given 48 hours to produce protein, and treated with 5-Ph-IAA 1 hour before flow cytometry. Maximal downregulation of reporter occurred in clone Os.1.A10 at a roughly 11% decrease; this would be insufficient to study the effect of complete protein oblation.

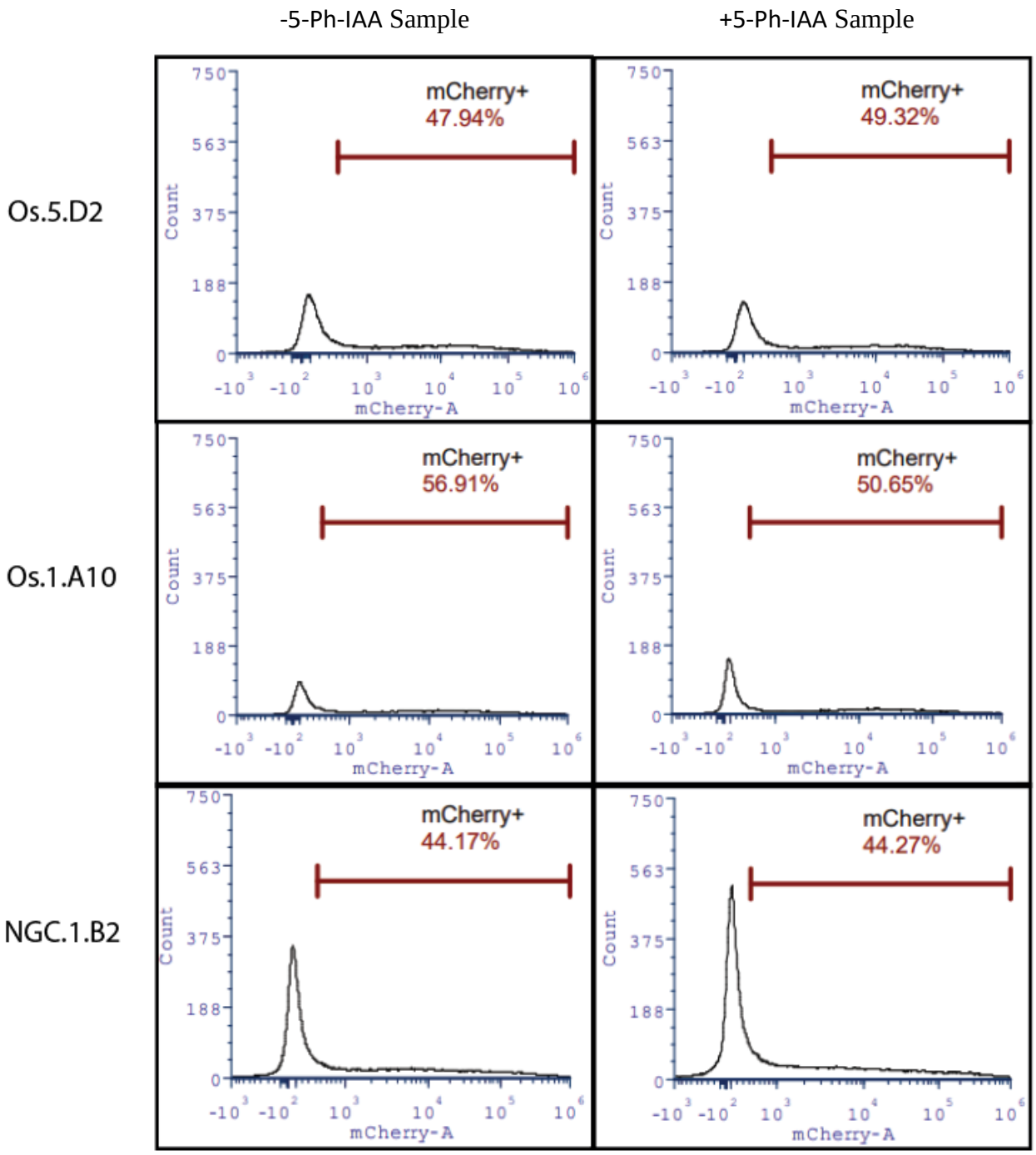

-5-Ph-IAA Sample
+5-Ph-IAA Sample
Os.5.D2
Os.1.A10
NGC.1.B2
mCherry+
47.94%
mCherry+
49.32%
mCherry+
56.91%
mCherry+
50.65%
mCherry+
44.17%
mCherry+
44.27%
Count
mCherry-A
750
563
375
188
0

Figure 4) Transient plasmid expression at maximal amounts fail to completely oblate reporter expression in the presence of auxin. Experiments were done in wild-type HCT 116 cells. Reporter plasmid was cotransfected with a CMV-empty control or the OsTIR1-F74G plasmid. The OsTIR1-F74G group + 5-Ph-IAA treatment caused a 19.1% decrease in reporter expression compared to the 5-Ph-IAA untreated group. However, the OsTIR1-F74G + 5-Ph-IAA condition still had roughly 43% of cells still expressing the fluorescent reporter, which would not allow study of a “knockout” condition.

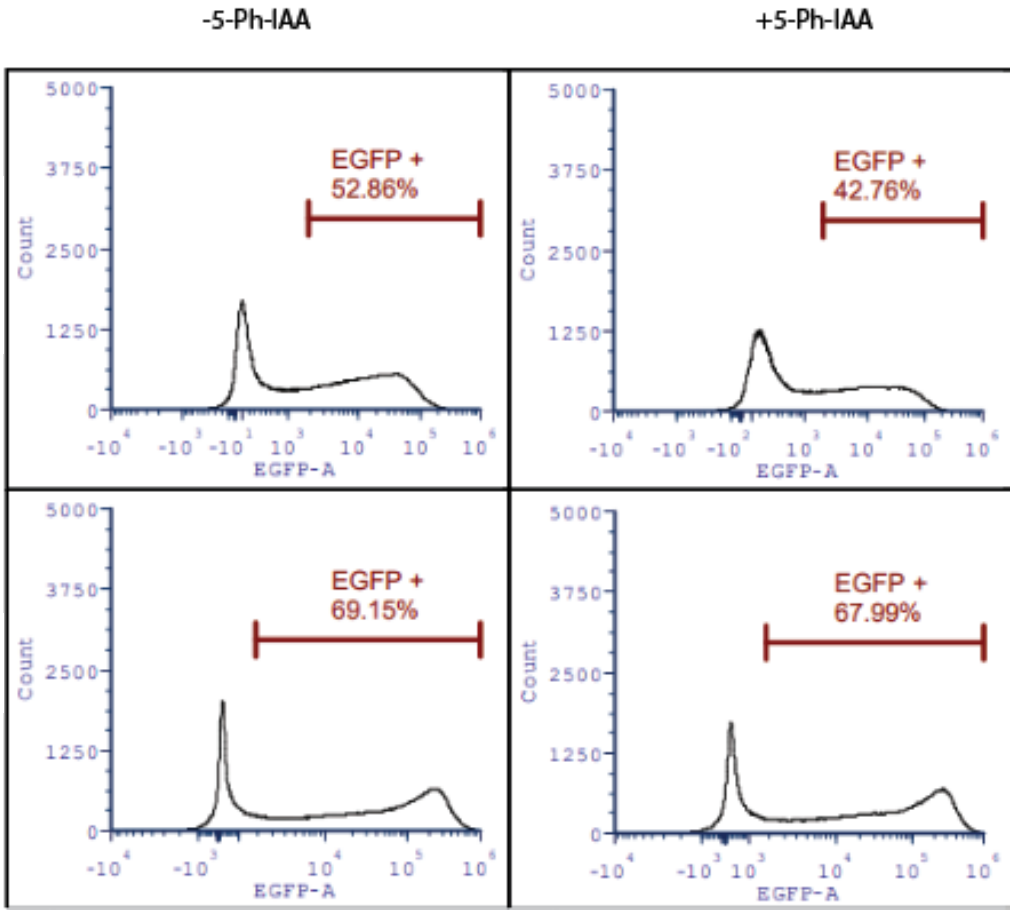

-5-Ph-IAA
+5-Ph-IAA
OsTIR1-F74G (pMK381) + mAID-EGFP (pAY5)
mAID-EGFP (pAY5)
EGFP +
52.86%
EGFP +
42.76%
EGFP +
69.15%
EGFP +
67.99%
Count
EGFP-A

Figure 5) Titration of OsTIR1-F74G plasmid reveals strong reporter basal degradation at higher concentrations. Wild-type HCT 116 cells were cotransfected with 500ng pAY5 and the CMV-empty control plasmid or the OsTIR1-F74G plasmid. No groups were treated with auxin. The maximum dose of OsTIR1-F74G plasmid (1,000ng) decreased the reporter fluorescence 92.5% compared to the lowest dose (100ng). Future experiments used 100ng of OsTIR1-F74G plasmid to minimize basal degradation effects.

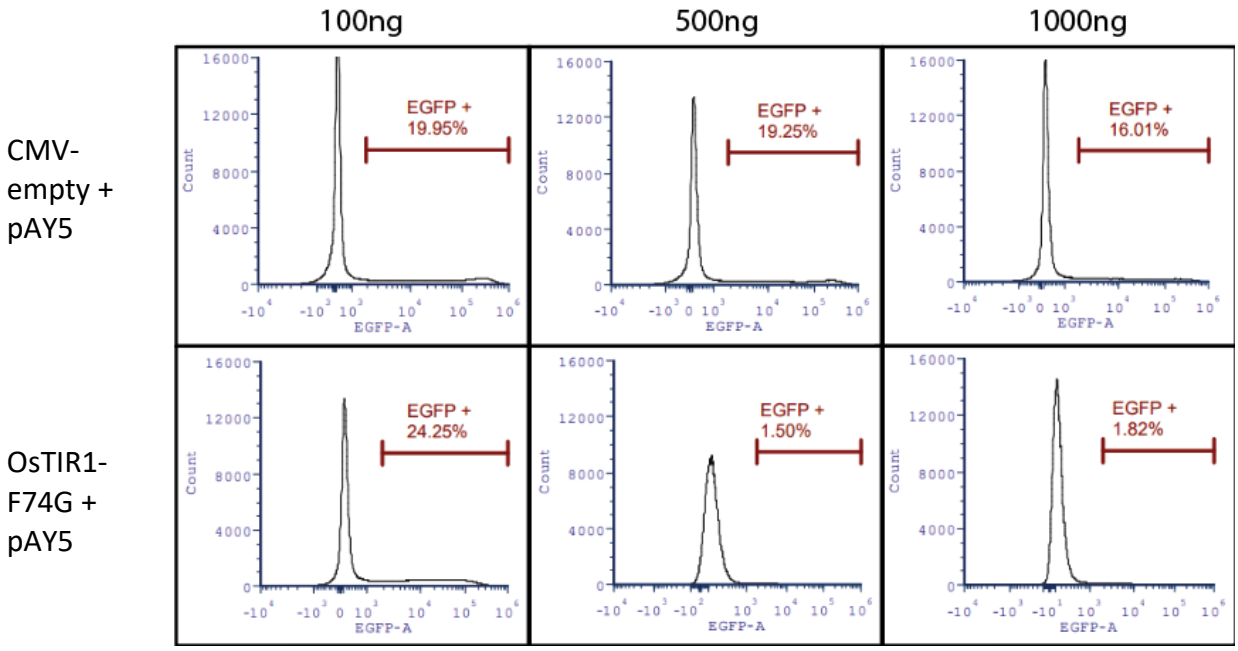
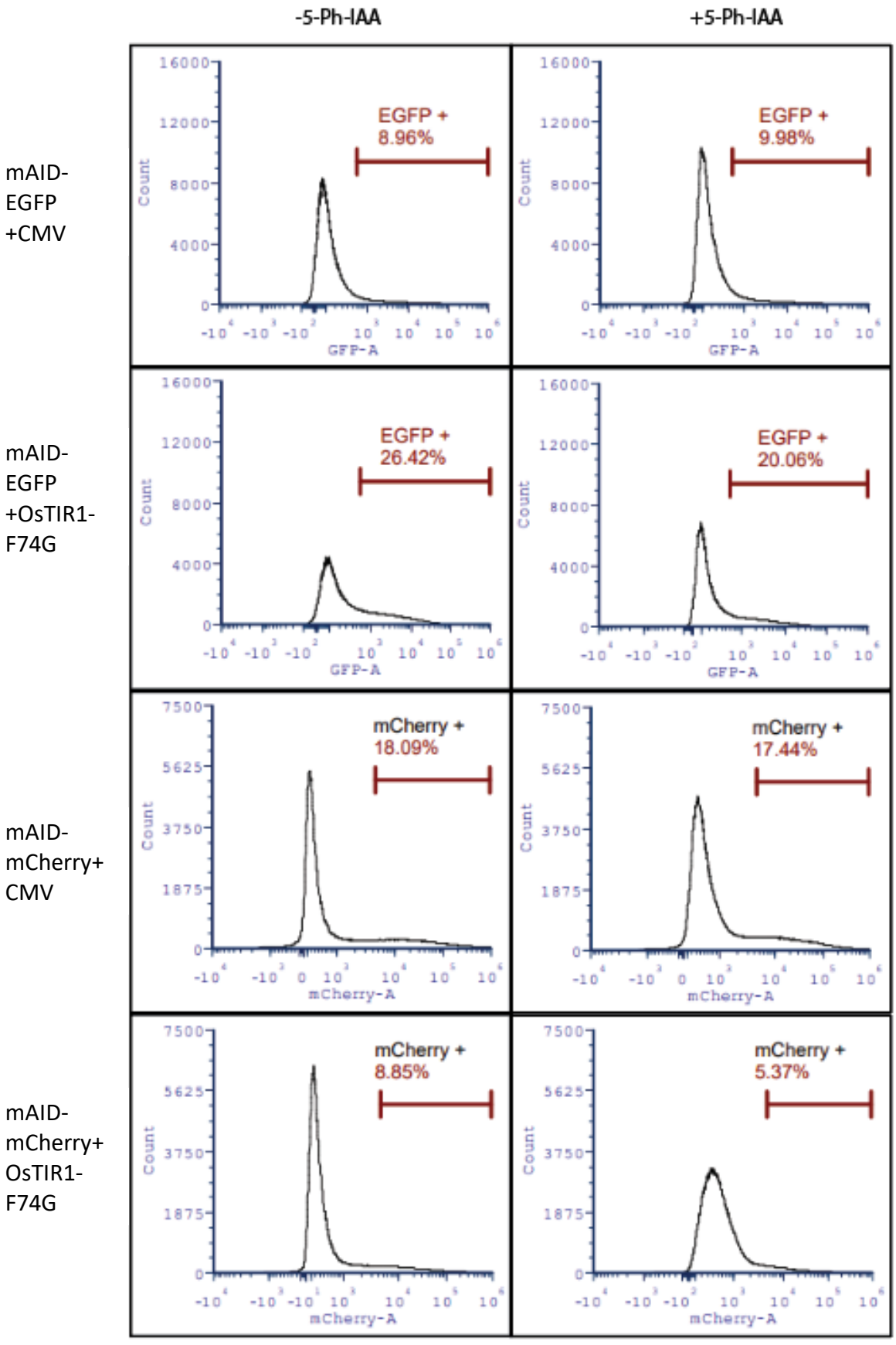

100ng
500ng
1000ng
CMV-empty + pAY5
OsTIR1-F74G + pAY5
EGFP + 19.95%
EGFP + 19.25%
EGFP + 16.01%
EGFP + 24.25%
EGFP + 1.50%
EGFP + 1.82%
Count
EGFP-A

Figure 6) Addition of auxin does not completely oblate protein expression. Wild-type HCT 116 cells were cotransfected with a fluorescent reporter (pAY5 or mAID-mCherry) and either a CMV-empty control or the OsTIR1-F74G plasmid. Groups were then treated with 5-Ph-IAA or a DMSO control for 24 hours before flow cytometry. OsTIR1-F74G-treated groups experienced a 24.1% (EGFP) or 39.3% (mCherry) decrease in reporter fluorescence. However, complete protein oblation was not achieved. Additionally, even a 39.3% decrease in protein amount would not provide a “knockout” situation.


-5-Ph-IAA
+5-Ph-IAA
mAID-EGFP +CMV
mAID-EGFP +OsTIR1-F74G
mAID-mCherry+ CMV
mAID-mCherry+ OsTIR1-F74G
EGFP + 8.96%
EGFP + 9.98%
EGFP + 26.42%
EGFP + 20.06%
mCherry + 18.09%
mCherry + 17.44%
mCherry + 8.85%
mCherry + 5.37%
Count
GFP-A
mCherry-A

REFERENCES

1. Yesbolatova, A. *et al.* The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nat Commun* **11**, 5701 (2020).

2. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).

3. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).

4. Jiang, F. & Doudna, J. A. CRISPR-Cas9 Structures and Mechanisms. *Annu Rev Biophys* **46**, 505–529 (2017).

5. Dana, H. *et al.* Molecular Mechanisms and Biological Functions of siRNA. *Int J Biomed Sci* **13**, 48–57 (2017).

6. Das, A. T., Tenenbaum, L. & Berkhout, B. Tet-On Systems For Doxycycline-inducible Gene Expression. *Curr Gene Ther* **16**, 156–167 (2016).

7. Nishijima, H., Yasunari, T., Nakayama, T., Adachi, N. & Shibahara, K. Improved applications of the tetracycline-regulated gene depletion system. *Biosci Trends* **3**, 161–167 (2009).

8. Kanke, M. *et al.* Auxin-inducible protein depletion system in fission yeast. *BMC Cell Biol* **12**, 8 (2011).

9. Natsume, T., Kiyomitsu, T., Saga, Y. & Kanemaki, M. T. Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Rep* **15**, 210–218 (2016).

10. Saito, Y. & Kanemaki, M. T. Targeted Protein Depletion Using the Auxin-Inducible Degron 2 (AID2) System. *Current Protocols* **1**, e219 (2021).

11. Pierce™ BCA Protein Assay Kit.

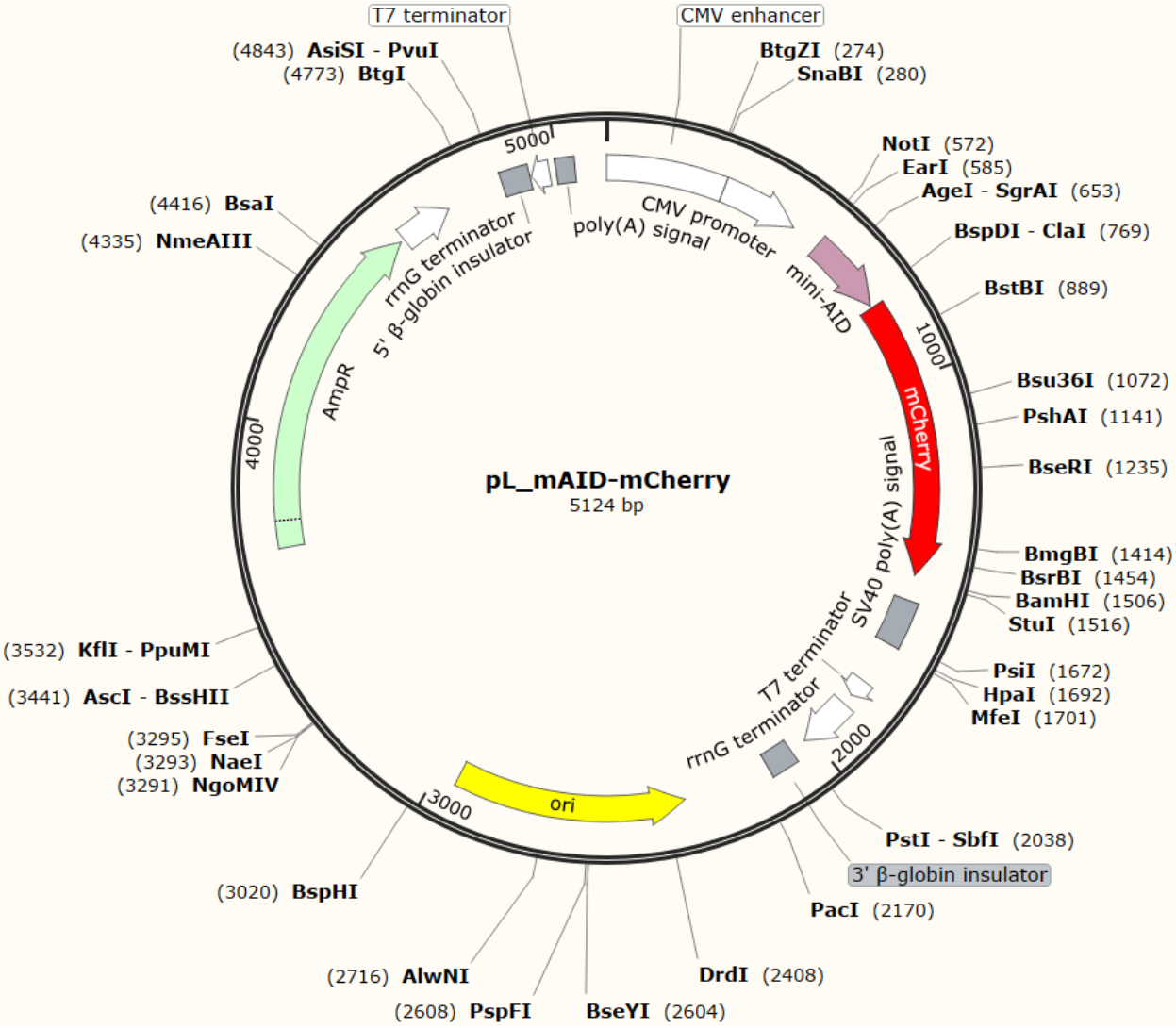
https://www.thermofisher.com/order/catalog/product/23225.

12. Lipofectamine™ 3000 Transfection Reagent.

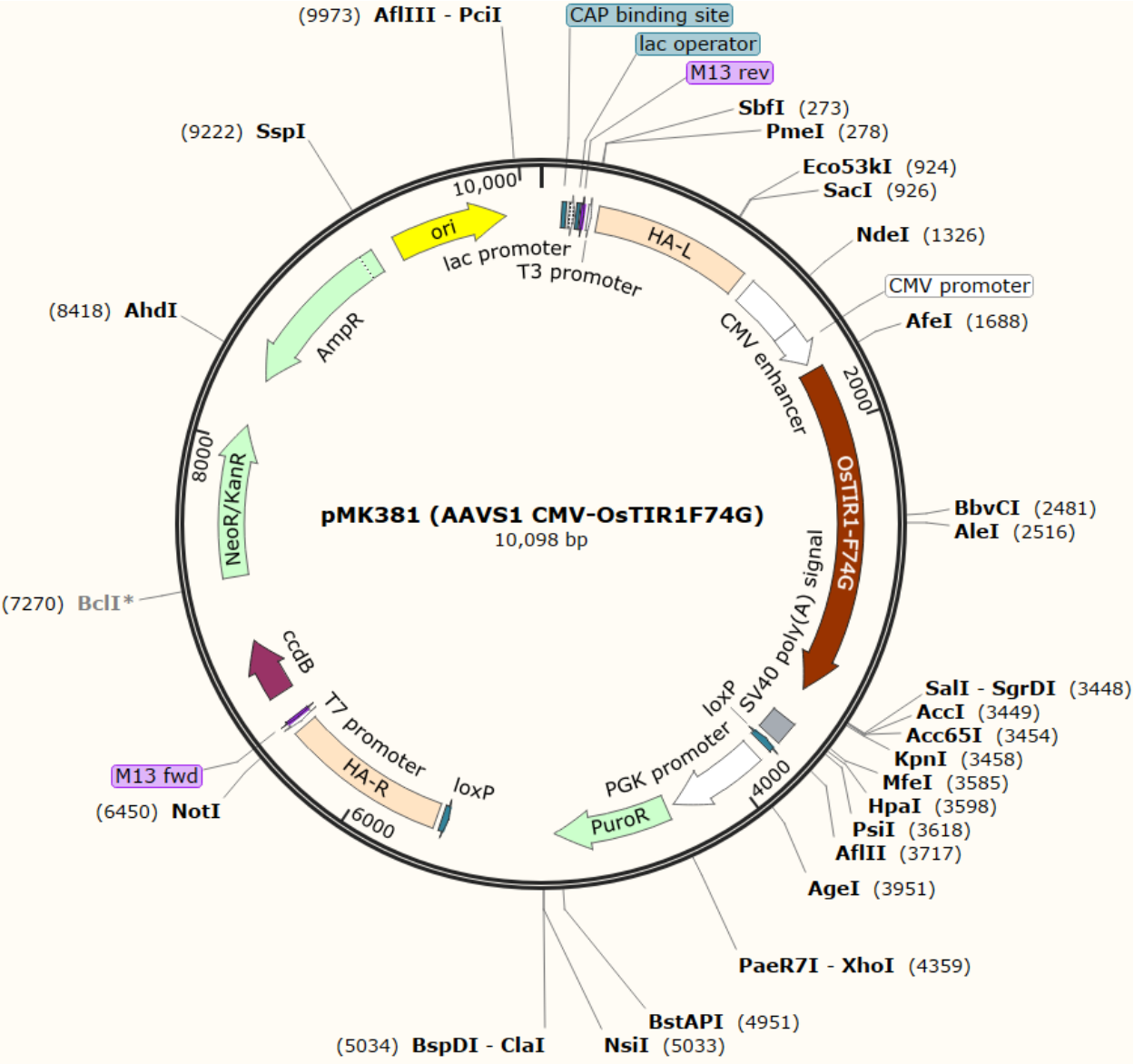
https://www.thermofisher.com/order/catalog/product/L3000150?SID=srch-srp-L3000150.

APPENDIX

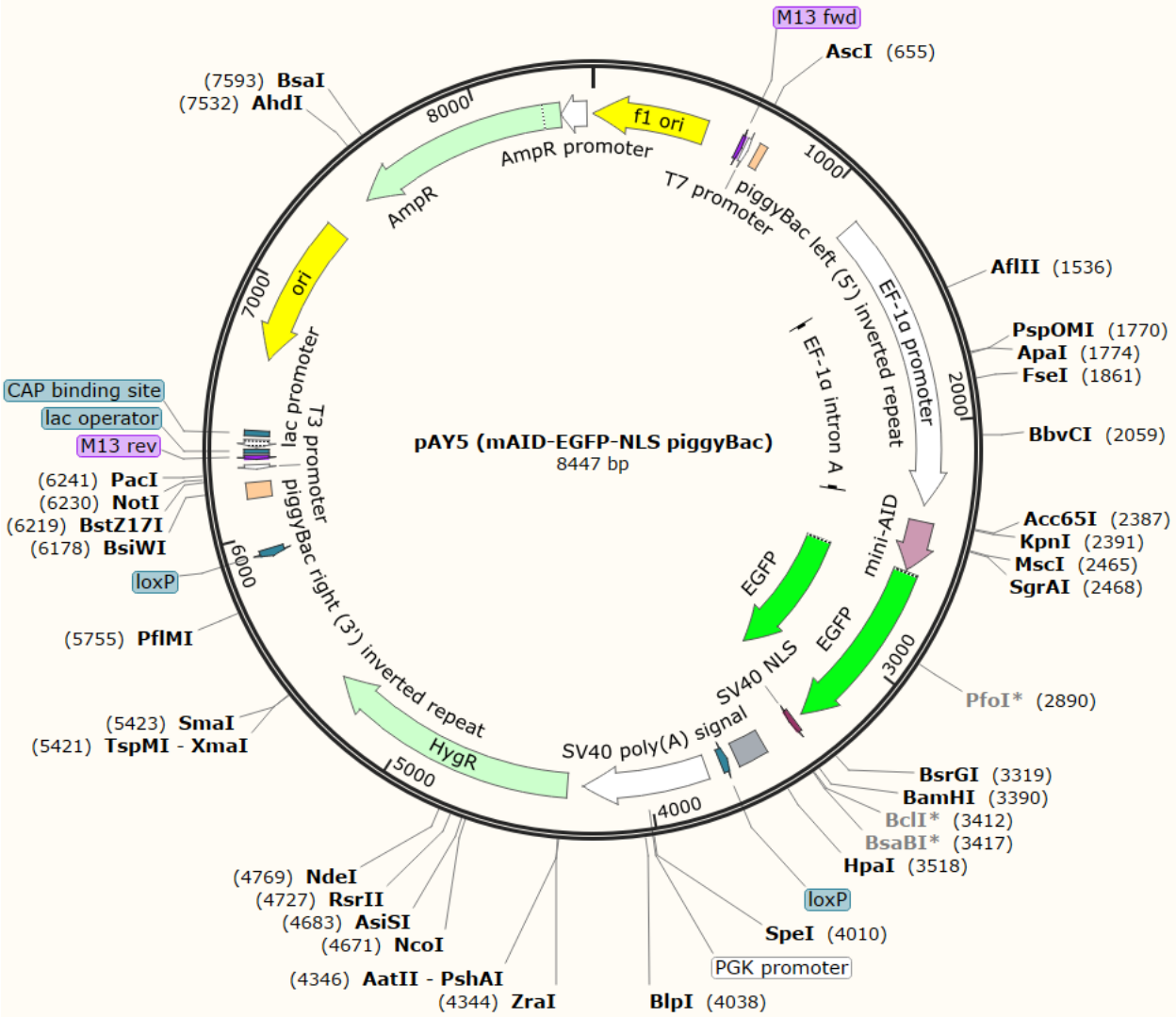
Appendix 1. Plasmid map of mAID-mCherry. The plasmid was synthesized by Twist Bioscience and uses the pTwist CMV vector. Sequences are mammalian codon optimized.



Appendix 2. OsTIR1-F74G (pMK381) plasmid map. The plasmid was received from AddGene. It was deposited by the Kanemaki lab[1].



Appendix 3. Plasmid map of mAID-EGFP (pAY5). The plasmid was received from AddGene. It was deposited by the Kanemaki lab[1].



Chapter 5: An Investigation into the Aryl Hydrocarbon Receptor Repressor as a Repressor and Transcription Factor

ABSTRACT

The aryl hydrocarbon receptor (AHR) is a ligand-activated transcription factor involved in a myriad of biological processes including innate immunity, development, xenobiotic metabolism, and responses to external stimuli[1–4]. When activated by one of its many ligands, the AHR up-regulates a variety of genes known for their capacity to metabolize xenobiotics, such as the cytochromes P450 (CYP) 1A1, 1A2, and 1B1. The AHR also upregulates the AHR repressor (AHRR). The AHRR is commonly thought to be a main mechanism of AHR signaling regulation and cessation. However, some evidence suggests the AHRR plays an additional role in downstream AHR signaling or in signaling by an independent pathway[5–7]. In this chapter, we investigate AHRR's ability to act as a repressor and explore its role as an essential component of an additional wave of gene regulation. To these ends, we performed induction experiments with a novel AHRR null cell line and observed only a modest degree of modulation of the induced CYP1A1 transcriptional response. Similarly, in complementary AHRR overexpression experiments, we only observed a ~50% decrease in the CYP1A1 response compared to wild-type cells. These experiments suggest the AHRR functions as an inefficient repressor of the system. Some RNA-seq experiments using transient overexpression of AHRR and AHR/AHRR chimeric proteins also suggest the AHRR regulates genes without AHR pathway activation. Taken together, this evidence suggests a biologically important function of the AHRR may be as a secondary transcription factor in the AHR pathway. The authors additionally report a constitutively active form of AHR that rivals or surpasses the abilities of published systems[8–12].

INTRODUCTION

The aryl hydrocarbon receptor (AHR) is a ligand-activated transcription factor of the Per-ARNT-Sim (PAS) protein family (Figure 1)[1,4]. It plays a role in biological functions as diverse as environmental sensing, regulation of xenobiotic metabolism, vascular development, immune signaling, and cancer development[2,13–19]. In a well-accepted model of AHR signal transduction, the soluble receptor binds ligand that has diffused into the cytosol, the protein-ligand complex then translocates to the nucleus where AHR-chaperones dissociate (e.g., Hsp90 and AIP). The AHR is then free to dimerize with its partner, another PAS protein known as the AHR nuclear translocator (ARNT) positioning alpha helices to bind DNA. The PAS protein heterodimer then binds at "dioxin responsive enhancers" (DREs) and causes upregulation of batteries of genes which include the cytochromes P450 (CYP) 1A1, 1A2, 1B1, and the AHR repressor (AHRR) (Figure 2).

The AHRR was discovered in the 1990s and shares many sequence and structural similarities with the AHR (Figure 3). Both members of the PAS protein family, the AHR and AHRR share nearly identical amino acid sequences within their PASA domains, and both have additional regions of similarity within their amino terminal (N-term) nuclear localization signals (NLS) and their basic-helix-loop-helix (bHLH) DNA binding domains. Major differences exist primarily at their carboxy terminal (C-term) halves, where the AHR is known to have a functional transactivation domain (TAD) while the AHRR harbors what has been predicted to function as a transrepression domain (TRD). Another major between these two proteins in this half is that while the AHR has a PASB domain responsible for chaperone and ligand binding, the AHRR has only a small alpha-helix remnant of this region which does not appear to bind chaperones or

ligands. Importantly, both the AHR and AHRR are highly conserved across hundreds of species in the phylum Chordata.

One view of AHR signaling is that it serves primarily as a xenobiotic sensor and that its regulated transcriptional targets serve to clear ligands. Cytochrome P450s are widely recognized as a class of enzyme that plays important roles in the metabolism of many hydrophobic compounds, both endogenous (endobiotics) and exogenous (xenobiotics)[20,21]. The observation that the CYPs upregulated by AHR, i.e., CYP1A1, CYP1B1 and CYP1A2, have been shown to metabolize many of the ligands responsible for AHR pathway activation[22,23] has led to the idea that these enzymes also result in the pathway's returns to its basal state (Figure 2). The more recently discovered AHRR is commonly thought to be a mechanism of feedback repression for this pathway[24]. Once AHRR is upregulated in response to AHR activation, the AHRR competes for dimerization with ARNT and also binds dioxin response elements (DREs) upstream of AHR-regulated genes where it is thought to act as a repressor.

In our view, the evidence that the AHRR acts primarily as a repressor of the AHR pathway is weak. In part, our skepticism is related to a confusing array of data documenting repressor activity and describing its mechanism of action. It has been suggested that the AHRR's competition for ARNT and DRE binding is one mechanism by which the AHRR feedback represses AHR signaling. Additionally, the replacement of the AHR-ARNT activation complex on DREs with an AHRR-ARNT repression complex is thought to transrepress target gene expression, further inhibiting the pathway. While the action of the AHRR-ARNT-DRE interaction is a widely held model[24], some laboratories have reported that the AHRR does not

require its DNA binding properties to function as a repressor[25,26]. Thus a second model of repression has arisen, termed transrepression, and postulates that the AHRR downregulates DRE driven genes as part of a protein complex made up of AHR-ARNT dimers on DREs, with the AHRR interacting with this dimer in a DNA independent manner to repress (Figure 2)[25].

Interestingly, much of the published work describing AHRR mediated repression, suggests that such repression is weak. For example, mice with massive overexpression of AHRR through a transgene, reduced target gene expression is only reduced ~ 25-40% in most tissues[27]. Similarly, examination of the repression of CYPs in skin fibroblasts indicated that the major contribution to downregulated response was independent of AHRR[28]. Finally, the majority of cell culture experiments to characterize AHRR were done with transient plasmid expression where overexpression of AHRR can be orders of magnitude beyond physiological levels[24,29]. This observation leads to the possibility that the ectopic expression of repressor may have overrepresented its abilities due to abnormally high amounts of protein in the system.

Due to the uncertainty regarding AHRR's mechanism of repression, and questions related to this protein's potency as a repressor our aims in this chapter were two-fold. The first aim was to characterize the AHRR's capacity to downregulate AHR pathway signaling in a manner closer to endogenous signals like those outlined in Chapter 3. In part, this idea was driven by the idea that signaling may be repressed more efficiently by the AHRR than are responses to xenobiotics which might often act as superagonists. Our focus here was not to elucidate the mechanism of repression, rather to provide a satisfying answer to the question; "Can the AHRR repress AHR signaling to a state near basal levels when endogenous signaling pathways are initiated?" The

second aim was to explore the hypothesis that the AHRR functions as an auxiliary transcription factor that regulates a secondary battery of genes that are uniquely AHRR responsive. Our rationale was that if the AHRR were only a repressor for the pathway in question, then AHRR overexpression should only function at genes that AHR regulates. While this work was in progress, another laboratory provided some support for this idea, using ChiP assays to suggest the AHR and AHRR function at non-overlapping sets of genes[30]. Additionally, in some experiments high AHRR expression levels did not appear to inhibit CYP1A1 expression[28,31]. To state our hypothesis clearly, we predict that AHRR functions as a transcription factor at sites distinct from AHR-regulated genes, but it relies on the receptor for its initial expression.

A growing body of evidence supports AHRR as a tumor suppressor, immune system modulator, and biomarker for smoking, atherosclerosis, and cancer survival outcomes[5–7,32–39]. Several papers and reviews exist for more information on the general characteristics and known function of the AHRR[3,40–43]. Answers to the two goals of this chapter will further inform thinking about a protein whose function has long remained uncertain within biological contexts.

MATERIALS AND METHODS

Biochemicals: D-amino acid oxidase (DAO, cat. no. A5222) and D-tryptophan (cat. no. T9753) were purchased from Sigma Aldrich (St. Louis, MO). Lipofectamine 3000 (cat. no. L3000015) and 4',6-diamidino-2-phenylindole (DAPI, cat. no. D1306) were purchased from ThermoFisher (Waltham, MA). Dulbecco's Modified Eagle Medium (DMEM) cell culture media (cat. no. 11960044) was purchased from Gibco (Waltham, MA), along with fetal bovine serum (FBS), L-glutamine (cat. no. 25030081), and 0.25% trypsin-EDTA (cat. no. 25200056).

Cell culture and induction: The HCT 116 (cat. no. CCL-247) cells were obtained from the American Type Culture Collection (ATCC, Manassas, VA). Cells were cultured in DMEM with 10% FBS and 2mM L-glutamine (complete DMEM). Cells were kept at 5% CO_2 and 37ºC in a humidified incubator. For induction of the AHR pathway, cells were plated at an initial density of 25,000 cells per well in a 96-well plate or 500,000 cells per well in a 6-well plate and given 24 hours to adhere. Stock solutions of DAO and D-TRP were diluted to 2U/mL and 21uM, respectively, using complete DMEM. Then, cell culture media was removed from cells plated 24 hours previously, and the induction mixture was pipetted onto cells. Cells were given an additional 24 hours under treatment before sorting or RNA collection.

PCR primers, gRNAs, and plasmids: Primers for PCR screening and gRNAs were all from Integrated DNA Technologies (IDT, Coralville, IA) (Table1). Plasmids were designed in SnapGene software from Dotmatics (Boston, MA). The AHRR-mCherry_WT, AHRR-mCherry_AA, AHRR_WT, and AHRR_AA were synthesized by GenScript (Piscataway, NJ). GenScript also cloned the following into a P2A-EGFP vector from a previously synthesized

order: pJCP_AHR_H-A_AHRR_B-TAD_P2A_EGFP (Chimera-1-EGFP), pJCP_AHR_H-B_AHRR_TAD_P2A_EGFP (Chimera-2-EGFP), pJCP_AHRR_H-A_AHR_B-TAD_P2A_EGFP (Chimera-3-EGFP), pJCP_AHRR_H-B_AHR_TAD_P2A_EGFP (Chimera-4-EGFP). The P2A site is a viral ribosomal skip site that causes a self-cleaving peptide, allowing cleavage between the chimera constructs and the EGFP marker. The mCherry and AHRR-mCherry plasmids were synthesized through Twist Bioscience (South San Francisco, CA). The CMV control plasmid has no ORF in front of a CMV promoter; it was a construct from the Bradfield lab (pL1966). See appendix for plasmid maps.

Cell line generation, fluorescence assisted cell sorting (FACS), and NGS screening: Cells were edited using nucleofection via the Lonza (Basel, Switzerland) 4D Nucleofector system. To make the nucleofection mixture, gRNA, Cas9 enzyme, and electroporation enhancer were mixed to 20uM, 19.5uM, and 20uM, respectively. HCT 116 cells of a low passage number (less than 10 passages) were counted, washed, resuspended in Lonza SE nucleofection buffer (cat. no. V4XC-1032), and the nucleofection mixture was added. Cells were then nucleofected in the 4D system using the EN-113 pulse code. Nucleofected cells were then plated a 96-well plate for recovery and growth. Two days post-nucleofection, cells were collected from plates via trypsinization, washed with wash buffer (WB, 2% FBS in PBS), and resuspended in WB with 1ug/uL DAPI. Resuspended cells were taken to the UW Carbone Cancer Center Flow Cytometry Laboratory for live/dead sorting on the BDFACSAria Cell Sorter. Single cells were deposited in each well of a 96-well plate. Approximately three weeks after single cell sorting, cells were split between another 96-well plate and a PCR plate for lysis. The aliquots for PCR were boiled at 95ºC for 15 minutes to lyse cells. Crude cell lysate was then used as template for a PCR reaction over the

region containing the gRNA cut sites. Primers OL2240 and OL2241 were used with the Takara (San Jose, CA) PrimeSTAR GXL DNA Polymerase (cat. no. R050B) kit for PCR reactions (Table 1). The subsequent amplified samples were sent to the UWBC Next Gen Sequencing Core for NGS sequencing on the Illumina MiSeq 2x250 Nano flow cell. The software CRISPResso2 was used to analyze the NGS data[44]. Clones with two distinct edits, i.e. those with one mutation in one allele and a separate mutation in the other, were chosen for expansion. Expanded clones were used for subsequent experiments and also banked in liquid nitrogen.

Lipofection and cell sorting: Cells were plated in a 96-well plate at 25,000 cells per well or in a 6-well plate at 500,000 cells per well if destined for cell sorting. After 24 hours, lipofection reagents were mixed according to the Lipofectamine3000 protocol and added to wells. Lipofected cells were given 24 hours and induced, if applicable, then given another 24 hours before endpoint assay. If unsorted, RNA was obtained using the Promega ReliaPrep kit (see below) and used for RT-qPCR. For cell sorting, cells at 48 hours post-lipofection were collected from plates via trypsinization and washed with wash buffer (WB, 2% v/v FBS in PBS). Then, cells were resuspended in WB with 1ug/uL DAPI and filtered. Cells were taken to the UW Carbone Cancer Center Flow Cytometry Laboratory. The flow operator ran the cells through the BD FACSAria Cell Sorter and collected fluor-positive cells in a bulk population. Then, cells positive for fluorescent markers by sorting were made immediately into RNA for downstream protocols.

RNA preparation, RT-qPCR, and RNA-seq: The Promega (Madison, WI) ReliaPrep RNA Miniprep System for cells (cat. no. Z6012) was used for all RNA preparations. The RT-qPCR

was performed with the Promega GoTaq Probe 1-Step RT-qPCR System (cat. no. A6121). Predesigned primer and probe assays were purchased from Integrated DNA Technologies (IDT, Coralville, IA). The probe dye for CYP1A1 (Hs.PT.58.219047) and AHRR (Hs.PT.58.23243692) was FAM. The probe dye for HPRT1 (Hs.PT.58v.45621572) was HEX. The RT-qPCR reactions were analyzed on an Applied Biosystems (Bedford, MA) QuantStudio Flex 7 machine. Samples for RNA-seq were prepared as above and submitted to the University of Wisconsin Biotechnology Center (UWBC) Gene Expression Core (GEC) for quality control and library preparation. Prepared libraries were then submitted to the UWBC Next-Gen Sequencing Core and sequenced on the Illumina (San Diego, CA) NovaSeq6000 at a depth of 30 million reads per sample. Differential gene expression (DGE) analysis was done by the UWBC Bioinformatics Resource Core (BRC), as well as with Qlucore (New York, NY) Omics Explorer. Samples were normalized using the trimmed mean of M-values (TMM) method[45,46].

Data analysis: Results of RT-qPCR were analyzed using the ddCt method[47]. In situations where genes had high initial Cts (~35 or above), the percent increase over the minimum response was used to adjust for the low baseline of these genes. The calculation used was as follows: $100 - ((dCt/dCt_{max})*100)$ where dCt is the difference between the gene of interest (GOI) and the housekeeping gene (HK) and dCt_{max} is the maximum dCt for a given experimental control group. Statistical analyses and graphs were made using GraphPad Prism (San Diego, CA). Data was commonly analyzed using 2-way ANOVA and multiple unpaired t-tests or 1-way ANOVA and one-sample t-test with hypothetical mean set to the stated value. Sequence alignments were done in the SnapGene software described previously with transcript sequences from the Ensembl

database. Protein alignments were done in PyMol (Schrodinger, Mannheim, Germany) using the Protein Data Bank (PDB)[48,49].

RESULTS

Generation of a functional AHRR knockout cell line reveals low-efficiency repressive action of the protein in response to endogenous ligands. The AHRR is a PAS family protein (Figure 1) thought to function as a repressor for the AHR pathway (Figure 2). The idea that AHRR was a repressor arose from the observation of its striking similarities with the AHR in protein sequence and structure (Figure 3). To study the AHRR's repressive action on AHR signaling induced by an endogenous ligand, gRNAs were designed to cut within its exon 2, near the initiating methionine (Figure 4A). Two different guides were used and these produced clones T3.7.H8 and T2.3.G12. Mutations at the target site were confirmed via NGS (Figure 4B). Effects on the protein sequence were predicted using SnapGene software (Figure 4C). No downstream methionines were found C-terminal to the edited sequence reducing the possibility for a later initiation event and expression of a truncated protein.

The two "knockout clones" and a no-guide control (NGC) clone were treated with 2U/mL DAO and 21uM D-TRP to induce AHR pathway genes (Chapter 3). Twenty-four hours after initial treatment cells were collected and made into RNA using the Promega ReliaPrep Miniprep Kit. The RNA was then used in an RT-qPCR reaction, and fold-change between untreated and treated was calculated for each cell line. A common AHR pathway gene, CYP1A1, was used as a readout of the AHR pathway's activation. The housekeeping gene was HPRT1. Clone responses for CYP1A1 were significantly increased compared to the wild-type response (Figure 5). P-values were in the 10^{-3} to 10^{-6} range.

Transient overexpression of an AHRR-mCherry construct supports marginal repressive capabilities of AHRR. HCT 116 cells were lipofected with mCherry or cotransfected with AHRR and mCherry plasmid constructs 24 hours after cells were plated at 500,000 cells per well in a 6-well plate. Cells were induced 24 hours after lipofection with DAO plus D-TRP. Then, 24 hours after induction (and 48 hours post-lipofection) cells were collected for FACS, and mCherry-positive cells were sorted to be immediately made into RNA with the Promega ReliaPrep RNA mini-prep kit. The RNA was used for RT-qPCR with CYP1A1 as the gene of interest and HPRT1 as the housekeeper. The mCherry-only group was induced to roughly 25-fold. The AHRR plus mCherry cotransfection was induced 15-fold (Figure 6). This drop in induction due to addition of AHRR was statistically significant ($p = 0.007$ with Welch's t-test).

A DNA-binding domain (DBD) mutant of AHR provides the logic for an AHRR DBD mutant. All data is unpublished from Bernice Lin (thesis, UW-Madison 2005). To generate an AHR DNA DBD mutant, site directed mutagenesis was employed to insert alanine residues on the C-term of the DNA contacting "basic" alpha helix of the protein (Figure 7A). Nuclear localization studies were then undertaken to analyze the cellular localization of the mutant AHR and to demonstrate translocation to the nucleus. These experiments clearly show that insertion of two alanine residues (similar to the AHRR mutant described below) results in an AHR protein that localizes to the nucleus, and it does so regardless of ligand (Figure 7B). In fact, addition of a single alanine was able to cause this constitutive nuclear localization. Gel shift assays confirmed the absence of DNA binding in these mutants (Figure 7C). Given the similarity of the protein sequence between AHR and AHRR in this region, these data were used to support the generation of a similar DBD mutation in AHRR (below).

A DBD mutation in AHRR (unable to bind DNA) functions similarly to the wild-type AHRR as a repressor. Using the AHR strategy described above, an AHRR DBD mutant was created by inserting two alanine residues after R37 in the protein sequence (Figure 8A). This "AA mutant" was fused with mCherry and also created as a wild-type ("fluor-less") version. The AA-mutants were transiently overexpressed along with the non-mutated control versions. Plasmids without a fluor were cotransfected with the mCherry plasmid to allow sorting of transfected cells. Then, cells were induced and 24 hours later sorted for live, mCherry-positive cells. Immediately after sorting, RNA was made for RT-qPCR. Statistical significance was determined using a one-sample t-test with the hypothetical mean set to 1.

Both AHRR-mCherry fusions (i.e., the AA mutant and its AHRR wildtype) yield similar repression of CYP1A1 expression of about 60% of that observed with an mCherry alone control. No difference was observed when comparing the AA vs wild-type AHRR as fusions. For those constructs that were not fused to mCherry, we observed less repression (40-60%) but we did observe a trend towards the wild-type AHRR being slightly more potent as a repressor compared to the AA-mutant. This was not statistically significant (Figure 8B). Follow up experiments with these same samples in RNA-seq (see below) did not display a difference between the WT and AA constructs. In general, these data are consistent with the published literature[25].

The RNA from the mCherry control, AHRR-mCherry_WT, and AHRR-mCherry_AA from these experiments was also used for an RNA-seq experiment. Data from the UWBC BRC was analyzed using Qlucore software. The samples were normalized using the TMM method, and the

resulting values were used as x in 2^x to remove the log transformation. Then, these values for several AHR pathway genes were visualized in GraphPad Prizm (Figure 8C). Sixteen targets were identified by virtue of their repression by AHRR and AHRR_AA. Four housekeeping controls were also monitored, interestingly, two of these (AHR and ARNT) were significantly upregulated by the repressor (Figure 8D).

Chimeras of AHR and AHRR provide a tool to demonstrate DNA interactions in vivo. Alignments of the AHR and AHRR amino acid sequences and resultant structures were generated in Figure 3. The Ensembl transcripts for AHR and AHRR (ENST00000242057.9 and ENST00000684583.1, respectively) were translated in SnapGene to generate protein sequences which were then locally aligned using the Smith-Waterman algorithm (Figure 3A). Protein domains are indicated by boxes above and below the protein sequence. The AHR and AHRR bHLH and PASA domains from recent crystallographic analysis were aligned in PyMol (Figure 3B). These models were used to create AHR-AHRR chimeras to study repressor function. Chimeric junction sites were identified as residues conserved in both proteins, located between the PASA and PASB or PASB and TAD/TRD, and located in or close to regions of similarity between AHR and AHRR. The first junction site was a glutamate residue conserved in both proteins and located at the junction between the defined PASA and B domains (AHR aa279 and AHRR aa280). The second junction site was also a glutamate conserved in both proteins. This glutamate was after the AHR's defined PASB domain (AHR aa390 and AHRR aa374) and was chosen for its position at the terminal end of a region of similarity. Four chimeras were made using these sites (Figure 9). Chimera-1 is the bHLH and PASA of AHR with the PASB and TRD of AHRR; Chimera-2 is AHR from bHLH to PASB with the AHRR's TRD; Chimera-3 is

AHRR's bHLH and PASA with the AHR's PASB and TAD; Chimera-4 is AHRR from bHLH to PASB with the AHR's TAD (Figure 9A).

Rationale for chimera fusions: Chimera-1 was predicted to repress constitutively based upon the concept that it harbors the dimerization and DRE binding domains of the AHR, is missing the ligand/chaperone binding domains of the AHR and now harbors the TRD of AHRR. Chimera-2 was predicted to repress in a ligand-dependent manner as it harbors the AHRR TRD with the entire signaling domains of the AHR. Chimera-3 was predicted to be a ligand-dependent activator that would identify transcriptional AHR targets in addition to targets that were unique to the AHRR-ARNT complex. Chimera-4 was predicted to constitutively activate the AHR targets as well as targets unique to the AHRR-ARNT dimer (Figure 9B). See appendix 5 to 8 for plasmid maps.

A chimera of AHR and AHRR is constitutively active. Each chimera was constructed with a P2A cleavage site immediately N-terminal to an in-frame EGFP cassette so that EGFP could be used to monitor expression and be used as a fluor in cell sorting (Appendix 5 to 8). This cleavage also allowed the removal of the EGFP from the chimera so additional steric influences had minimal effect on chimera function. Cells were transiently transfected with each of the four Chimera-P2A-EGFP plasmids and allowed 48 hours for maximal plasmid expression. At day two after lipofection, cells were sorted for live, EGFP-positive cells. The collected cells were immediately subjected to RNA isolation which was later used for RT-qPCR measurements of CYP1A1 mRNA expression. In our initial expression study where we monitored expression in the absence of endogenous ligand activation, we observed that Chimera-1, Chimera-3-EGFP, and

Chimera-4-EGFP all had constitutive activity (Figure 10). Chimera-4-EGFP had the highest CYP1A1 expression, approximately 35-fold over the control cells, while Chimera-3 was the next most active followed by Chimera-1 (~25- and 8-fold, respectively). While difficult to visualize in a dot plot, Chimera-2-EGFP had statistically significant repressive capabilities as compared to the control. All statistics are from a one-sample t-test with hypothetical mean set to 1.

Chimera-4 upregulates multiple AHR pathway genes in an RNA-seq experiment. In an effort to identify the most comprehensive number of AHR and AHRR target genes, we next performed an RNA-seq experiment that compared cells transfected with Chimera-2 to those expressing Chimera-4. Chimera-4-P2A-EGFP and Chimera-2-P2A-EGFP were overexpressed in HCT 116 cells, and 48 hours post-lipofection EGFP-positive cells were sorted and immediately used to generate RNA. The RNA was sent to the UWBC for RNA-seq library preparation and sequencing. In our initial analysis, we performed differential gene expression between Chimera-2 and Chimera-4 using Qlucore software. Differentially expressed genes were visualized in a heatmap with the following statistical cutoffs: $q = 0.007$ and fold-change = 2.9 (Figure 11). Genes were ordered using hierarchical clustering. Four of the seven TRP Response Set (TRS) genes identified in Figure 8C were upregulated in this experiment, CYP1A1, ALDH1A3, TIPARP, and AHRR. In addition, AC015712.2 from the putative genes set was also upregulated. These genes clustered close together. Additionally, TLR9 clustered next to these AHR pathway genes and is of interest because of its potential role in IBD.

DISCUSSION

The AHRR functions as a repressor of the AHR pathway. The AHRR is a PAS family protein that is transcriptionally upregulated by the AHR-ARNT dimer's interaction at DREs (Figure 2). The AHRR is thought to act as a feedback inhibitor of the AHR pathway, with a leading hypothesis being that this repression is related to the high degree of structural similarity between the proteins. This structural similarity is thought to allow the AHRR to 1) reduce signaling through competition for ARNT and DRE interactions and 2) reduce signaling through active repression through the AHRR's TRD (Figure 3)[24,26,29]. A more recent idea suggests that the AHRR functions in a DRE-independent fashion where the AHRR acts more as a "corepressor" interacting with the AHR-ARNT dimer in a unique way, perhaps forming a trimer with the AHR-ARNT complex (Figure 1)[25].

The literature related to AHRR's molecular role is incomplete and leaves many important questions to be answered. Among these are whether the AHRR's repressor activity is of significant magnitude to be of biological importance, or even whether it is a repressor at all. In this regard, most reports of AHRR's repressor activity describe rather modest influences on AHR signal transduction. Moreover, most demonstrations of repression require transfection of AHRR in cell culture, or transgenesis in mouse models, where levels of the protein may reach super-physiological levels[24,25,27]. One final observation we make regarding prior research is that most studies employed the upregulation of signaling using potent xenobiotic ligands with little relevance to endogenous AHR signaling. For example, most studies employ chlorinated superagonists such as 2,3,7,9-tetrachlorodibenzo-p-dioxin (TCDD) that display a remarkably high affinity for the AHR and biological half-lives that are measured in months not minutes.

To examine AHRR's mechanism of action under conditions that are more reflective of endogenous signaling, we employed the I3P dependent activation mechanism explored in Chapter 3 and use gene editing as a novel method to eliminate expression of genomic AHRR. Two AHRR-knockout clones made using CRISPR/Cas9 were verified with PCR and NGS sequencing (Figure 4). Both clones increase CYP1A1 upregulation by DAO-DTRP to a greater degree than their wild-type counterpart (Figure 5). While the corresponding repression was modest (at most 3.5 times greater than the NGC controls) it was reproducible across both independent clones. In a parallel set of experiments, AHRR overexpression generated via a CMV-controlled system showed a roughly 50% decrease in induction compared to a "no-AHRR" control (Figure 6). While statistically significant, the pathway was never downregulated to a basal level or approached an uninduced state. These experiments lead us to suggest the AHRR is a weak repressor at classic AHR targets, such as CYP1A1, and suggest this protein may have a more biologically relevant function as a co-modulatory factor in the AHR pathway.

Alternate mechanisms of AHRR action: The weakness of AHRR's repressor is in keeping with our idea that the AHR may be regulating genes that may lie outside that set directly driven by AHR-ARNT dimers binding at DREs. Two possible mechanisms may explain such a path. First, the AHRR's "basic region" (within its bHLH) might recognize slightly different genomic sequences than the corresponding basic region of the AHR. This could lead to a unique set of DREs being influenced by AHRR-ARNT as opposed to AHR-ARNT. We do note that our structural analysis argues against such a model, but crystallographic structures can be oversimplifications due to the high levels of expression required (Figure 3). Second, the AHRR might interact with the AHR-ARNT complex as a "co-modulator," akin to the way many

coactivators interact with cognate DNA binding transcription factors. Such a model has some support from experiments indicating that AHRR may interact directly with the AHR in the AHR-ARNT complex[41]. While such a model is in conflict with many published models of PAS dimerization, we were driven to formally test the idea that AHRR may influence gene expression through mechanisms that do not require its binding to DNA[2].

To examine the importance of AHRR direct interactions with DNA (the second model above), we designed AHRR alanine mutants to block DRE contacts (Figure 7). Our approach was guided by experiments with the AHR, where insertion of alanines upstream of the DNA binding domain prevents DRE interactions (Figure 7A). These DRE binding domain mutants ("DBD") of AHR translocate to the nucleus but do not bind the DRE because the contacts within the alpha helix are pushed beyond the major groove (Figure 7B and C). The structural similarity between AHR and AHRR in this region gave us confidence to employ the mutations to create an AHRR-DBD mutant (Figure 8A). To this end, double alanine insertion mutants (AHRR_AA) and its wild-type counterparts (AHRR_WT) were constructed. An additional set of mutants were also fused to an in-frame mCherry cassette to allow monitoring of expression and simplify cell sorting protocols. When transfected into HTC-116 cells, both classes of mutants were able to repress CYP1A1 induction. While the mCherry fusions were most repressive and indicate that the AA insertions have little impact on repression, we did observe a small contribution of DNA binding to the repressor's activity for constructs without mCherry. Taken in sum, these results suggest the AHRR does not require DRE-binding for repression, but the ability to interact with DNA may contribute partially to repression. This supports some published literature and also suggests that

the AHRR may repress AHR signaling by multiple mechanisms, one of which does not require DRE binding[25].

The AHR and AHRR protein structure and domain conservation enables creation of informative chimeras. The AHR and AHRR share a protein domain structure conserved across the PAS protein family. These family members use common modes of heterodimerization between PAS partners to accomplish circadian, hypoxia and AHR signaling (Rojas et al., In Preparation). The PASA domains are characterized by a secondary protein structure of an α' helix, two β sheets, four α helices, and three additional β sheets that form a "cornucopia"-like tertiary structure (Figure 1). This domain typically supports PASA-PASA and PASB-PASB interactions and allows independence of unstructured C-terminal regions to "float" for interactions with the transcriptional machinery of the cell. The α-helix unique to the PASA domain, known as α', is of primary importance for dimerization between the PASA partners (Rojas in preparation). The PASB domains share the same potential for dimeric interactions, although we now know the mechanisms of these interactions are unique from those driving PASA-PASA (Figure 1).

The AHRR is a somewhat unique PAS family member in that this protein lacks a full PASB domain (Figure 1B and C). When the PASB of the AHRR is referred to in this document, it is used when discussing a single alpha helix in the same region of AHR and AHRR. Only the receptor has a true PASB domain of two beta sheets, four alpha helices and three beta sheets as described above for A domains. Our interpretation of this structural feature is that lack of a PASB domain which encodes chaperone and ligand binding domain (PASB) leads to the

AHRR's constitutive activity. We used this structural insight to swap portions of AHR with parts of AHRR to produce chimeras with predicted activities. We then expressed these chimeras via transient transfection to determine their function in the absence of biochemical induction. This initial experiment identified Chimera-4 as a constitutive CYP1A1 activator (Figure 10). This chimera's function, at or above the levels of other published constitutive systems, makes it a promising model for further constitutive AHR research[8–11].

Chimeras aid in identifying novel AHR/AHRR responsive genes. To more formally define the gene sets that arose from this and prior experiments, we employed the list comparison strategies outlined in Chapter 3 Figure 10. First, we developed a list of "AHR responsive genes for endogenous ligands derived from TRP" (TRP Response Set, or TRS) by comparing the overlap of targets identified in all three of our prior RNA-seq experiments that measured the transcriptome of I3P, DAO+D-TRP and also UV treated HCT-116 cells (Chapter 3 Figure 10). To add support for membership in the TRS list, we assumed that those genes should also be repressed by AHRR as well as the AHRR_AA mutant. In support of this, the first seven genes in Figure 8C (CYP1A1, AHRR, ALDH1A3, NQO1, TIPARP, PPARG and CDK6) define our initial set, and all indicate significant inhibition when AHRR-mCherry_AA is expressed. While the repression by the AHRR-mCherry construct was not significant at $p < 0.05$, its modest repression was reproduced at all targets. This RNA-seq set also revealed a second putative set of AHR/AHRR target genes that were significantly upregulated by DAO+D-TRP and repressed by the AHRR with the exact same structure activity as the TRS (i.e. repressed by AHRR with greater repression by AHRR_AA). We refer to this set of genes as the Putative Response Set (PRS). See Appendix 11 for formalized TRS and PRS lists. Interestingly, of all fourteen targets

identified in this experiment, GDF is the one that is upregulated in response to AHRR. This suggests that the AHR system may actually be activating a repressor of GDF15. Finally, a set of housekeeping genes was also analyzed, and this set was not significantly downregulated by the expression of AHRR or DAO-D-TRP (Figure 8D).

Chimera comparisons provide additional insight into AHR signaling. As predicted, in the absence of induction by endogenous ligand, Chimera-4 provided a potent constitutively active protein at the DRE regulated target CYP1A1, while Chimera-2 provided a constitutive repressor at this locus. Chimeras 1 and 3 were intermediate in activity and although interesting were not considered further in the current study. Based on the potencies and unique characteristics of Chimera-2 and Chimera-4, we performed differential gene expression on the global gene expression profiles induced by these chimeras in an effort to confirm the AHR-responsive genes described above and add to the list of DRE regulated genes in our HCT 116 model system. Our rationale was that a comparison of activated versus repressed genes would be an optimally sensitive approach to identify novel genes that are driven by DREs, as well as AHR/AHRR, in this system. This experiment also holds potential to identify targets that are AHRR responsive through some co-modifier mechanism.

The results from this RNA-seq experiment comparing gene expression between Chimera-4 and Chimera-2 are visualized in a heatmap (Figure 11). As a positive control, we note that four of the previously described TRS genes are observed in this list: CYP1A1, ALDH1A3, TIPARP, and AHRR. Using this as a criterion to identify additional novel AHR/AHRR regulated targets, we propose that genes TLR9, NXF3, TPT1P6, RN7SL589P, AC008065.1, and AC015712.2 form

PRS2 and represent previously unidentified physiological targets of this pathway (Appendix 12). Interestingly, AC015712.2 (ENSG00000259583) was also identified as a potential target in our initial PRS list from Figure 8C. A second interpretation of this experiment is that the AHRR's DRE binding domain can be confirmed to interact with the known genomic DREs in vivo, as Chimera-4 upregulates many of the known targets from the TRS list with a magnitude of response similar to that observed when AHR is activated by endogenous ligand.

Two genes from our PRS2 set deserve note. The first, AC015712.2, is predicted to be an anti-sense RNA for ALDH1A3 (ENSG00000259583). This may represent additional feedback inhibition on this pathway and adds further support for the importance of this dehydrogenase in AHR signaling. The second gene of interest that caught our attention is toll-like receptor 9 (TLR9). The TLR9 gene clustered closely to the TRS genes in the chimera response heatmap, suggesting it is of biological significance. There is some evidence that toll-like receptors can upregulate the expression of AHR, providing an enhanced pathway response due to more receptor flooding the system[50]. Also, TLRs are involved in the innate immune system, which is of particular relevance to our intestinal cell culture model and IBD[23,51–53]. The regulation of TLR9 by the AHRR may be one way the repressor modulates AHR inflammatory signaling. For these reasons, further investigation into TLR9 and AC015712.2 and its connection to the AHR pathway is warranted.

The AHRR has a complex function in the AHR pathway. Our experiments suggest the AHRR can repress the AHR pathway in a DNA-independent fashion, and possibly a DNA-dependent fashion. This evidence suggests a dualistic model of AHRR repression. One mechanism of

repression follows “canonical” thinking and holds that the AHRR represses AHR pathway genes via binding DNA. The other mechanism of repression has been termed “transrepression” by Evans et al.; this hypothesis states that AHRR repression is mediated by DNA-independent, protein-protein interactions[25]. Which mechanism predominates may be dependent upon the gene being regulated and whether the AHR-ARNT heterodimer is already occupying a certain DRE. However, in all the experiments we conducted, the AHRR modestly repressed AHR pathway induction, and a return to basal levels was never seen. Subsequent experiments done to test whether AHRR may function as a transcription factor confirms AHRR binds AHR-controlled genes as well as putative AHRR-only genes. An important result of these experiments was the generation of a constitutively active chimera that performs at or above published levels of similar constructs[8–10]. Taken together, our experiments suggest AHRR may have a larger role in AHR pathway signaling as a later wave of transcriptional control. Additionally, TLR9 and AC015712.2 were identified as genes involved in AHR physiology with possible ties to the regulation of AHR levels and subsequent downstream effects on the pathway. This would fit into the role of the AHR pathway in the innate immune system, particularly within intestinal cells were AHR is involved in IBD.

TABLES

Table 1. Sequences of PCR primers and gRNAs.

Name	Sequence	Direction
OL2240	acactctttccctacacgacgctcttccgatct NNNNNN ACGAGGAGGAGCAGGAGGTG	5’ > 3’
OL2241	gtgactggagttcagacgtgtgctcttccgatct GGGTGCAACAACCCACACAAG	3’> 5’
AHRR.ex2.target.3 (T3)	CGAGGACGATGATCCCGCCGGGG	5’ > 3’
AHRR.ex2.target.2 (T2)	CCGGGGGAGTGCACGTACGCGGG	5’ > 3’

FIGURES

Figure 1) Insights into PAS domain structures and interactions from crystal structures shed light on AHR pathway biology. A) The actual protein secondary structure of a PASA domain can be portrayed in a simplified but accurate manner. The N-term of the PASA domain is an alpha helix, termed α', which interacts with the α' of another protein's PASA to form a dimer. Downstream of this helix is the highly conserved PAS domain structure. This consists of two β sheets, three α helices, and three β sheets; all these secondary structures together form a cornucopia-shaped domain that is hollow on the inside. The organization of sheets and helices form a distinct sidedness to the domain since helices form one portion and sheets the other. PASB domains are identical in structure but lack the α' dimerization helix. **B) PASA and PASB domains interact in distinct but different manners.** The two PAS domains interacting, whether A or B, are termed the α partner (e.g. AHR, AHRR) and the β partner (e.g. ARNT). This is not to be confused with the α or β of a protein structure, which is a separately defined nomenclature. On the left, two PASA domains dimerize utilizing their α' helices. On the right, two PASB domains interact. The α partner is tucked beneath the β partner with its α helix backbone against the β partner's β sheet side. **C) A cartoon model of AHR-ARNT and AHRR-ARNT dimerization influences thinking on protein function.** Unpictured for both AHR and AHRR are their TAD and TRD, respectively. On the left, AHR and ARNT dimerize to form a heterodimer capable of binding DNA. The PASB of ARNT keeps the PASB of AHR occupying a small space between the dimerized PASA domains beneath them. On the right, AHRR and ARNT also dimerize into a complex able to bind DNA. The AHRR lacks a true PASB domain, so the PASB of ARNT is left without a partner.

Figure 1A)

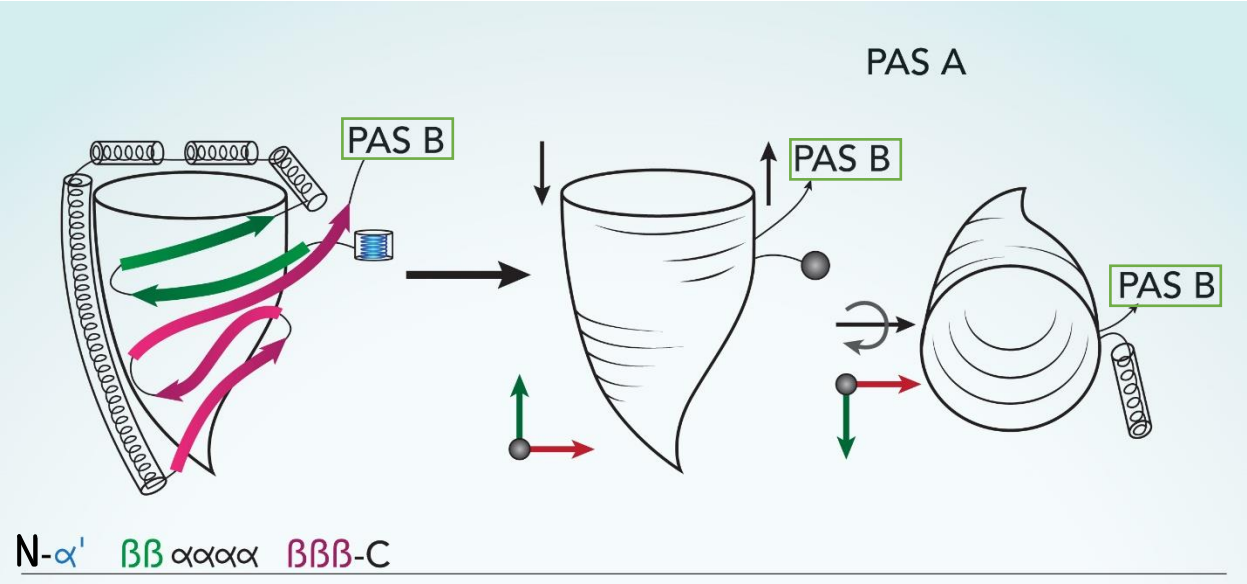


Figure 1B)

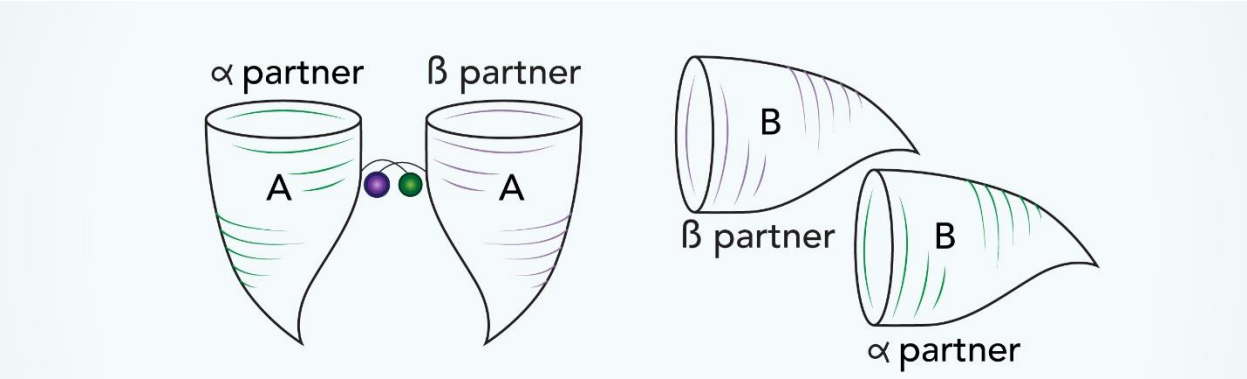


Figure 1C)

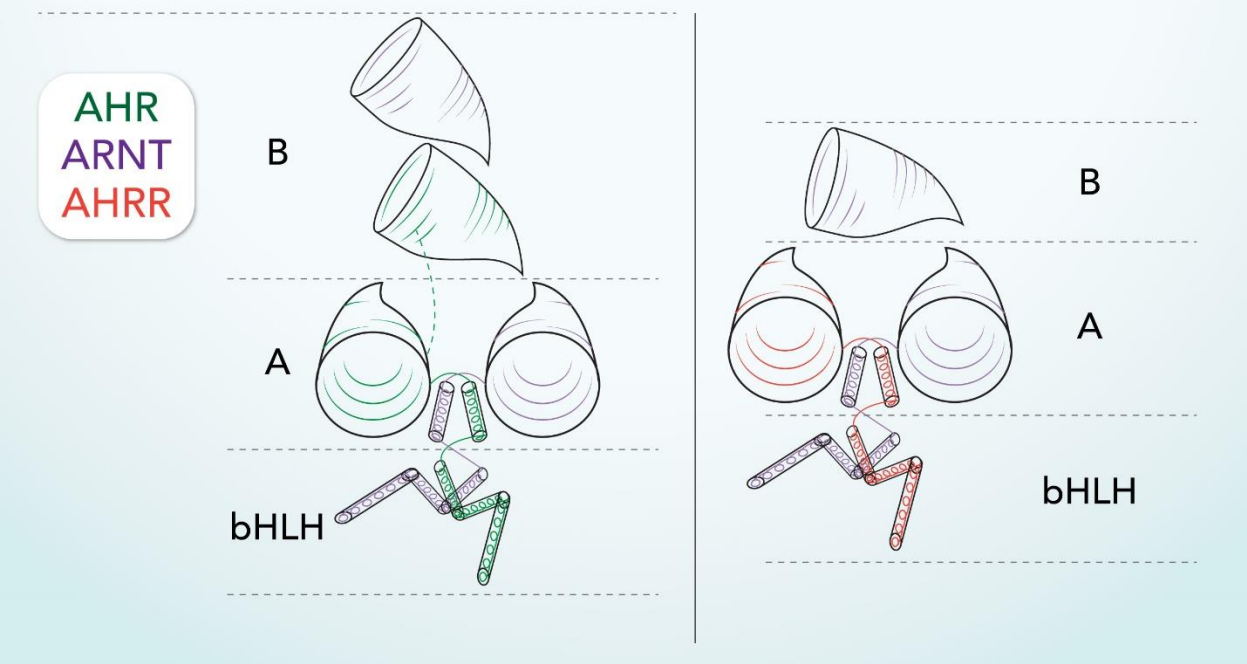


Figure 2) The AHR signaling pathway with an update to PAS protein modeling. The basic helix-loop-helix (bHLH), PASA, and PASB protein domains have been modeled to reflect crystal structure insights. The AHRR lacks a PASB domain. **1)** A hydrophobic, planar aromatic ligand diffuses across the cell membrane into the cytoplasm. **2)** Ligand binds in the AHR PASB domain, causing a slight conformational shift and revealing a nuclear localization sequence (NLS). **3)** The AHR-chaperone-ligand complex translocates to the nucleus. **4)** Chaperone proteins (Hsp90, ARA9, and p23) release the AHR and free up dimerization domains. **5)** The AHR dimerizes with its partner, ARNT. **6)** The AHR-ARNT heterodimer binds DNA at a dioxin response element (DRE). **7)** Transcription of downstream genes is affected; classically, the cytochromes P450 (CYPs) CYP1A1, CYP1A2, and CYP1B1, as well as the AHR repressor (AHRR), are all induced. **8A)** The AHRR translocates to the nucleus in a ligand-independent manner and dimerizes with ARNT, the same partner as AHR. **8B)** The CYPs bind the AHR ligand. **9A)** The AHRR-ARNT heterodimer represses AHR signaling at DRE-driven genes. **9B)** The CYPs metabolize the ligand, thereby removing the signal inducing AHR pathway expression.

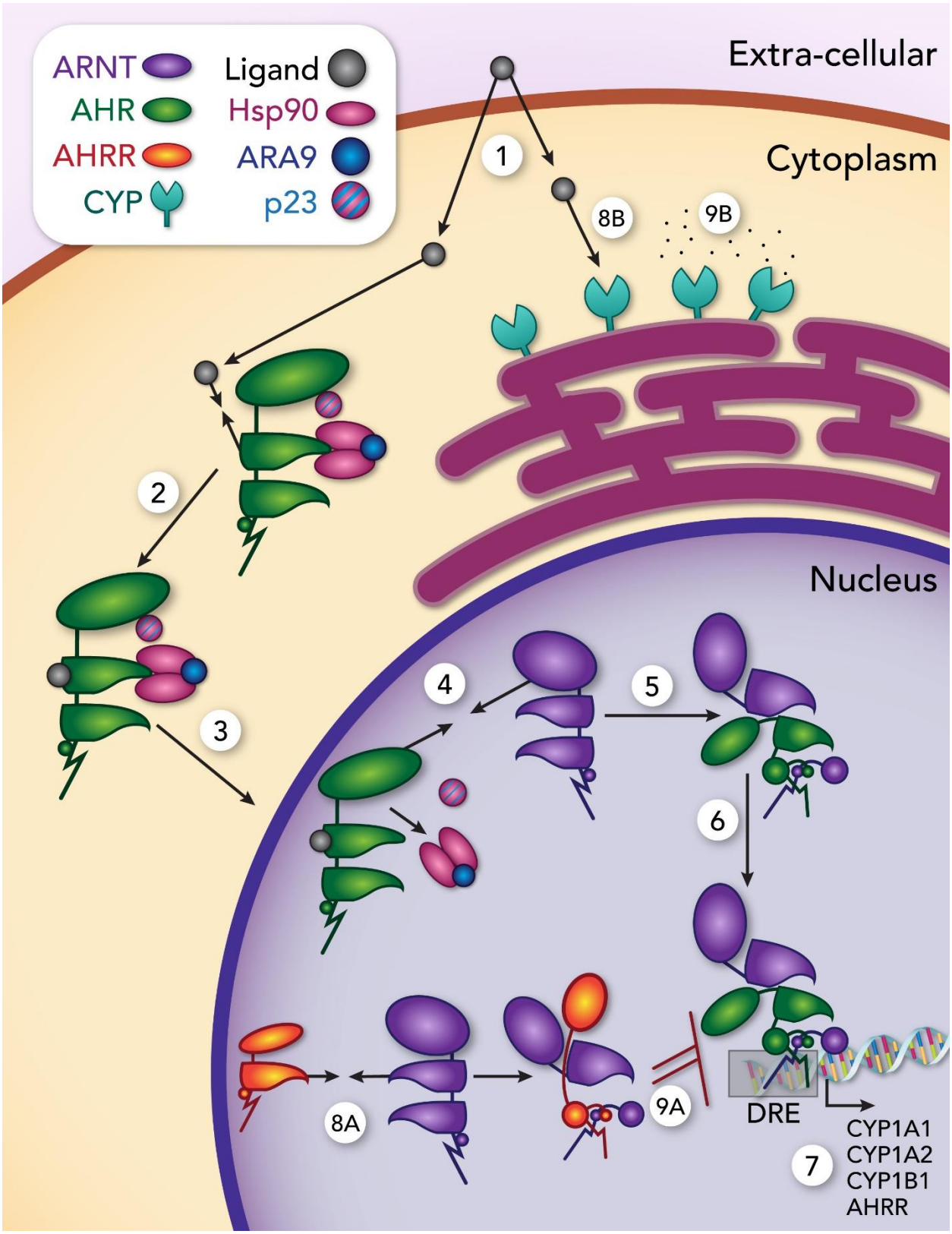

ARNT
AHR
AHRR
CYP
Ligand
Hsp90
ARA9
p23
Extra-cellular
Cytoplasm
Nucleus
1
2
3
4
5
6
7
8A
8B
9A
9B
DRE
CYP1A1
CYP1A2
CYP1B1
AHRR

Figure 3) The AHR and AHRR are similar in protein primary sequence and structure. A) Primary protein sequence alignment. Top line of text is AHR protein sequence. Bottom line of text is AHRR protein sequence. Middle line of text is shared residues between proteins; a + denotes similarity but not an exact match. Corresponding protein domain is as-indicated above or below AHR or AHRR protein sequence, respectively. **B) PyMol structures and alignment.** The AHR-ARNT heterodimer (PDB 5V0L) bHLH and PASA interacting with DNA is modeled in the first panel; the basic DNA-binding domain (DBD) is highlighted in burgundy[48]. The AHRR-ARNT heterodimer (PDB 5Y7Y) is modeled in the second panel; the DBD is highlighted in scarlet[49]. The third panel depicts an alignment of the bHLH and PASA domains of AHR and AHRR when each is dimerized with ARNT.

Figure 3A)


AHR
basic region
HLH
AHRR
basic region
HLH
PAS A
PAS A
PAS A
PAS A
PAS A
PAS B
PAS A
PAS B?

Figure 3B)

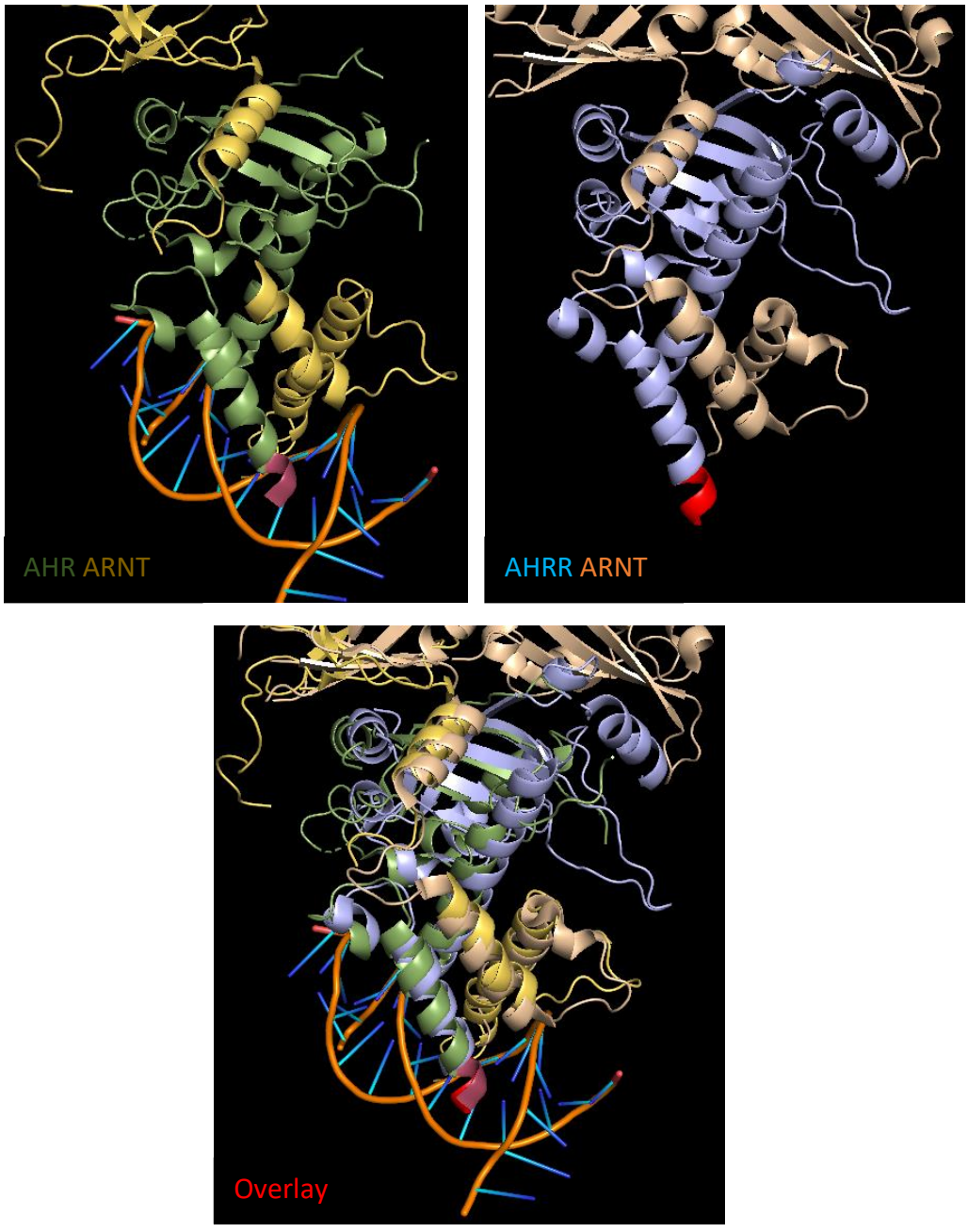


Figure 4) Design of AHRR CRISPR knockout cell lines and AHRR alanine mutants. A) Map of the AHRR gene and its corresponding protein product. The protein sequence is above, with untranslated regions (UTRs) denoted as horizontal lines; exons corresponding to protein section are denoted by numbers. The bHLH and PASA are as indicated. The DNA sequence is below; exons are boxes while introns are horizontal lines. The5' and 3' UTRs are denoted in red. Dashed lines denote which exon sections map to the protein regions. Stars mark DREs. Box marks exons of interest. **B) Mutations to AHRR exons 2 and 3 with their effects on DNA sequence.** The protein sequence of exons 2 and 3 is the first row in black; basic regions that are important in DNA binding are denoted in cyan italics. Below the protein sequence is the DNA sequence of the exons. In bold are the PAM sequences while sequences of the basic regions are in italics. The two gRNAs used to create AHRR knockout clones are in navy and orange; cut site is marked by ^ and direction of gRNA is denoted by >. Below each gRNA is the clone it created with the NGS sequence of both alleles; an insertion is marked by a red letter while a deletion is a -. Downstream of the gRNAs is the double alanine insertion to make a mutant unable to bind DNA. **C) Effects of mutations on protein structure.** Protein domains are indicated above wild-type sequence (WT); basic domains are in cyan. Two and three above the sequence marks the place where exon 2 turns into exon 3. Sequences in red are different from WT. Asterisk denotes stop codon while > indicates further protein sequence.

Figure 4)

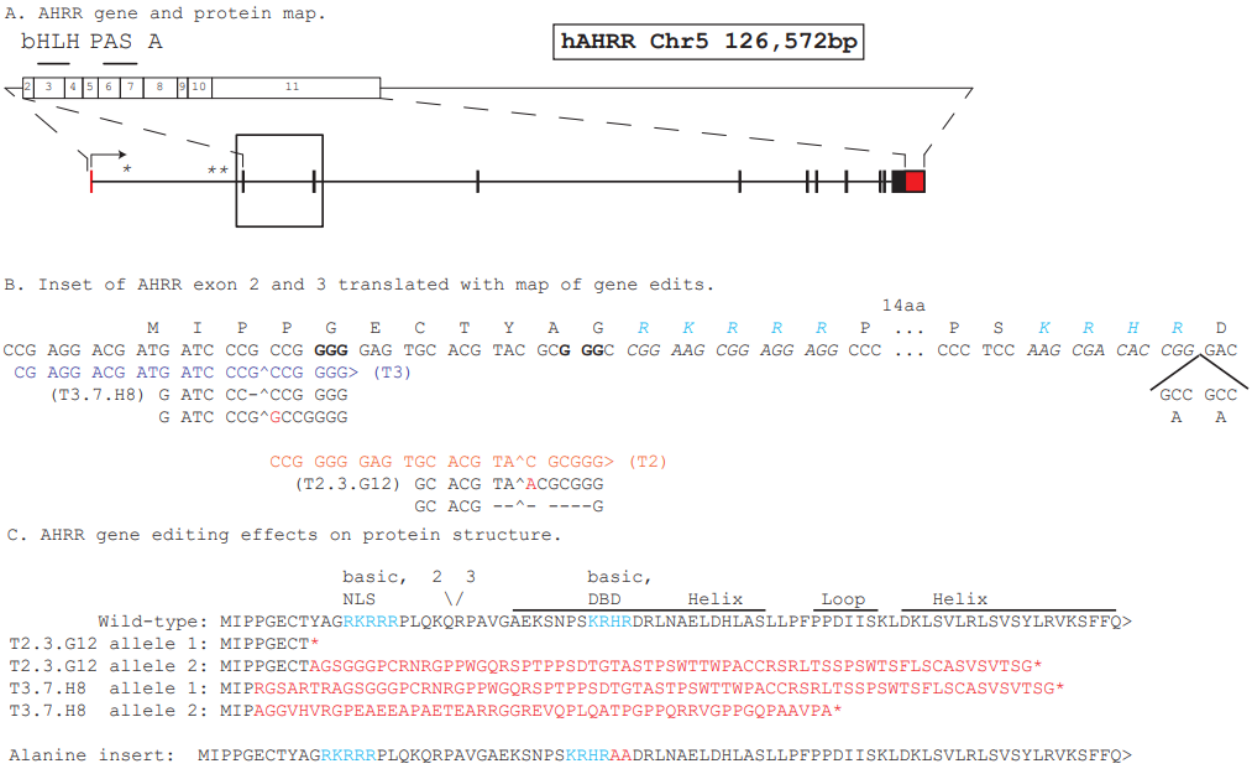

A. AHRR gene and protein map.
bHLH PAS A
hAHRR Chr5 126,572bp
2 3 4 5 6 7 8 9 10 11
*
**
B. Inset of AHRR exon 2 and 3 translated with map of gene edits.
14aa
M I P P G E C T Y A G R K R R R P ... P S K R H R D
CCG AGG ACG ATG ATC CCG CCG GGG GAG TGC ACG TAC GCG GGC CGG AAG CGG AGG AGG CCC ... CCC TCC AAG CGA CAC CGG GAC
CG AGG ACG ATG ATC CCG^CCG GGG> (T3)
(T3.7.H8) G ATC CC-^CCG GGG
G ATC CCG^GCCGGGG
GCC GCC
A A
CCG GGG GAG TGC ACG TA^C GCGGG> (T2)
(T2.3.G12) GC ACG TA^ACGCGGG
GC ACG --^- ----G
C. AHRR gene editing effects on protein structure.
basic, 2 3 basic,
NLS \/ DBD Helix Loop Helix
Wild-type: MIPPGECTYAGRKRRRPLQKQRPAVGAEKSNPSKRHRDRLNAELDHLASLLPFPPDIISKLDKLSVLRLSVSYLRVKSFFQ>
T2.3.G12 allele 1: MIPPGECT*
T2.3.G12 allele 2: MIPPGECTAGSGGGPCRNRGPPWGQRSPTPPSDTGTASTPSWTTWPACCRSRLTSSPSWTSFLSCASVSVTSG*
T3.7.H8 allele 1: MIPRGSARTRAGSGGGPCRNRGPPWGQRSPTPPSDTGTASTPSWTTWPACCRSRLTSSPSWTSFLSCASVSVTSG*
T3.7.H8 allele 2: MIPAGGVHVRGPEAEEAPAETEARRGGREVQPLQATPGPPQRRVGPPGQPAAVPA*
Alanine insert: MIPPGECTYAGRKRRRPLQKQRPAVGAEKSNPSKRHRAADRLNAELDHLASLLPFPPDIISKLDKLSVLRLSVSYLRVKSFFQ>

Figure 5) AHRR knockout clones are hyper-inducible compared to wild-type HCT 116 cells. The wild-type (WT) and knockout (KO) clones were induced with DAO+D-TRP treatment for 24 hours before being made into RNA and used for RT-qPCR. Both KO clones had CYP1A1 induction increased compared to the WT induced group. A one-sample t-test was used to compare DAO+D-TRP treated groups of the clones. Hypothetical mean was set to the mean of NGC.1.F2 (17.78). Numbers indicate corrected p-values.

Figure 5)

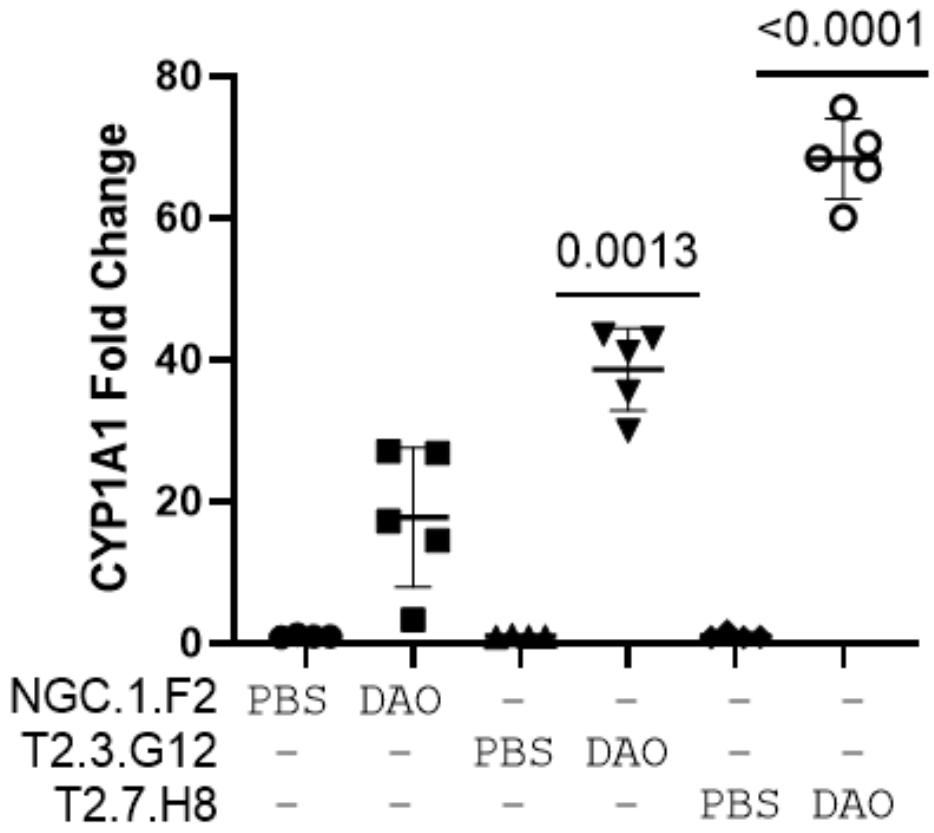

CYP1A1 Fold Change
80
60
40
20
0
<0.0001
0.0013
NGC.1.F2 PBS DAO – – – –
T2.3.G12 – – PBS DAO – –
T2.7.H8 – – – – PBS DAO

Figure 6) Overexpression of AHRR via transient transfection causes modest repression in induced cells. Plasmids were transfected into cells 48 hours before cell sorting for mCherry-positive cells. At 24 hours before sorting, cells were induced with DAO+D-TRP. Stars indicate statistical significance (p-value = 0.007) with Welch's t-test comparing AHRR to mCherry overexpression.

Figure 6)

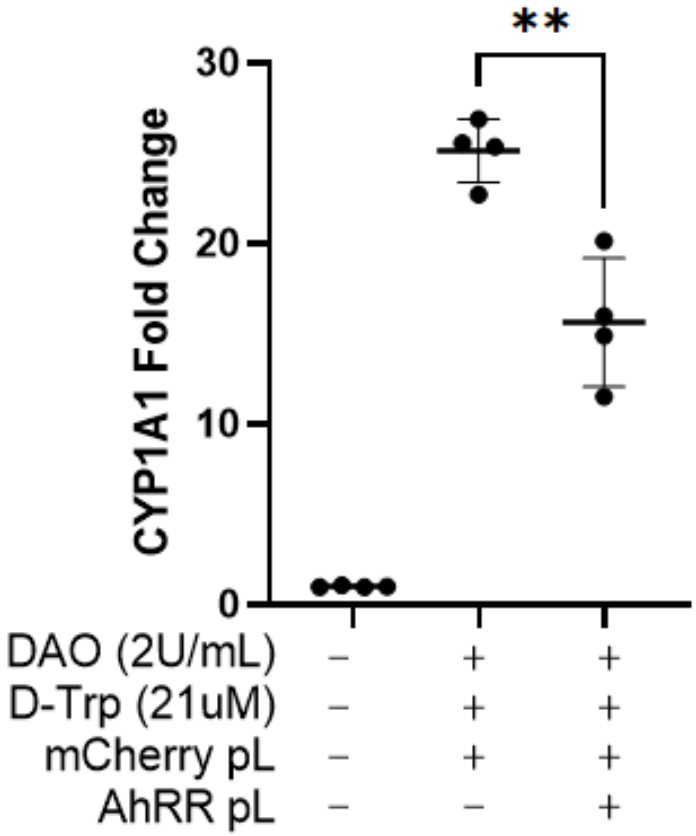

**
30
20
10
0
CYP1A1 Fold Change
DAO (2U/mL) − + +
D-Trp (21uM) − + +
mCherry pL − + +
AhRR pL − − +

Figure 7) Design and function of an AHR DNA binding domain (DBD) mutant informs a strategy to create an AHRR DBD mutant. Unpublished data from Bernice Lin. **A) Design of an AHR DBD mutant.** Binding and protein domains are as indicated. Boxes indicate exons; nuclear localization signal (NLS) and nuclear export signal (NES) are underlined in expanded sequence. Alanines were inserted directly after the second basic region with a sequence of KRHR. **B) Nuclear localization imaging of AHR WT versus DBD mutant.** Wild-type AHR remains cytosolic until BNF treatment causes translocation to the nucleus. The AHR mutant with two alanines inserted is constitutively nuclear. **C) Results of AHR WT versus DBD mutant gel shift assay.** Insertion of a single alanine is sufficient to prevent DNA binding, and insertion of a second alanine has the same result.

Figure 7A)

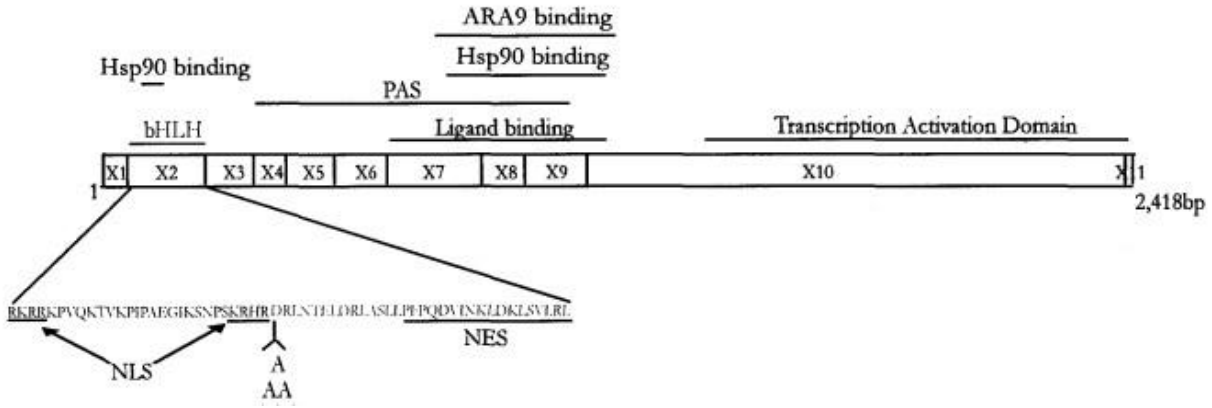

ARA9 binding
Hsp90 binding
Hsp90 binding
PAS
bHLH
Ligand binding
Transcription Activation Domain
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X
1
1
2,418bp
NLS
A
AA
NES

Figure 7B)

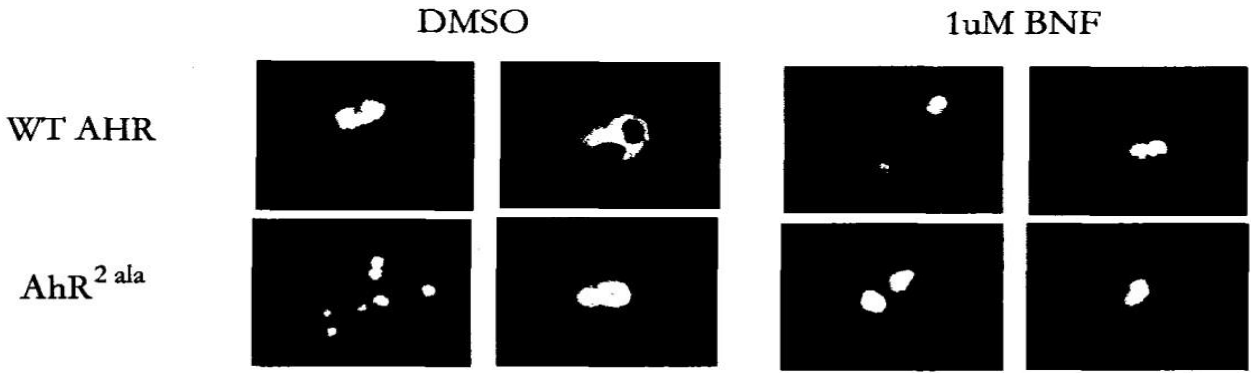


Figure 7C)

ALA Mutant	+ BNF	+ ARNT	DRE Binding
WT AHR	-	-	NO
WT AHR	+	-	NO
WT AHR	-	+	NO
WT AHR	+	+	YES
1 ALA	+	+	NO
2 ALA	+	+	NO

Figure 8) A mutated AHRR-mCherry construct without capability to bind DNA functions similarly to the wild-type fusion. A) Design of the AHRR alanine mutant. For broader genomic context see Figure 4. Two alanine residues were inserted after the second basic region (KRHR), which is conserved between AHR and AHRR. **B) RT-qPCR data of wild-type and mutant overexpressions.** Plasmids were transfected into HCT 116 cells 48 hours before cell sorting, and cells were treated with DAO+D-TRP 24 hours before sorting. RNA was immediately made from sorted mCherry-positive cells. The lanes are as follows: 1 is the non-AHRR control; 2 is AHRR-mCherry_WT; 3 is AHRR-mCherry_AA; 4 is AHRR_WT, and 5 is AHRR_AA. Overexpression of any AHRR construct caused modest repression. Wild-type constructs repressed slightly better than the alanine mutants, which is consistent with published literature[25]. Stars indicate statistical significance according to a one-sample t-test with hypothetical mean set to 1. **C) Comparison of TMM-normalized RNA-seq values for WT versus AA mutants across multiple AHR pathway genes.** Cells were induced with DAO+D-Trp and transiently transfected with one of mCherry, AHRR-mCherry_WT, or AHRR-mCherry_AA plasmids. After sorting for mCherry-positive cells 48 hours after initial transfection and 24 hours after DAO treatment, RNA was prepared and sent for RNA-seq. TMM-normalized values were generated using Qlucore. These values were then used as x in 2^x to remove the log scaling. The first seven genes are known to be in the AHR pathway. Multiple unpaired Welch's t-tests between the WT and AA groups reveals there are no statistically significant differences. **D) No significant downregulation occurs in housekeeping genes when TMM-normalized RNA-seq values are compared.** The same analysis as in C was run on four housekeeping genes. No significant downregulation occurred in any of the four genes. Values for GAPDH were divided by ten for scaling convenience. The values for AHR and ARNT were

significantly upregulated. Statistical significance is from multiple unpaired Welch's t-tests between the Ctrl and AA groups; no difference was found between the WT and AA groups.

Figure 8A)

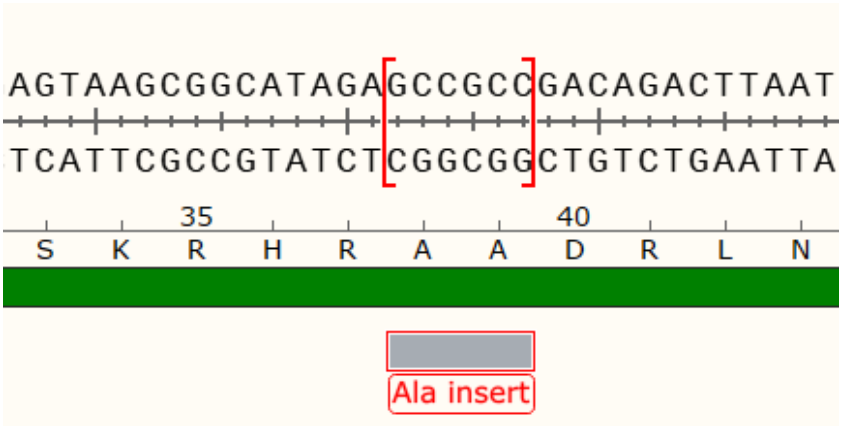

AGTAAGCGGCATAGAGCCGCCGACAGACTTAAT
TCATTCGCCGTATCTCGGCGGCTGTCTGAATTA
35
40
S K R H R A A D R L N
Ala insert

Figure 8B)

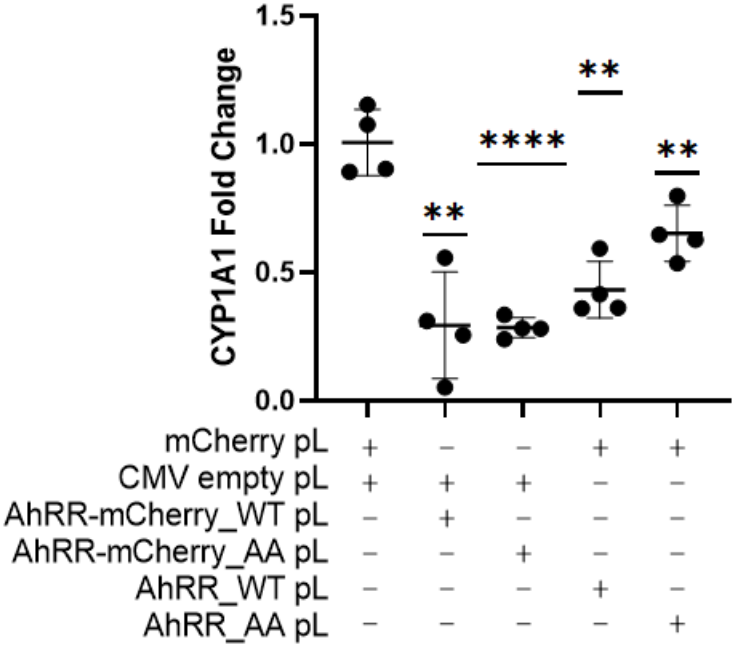

CYP1A1 Fold Change
1.5
1.0
0.5
0.0
**
****
**
**
mCherry pL + − − + +
CMV empty pL + + + − −
AhRR-mCherry_WT pL − + − − −
AhRR-mCherry_AA pL − − + − −
AhRR_WT pL − − − + −
AhRR_AA pL − − − − +

Figure 8C)

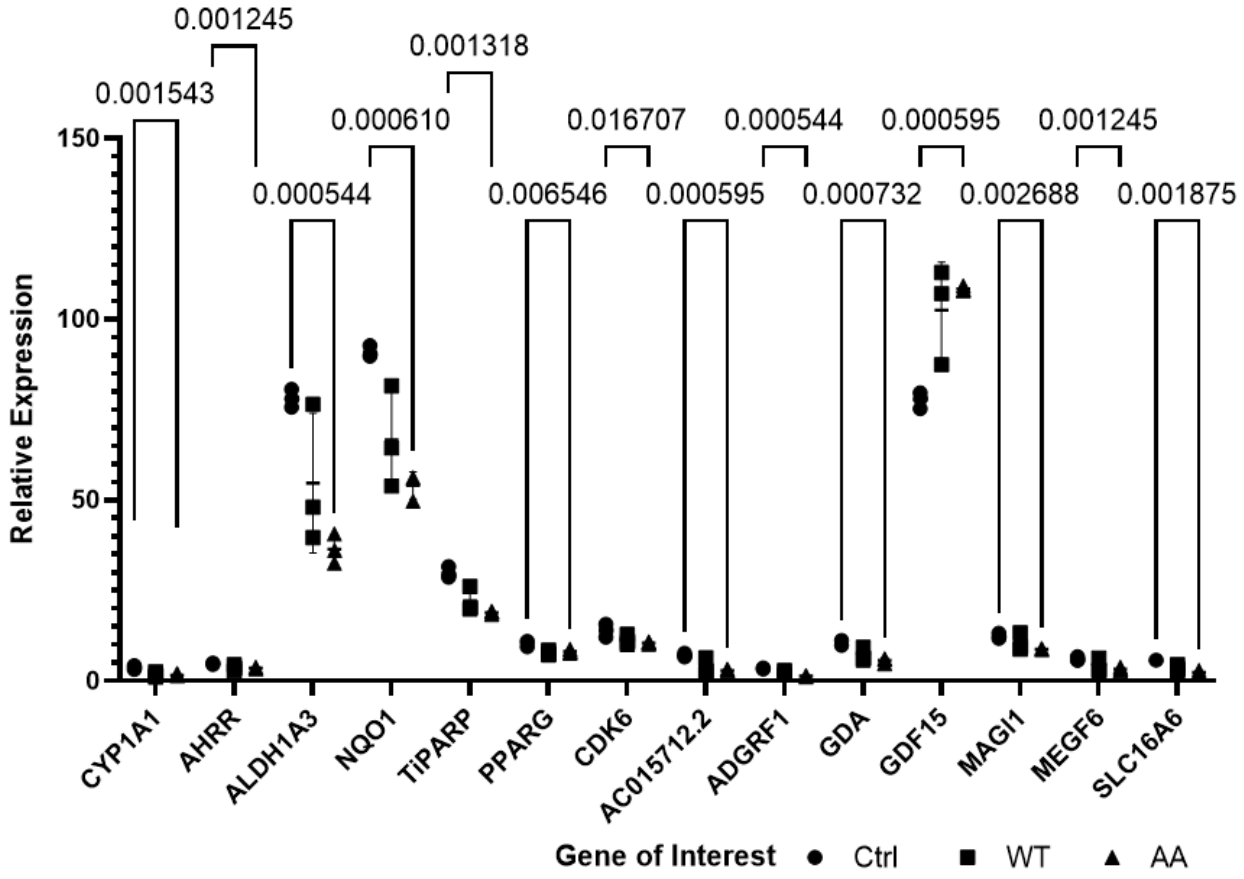

0.001245
0.001543
0.001318
0.000610
0.016707
0.000544
0.000595
0.001245
0.000544
0.006546
0.000595
0.000732
0.002688
0.001875
150
100
50
0
Relative Expression
CYP1A1
AHRR
ALDH1A3
NQO1
TiPARP
PPARG
CDK6
AC015712.2
ADGRF1
GDA
GDF15
MAGI1
MEGF6
SLC16A6
Gene of Interest
Ctrl
WT
AA

Figure 8D)

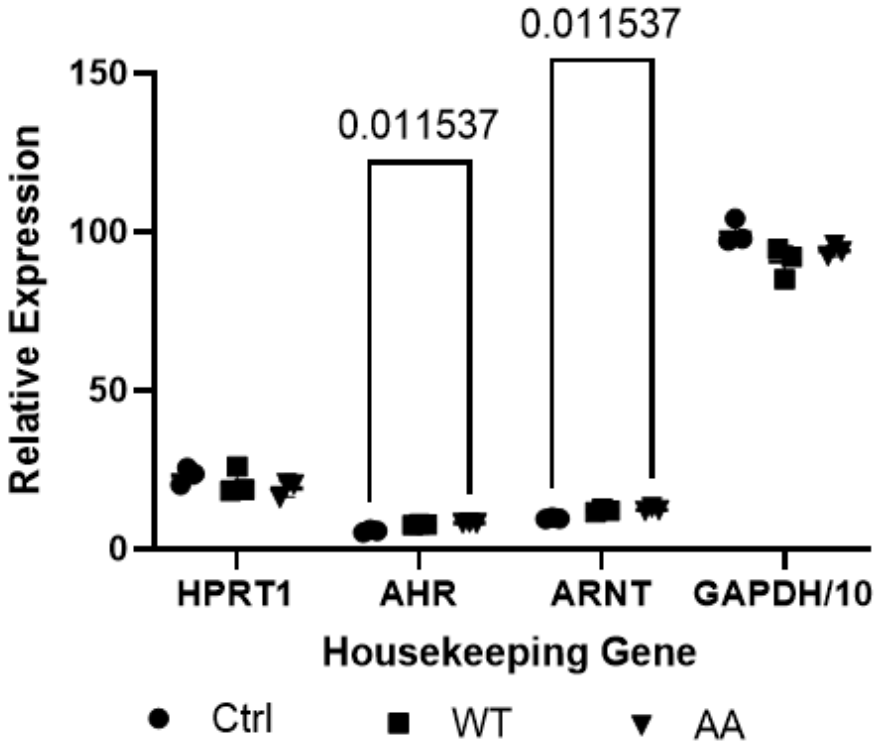

0.011537
0.011537
150
100
50
0
Relative Expression
HPRT1
AHR
ARNT
GAPDH/10
Housekeeping Gene
Ctrl
WT
AA

Figure 9) AHR and AHRR chimeras as a way to study AHRR function. A) Design of AHR and AHRR chimera constructs. Domains of AHR are blue while AHRR domains are orange. The PASB of AHRR is not a true PASB domain and comprises only the sequence of a single alpha helix; PASB is used as a convenient naming convention to refer to the same portion of protein between AHR and AHRR. Wild-type AHR and AHRR sequences are uppermost. Chimeras are as follows: Chimera-1 – AHR bHLH and PASA, AHRR PASB and TRD; Chimera-2 – AHR bHLH to PASB, AHRR TRD; Chimera-3 – AHRR bHLH and PASA, AHR PASB and TAD; Chimera-4 – AHRR bHLH to PASB, AHR TAD. **B) Predictions of chimera function.** The above description of whether a chimera domain is from AHR or AHRR is summarized here. Additionally, predictions of chimera function are in the last column. Predictions are based on the ligand-binding, activating function of the AHR PASB domain and its absence in AHRR.

Figure 9A)

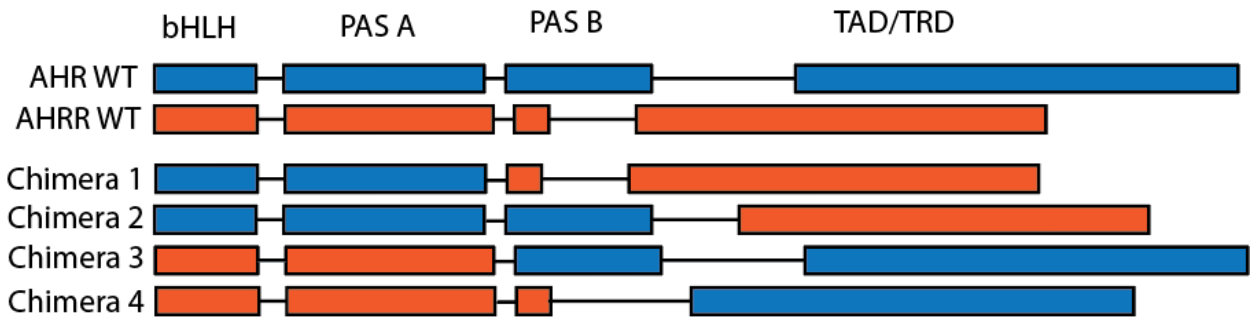

bHLH
PAS A
PAS B
TAD/TRD
AHR WT
AHRR WT
Chimera 1
Chimera 2
Chimera 3
Chimera 4

Figure 9B)

Chimera	bHLH	PASA	PASB	TAD/TRD	Function Prediction
1	AHR	AHR	AHRR	AHRR	Constitutive repression
2	AHR	AHR	AHR	AHRR	Ligand-dependent repression
3	AHRR	AHRR	AHR	AHR	Ligand-dependent activation
4	AHRR	AHRR	AHRR	AHR	Constitutive activation

Figure 10) Chimera overexpressions identify a constitutively active form of AHR and AHRR. Chimera-P2A-EGFP constructs were overexpressed via transient transfection and given two days before single-cell sorting. EGFP-positive cells were immediately made into RNA post-sorting. Then, RNA was used for RT-qPCR for CYP1A1 expression. Chimera-4 had the highest induction at 35-fold. Chimera-3 also had some constitutive activity, but this was to a lesser degree than Chimera-4. Chimera-2 significantly repressed in this experiment. Cells were never induced. Asterisks denote statistical significance according to a one-sample t-test with hypothetical mean set to 1.

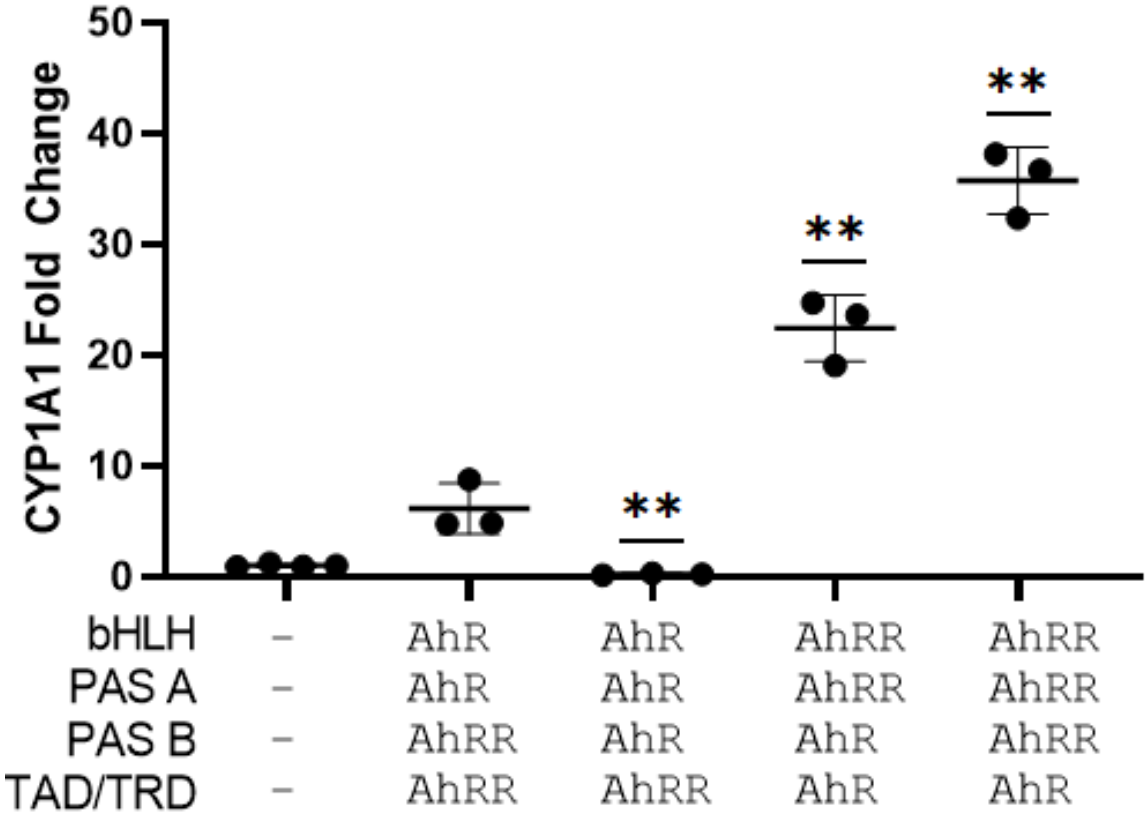
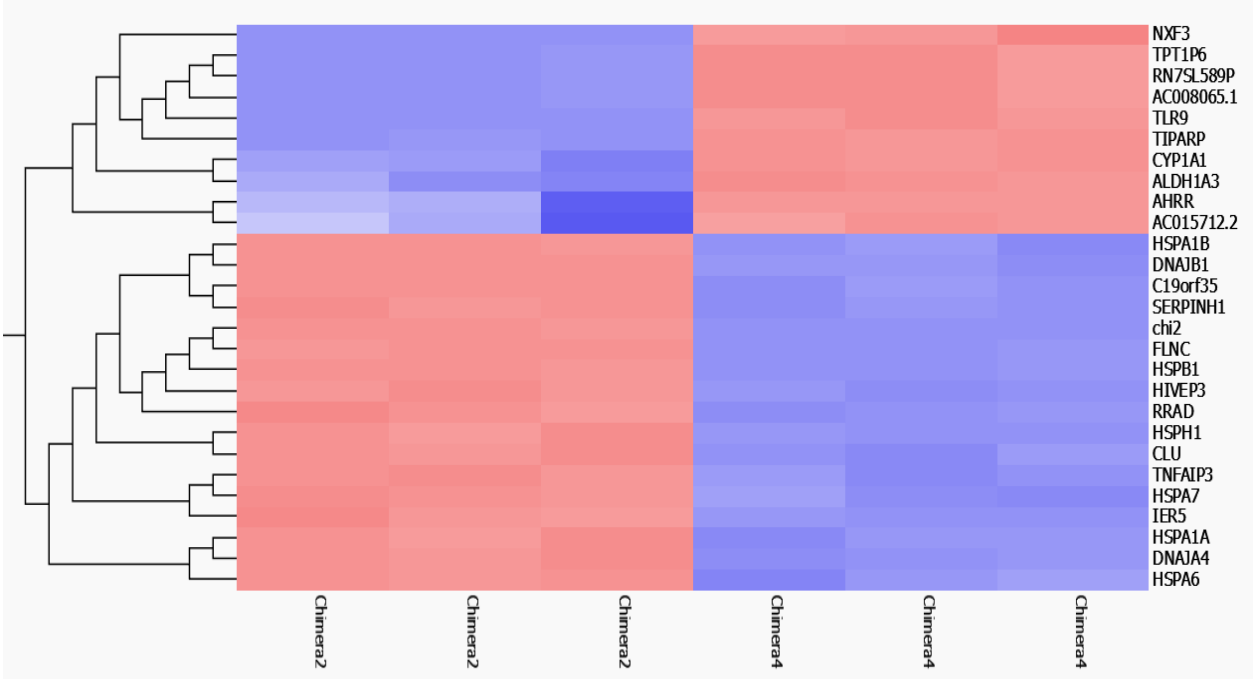

CYP1A1 Fold Change
50
40
30
20
10
0
**
**
**
bHLH – AhR AhR AhRR AhRR
PAS A – AhR AhR AhRR AhRR
PAS B – AhRR AhR AhR AhRR
TAD/TRD – AhRR AhRR AhR AhR

Figure 11) RNA-seq reveals Chimera-4 constitutively activates many AHR pathway genes when compared to Chimera-2. RNA from sorted, EGFP-positive, uninduced cells transfected with Chimera-2- or Chimera-4-P2A-EGFP was sent to the UWBC for RNA-seq library prep, sequencing, and analysis. Heatmap was generated in Qlucore; a q-value of 0.007 and fold-change value of 2.9 were used to pare the list of DEGs to a manageable final number of 27. Upregulated genes are colored red while downregulated genes are in blue. Darkness of color denotes the magnitude of up- or downregulation. The gene symbol is indicated on the right, and genes were organized with hierarchical clustering. Common AHR pathway genes (CYP1A1, ALDH1A3, AHRR, and TIPARP) cluster together, as expected. TLR9 is a novel gene identified in this experiment with possible implications on AHR biology. Downregulated genes include many heat-shock proteins that possibly interact with the p53 pathway.

Figure 11)


NXF3
TPT1P6
RN7SL589P
AC008065.1
TLR9
TIPARP
CYP1A1
ALDH1A3
AHRR
AC015712.2
HSPA1B
DNAJB1
C19orf35
SERPINH1
chi2
FLNC
HSPB1
HIVEP3
RRAD
HSPH1
CLU
TNFAIP3
HSPA7
IER5
HSPA1A
DNAJA4
HSPA6
Chimera2
Chimera2
Chimera2
Chimera4
Chimera4
Chimera4

REFERENCES

1. Gu, Y. Z., Hogenesch, J. B. & Bradfield, C. A. The PAS superfamily: Sensors of environmental and developmental signals. *Annu. Rev. Pharmacol. Toxicol.* **40**, 519–561 (2000).

2. Vazquez-Rivera, E. *et al.* The aryl hydrocarbon receptor as a model PAS sensor. *Toxicol. Rep.* **9**, 1–11 (2022).

3. Hahn, M. E. Aryl hydrocarbon receptors: diversity and evolution. *Chem. Biol. Interact.* **141**, 131–160 (2002).

4. 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).

5. Kanno, Y., Takane, Y., Izawa, T., Nakahama, T. & Inouye, Y. The inhibitory effect of aryl hydrocarbon receptor repressor (AhRR) on the growth of human breast cancer MCF-7 cells. *Biol. Pharm. Bull.* **29**, 1254–1257 (2006).

6. Zudaire, E. *et al.* The aryl hydrocarbon receptor repressor is a putative tumor suppressor gene in multiple human cancers. *J. Clin. Invest.* **118**, 640–650 (2008).

7. Ishihara, Y., Tsuji, M. & Vogel, C. F. A. Suppressive effects of aryl-hydrocarbon receptor repressor on adipocyte differentiation in 3T3-L1 cells. *Arch. Biochem. Biophys.* **642**, 75–80 (2018).

8. McGuire, J., Okamoto, K., Whitelaw, M. L., Tanaka, H. & Poellinger, L. Definition of a dioxin receptor mutant that is a constitutive activator of transcription: delineation of overlapping repression and ligand binding functions within the PAS domain. *J. Biol. Chem.* **276**, 41841–41849 (2001).

9. Ito, T. *et al.* A Constitutively Active Arylhydrocarbon Receptor Induces Growth Inhibition of Jurkat T Cells through Changes in the Expression of Genes Related to Apoptosis and Cell Cycle Arrest*. *J. Biol. Chem.* **279**, 25204–25210 (2004).

10. Andersson, P. *et al.* A constitutively active dioxin/aryl hydrocarbon receptor induces stomach tumors. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 9990–9995 (2002).

11. Köhle, C. *et al.* Conditional expression of a constitutively active aryl hydrocarbon receptor in MCF-7 human breast cancer cells. *Arch. Biochem. Biophys.* **402**, 172–179 (2002).

12. Brunnberg, S. *et al.* The constitutively active Ah receptor (CA-Ahr) mouse as a potential model for dioxin exposure--effects in vital organs. *Toxicology* **224**, 191–201 (2006).

13. Stockinger, B., Shah, K. & Wincent, E. AHR in the intestinal microenvironment: safeguarding barrier function. *Nat. Rev. Gastroenterol. Hepatol.* **18**, 559–570 (2021).

14. Esser, C. & Rannug, A. The aryl hydrocarbon receptor in barrier organ physiology, immunology, and toxicology. *Pharmacol. Rev.* **67**, 259–279 (2015).

15. Rannug, A. How the AHR Became Important in Intestinal Homeostasis—A Diurnal FICZ/AHR/CYP1A1 Feedback Controls Both Immunity and Immunopathology. *Int. J. Mol. Sci.* **21**, 5681 (2020).

16. Rannug, A. & Fritsche, E. The aryl hydrocarbon receptor and light. *Biol. Chem.* **387**, 1149–1157 (2006).

17. Murray, I. A., Patterson, A. D. & Perdew, G. H. Aryl hydrocarbon receptor ligands in cancer: friend and foe. *Nat. Rev. Cancer* **14**, 801–814 (2014).

18. Neavin, D. R., Liu, D., Ray, B. & Weinshilboum, R. M. The Role of the Aryl Hydrocarbon Receptor (AHR) in Immune and Inflammatory Diseases. *Int. J. Mol. Sci.* **19**, (2018).

19. Nebert, D. W., Dalton, T. P., Okey, A. B. & Gonzalez, F. J. Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. *J. Biol. Chem.* **279**, 23847–23850 (2004).

20. Nebert, D. W. & Gonzalez, F. J. P450 Genes: Structure, Evolution, and Regulation. *Annu. Rev. Biochem.* **56**, 945–993 (1987).

21. Gonzalez, F. J. & Korzekwa, K. R. Cytochromes P450 Expression Systems. *Annu. Rev. Pharmacol. Toxicol.* **35**, 369–390 (1995).

22. Wincent, E. *et al.* Inhibition of cytochrome P4501-dependent clearance of the endogenous agonist FICZ as a mechanism for activation of the aryl hydrocarbon receptor. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4479–4484 (2012).

23. Schiering, C. *et al.* Feedback Control of AHR Signaling Regulates Intestinal Immunity. *Nature* **542**, 242–245 (2017).

24. Mimura, J., Ema, M., Sogawa, K. & Fujii-Kuriyama, Y. Identification of a novel mechanism of regulation of Ah (dioxin) receptor function. *Genes Dev.* **13**, 20–25 (1999).

25. Evans, B. R. *et al.* Repression of Aryl Hydrocarbon Receptor (AHR) Signaling by AHR Repressor: Role of DNA Binding and Competition for AHR Nuclear Translocator. *Mol. Pharmacol.* **73**, 387–398 (2008).

26. Oshima, M., Mimura, J., Yamamoto, M. & Fujii-Kuriyama, Y. Molecular mechanism of transcriptional repression of AhR repressor involving ANKRA2, HDAC4, and HDAC5. *Biochem. Biophys. Res. Commun.* **364**, 276–282 (2007).

27. Vogel, C. F. A. *et al.* Transgenic Overexpression of Aryl Hydrocarbon Receptor Repressor (AhRR) and AhR-Mediated Induction of CYP1A1, Cytokines, and Acute Toxicity. *Environ. Health Perspect.* **124**, 1071–1083 (2016).

28. Tigges, J. *et al.* Aryl hydrocarbon receptor repressor (AhRR) function revisited: repression of CYP1 activity in human skin fibroblasts is not related to AhRR expression. *J. Invest. Dermatol.* **133**, 87–96 (2013).

29. Baba, T. *et al.* Structure and Expression of the Ah Receptor Repressor Gene. *J. Biol. Chem.* **276**, 33101–33110 (2001).

30. 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).

31. Bernshausen, T., Jux, B., Esser, C., Abel, J. & Fritsche, E. Tissue distribution and function of the Aryl hydrocarbon receptor repressor (AhRR) in C57BL/6 and Aryl hydrocarbon receptor deficient mice. *Arch. Toxicol.* **80**, 206–211 (2006).

32. Bergens, M. A. *et al.* Smoking-associated AHRR demethylation in cord blood DNA: impact of CD235a+ nucleated red blood cells. *Clin. Epigenetics* **11**, 87 (2019).

33. Burris, H. H. *et al.* Offspring DNA methylation of the aryl-hydrocarbon receptor repressor gene is associated with maternal BMI, gestational age, and birth weight. *Epigenetics* **10**, 913–921 (2015).

34. Cole, J. W. & Xu, H. Aryl Hydrocarbon Receptor Repressor Methylation: A Link Between Smoking and Atherosclerosis. *Circ. Cardiovasc. Genet.* **8**, 640–642 (2015).

35. Kanno, Y., Takane, Y., Takizawa, Y. & Inouye, Y. Suppressive effect of aryl hydrocarbon receptor repressor on transcriptional activity of estrogen receptor alpha by protein-protein interaction in stably and transiently expressing cell lines. *Mol. Cell. Endocrinol.* **291**, 87–94 (2008).

36. Kemp Jacobsen, K., Johansen, J. S., Mellemgaard, A. & Bojesen, S. E. AHRR (cg05575921) methylation extent of leukocyte DNA and lung cancer survival. *PloS One* **14**, e0211745 (2019).

37. Li, Y. *et al.* Poor Prognosis of Gastric Adenocarcinoma with Decreased Expression of AHRR. *PLOS ONE* **7**, e43555 (2012).

38. Reynolds, L. M. *et al.* DNA Methylation of the Aryl Hydrocarbon Receptor Repressor Associations With Cigarette Smoking and Subclinical Atherosclerosis. *Circ. Cardiovasc. Genet.* **8**, 707–716 (2015).

39. Kumar, P. *et al.* Expression analysis of aryl hydrocarbon receptor repressor (AHRR) gene in gallbladder cancer. *Saudi J. Gastroenterol. Off. J. Saudi Gastroenterol. Assoc.* **27**, 54–59 (2021).

40. Haarmann-Stemmann, T. & Abel, J. The arylhydrocarbon receptor repressor (AhRR): structure, expression, and function. *Biol. Chem.* **387**, 1195–1199 (2006).

41. Hahn, M. E., Allan, L. L. & Sherr, D. H. Regulation of constitutive and inducible AHR signaling: complex interactions involving the AHR repressor. *Biochem. Pharmacol.* **77**, 485–497 (2009).

42. Vogel, C. F. A. & Haarmann-Stemmann, T. The aryl hydrocarbon receptor repressor – More than a simple feedback inhibitor of AhR signaling: Clues for its role in inflammation and cancer. *Curr. Opin. Toxicol.* **2**, 109–119 (2017).

43. Karchner, S. I. *et al.* The Active Form of Human Aryl Hydrocarbon Receptor (AHR) Repressor Lacks Exon 8, and Its Pro185 and Ala185 Variants Repress both AHR and Hypoxia-Inducible Factor. *Mol. Cell. Biol.* **29**, 3465–3477 (2009).

44. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).

45. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).

46. Oshlack, A., Robinson, M. D. & Young, M. D. From RNA-seq reads to differential expression results. *Genome Biol.* **11**, 220 (2010).

47. Livak, K. J. & Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* **25**, 402–408 (2001).

48. 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).

49. Sakurai, S., Shimizu, T. & Ohto, U. The crystal structure of the AhRR-ARNT heterodimer reveals the structural basis of the repression of AhR-mediated transcription. *J. Biol. Chem.* **292**, 17609–17616 (2017).

50. Kado, S. *et al.* Aryl hydrocarbon receptor signaling modifies Toll-like receptor-regulated responses in human dendritic cells. *Arch. Toxicol.* **91**, 2209–2221 (2017).

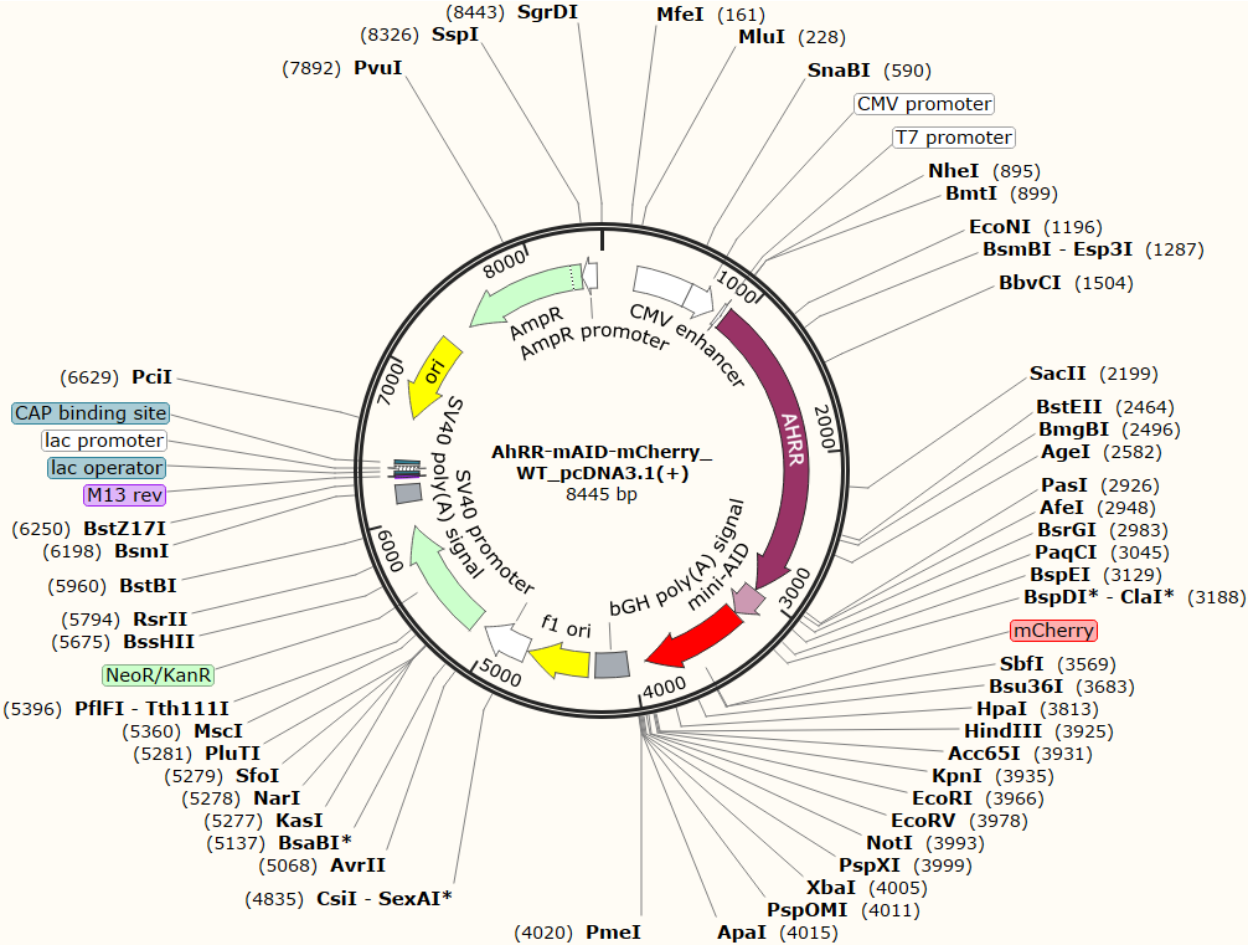
51. Takeda, K. & Akira, S. TLR signaling pathways. *Semin. Immunol.* **16**, 3–9 (2004).

52. Sasabe, J. & Suzuki, M. Emerging Role of D-Amino Acid Metabolism in the Innate Defense. *Front. Microbiol.* **9**, 933 (2018).

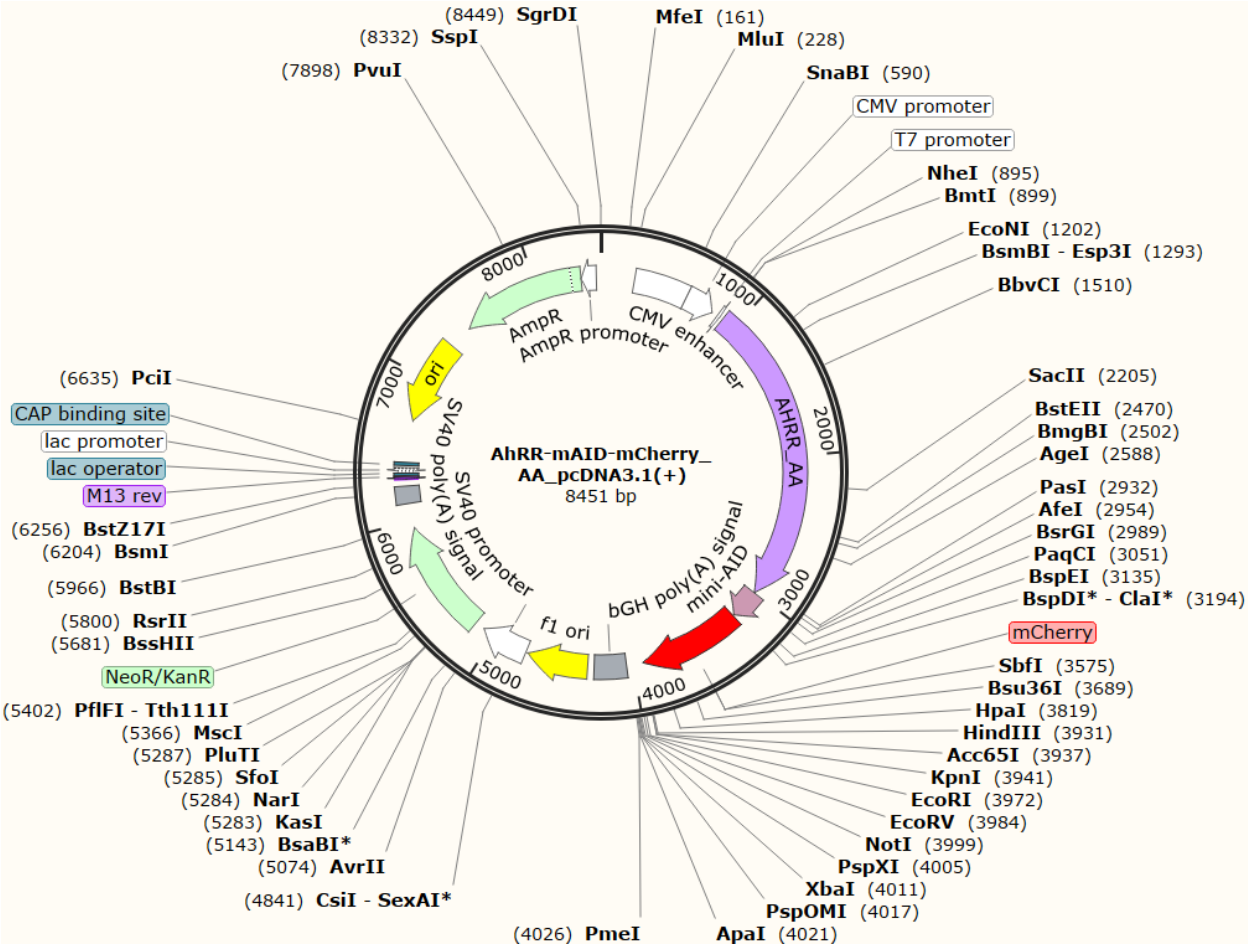
53. Yamada, T. *et al.* Constitutive aryl hydrocarbon receptor signaling constrains type I interferon–mediated antiviral innate defense. *Nat. Immunol.* **17**, 687–694 (2016).

APPENDIX

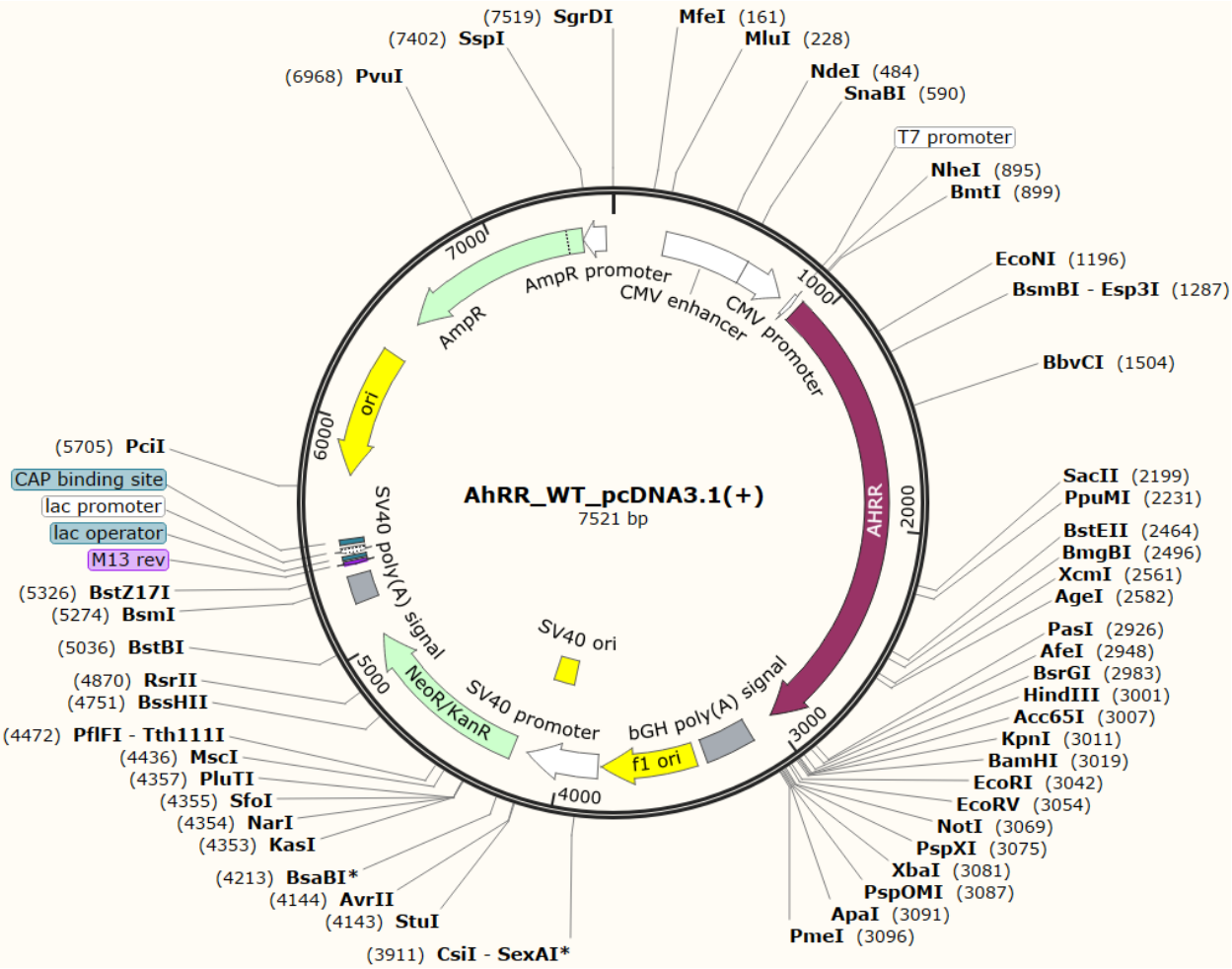
Appendix 1. AHRR-mCherry_WT plasmid map. Synthesized by GenScript using the pcDNA3.1(+) vector. The AHRR-mCherry insert sequences are identical to the sequences of AHRR, mAID, and mCherry in plasmids synthesized by Twist Bioscience.



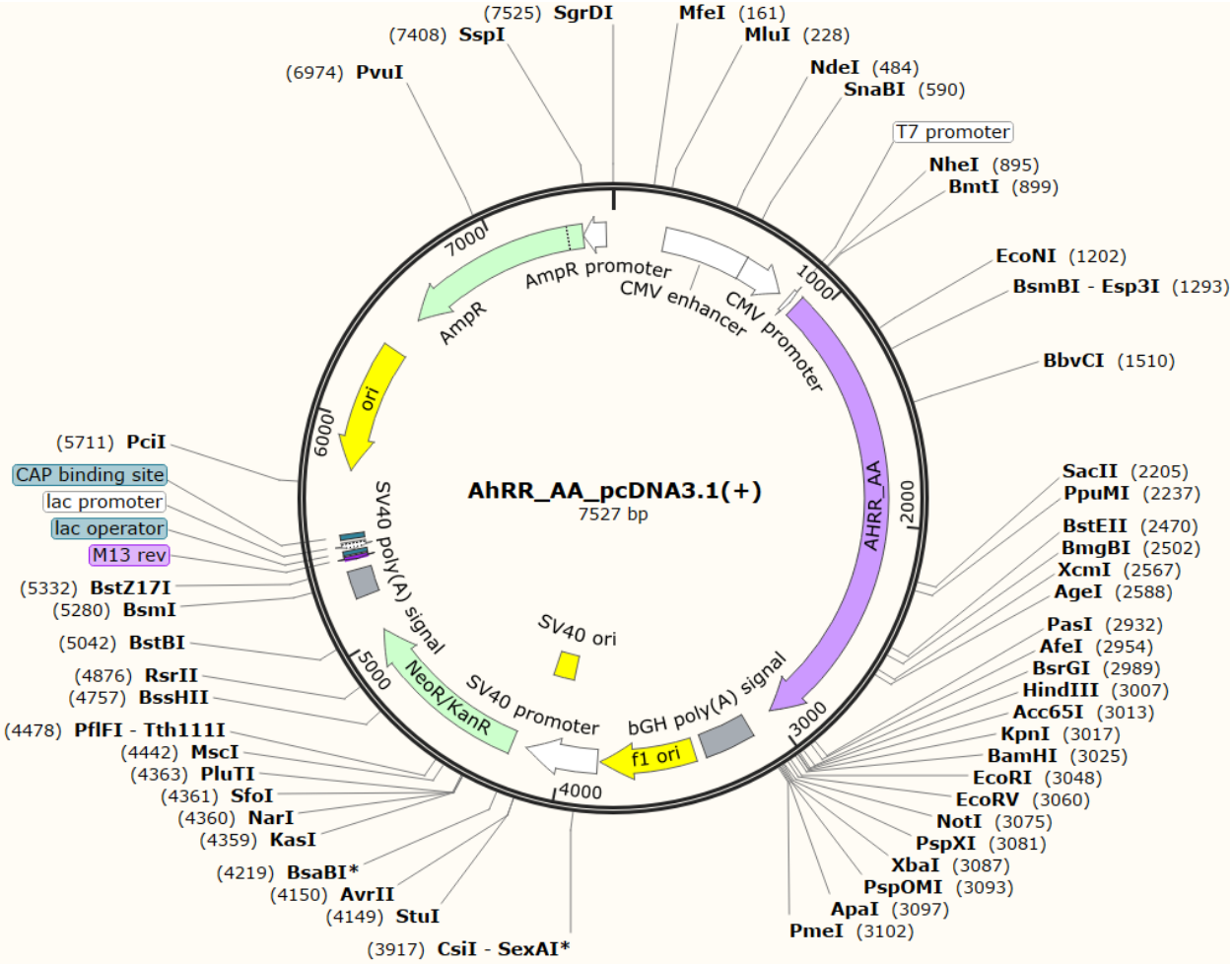
Appendix 2. AHRR-mCherry_AA plasmid map. Synthesized by GenScript using the pcDNA3.1(+) vector.



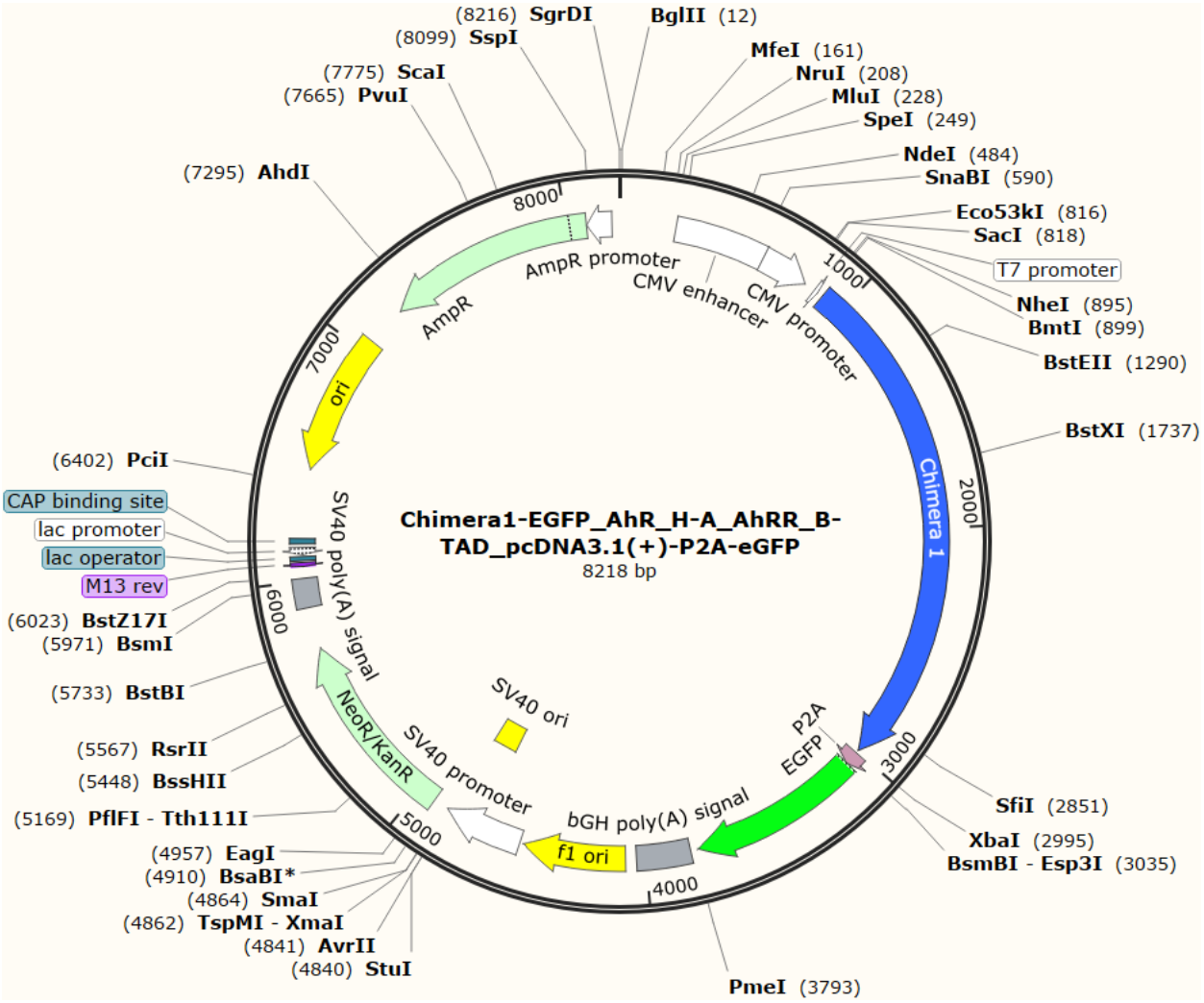
Appendix 3. AHRR_WT plasmid map. Synthesized by GenScript using the pcDNA3.1(+) vector.



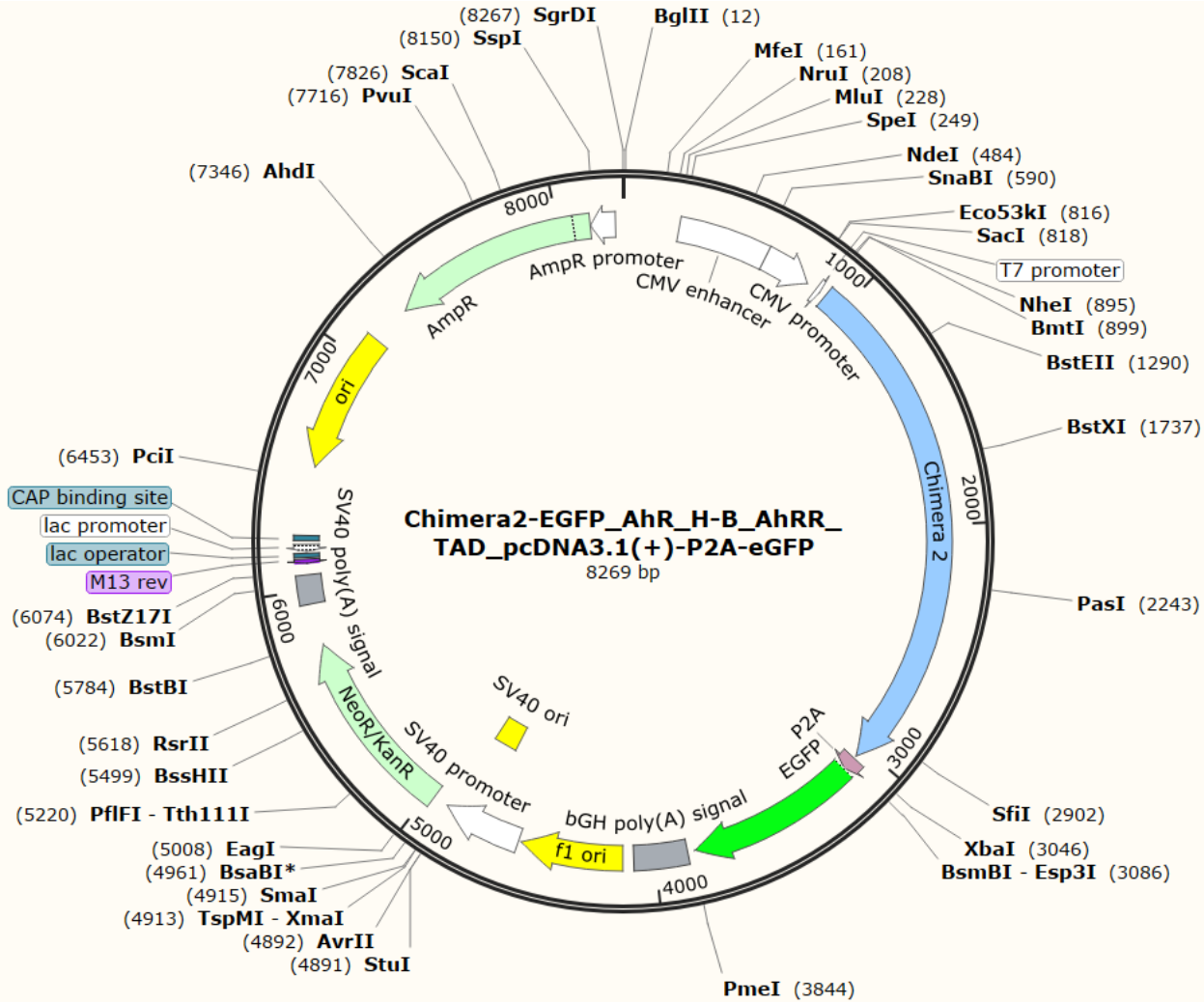
Appendix 4. AHRR_AA plasmid map. Synthesized by GenScript using the pcDNA3.1(+) vector.



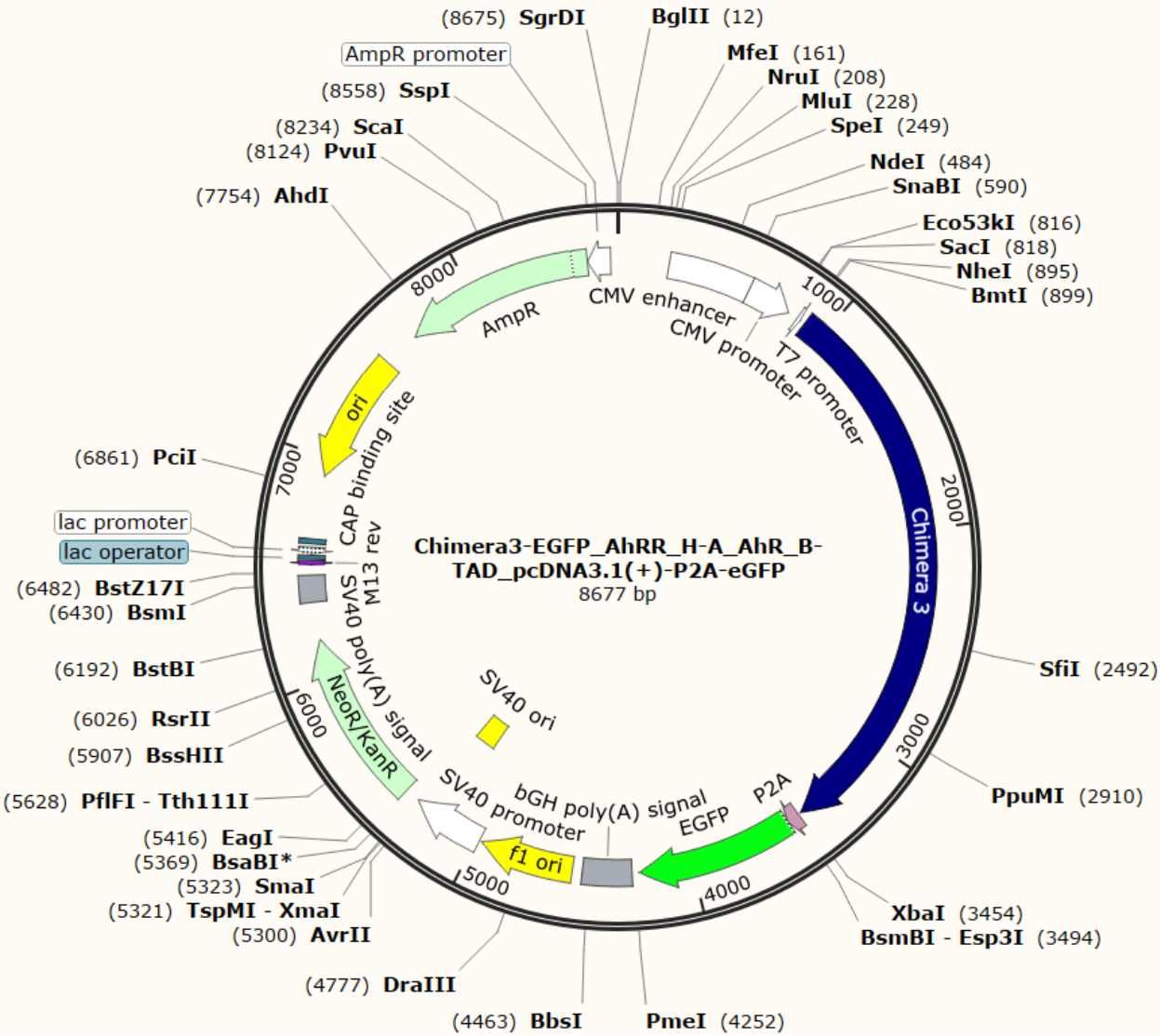
Appendix 5. Plasmid map of pJCP_AHR_H-A_AHRR_B-TAD_P2A_EGFP (Chimera-1-EGFP).



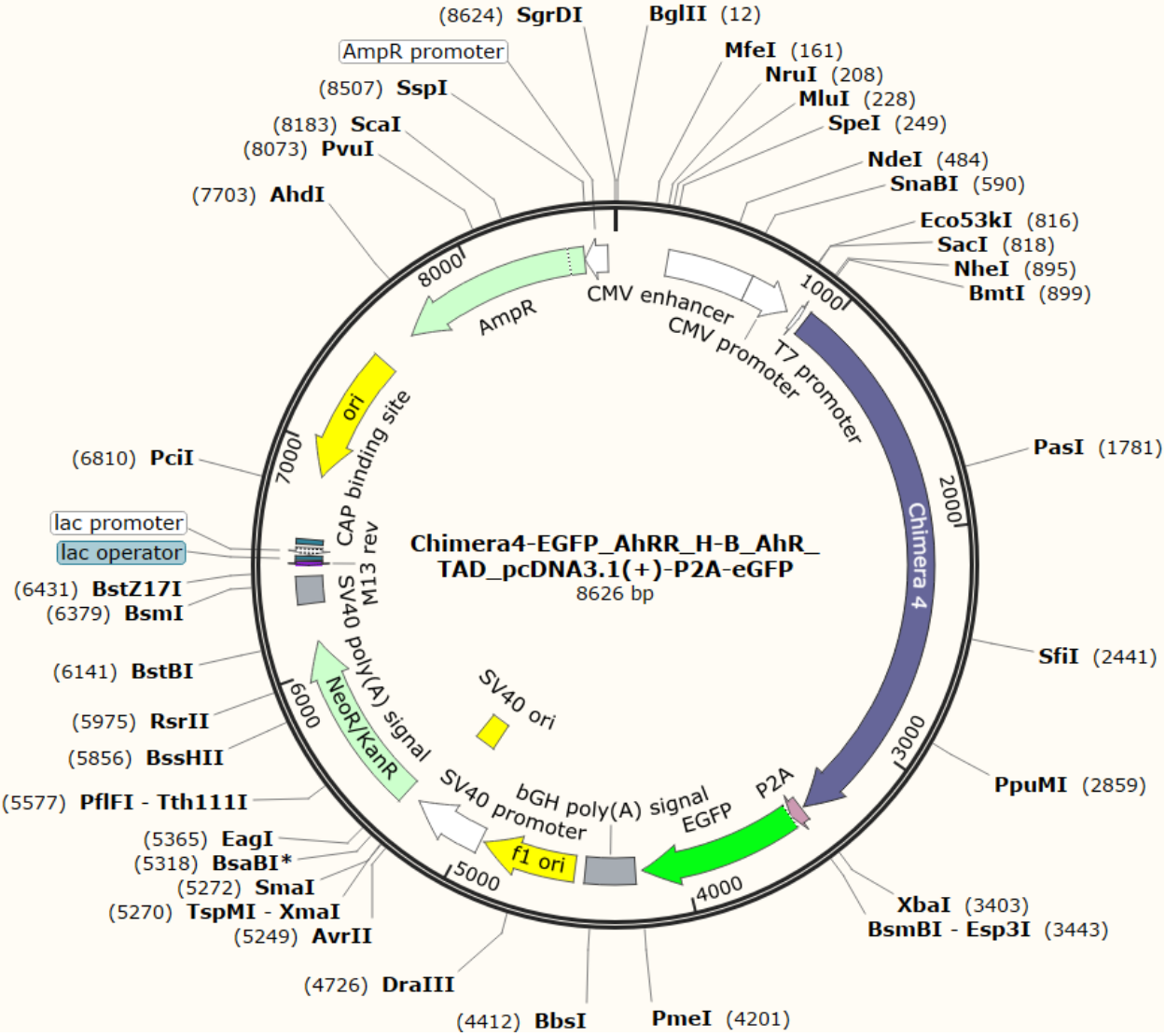
Appendix 6. Plasmid map of pJCP_AHR_H-B_AHRR_TAD_P2A_EGFP (Chimera-2-EGFP).



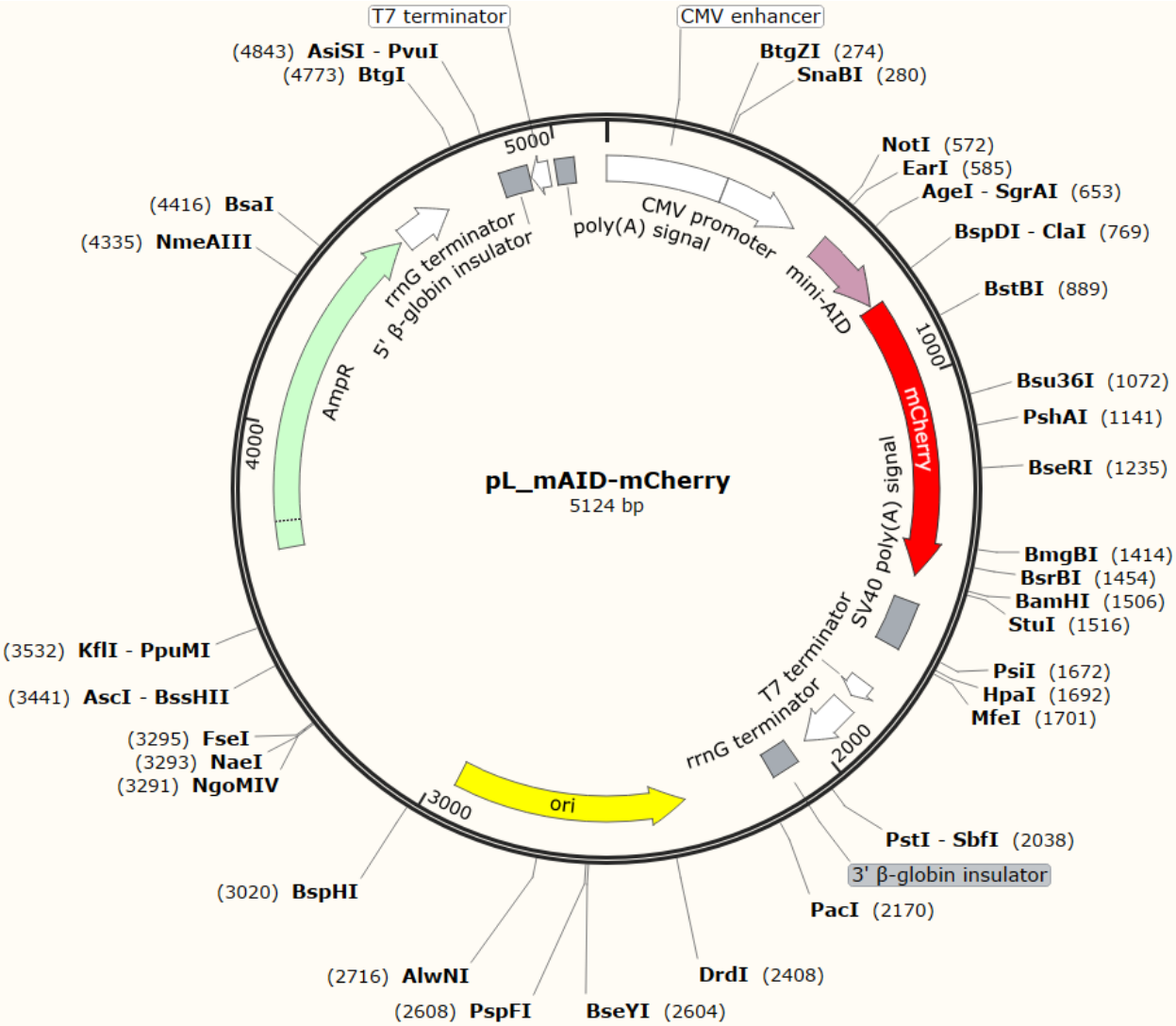
Appendix 7. Plasmid map of pJCP_AHRR_H-A_AHR_B-TAD_P2A_EGFP (Chimera-3-EGFP).



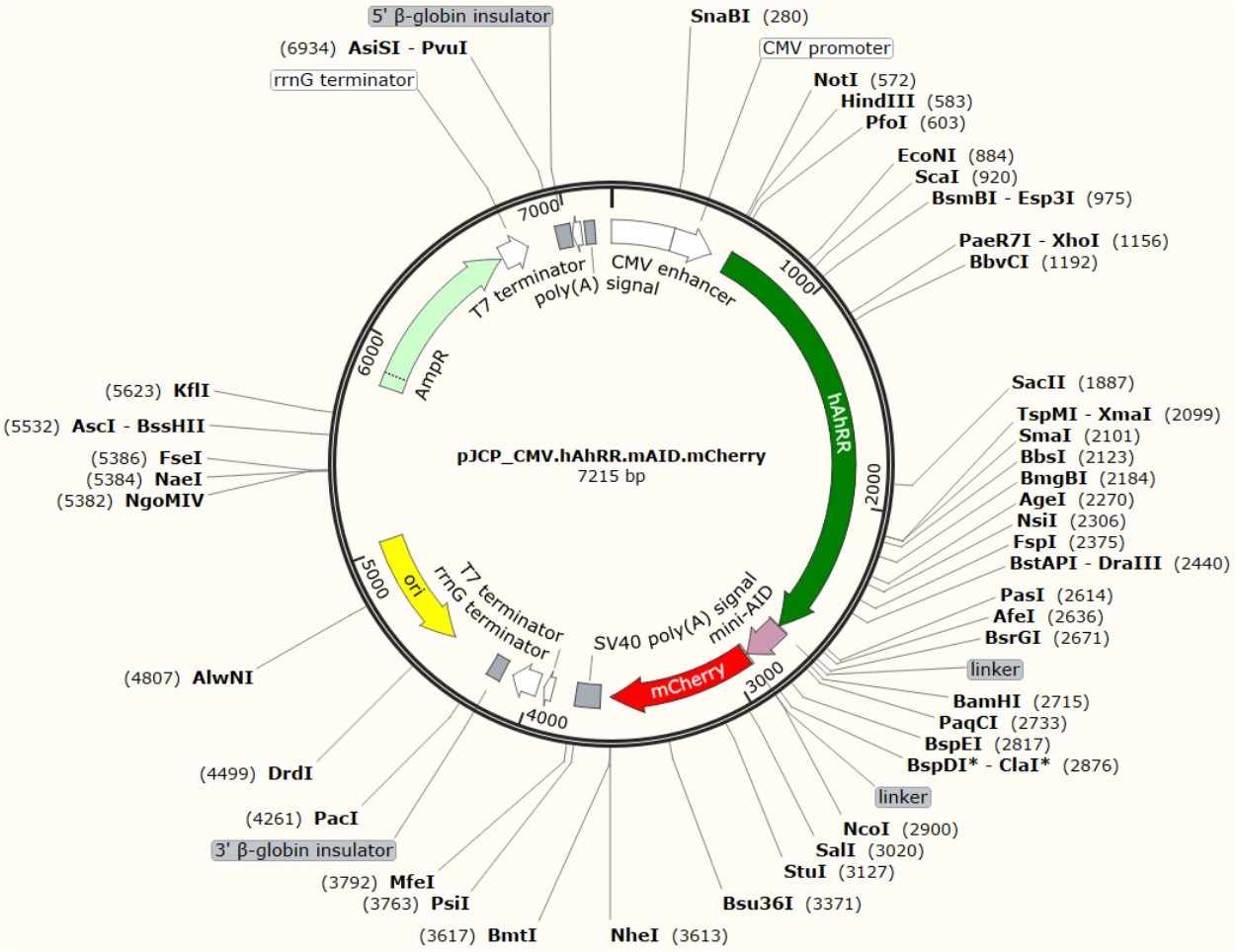
Appendix 8. Plasmid map of pJCP_AHRR_H-B_AHR_TAD_P2A_EGFP (Chimera-4-EGFP).



Appendix 9. Plasmid map of mCherry. Synthesized by Twist Bioscience using the pTwist CMV vector.



Appendix 10. Plasmid map of AHRR-mCherry. Synthesized by Twist Bioscience using the pTwist CMV vector.



Appendix 11. Gene lists: TRP Response Set (TRS), Putative Response Set (PRS), PRS2. The TRS and PRS genes are from a set defined and created in Chapter 3. Genes from the TRS (CYP1A1, AHRR, ALDH1A3, and TIPARP) and the PRS (AC015712.2) overlap with the PRS2, giving them greater weight as true AHR pathway genes. Genes in the PRS2 that overlap with the TRS or PRS are in italics.

TRS	**PRS**	**PRS2**
CYP1A1	GDF15	NXF3
AHRR	*AC015712.2*	TPT1P6
ALDH1A3	ADGRF1	RN7SL589P
NQO1	FTH1	AC008065.1
TIPARP	GDA	TLR9
PPARG	MAGI1	*CYP1A1*
CDK6	SLC16A6	*AHRR*
	MEGF6	*ALDH1A3*
		TIPARP
		AC015712.2

Future Directions and Impact

FUTURE DIRECTIONS

The work contained in this thesis, as well as unfinished work in the lab which did not meet the level of completion required for the thesis, suggest several opportunities for ensuing graduate students.

The first future direction lies within the chimera experiments of Chapter 5. The chimeras in that chapter were created using the wild-type AHR and AHRR gene sequences. Codon optimization for human expression systems would potentially increase the potency of these chimeras, providing a system potentially better than current options. Additionally, the creation of a fifth chimera presents an interesting opportunity to study the DNA binding preferences between AHR and AHRR. The proposed design of Chimera 5 would be to take the wild-type AHR sequence but substitute AHRR's PAS B domain for that of the receptor using the junction points defined in this thesis. Chimera 5 would be predicted to be another constitutively active method of inducing the AHR pathway. However, since the DNA-interacting elements belong to AHR, affected loci would be AHR-specific. Differential gene expression analysis between Chimera 4 and Chimera 5 may reveal loci specific to AHRR, those specific to AHR, and ones that overlap between the proteins. This would shed more light on AHRR's function as a transcription factor, as well as provide the potential for the identification of novel AHR pathway genes independent of ligand effects.

A second future direction would interrogate the importance of nuclear localization in AHRR function. A plasmid was created that inserted a methionine into the AHRR sequence after the basic region with the amino acid sequence of KRHR (where the AA mutants had alanine

residues inserted). All sequence before the added methionine was deleted, resulting in an AHRR with a truncated N-term (noNLS mutant). The plasmid insert had mCherry fused to the AHRR mutant with a P2A-EGFP at its C-term, allowing fluorescent microscopy to be used to study the construct's localization (Appendix 1). A transient transfection with this plasmid confirmed the noNLS-AHRR-mCherry mutant does not localize to the nucleus, even though the EGFP marker suffuses the entire cell (Figure 1). Subsequent experiments with this construct could answer fundamental questions about AHRR's function, such as whether nuclear localization is necessary for repression.

A third future direction would expand on the experiments with Chimera 4 and create both a noNLS version of Chimera 4 and an AA mutant Chimera 4. These variations on our constitutively active chimera would answer questions about the role of DNA binding in AHR pathway activation, how cellular localization affects function, and would investigate the AHRR's non-canonical, protein-protein interaction capabilities.

The fourth and final future direction of this work stems from the tryptophan metabolism work of Chapter 3. Our RNA-seq analysis of the multiple induction mechanisms suggested that the core battery of AHR pathway genes includes CYP1A1, NQO1, and ALDH1A3 but omits CYP1A2 and CYP1B1. As NQO1 and ALDH1A3 are also involved in metabolism and oxidative stress, these genes may be functionally important in the AHR pathway and are worthy of further study. Creation of knockout cell lines with deletions of CYP1A1, NQO1, and ALDH1A3, individually and in combination, may provide additional insight into the metabolic arm of AHR signaling.

Additionally, creation of a triple knockout mouse with these genes may provide evidence for their biological necessity.

IMPACT

This thesis adds value to the space of AHR biology in several ways. One contribution to the literature is the suggestion that AHRR can repress the AHR pathway in a DNA-independent manner, which supports a small body of largely ignored literature on the mechanism of AHRR function. Our data also suggests AHRR is a modest repressor, and an overlooked function of this protein may be as an important secondary wave of transcriptional control in the AHR pathway.

Another contribution of this thesis is the creation of a new technique to screen for novel AHR pathway genes. By doing differential gene expression analysis on Chimera 2 versus Chimera 4, thereby finding the maximal difference in expression, we have created the reagents and bioinformatic workflow to identify ligand-independent AHR pathway genes in any cell line of interest that can be transfected. This would allow identification of novel AHR pathway genes in multiple tissue types, an important step in examining the tissue-specific biology of this enigmatic pathway.

Lastly, this thesis contributes Chimera 4, a constitutive AHR pathway activator. This chimera, in an unoptimized form, induces the AHR pathway that appears to be at or above the magnitude of published systems. This construct could be used in AHR overexpression experiments or for the creation of AHR knock-in cells and organisms, allowing study of this pathway without the need for ligand treatment and ligand-specific effects. Optimization of Chimera 4 would possibly give this construct an even greater ability to induce the pathway, providing a powerful, ligand-independent tool to study AHR biology.

FIGURES

Figure 1) Lack of nuclear localization in a noNLS AHRR mutant. Cells were transfected with noNLS-AHRR-mCherry-P2A-EGFP construct under a CMV promoter and imaged two days post-nucleofection. **A) Bright field image of transfected cells. B) Transfected cells viewed in the EGFP channel.** Free EGFP suffuses the entire cell, fluorescing in the nuclear and cytosolic compartments. **C) Transfected cells viewed in the mCherry channel.** The mCherry is fused to the AHRR protein in this construct. The mCherry fluoresces in the cytosolic compartment but does not appear to be nuclear. **D) Overlay of all three previous images.**

Figure 1A)

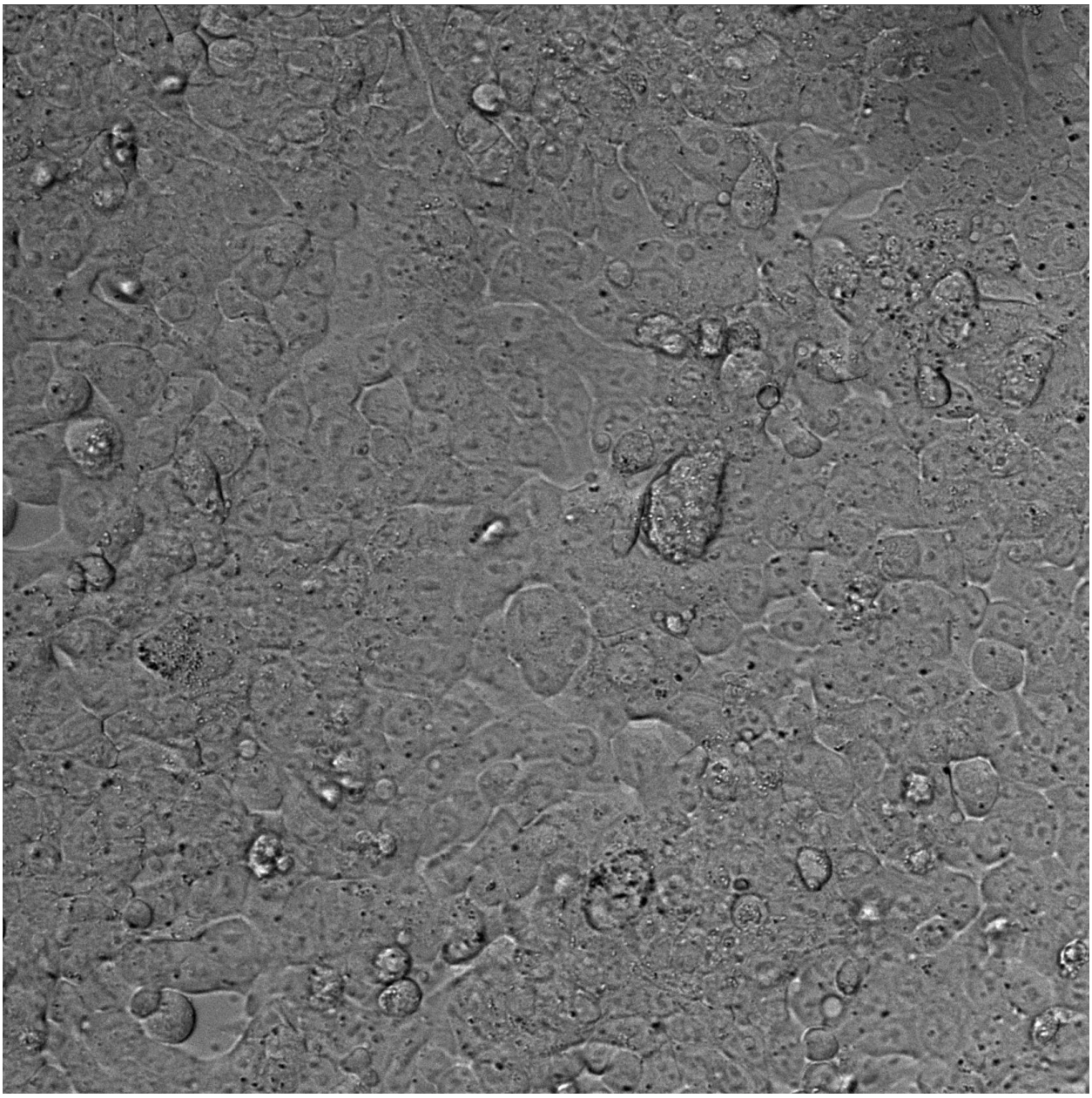


Figure 1B)

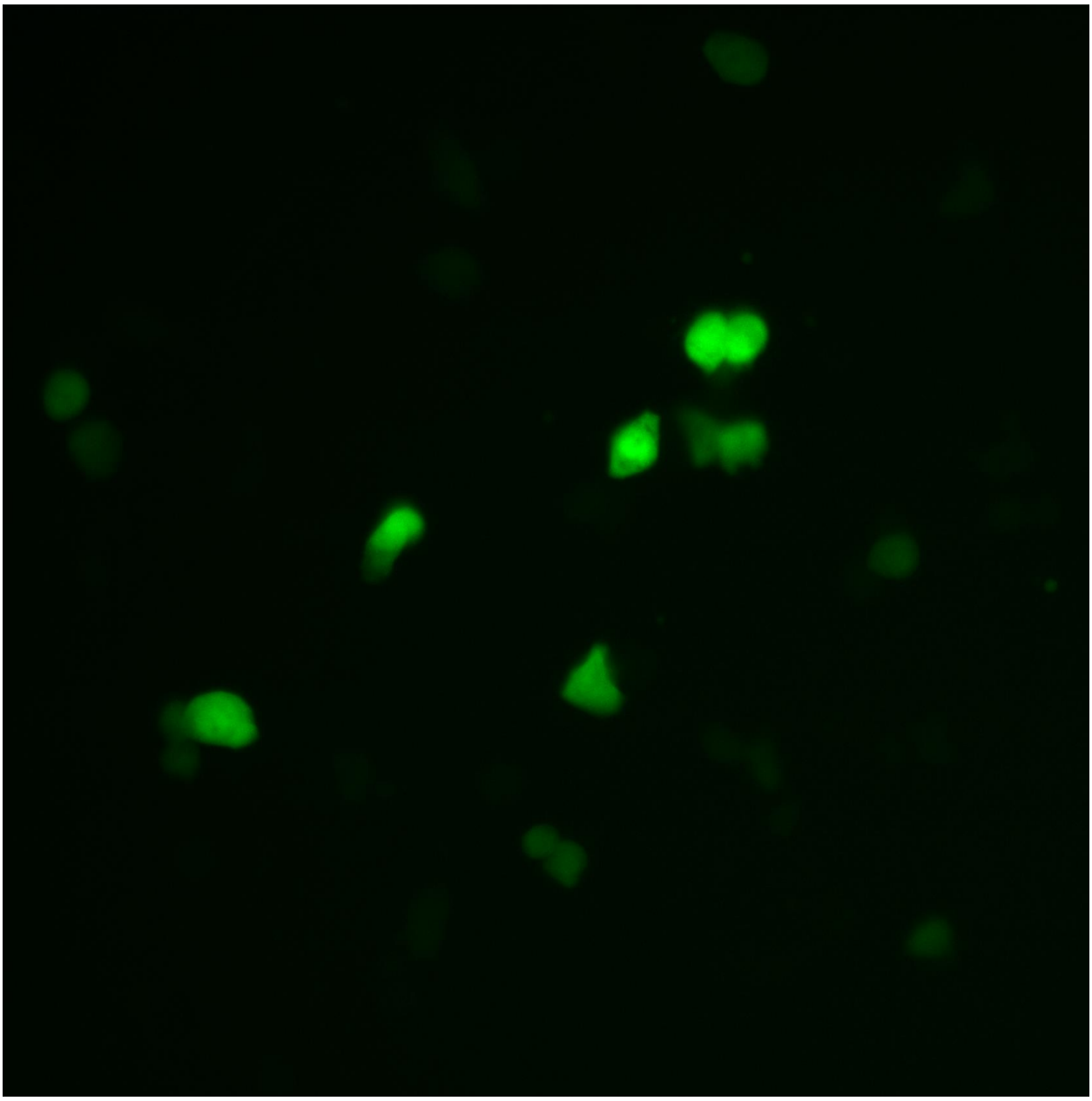


Figure 1C)

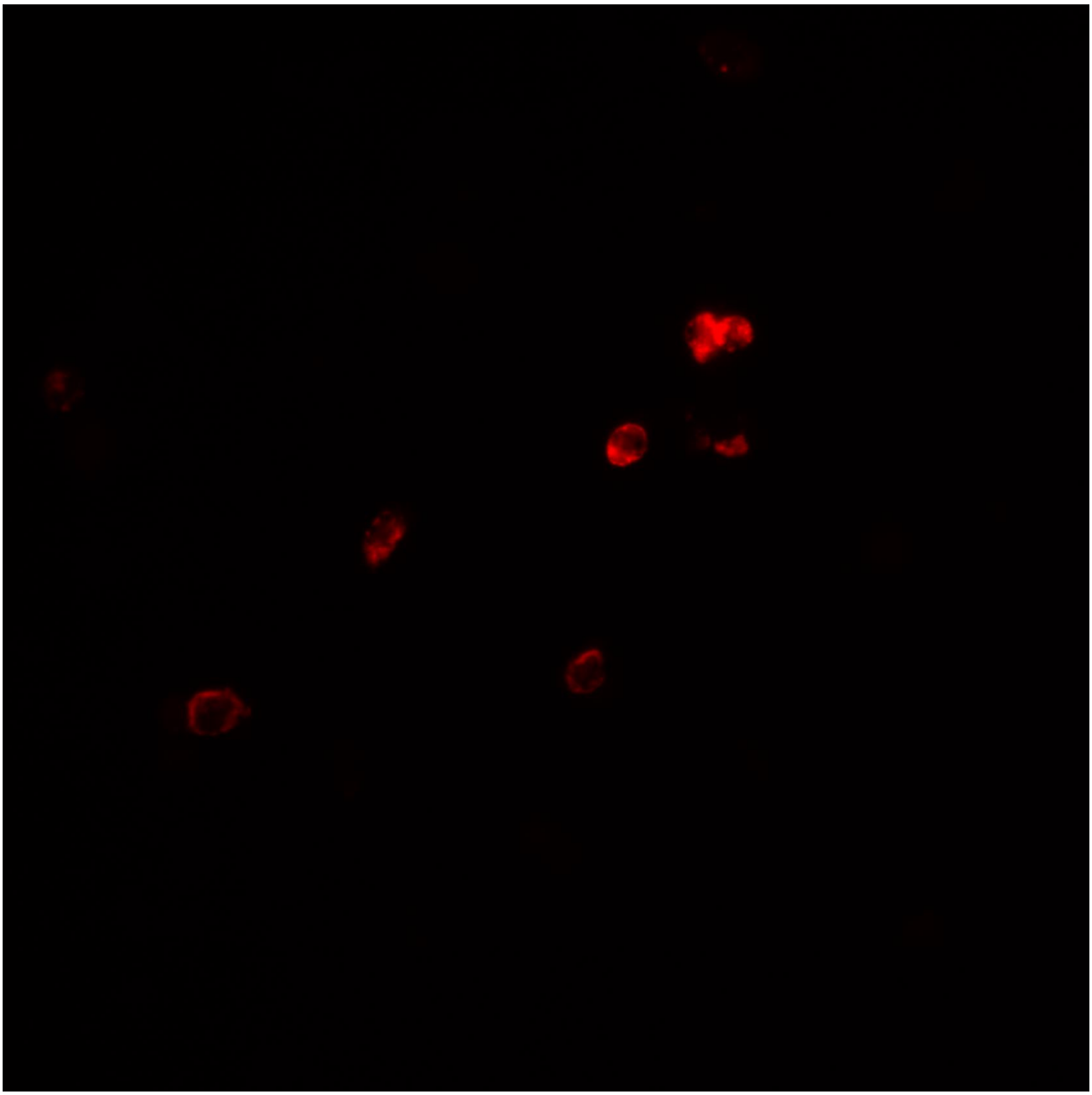
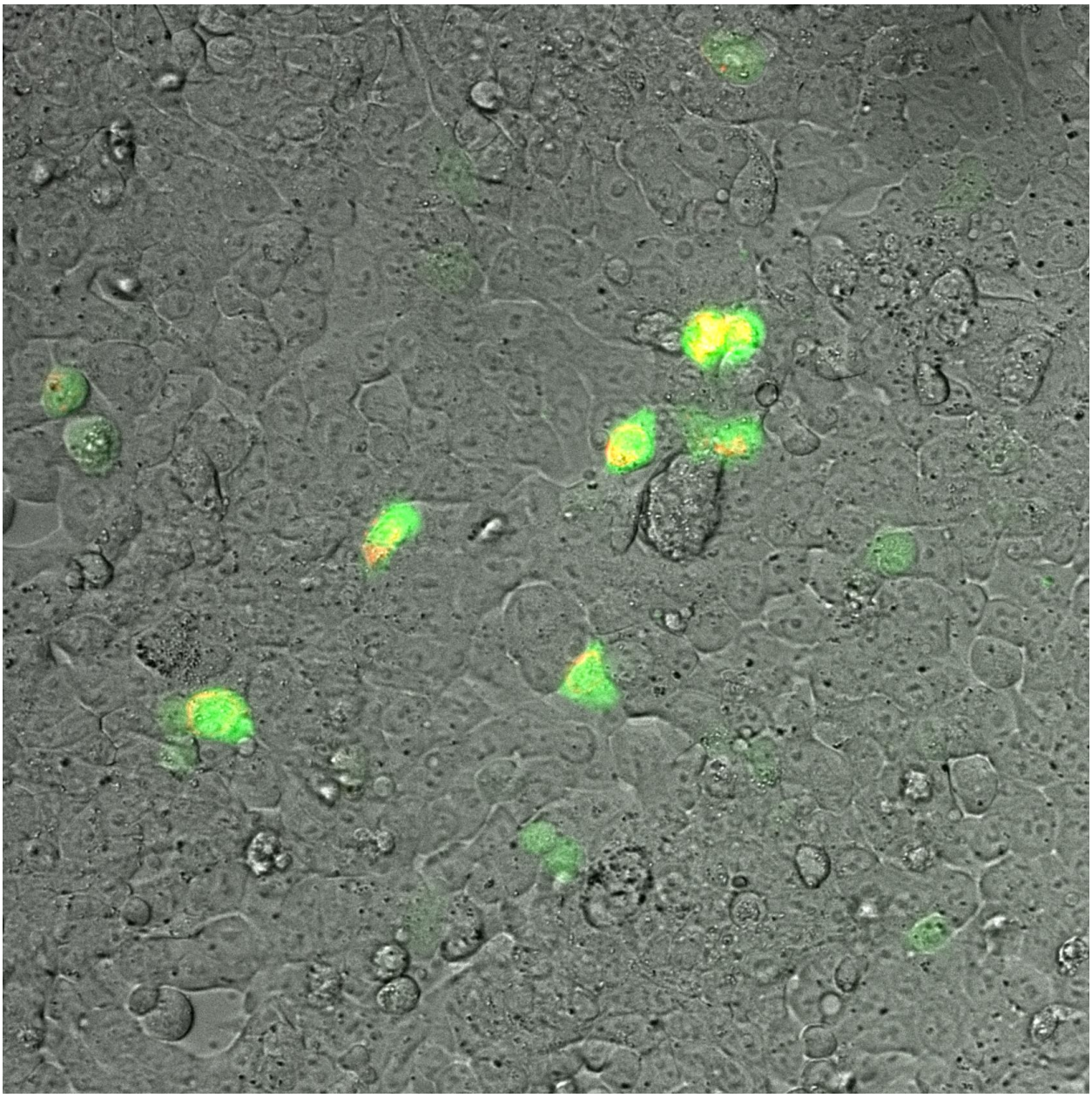


Figure 1D)



APPENDIX

Appendix 1. Plasmid map of noNLS_AHRR-mCherry-P2A-EGFP