Functional genomic analysis of non-canonical DNA regulatory elements of the aryl hydrocarbon receptor

Shayan Shahriar*[$], Tajhal D Patel[^$], Manjula Nakka*, Sandra L Grimm*[#^], Cristian Coarfa*[#^], Daniel A Gorelick*[#]

*Center for Precision Environmental Health

[#]Department of Molecular & Cellular Biology

^Dan L Duncan Comprehensive Cancer Center

Baylor College of Medicine, Houston, Texas, USA

[$]These authors contributed equally to this manuscript

Authors for correspondence: CC (coarfa@bcm.edu), DAG (gorelick@bcm.edu)

ABSTRACT

The aryl hydrocarbon receptor (AHR) is a ligand-dependent transcription factor that is activated by environmental toxicants, like halogenated and polycyclic aromatic hydrocarbons, and then binds to DNA and regulates gene expression. AHR is involved in various physiological processes, including liver and immune system function, cell cycle regulation, oncogenesis, and metabolism. In the canonical pathway, AHR binds to a consensus DNA sequence (GCGTC), termed the xenobiotic response element (XRE), recruits protein coregulators, and regulates target gene expression. Emerging evidence suggests that AHR may regulate gene expression via an additional pathway, by binding to a non-consensus DNA sequence (GGGA) termed the non-consensus XRE (NC-XRE). The prevalence of NC-XRE motifs in the genome is not known. Studies using chromatin immunoprecipitation and reporter genes provide indirect evidence of AHR-NC-XRE interactions, but direct evidence for an AHR-NC-XRE interaction that regulates transcription in a natural genomic context is lacking. Here, we analyzed AHR binding to NC-XRE DNA on a genome-wide scale in mouse liver. We integrated ChIP-seq and RNA-seq data and identified putative AHR target genes with NC-XRE motifs in regulatory regions. We found that NC-XRE motifs are present in 82% of AHR-bound DNA, which are significantly enriched relative to random genomic regions. These AHR-bound regions include, but are not limited to, promoters and enhancers of AHR target genes. To obtain direct evidence of AHR regulation via this non-canonical pathway, we performed functional genomics on the mouse *Serpine1* gene, a putative AHR target via NC-XRE. Deleting NC-XRE motifs from the *Serpine1* promoter reduced the upregulation of *Serpine1* by TCDD, an AHR ligand. We conclude that AHR upregulates *Serpine1* via NC-XRE DNA. Taken together, our results provide the first direct evidence of AHR-mediated gene regulation via NC-XRE in a natural genomic context. These findings enhance our understanding of AHR-bound DNA regions and their influence on target gene expression. Our results will also improve our ability to identify AHR target genes and their physiological relevance.

INTRODUCTION

The toxic effects of secondhand cigarette smoke, fossil fuels, and byproducts of industrial combustion are caused by halogenated and polycyclic aromatic hydrocarbons such as TCDD. Understanding the signaling pathways by which such compounds cause toxicity is crucial for predicting tolerable exposure levels and for reversing the effects of adverse exposure. Aromatic hydrocarbons activate aryl hydrocarbon receptors (AHR), a ligand-dependent transcription factor. Upon activation, AHR translocates to the nucleus, where it recruits protein coregulators and directly regulates gene expression (Denison et al., 2011). AHR forms a dimer with the aryl hydrocarbon receptor nuclear translocator protein (ARNT1, also known as HIF1β) to regulate transcription (Hankinson, 2005; Reyes et al., 1992). The AHR-ARNT1 complex binds to a consensus DNA sequence (GCGTG) termed the xenobiotic response element (XRE) (Denison et al., 1988; Yao and Denison, 1992).

Recent evidence complicates these findings. AHR may regulate gene expression by binding to non-consensus XRE sequences. AHR was shown to physically interact with RelB, a member of the NF-κB family, and bind to a distinct RelB/AHR responsive element (RelBAhRE) within the IL-8 promoter. Furthermore, TCDD induced a time-dependent recruitment of AHR to the RelBAhRE site (nucleotides GGGTGCAT) (Vogel et al. 2007). A chromatin immunoprecipitation study found that approximately half of TCDD-AHR complexes in mouse liver were bound to DNA lacking a consensus XRE (Dere et al., 2011). Exploring the genome more closely, Huang and colleagues identified a non-consensus DNA sequence to which AHR binds (Huang and Elferink, 2012). Chromatin immunoprecipitation (ChIP), gel shift (electrophoretic mobility shift assay) and reporter assays support the hypothesis that AHR can bind to a non-consensus XRE DNA sequence characterized by the nucleotides GGGA (referred to as the NC-XRE motif) (Huang and Elferink, 2012; Jackson et al., 2014; Joshi et al., 2015; Wilson et al., 2013).

However, these studies have limitations. They did not analyze AHR-NC-XRE interactions on a genome-wide scale. Additionally, chromatin immunoprecipitation (ChIP) identifies where AHR binds to DNA, but does not reveal whether transcription actually occurs. Gel shift and reporter gene assays sometimes fail to reflect what occurs in a natural genomic context. Direct evidence for an AHR-NC-XRE interaction that regulates transcription in a natural genomic context is lacking.

Here, we sought to analyze AHR binding to non-canonical DNA elements on a genome-wide scale and, in the case of a single target gene, directly test whether NC-XRE motifs are required for AHR-dependent target gene expression.

RESULTS

Frequency and distribution of non-canonical AHR binding motifs

To determine the genome-wide distribution of AHR binding to DNA following TCDD exposure, we analyzed published AHR ChIP-seq data from livers of C57BL/6 male and female mice (28 days old) following 2 hour exposure to 30 µg/kg TCDD (GEO accession numbers GSE97634, GSE97636) (Fader et al., 2017; Fader et al., 2019). This data was previously assayed for AHR binding to consensus XRE DNA [GCGTG] but was not interrogated for AHR binding to non-consensus DNA such as NC-XRE [GGGA] or RelBAhRE [GGGTGCAT]. Given that the XRE, NC-XRE, and RelBAhRE motifs differ in length, which influences their frequency of occurrence across the genome, we assessed the likelihood of finding XRE, NC-XRE or RelBAhRE motifs in AHR ChIP peaks versus in a random region of DNA. Both XRE and NC-XRE motifs were significantly enriched in AHR-bound regions, whereas RelBAhRE was not (Fig. 1A).

We also found consistent patterns of AHR-bound motif enrichment in the livers of male and female mice following TCDD exposure, based on AHR ChIP-seq data. In males, 29.74% of AHR peaks contained only NC-XRE motifs, compared to 25.79% in females. Peaks containing only XRE motifs were slightly more frequent in females (11.75%) than in males (8.99%). The majority of peaks—51.72% in males and 53.96% in females—contained both XRE and NC-XRE motifs, indicating frequent co-occurrence. Despite these minor differences, the overall distribution of motif combinations was remarkably similar between sexes. RelBAhRE motifs were rare in both groups, appearing in fewer than 1% of peaks (Fig. 1B and Supplementary Fig. 1A). These findings indicate that AHR binding to canonical and non-canonical DNA motifs occurs at comparable frequencies in male and female mouse livers, supporting a robust and sex-independent pattern of AHR-DNA interactions in response to TCDD. They also highlight the substantial role of NC-XRE motifs in mediating AHR binding. Because of the low number of RelBAhRE in AHR peaks, we only focused on the XRE and NC-XRE motifs in the subsequent studies.

We next examined how AHR binding sites containing XRE and NC-XRE motifs are distributed across the genome. To do this, we analyzed AHR ChIP-seq peaks in adult mouse liver (GEO accession numbers GSE97634, GSE97636) and overlaid them with genomic regions annotated using a 10-state ChromHMM model (van der Velde et al. 2021). We generated this model using eight histone modifications profiled via publicly available ChIP-seq datasets (GEO accession numbers: GSM1000140, GSM918718, GSM1000150, GSM1000151, GSM1000152, GSM1000153, GSM769014, GSM769015, GSM751035) (Mouse ENCODE Consortium et al. 2012; Tennant et al. 2013). These included histone marks associated with active promoters

(H3K4me3, H3K9ac), enhancers (H3K4me1, H3K27ac), gene bodies (H3K36me3), polycomb-mediated repression (H3K27me3), and heterochromatin (H3K9me3) (Supplementary Figure 2E).

We found that AHR binding was enriched in active promoters (e.g. active TSS) and enhancers (Figure 1C, 1D). Specifically, AHR peaks containing both XRE and NC-XRE motifs were predominantly located in active transcription start sites (TSS) and active enhancers. On the other hand, most NC-XRE-only peaks were found in the enhancer regions, including both active and poised enhancers. Active enhancers have high chromatin accessibility and confer strong enhancer activity whereas poised enhancers are marked by intermediate levels of enhancer-associated histone modifications and are thought to be primed for activation. We also noticed that AHR binding was rare in repressive chromatin states, such as those enriched for H3K27me3 (polycomb repression) or H3K9me3 (heterochromatin), as well as in the quiescent regions lacking significant histone modification signals. These patterns were consistent across male and female samples, supporting a sex-independent mechanism of AHR-DNA interaction (Figure 1C, 1D). Notably, promoter- and enhancer-associated regions comprise only ~5.3% of the mouse genome, whereas repressive and quiescent regions account for approximately 84% (van der Velde et al. 2021). This highlights a strong preference of AHR for binding within the active regulatory elements, consistent with its function as a transcription factor. Specifically, NC-XRE motifs may play a particularly important role in enhancer-associated AHR binding.

Repeated NC-XRE motifs associated with AHR target genes

The preceding analysis examined the presence of NC-XRE motifs in all genomic regions where AHR was bound to DNA. However, transcription factors like AHR may bind DNA without necessarily affecting gene expression. Therefore, to characterize functional AHR binding events, we sought to identify AHR ChIP-seq peaks associated with AHR target genes. For this, we integrated the AHR ChIP-seq data with published RNA-seq data from mouse liver (Fader et al., 2017; Nault et al., 2015; Nault et al., 2018). To maximize identification of AHR target genes, we included RNA-seq data from mouse livers exposed to various doses of TCDD, a potent AHR ligand, across multiple durations of exposure (Tables 1, 2). We defined AHR target genes as differentially expressed genes (RNA-seq, fold difference in TCDD vs vehicle exceeding 1.5x to cover both up- and downregulation, FDR<5%) containing AHR binding site (ChIP peak) within 10 kb of a gene body. The 10 kb parameter allows us to include promoter and gene proximal regulatory DNA peaks, important because our analysis of ChIP-seq data suggested that AHR binds NC-XRE at enhancers. Among the differentially expressed genes in TCDD vs vehicle, 520 had AHR binding peaks at XRE, NC-XRE or XRE+NC-XRE motifs within 10 kb of a gene body. 28 of these genes had AHR binding peaks at NC-XRE but not at XRE or XRE+NC-XRE motifs (Supplementary Table 1).

The preceding analysis is for any single instance of XRE or NC-XRE DNA sequence. However, for some transcription factors, repeated DNA binding sites produce a more robust response (Klein-Hitpass et al., 1988). To explore the functional importance of repeated runs of XRE or NC-XRE motifs, we sought to examine the association between repeated runs of XRE or NC-XRE motifs and AHR target gene expression.

We defined tandem XRE and NC-XRE motifs based on two parameters: the number of repeats and the maximum allowed spacer length (in base pairs) between adjacent motifs, as illustrated in Figure 2A. The number of repeats refers to the minimum number of XRE or NC-XRE motifs within a given region, while the spacer defines the maximum distance permitted between any two motifs in a repeat run. Using this framework, we analyzed AHR target genes for varying numbers of XRE or NC-XRE repeats and spacer lengths. Notably, the previously characterized AHR target gene *Serpine1* (also known as *Pai1*) contains five NC-XRE motifs within its putative promoter region. This observation aligns with previous findings that AHR upregulates *Serpine1* via NC-XRE-mediated binding (Huang and Elferink, 2012), prompting us to further investigate the regulatory significance of five-repeat NC-XRE clusters.

Fig 2B shows the number AHR target genes with AHR peaks containing 5 NC-XRE motifs separated by up to 100 basepairs (bp), compared to target genes found near AHR peaks with 5 XRE motifs. Notably, we observed a ninefold increase in the number of AHR target genes linked to peaks with five NC-XRE motifs separated by 50 bp or less, compared to those with five XRE motifs under the same spacing constraint. This subset includes *Serpine1*, a well-established AHR target regulated via NC-XRE motifs, as previously discussed (Fig 2B).

Further analysis revealed that while 82% of all AHR peaks contain at least one NC-XRE motif, only 5.2% harbor a cluster of five NC-XRE motifs spaced by 50 bp or less (Figure 2C, Supplementary Figure 2A). We next characterized the genomic distribution of these 5x NC-XRE clusters. Of the 426 identified clusters, over 71% were located at or near transcription start sites (TSS), indicating a strong promoter association. An additional 27.8% were found in enhancer regions, with 21% in active enhancers and 6.8% in poised enhancers. These clusters were rarely found in repressed chromatin states (e.g., heterochromatin, polycomb-repressed, or quiescent regions) or within gene bodies associated with transcriptional activity (transcription or weak transcription) (Figure 2D, Supplementary Figures 2C and 2D). Similarly, clusters of five XRE motifs were predominantly located in promoters and enhancers, with a small fraction (6.5%) found within gene bodies. We compared the binding strengths of different histone marks at AHR peaks containing 5x XRE or NC-XRE (50 bp spacer). NC-XREs are enriched at AHR binding sites proximal to transcription start sites (H3K4me3, H3K27ac) and enhancers (H3K27ac) (Fig 2E, Supplementary Fig 2E). Consistent with the broader genomic distribution of

NC-XREs, 5X NC-XRE clusters were enriched in enhancer regions compared to 5x XRE clusters. This also suggests that NC-XREs may be more potent when they occur in tandem repeats.

Putative AHR-NC-XRE target pathways

AHR ChIP-seq peaks containing five NC-XRE motifs spaced by 50 base pairs or less were found near 28 AHR target genes (defined by differential expression in TCDD-treated mouse liver at 2 hours vs. DMSO, fold change ≥1.5, FDR < 5%). Among these, 22 genes were upregulated in response to TCDD exposure. These differentially upregulated genes with 5 or more NC-XRE motifs are shown in Fig 3A. As validation, the known AHR target gene *Serpine1* (aka *Pai1*) contains 5 NC-XRE motifs in the putative promoter, consistent with results that AHR upregulates *Serpine1* via NC-XRE DNA (Huang and Elferink, 2012). AHR binding was also detected at NC-XRE motifs near genes like *Tiparp*, another known AHR target gene induced by TCDD (Ma et al. 2001) and reported to suppress hepatic gluconeogenesis (Diani-Moore et al. 2010). These findings suggest that NC-XRE motifs may contribute to the regulation of *Tiparp* expression, pointing to a potential novel mechanism underlying TCDD-induced hepatotoxicity.

To further explore the biological relevance of the 22 upregulated AHR target genes containing 5X NC-XRE motifs (50 bp spacers), we performed an over-representation analysis (ORA) to identify significantly related Gene Ontology (GO) biological processes (FDR < 5%). This analysis revealed 14 pathways (Figure 3B). However, many of these pathways are broad and involve large gene sets—for example, the GO term "embryo development" includes over 1,000 genes. To contextualize these findings, we conducted a gene set enrichment analysis (GSEA) using a global ranked list of genes based on differential expression in mouse liver following 2-hour TCDD treatment compared to DMSO. This approach allowed us to assess whether the identified pathways were enriched across the entire transcriptome. 13 out of the 14 pathways identified in the ORA were also significantly enriched in the GSEA (FDR < 25%) (Figure 3C).

The enriched pathways span a range of biological functions, including development, morphogenesis, oxygen response, and enzymatic activity—processes that align with known physiological roles of AHR. These findings reveal a compelling link between AHR-NC-XRE target gene expression and key biological pathways, suggesting that AHR may regulate these biological processes, at least in part, through NC-XRE-mediated transcriptional control.

Putative transcription factors regulated via NC-XRE motifs

Transcription factors bind regulatory DNA in a complex composed of multiple different transcription factors. For example, AHR is associated with the transcription factor estrogen receptor alpha at regulatory DNA (Beischlag and Perdew, 2005; Klinge et al., 2000; Madak-Erdogan and Katzenellenbogen, 2012). Another transcription factor, ARNT, is thought to be an obligate co-regulator of AHR (Lo and Matthews, 2012; Reyes et al., 1992).

Previous studies have focused on transcription factor binding at XRE motifs. To identify transcription factors that might form a complex with AHR at NC-XRE motifs, we identified known transcription factor DNA binding sites flanking NC-XRE motifs. We focused on AHR ChIP peaks containing at least 2 NC-XRE motifs within 25 basepairs of each other. The top 12 most frequently occurring known transcription factor DNA binding sites are shown in Figure 4A.

As a complementary approach to identify transcription factors that might form a complex with AHR at NC-XRE motifs, we took the 22 upregulated genes mentioned in Figure 3A and asked whether any combination of these genes is known to be targets of additional transcription factors, besides AHR. We queried the CollecTRI database—a curated resource of transcriptional regulators and their experimentally validated targets, commonly used to infer regulon activity (Müller-Dott et al. 2023). A regulon is defined as a group of genes regulated by a transcription factor through direct binding to specific regulatory DNA motifs. This analysis revealed 20 significantly overrepresented regulons (FDR < 5%), indicating that multiple transcription factors may form a transcription complex with AHR at NC-XRE DNA (Fig 4B).

To further investigate potential transcription factors that might form a complex with AHR at NC-XRE motifs, we searched for transcription factors that physically bind to the same genomic regions as AHR. We focused on AHR ChIP-seq peaks containing five NC-XRE motifs spaced by 50 base pairs or less. Using the CistromeDB Toolkit, we calculated GIGGLE scores to identify ChIP-seq datasets from the Cistrome Data Browser that significantly overlap with AHR-NC-XRE peak regions (Fig 4C, 4D, Supplementary Fig 4C) (Zheng et al. 2019; Mei et al. 2017). The GIGGLE score integrates statistical significance (via Fisher's exact test) and enrichment (odds ratio) to determine how likely the observed overlap is beyond chance (Layer et al. 2018). A higher GIGGLE score between two different ChIP-seq datasets (targeting different proteins for the pull down) indicates a strong and statistically meaningful overlap in their peak regions, suggesting that the two different proteins of interest tend to bind in similar genomic locations.

As expected, GIGGLE scores confirmed strong overlap between our AHR-NC-XRE ChIP peak set and existing AHR ChIP-seq datasets, including those originally used to define the 5X NC-XRE peaks. This served as an internal control, validating the sensitivity of our overlap analysis (Figure 4C; Supplementary Figure 4B).

Beyond AHR itself, we identified 23 transcription factors with ChIP-seq peaks from mouse liver or from cultured mouse hepatocytes that significantly overlapped with the 5X NC-XRE (50 bp spacer) AHR peaks (Figure 4D). This suggests that these transcription factors may co-occupy NC-XRE-rich DNA regions alongside AHR. Additionally, we observed significant overlap with several transcription factors identified through ORA and motif enrichment analyses in ChIP-seq datasets derived from non-liver mouse tissues (Supplementary Figure 4C), pointing to broader regulatory interactions that may extend beyond liver-specific contexts.

We found several motifs enriched in peaks containing both XRE and NC-XRE motifs. As expected, the ARNT motif was enriched in both, supporting the previously known association with XRE motifs and suggesting that the AHR-ARNT complex could be associated with NC-XRE motifs (Fig 4A). However, ARNT was not enriched in our regulon ORA involving AHR peaks containing 5X NC-XRE motifs. This observation aligns with the findings of Huang and Elferink, who originally observed AHR bound to the NC-XRE motif in the promoter region of *Serpine1*, but did not detect ARNT (Huang and Elferink, 2012). This might indicate that AHR can function independently of ARNT at NC-XRE binding sites.

This analysis also provides important insights into previously unrecognized AHR-mediated regulatory pathways. AHR is known to function through activation complexes involving various coregulators, and our findings suggest that several transcription factors—previously not linked to AHR—may co-occupy AHR-bound genomic regions. Among these, ARNTL (also known as BMAL1), a basic helix-loop-helix (bHLH) transcription factor like AHR, emerged as a particularly intriguing candidate. Although no direct interaction between AHR and ARNTL has been reported to date, ARNTL is well known for forming heterodimers with CLOCK to regulate circadian rhythm genes such as *Period* (*Per1, Per2, Per3*) and *Cryptochrome* (*Cry1, Cry2*) (Takumi, Matsubara, et al. 1998; Takumi, Taguchi, et al. 1998; Kume et al. 1999; Hogenesch et al. 1998). We observed co-occupancy of AHR ChIP-seq peaks containing 5X NC-XRE motifs with both ARNTL and CLOCK binding sites (Fig 4D). We also saw that two ARNTL target genes, *Noct* and *Serpine1*, are significantly overrepresented in the CollecTRI ORA analysis. These two target genes are also upregulated in mouse liver following TCDD exposure (Fig 3A, Fig 4B). Moreover, these AHR-5x NC-XRE peak regions also overlapped with ChIP-seq peaks for *Per2*, *Cry1*, and *Cry2*, suggesting that AHR may be involved in regulating components of the circadian clock. These findings point to a novel role for AHR in

circadian biology—potentially mediated through non-canonical DNA binding at NC-XRE motifs—and open new avenues for understanding how environmental signals might intersect with circadian regulation.

Several transcription factor motifs were specifically enriched in the regions flanking NC-XREs, but not in regions flanking canonical XREs. Among these, binding sites for TGIF1 and SMAD3 were particularly prominent (Figure 4A). Consistent with this, multiple targets of SMAD family transcription factors were upregulated following TCDD exposure in mouse liver (Figures 3A, 4A). Furthermore, both SMADs and TGIF1 showed significant co-occupancy with AHR-NC-XRE peaks (Supplementary Figure 4C). These findings point to a potential crosstalk between AHR and other known signaling pathways, in this particular case, the TGF-β/SMAD signaling pathway at NC-XRE-enriched regulatory regions. Notably, such crosstalk has not been previously reported, highlighting its potential as a novel and meaningful observation. It suggests that AHR may integrate environmental signals with developmental and homeostatic pathways through non-canonical DNA binding mechanisms, expanding the scope of AHR's regulatory network.

Our analysis also found many other transcription factors working together with AHR at NC-XRE sites, forming a larger network than previously known. For example, EP300, which is already known to interact with the AHR:ARNT complex (Kobayashi et al. 1997), showed strong peak overlap with AHR at 5x NC-XRE sites. We also saw overlap with other known co-regulators like E2F and HNF4A (Marlowe et al. 2004; Cholico et al. 2022). Although these proteins were already linked to AHR, their potential binding the at NC-XRE-rich regions has not been studied. Even more interesting, we found transcription factors, which have not been previously linked to AHR as a potential coregulator. These include factors involved in oncogenic signaling (e.g., MYC), and core transcriptional machinery (e.g., POLR2A, POLR2B). Their presence at the AHR-NC-XRE sites suggests that AHR may regulate cell cycle, cancer, and RNA production using a previously uncharacterized non-canonical pathway.

Testing NC-XRE function in a natural genomic context

Our bioinformatics analysis supports the hypothesis that AHR regulates target gene expression via NC-XRE DNA. To functionally test this hypothesis at a specific AHR target gene, we focused on *Serpine1*. ChIP studies in mouse liver identified an AHR binding site ~150 bp upstream of the transcription start site, which contains a run of 5 NC-XRE motifs (Fig 4A) (Huang and Elferink, 2012). Indirect evidence suggests that AHR upregulates *Serpine1* expression via NC-XRE motifs, but this has never been tested directly.

We cultured mouse Hepa1-6 liver cells, used CRISPR-Cas9 to delete the 5x NC-XRE motif in the *Serpine1* promoter, generated clonal mutant cell lines, and directly tested whether the 5x NC-XRE motif contributes to TCDD-dependent *Serpine1* expression. To control for heterogeneity in gene expression between wildtype cells (Westermann et al. 2022; Giuliano et al. 2019), we first derived a population of wild-type Hepa1-6 cells from a single cell (see methods). These clonal wild-type cells were used for all subsequent experiments. We transfected Hepa1-6 cells with Cas9 protein together with a guide RNA targeting the NC-XRE sequence, picked single cells and established multiple cell lines containing different deletion mutations in the target NC-XRE (Fig 4A).

We exposed mutant and wild-type cells to TCDD or vehicle for 2 hours and assayed *Serpine1* expression by RT-qPCR. In all clones, deletion of the NC-XRE reduced upregulation of *Serpine1* in response to TCDD (Fig 4B). To test whether AHR function was normal, we also examined expression of *Cyp1a1*, a prototypical AHR target gene thought to be regulated by XRE motifs. All cell lines exhibited increased *Cyp1a1* expression following TCDD exposure, and there was no difference in mean *Cyp1a1* expression levels between wild-type and any mutant cell line (Fig 4C). Our results suggest that AHR transcription factor function is normal in NC-XRE mutant cell lines. We conclude that AHR upregulates *Serpine1* via this NC-XRE sequence 150 bp upstream of the transcription start site.

DISCUSSION

We performed a genome-wide assessment of NC-XRE motifs and find they are prevalent in regulatory regions associated with AHR target genes. We also provide functional evidence that NC-XRE motifs in the *Serpine1* promoter are necessary for full upregulation of *Serpine1* in response to TCDD. However, we have only a basic understanding of the physiologic relevance of AHR signaling at NC-XRE motifs. Only two target genes (*Serpine1* and *Cdkn1a*) are known to be directly regulated by AHR at NC-XRE DNA, and these interactions were explored in only one cell type, hepatocytes, in adulthood (Huang and Elferink, 2012; Jackson et al., 2014). The degree to which other AHR target genes are regulated by non-consensus DNA sequence, in additional tissues and cell types and at different stages of organismal development, is not known.

To address this, we performed integrative pathway and transcription factor analyses of AHR target genes associated with 5x NC-XRE clusters. We identified 22 genes upregulated by TCDD that are proximal to these clusters, and pathway enrichment analyses revealed significant associations with biological processes such as development, morphogenesis, oxygen response, and enzymatic activity. These findings suggest that AHR may regulate a broader array of physiological processes through NC-XRE-mediated transcriptional control than previously appreciated.

ARNT1 is considered an obligate binding partner of AHR (Reyes et al., 1992), but these studies focused on AHR-ARNT1 interactions at XRE DNA. Whether ARNT1 binds AHR at NC-XRE DNA is not known. Studies suggest that under certain conditions AHR can act independently of ARNT1 to upregulate gene expression. In mouse liver, AHR was bound to NC-XRE in the promoter of the *Serpine1* gene following TCDD exposure, but no ARNT1 was detected (Huang and Elferink, 2012). In a rat hepatoma cell line containing a Serpine1:luciferase reporter construct, TCDD increased luciferase levels but knockdown of ARNT1 had no effect (Huang and Elferink, 2012). Our analysis supports this observation. Although ARNT motifs were enriched in AHR peaks containing both XRE and NC-XRE motifs, ARNT did not show co-occupancy at AHR peaks containing 5x NC-XRE motifs. This raises the possibility that AHR may function independently of ARNT at NC-XRE sites, potentially through interactions with alternative co-regulators.

A limitation of the previous studies is that none considered ARNT2. ARNT2 shares significant sequence homology with ARNT1 and was shown to interact with AHR and XRE DNA *in vitro* (Hirose et al., 1996; Rowatt et al., 2003; Tanguay et al., 2000), but to our knowledge there is absence of evidence whether ARNT2 interacts with AHR *in vivo*. DNA binding, coregulator recruitment and target gene expression could be different between AHR-ARNT1 and AHR-ARNT2 transcriptional complexes. Given our findings that ARNT1 does not

appear to co-occupy AHR-bound NC-XRE regions, it is plausible that ARNT2 may serve as an alternative co-regulator in these contexts. Notably, we were unable to assess ARNT2 co-localization at AHR-NC-XRE peaks due to the absence of publicly available ARNT2 ChIP-seq datasets. As such, the potential involvement of ARNT2 in AHR-NC-XRE-mediated transcription remains an open question that warrants further study. This raises the intriguing possibility that AHR may form distinct transcriptional complexes depending on the DNA motif it binds—partnering with ARNT1 at XREs and potentially with ARNT2 at NC-XREs. Such differential dimerization could influence target gene specificity, chromatin recruitment, and transcriptional outcomes, warranting deeper investigation into the role of ARNT2 in non-canonical AHR signaling.

Our integrative analysis uncovered a broader network of transcriptional regulators co-localizing at AHR-NC-XRE sites, including factors involved in circadian rhythm (e.g., Arntl/Bmal1, Clock), TGF-β/SMAD signaling (e.g., SMAD3, TGIF1), and oncogenic or epigenetic regulation (e.g., MYC, EP300, BCL6). These findings suggest that AHR may orchestrate complex transcriptional programs through non-canonical DNA binding, extending its influence into diverse biological pathways.

We appreciate that DNA regulatory elements may be incompletely conserved between species. Showing that AHR binds NC-XRE DNA to regulate gene expression in any species is important foundational knowledge. While the specific number or configuration of NC-XRE sequence at a given promoter or enhancer may differ between strains of animals or between species, it is extremely unlikely that AHR regulates gene expression by binding to NC-XRE in mice but not in humans. Considering that AHR binds identical XRE DNA sequence in zebrafish, mice and humans (Andreasen et al., 2002; Denison et al., 1988; Tanguay et al., 1999; Yao and Denison, 1992), we argue that the same is likely true for NC-XRE.

METHODS

Re-analysis of published murine liver ChIP-Seq datasets

ChIP-Seq datasets of AHR in murine livers after 2hrs of TCDD treatment in male (GSE97634) and female (GSE97636) mice were downloaded from NCBI Gene Expression Omnibus (GEO) (Table 1) (Fader et al., 2017; Fader et al., 2019; Nault et al., 2016). Data were trimmed using trim_galore (Krueger 2015), mapped to the mouse genome build UCSC mm10 using bowtie2 (Langmead and Salzberg 2012), and peaks were called using MACS2 using the provided controls (Langmead and Salzberg, 2012; Zhang et al., 2008). Peaks were merged using bedtoools into male, female and male + female merged datasets (Quinlan and Hall, 2010). Each peak was searched for AHR motifs using an in-house script. Each peak was searched for AHR motifs using an in-house script. To test the likelihood of finding AHR motifs anywhere in the genome, random DNA peaks were generated matching the chromosome and size distribution as previously described (Coarfa et al., 2020), then searched for AHR motifs. Enrichment of AHR motifs compared to the random controls was assessed using a Fisher's exact test, with significance achieved at p-value<0.05. Venn diagrams of AHR motifs in the merged AHR peaks were derived using *eulerr* package (J 2024) in R studio (RStudio Team (2019) n.d.) and pie charts were made using using GraphPad Prism version 10.2.2, MacOS, GraphPad Software, Boston, Massachusetts, USA, www.graphpad.com.

To investigate the genomic distribution of AHR binding sites, we used ChromHMM, a chromatin state annotation tool (Ernst and Kellis 2012). We utilized ChIP-Seq datasets for eight histone modifications—H3K27ac, H3K27me3, H3K36me3, H3K79me2, H3K9ac, H3K4me3, H3K4me1, and H3K9me3—along with an input control, all derived from 8-week-old mouse liver tissue. These datasets were downloaded from the NCBI Gene Expression Omnibus (GEO; accession numbers: GSM1000140, GSM918718, GSM1000150, GSM1000151, GSM1000152, GSM1000153, GSM769014, GSM769015, GSM751035) (Table 1) (Mouse ENCODE Consortium et al. 2012; Tennant et al. 2013). Raw reads were trimmed using *trim_galore* (Krueger 2015) and aligned to the UCSC mm10 mouse genome build using *bowtie2* (Langmead and Salzberg 2012), The aligned reads were converted into binned signal representations alongside genomic windows of signal abundance, which were then used to train a hidden Markov model to segment the genome into 10 chromatin states (Ernst and Kellis 2017) using the ChromHMM tool (Ernst and Kellis 2012). We overlaid these chromatin states onto AHR peaks containing XRE and NC-XRE motifs, as shown in Figures 1C, 1D, and 2D, using a custom in-house script. Additionally, we used the same histone ChIP-Seq datasets to compute signal distributions across genomic regions (Figure 2E) using the HOMER software (Heinz et al. 2010). All figures were generated using GraphPad Prism version 10.2.2.

Runs of AHR motifs, e.g. sequences of successive motifs within a maximum distance from each other, were determined using an in-house Python script and bedtoools. Flanking sequences were determined using the bedtools software (Quinlan and Hall, 2010). Enriched motifs within the flanking sequences were determined using the HOMER software (Heinz et al., 2010).

Re-analysis of published murine liver AHR RNA-Seq datasets and integration with ChIP-seq datasets

Several RNA-seq datasets of mouse liver after treatment with TCDD at different doses and for multiple timepoints were downloaded from NCBI GEO: GSE109863, GSE62902, and GSE87519 (Fader et al., 2017; Nault et al., 2015; Nault et al., 2018) (Table 1). Data was trimmed using trim_galore (Krueger 2015), then mapped onto the mouse genome build UCSC mm10 using STAR (Dobin et al., 2013), and gene expression was quantified using feature_counts (Liao et al., 2014). Differentially expressed genes between TCDD and vehicle treatment were inferred using the EdgeR and RUVr R packages (Risso et al., 2014; Robinson et al., 2010), with significance achieved at fold change exceeding 1.5x and FDR-adjusted p-value<0.05. We used the ranked gene list to perform Gene Set Enrichment Analysis (Subramanian et al. 2005) using datasets from mouse molecular signature database version 7.5.1 (Castanza et al. 2023). Genomic locations (represented as bed files) and gene signatures were integrated using bedtools (Quinlan and Hall 2010). To capture both promoter and proximal enhancer effects, peaks were considered as associated with genes if the gene body was within 10,000 basepairs of a peak. Over Representation Analysis (ORA) was performed using a custom Python script developed by Rajapakshe and Coarfa (version 4.0), which applies the hypergeometric test to assess enrichment of gene sets from the gene ontology biological process gene set version 7.5.1 (Castanza et al. 2023) and CollecTRI version 2 regulons (Müller-Dott et al. 2023).

Cell culture

The mouse hepatoma cell line Hepa 1–6 was obtained from ATCC (Manassas, VA) (Darlington et al., 1980). Hepa 1-6 cells cells were cultured in 10 cm plates in Dulbecco's Modified Eagle's medium (DMEM) supplemented with 1% (v/v) Penicillin/Streptomycin and 10% (v/v) FBS in a humidified incubator (5% CO_2, 37°C). To derive wild-type clones for CRISPR mutagenesis, Hepa 1-6 cells were separated using BD FACSAria II cell sorter (Cytometry and Cell Sorting Core, Baylor College of Medicine) and one cell per well was plated into a 96-well plate. Cells grew for 2-3 weeks until colonies were visible and selected for further expansion. Wild-type clone 12 was used for subsequent studies.

CRISPR mutagenesis of Hepa 1-6 cells

We used CHOPCHOP (Labun et al., 2019) to design guide RNAs targeting the 5x NC-XRE motif region 150bp upstream of transcription start in the mouse Serpine1 gene (target sequence CAGCAAGTCACTGGGAGGGAGGG, PAM motif underlined). Synthetic single guide RNA—modified with 2'-O-methyl at 3 first and last bases, and 3'-phosphorothioate bonds between first 3 and last 2 bases—and purified SpCas9-2NLS protein was purchased from Synthego (Redwood City, CA). We mixed guide RNA and Cas9 (1.3:1 ratio) to form ribonucleoproteins according to Synthego's recommended protocol. Guide RNA and Cas9 ribonucleoprotein complexes were transfected into Hepa1-6 WT clone 12 cells (see above) using Lipofectamine™ CRISPRMAX™ Cas9 Transfection reagent (Thermo Fisher Scientific, Waltham, MA) according to the manufacturer's protocol. 100,000 Hepa-1-6 WT-12 cells were mixed with the ribonucleoprotein-transfection solution and split into two wells for genomic analysis and clonal expansion. Cells were incubated in a humidified incubator (5% CO_2, 37 °C) for 24 hours. We then changed the media and allowed the cells to incubate for another 3 days before genotyping or deriving single-cell colonies for clonal expansion.

To derive mutant colonies from single cells, we separated and plated transfected WT-12 cells using BD FACSAria II cell sorter as above. 48 single cell clones were expanded and genotyped, focusing on mutant cell lines from clones 2, 9, 10, 21, and 38.

Genotyping cell lines

Genomic DNA was extracted from cells using phenol-chloroform extraction. Cells were lysed with DNA lysis buffer (10 mM Tris, 200 mM NaCl, 5 mM EDTA, 1% SDS, 0.4 mg/ml proteinase K). Equal volume of phenol:choloroform:isoamyl alcohol was added for DNA extraction into upper phase followed by ethanol-sodium acetate precipitation of DNA. We used PCR to amplify the region flanking the NC-XRE motifs using primers 5'- AAGCCAGGCCAACTTTTCCT and 5'- CGGTCCTCCTTCACAAAGCT. PCR products were Sanger sequenced to identify mutations. To identify mutations from mosaic cell populations, we used ICE CRISPR Analysis Tool (Conant et al. 2022) to deconvolute Sanger sequencing results.

Mutant clones (all derived from WT clone 12), selected for further experiments, were clone 2 (54 bp deletion in NC-XRE region of Serpine1 promoter), clone 9 (32 bp deletion), clone 10 (8 bp deletion), clone 21 (28 bp deletion) and clone 38 (74 bp deletion). The deletions in each clone were further validated by next generation sequencing. We used PCR to amplify a 450 bp region of DNA encompassing the NC-XRE. After PCR amplification, DNA was quantified by Qubit dsDNA HS fluorometric quantitation (ThermoFisher Q32851). Cytation 5 imaging reader (BioTek; Santa Clara, CA) was used to read the samples. 500 ng of each amplicon

was sent for Amplicon E-Z Next Generation Sequencing (GENEWIZ from Azenta Life Sciences, South Plainfield, NJ), paired end Illumina MiSeq 2X250 bp, 50,000 reads per sample. Each clone contained > 99% of the expected mutation.

Quantitative reverse-transcription PCR

70-80 % confluent Hepa-1-6 parental WT12 cells and WT12 mutant clonal cell lines 2, 9, 10, 21, 38 were treated with 10 nM DMSO (n=3 per treatment group) and TCDD (n=3 per treatment group) for 2 hours each. Cells were then lysed in TRIzol followed by RNA extraction according to the manufacturer's protocol using Direct-zol RNA Miniprep Kit, including the optional on-column DNase digestion (cat no, 11-331; Zymo Research, Irvine, CA). RNA concentration was measured using NanoDrop 2000 Spectrophotometer (Thermo Scientific, Waltham, MA). Purity of the RNA (A260/A280) was ≥ 2 with the yield of 60ng/µl or higher. 1000 ng of the RNA were converted into cDNA using iScript Reverse Transcription Supermix for RT-qPCR (Biorad, Hercules, CA; cat no 1708841) in a total 20 µl volume and incubated for 5 min at 25°C followed by incubation at 46°C for 20 min, at 95°C for 1 min in C1000 Touch Thermal Cycler (Biorad, Hercules, CA).

To validate differential gene expression, we used real-time quantitative PCR (qPCR) by the PrimePCR Assay (Biorad, Hercules, CA) based PCR amplification using the CFX Opus 96 Real-Time PCR System (Biorad, Hercules, CA) via the $2^{-\Delta\Delta Ct}$ method under default reaction conditions (95°C for 2min; 95°C for 5 sec; 60°C for 30 sec. for 40 cycles). An Aliquot of 4 µl of 1:10 dilution of cDNA was mixed with 10 µL BIORAD SsoAdvanced Universal Probes Supermix (cat no 1725280) and target gene specific PrimePCR Assays (1 µL) and Nuclease free water up to a total volume of 20 µl (three technical replicates were run for each sample). cDNA samples were also used for measurement of Gapdh levels for normalization. Ct values were calculated by the Bio-Rad instrument using the CFX Maestro™ Software version 2.3.

The following PrimePCR Assays were used. Note that BioRad does not provide specific primer/probe sequences. BioRad provides the amplicon sequence with additional base pairs added to the beginning and/or end of the sequence (amplicon context sequence). This is in accordance with the minimum information for the publication of real-time quantitative PCR experiments (MIQE) guidelines (Bustin et al. 2011).

Serpine1, Mouse (qMmuCIP0037358)

Amplicon Context Sequence:

CTGTCGGGTTGTGCCGAACCACAAAGAGAAAGGATCGGTCTATAACCATCTCCG

TGGGGGCCATGCGGGCTGAGATGACAAAGGCTGTGGAGGAAGACGCCACTGTG

CCGCTCTCGTTTACCTCGATCCTGACCTTTTGCAGTGCCTGTGCT

***Gapdh*, Mouse (qMmuCEP0039581)**

TGGGAGTTGCTGTTGAAGTCGCAGGAGACAACCTGGTCCTCAGTGTAGCCCAAGATGCCCTTCAGT
GGGCCCTCAGATGCCTGCTTCACCACCTTCTTGATGTCA

***Cyp1a1*, Mouse (qMmuCEP0041980)**

CAGCCACCTAGATCATGCCTTCCATGTATGGACTTCCAGCCTTCGTGTCAGCCA

CAGAGCTGCTCCTGGCTGTCACCGTATTCTGCCTTGGATTCTGGGTGGTCAGAG

CCACAAGAACCTGGGTTCCCAAAGGCCTGAAGACTCC

The results shown were the average of three or more biologically independent experiments. The cumulative results were statistically analyzed by using Prism version 10 (GraphPad Software, Boston, MA).

ACKNOWLEDGMENTS

We dedicate this manuscript to our dear friend and colleague Professor Cornelis Elferink, who provided invaluable advice and encouragement as we pursued these studies. Professor Elferink, who discovered the association between AHR and NC-XRE DNA, died unexpectedly during the preparation of this manuscript. TDP, SLG and CC were partially supported by The Cancer Prevention Institute of Texas (CPRIT) [RP210227, RP200504], NIH/NCI P30 shared resource grant [CA125123], NIH/NIEHS center grants [P30 ES030285] and [P42 ES027725], and NIH/NIMHD [P50MD015496]. DAG supported by NIEHS R01 ES026337.

REFERENCES

Andreasen, E. A., Hahn, M. E., Heideman, W., Peterson, R. E. and Tanguay, R. L. (2002). The zebrafish (Danio rerio) aryl hydrocarbon receptor type 1 is a novel vertebrate receptor. *Mol. Pharmacol.* **62**, 234–249.

Beischlag, T. V. and Perdew, G. H. (2005). ER alpha-AHR-ARNT protein-protein interactions mediate estradiol-dependent transrepression of dioxin-inducible gene transcription. *J. Biol. Chem.* **280**, 21607–21611.

Bernshausen, T., Jux, B., Esser, C., Abel, J. and Fritsche, E. (2006). 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.

Carleton, J. B., Berrett, K. C. and Gertz, J. (2017). Multiplex Enhancer Interference Reveals Collaborative Control of Gene Regulation by Estrogen Receptor α-Bound Enhancers. *Cell Syst.* **5**, 333-344.e5.

Coarfa, C., Grimm, S. L., Katz, T., Zhang, Y., Jangid, R. K., Walker, C. L., Moorthy, B. and Lingappan, K. (2020). Epigenetic response to hyperoxia in the neonatal lung is sexually dimorphic. *Redox Biol.* **37**, 101718.

Darlington, G. J., Bernhard, H. P., Miller, R. A. and Ruddle, F. H. (1980). Expression of liver phenotypes in cultured mouse hepatoma cells. *J Natl Cancer Inst* **64**, 809–819.

Denison, M. S., Fisher, J. M. and Whitlock, J. P. (1988). Inducible, receptor-dependent protein-DNA interactions at a dioxin-responsive transcriptional enhancer. *Proc Natl Acad Sci USA* **85**, 2528–2532.

Denison, M. S., Soshilov, A. A., He, G., DeGroot, D. E. and Zhao, B. (2011). Exactly the same but different: promiscuity and diversity in the molecular mechanisms of action of the aryl hydrocarbon (dioxin) receptor. *Toxicol. Sci.* **124**, 1–22.

Dere, E., Lo, R., Celius, T., Matthews, J. and Zacharewski, T. R. (2011). Integration of genome-wide computation DRE search, AhR ChIP-chip and gene expression analyses of TCDD-elicited responses in the mouse liver. *BMC Genomics* **12**, 365.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.

Eren, M., Boe, A. E., Murphy, S. B., Place, A. T., Nagpal, V., Morales-Nebreda, L., Urich, D., Quaggin, S. E., Budinger, G. R. S., Mutlu, G. M., et al. (2014). PAI-1-regulated extracellular proteolysis governs senescence and survival in Klotho mice. *Proc Natl Acad Sci USA* **111**, 7090–7095.

Fader, K. A., Nault, R., Kirby, M. P., Markous, G., Matthews, J. and Zacharewski, T. R. (2017). Convergence of hepcidin deficiency, systemic iron overloading, heme accumulation, and REV-ERBα/β activation in aryl hydrocarbon receptor-elicited hepatotoxicity. *Toxicol. Appl. Pharmacol.* **321**, 1–17.

Fader, K. A., Nault, R., Doskey, C. M., Fling, R. R. and Zacharewski, T. R. (2019). 2,3,7,8-Tetrachlorodibenzo-p-dioxin abolishes circadian regulation of hepatic metabolic activity in mice. *Sci. Rep.* **9**, 6514.

Hankinson, O. (2005). Role of coactivators in transcriptional activation by the aryl hydrocarbon receptor. *Arch. Biochem. Biophys.* **433**, 379–386.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. and Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589.

Hirose, K., Morita, M., Ema, M., Mimura, J., Hamada, H., Fujii, H., Saijo, Y., Gotoh, O., Sogawa, K. and Fujii-Kuriyama, Y. (1996). cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix/PAS factor (Arnt2) with close sequence similarity to the aryl hydrocarbon receptor nuclear translocator (Arnt). *Mol. Cell. Biol.* **16**, 1706–1713.

Huang, G. and Elferink, C. J. (2012). A novel nonconsensus xenobiotic response element capable of mediating aryl hydrocarbon receptor-dependent gene expression. *Mol. Pharmacol.* **81**, 338–347.

Isoda, K., Sawada, S., Ayaori, M., Matsuki, T., Horai, R., Kagata, Y., Miyazaki, K., Kusuhara, M., Okazaki, M., Matsubara, O., et al. (2005). Deficiency of interleukin-1 receptor antagonist deteriorates fatty liver and cholesterol metabolism in hypercholesterolemic mice. *J. Biol. Chem.* **280**, 7002–7009.

Jackson, D. P., Li, H., Mitchell, K. A., Joshi, A. D. and Elferink, C. J. (2014). Ah receptor-mediated suppression of liver regeneration through NC-XRE-driven p21Cip1 expression. *Mol. Pharmacol.* **85**, 533–541.

Joshi, A. D., Mustafa, M. G., Lichti, C. F. and Elferink, C. J. (2015). Homocitrullination is a novel histone H1 epigenetic mark dependent on aryl hydrocarbon receptor recruitment of carbamoyl phosphate synthase 1. *J. Biol. Chem.* **290**, 27767–27778.

Klein-Hitpass, L., Kaling, M. and Ryffel, G. U. (1988). Synergism of closely adjacent estrogen-responsive elements increases their regulatory potential. *J. Mol. Biol.* **201**, 537–544.

Klinge, C. M., Kaur, K. and Swanson, H. I. (2000). The aryl hydrocarbon receptor interacts with estrogen receptor alpha and orphan receptors COUP-TFI and ERRalpha1. *Arch. Biochem. Biophys.* **373**, 163–174.

Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H. and Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* **47**, W171–W174.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359.

Liao, Y., Smyth, G. K. and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930.

Liu, X., Huang, T., Li, L., Tang, Y., Tian, Y., Wang, S. and Fan, C. (2015). CYP1B1 deficiency ameliorates obesity and glucose intolerance induced by high fat diet in adult C57BL/6J mice. *Am. J. Transl. Res.* **7**, 761–771.

Lopes, R., Korkmaz, G. and Agami, R. (2016). Applying CRISPR-Cas9 tools to identify and characterize transcriptional enhancers. *Nat. Rev. Mol. Cell Biol.* **17**, 597–604.

Lo, R. and Matthews, J. (2012). High-resolution genome-wide mapping of AHR and ARNT binding sites by ChIP-Seq. *Toxicol. Sci.* **130**, 349–361.

Madak-Erdogan, Z. and Katzenellenbogen, B. S. (2012). Aryl hydrocarbon receptor modulation of estrogen receptor α-mediated gene regulation by a multimeric chromatin complex involving the two receptors and the coregulator RIP140. *Toxicol. Sci.* **125**, 401–411.

Nault, R., Fader, K. A. and Zacharewski, T. (2015). RNA-Seq versus oligonucleotide array assessment of dose-dependent TCDD-elicited hepatic gene expression in mice. *BMC Genomics* **16**, 373.

Nault, R., Fader, K. A., Kirby, M. P., Ahmed, S., Matthews, J., Jones, A. D., Lunt, S. Y. and Zacharewski, T. R. (2016). Pyruvate Kinase Isoform Switching and Hepatic Metabolic Reprogramming by the Environmental Contaminant 2,3,7,8-Tetrachlorodibenzo-p-Dioxin. *Toxicol. Sci.* **149**, 358–371.

Nault, R., Doskey, C. M., Fader, K. A., Rockwell, C. E. and Zacharewski, T. (2018). Comparison of Hepatic NRF2 and Aryl Hydrocarbon Receptor Binding in 2,3,7,8-Tetrachlorodibenzo-p-dioxin-Treated Mice Demonstrates NRF2-Independent PKM2 Induction. *Mol. Pharmacol.* **94**, 876–884.

Niu, B., He, K., Li, P., Gong, J., Zhu, X., Ye, S., Ou, Z. and Ren, G. (2018). SIRT1 upregulation protects against liver injury induced by a HFD through inhibiting CD36 and the NF-κB pathway in mouse kupffer cells. *Mol. Med. Report.* **18**, 1609–1615.

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842.

Reyes, H., Reisz-Porszasz, S. and Hankinson, O. (1992). Identification of the Ah receptor nuclear translocator protein (Arnt) as a component of the DNA binding form of the Ah receptor. *Science* **256**, 1193–1195.

Risso, D., Ngai, J., Speed, T. P. and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902.

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140.

Rowatt, A. J., DePowell, J. J. and Powell, W. H. (2003). ARNT gene multiplicity in amphibians: characterization of ARNT2 from the frog Xenopus laevis. *J. Exp. Zool. B Mol. Dev. Evol.* **300**, 48–57.

Tanguay, R. L., Abnet, C. C., Heideman, W. and Peterson, R. E. (1999). Cloning and characterization of the zebrafish (Danio rerio) aryl hydrocarbon receptor. *Biochim. Biophys. Acta* **1444**, 35–48.

Tanguay, R. L., Andreasen, E., Heideman, W. and Peterson, R. E. (2000). Identification and expression of alternatively spliced aryl hydrocarbon nuclear translocator 2 (ARNT2) cDNAs from zebrafish with distinct functions. *Biochim. Biophys. Acta* **1494**, 117–128.

Turner, N., Lim, X. Y., Toop, H. D., Osborne, B., Brandon, A. E., Taylor, E. N., Fiveash, C. E., Govindaraju, H., Teo, J. D., McEwen, H. P., et al. (2018). A selective inhibitor of ceramide synthase 1 reveals a novel role in fat metabolism. *Nat. Commun.* **9**, 3165.

Wilson, S. R., Joshi, A. D. and Elferink, C. J. (2013). The tumor suppressor Kruppel-like factor 6 is a novel aryl hydrocarbon receptor DNA binding partner. *J. Pharmacol. Exp. Ther.* **345**, 419–429.

Yao, E. F. and Denison, M. S. (1992). DNA sequence determinants for binding of transformed Ah receptor to a dioxin-responsive enhancer. *Biochemistry* **31**, 5060–5067.

Yeager, R. L., Reisman, S. A., Aleksunes, L. M. and Klaassen, C. D. (2009). Introducing the "TCDD-inducible AhR-Nrf2 gene battery". *Toxicol. Sci.* **111**, 238–246.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B. D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–364.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137.

CAPTIONS

Figure 1. Enrichment and Distribution of NC-XRE, XRE, and RelBAhRE in AHR ChIP-seq Peaks. Following TCDD exposure, we analyzed mouse liver AHR ChIP-seq peaks from 5 male and 5 female mice for the presence of known AHR motifs: canonical xenobiotic response element (XRE), non-canonical xenobiotic response element (NC-XRE), and RelB AHR response element (RelBAHRE). (A) Likelihood of finding XRE, NC-XRE, or RelBAHRE motifs in AHR ChIP peaks versus in random genomic regions. The table shows the number of AHR ChIP peaks containing at least one XRE, NC-XRE, or RelBAHRE motif. Odds ratios and p-values derived from Fisher's exact test indicate significant enrichment of XRE and NC-XRE motifs in AHR peaks. (B) Proportion of AHR ChIP peaks containing XRE DNA sequence, NC-XRE, XRE + NC-XRE, or none in male and female mice. NC-XRE motifs are more prevalent than XRE motifs in both sexes. (C) Genomic distribution of AHR ChIP-seq peaks mapped to 10 chromatin states defined by ChromHMM using 8 histone modification states. The highest percentages of AHR peaks containing XRE and NC-XRE motifs are found in promoter and enhancer regions. (D) Enlarged view of panel C for detailed comparison of chromatin state distributions. This panel highlights the differences in the distribution of AHR peaks in various chromatin states. XRE and NC-XRE motifs are slightly more prominent in promoters and enhancers in female mice, whereas in male mice, these motifs are found at higher percentages in gene bodies and repressed DNA regions.

Figure 2. Analysis of NC-XRE Repeats and Their Genomic Context. (A) We categorized NC-XRE DNA motifs based on the number of repeated NC-XRE motifs and the distance separating those motifs (spacers) from each other. For example, 5 motifs separated by 50 basepairs (bp) would indicate 5 NC-XRE motifs where each spacer is up to 50 bp long. (B) Number of putative AHR target genes (from RNA-seq data) within 10 kb of AHR ChIP-seq peaks containing either 5 repeats of NC-XRE or XRE motifs separated by spacers ranging from 1 to 100 basepairs (bp). The size of the circles represents the number of AHR target genes. Highlighted circles denote that there are 9-fold more AHR target genes associated with AHR ChIP peaks containing 5 NC-XRE motifs versus 5 XRE motifs separated by 50 bp or less. (C) Frequency of repeated runs of NC-XRE motifs in AHR ChIP peaks. It shows the total number of AHR ChIP peaks containing different numbers of NC-XRE repeats (1-15) with a maximum of 50 bp spacer between any two motifs. (D) Genomic distribution of AHR ChIP-seq peaks containing 5 XRE or NC-XRE motifs separated by 50 bp or less. These are mapped to 10 chromatin states defined by ChromHMM using 8 histone modification states. The highest percentages of AHR peaks containing XRE and NC-XRE motifs are found in promoter and enhancer regions. NC-XRE motifs are prevalent in enhancers and promoters compared to XRE motifs but are absent in gene bodies (transcription, weak transcription, enhancer in gene). (E) Binding strength of indicated histone modifiers at AHR ChIP peaks containing 5 or more NC-XRE or XRE motifs separated by 50 bp or less. AHR-NC-XRE peaks are more associated with activating histone marks (H3K27ac, H3K9ac, H3K4me3, H3K4me1).

Figure 3. Analysis of AHR target genes. (A) Upregulated AHR target genes (fold change TCDD vs vehicle, 2-hour exposure, mouse liver) containing AHR ChIP peaks with 5 or more NC-XRE motifs separated by 50 base pairs or less. AHR target genes are defined as differentially expressed genes (RNA-seq, fold difference in TCDD vs vehicle > 1.5x, FDR < 5%) containing AHR ChIP peaks within 10 kb of a gene body. (B) Over-Representation Analysis (ORA) of Gene Ontology Biological Processes (GOBP) using the AHR target genes from panel A. Enriched pathways were ranked by their statistical significance (P-value) along with the number of hit genes and their names. For example, the embryo development GOBP pathway contains 8 upregulated genes (Atoh8, Notch1, Mafb, Lrp4, Maff, Hes1, Tsgz1, Sebox) that have 5 repeats of NC-XRE motifs with a

maximum of 50 bp spacers between them. (C) Gene Set Enrichment Analysis (GSEA) displaying normalized enrichment scores (NES) for mouse liver global RNA-seq (TCDD vs DMSO) ranked genes. Only the pathways identified in the ORA (panel B) are shown here. All except the 'response to oxygen-containing compound' pathway are enriched in the TCDD-exposed mouse liver.

Figure 4. Putative transcription factors acting with AHR at NC-XRE DNA. (A) Top 12 most frequently occurring known transcription factor motifs in the 50 basepairs flanking repeated NC-XRE and XRE motifs (2 or more motifs within 25 basepairs of each other) in genomic regions where AHR binds. Highlighted motifs are enriched flanking NC-XRE but not XRE. Significance defined as $p<10^{-11}$. (B) Over-Representation Analysis (ORA) using the CollecTRI comprehensive gene regulatory network to identify transcription factors known to upregulate AHR targets (described in Figure 3A), where the AHR target genes are associated with 5 NC-XRE repeats separated by 50 bp or less. Top represented transcription factors ranked by their statistical significance (-log10(P-value)). For example, FOXO4 has two hits (*Serpine1, Igfbp1*), indicating that the *Serpine1* and *Igfbp1* genes are (1) targets of the FOXO4 transcription factor protein, (2) are targets of AHR (RNAseq from TCDD in mouse liver), and (3) contain 5 NC-XRE repeats with 50 bp spacers in AHR ChIP peaks within 10 kb of the gene body (ChIP-seq from TCDD in mouse liver). (C, D) Giggle scores of AHR peaks with 5 or more XRE or NC-XRE repeats with up to 50 bp spacers compared to known transcription factor peaks from mouse liver or liver cell ChIP-seq. Each point represents how strongly XRE or NC-XRE containing AHR binding sites overlap with the known transcription factor binding sites. A higher giggle score indicates a higher likelihood of these transcription factors co-occupying the AHR peaks. For example, DNA sequence to which STAT5B binds is enriched in AHR ChIP peaks containing NC-XRE (5 or more repeats with 50 bp spacers). Positive control (C) shows high enrichment of AHR DNA binding motifs within AHR ChIP peaks.

Figure 5. Mutations in NC-XRE DNA in *Serpine1* promoter reduced TCDD-dependent expression of *Serpine1*. (A) Schematic of mouse *Serpine1* gene showing AHR ChIP peak ~150 bp upstream of transcription start site, which contains 5x repeated NC-XRE (nucleotides in red) and no XRE sequences. Hepa1-6 cell clones containing deletions in the NC-XRE region (dashed lines) are shown below the wild-type sequence. (B, C) Hepa1-6 wildtype or mutant cells were exposed to 10 nM TCDD or vehicle for 2 hours, followed by RNA extraction and qPCR for *Serpine1* (A) or *Cyp1a1* (B). Mutant cell lines with full or partial deletion of NC-XRE exhibit reduced increase in *Serpine1* following TCDD exposure compared to wild-type cells. Mutant cells showed similar increase in *Cyp1a1* following TCDD exposure as wild-type cells. Gene expression normalized to vehicle, 18S rRNA used as reference gene. One-way ANOVA with Tukey's multiple comparisons test, * $p < 0.05$, ** $p < 0.01$, ns, not significant ($p > 0.05$).

Supplementary Figure 1. Distribution of AHR ChIP-seq Peaks and Chromatin States. (A) Pie chart showing the distribution of AHR ChIP-seq peaks from mouse liver containing XRE, NC-XRE, and RelBAhRE motifs, corresponding to data in ***Figure 1B.*** The chart highlights that the proportion of XRE and NC-XRE motifs are similar between the sexes. The number of RelBAhRE motifs is miniscule compared to XRE and NC-XRE motifs. **(B)** Table detailing the genomic distribution of AHR ChIP-seq peaks across various chromatin states, corresponding to ***Figure 1(C-D)***. It includes the number of AHR peaks associated with each chromatin state for different motif variations (XRE/ NC-XRE for male/ female/ combined datasets). Chromatin states include Active TSS, Enhancer, Poised Enhancer, Weak Transcription, Repressed by Polycomb, Heterochromatin, and Quiescent regions.

Supplementary Figure 2. Detailed analysis of NC-XRE repeats and their genomic context. (A) Number of AHR ChIP peaks containing NC-XRE repeats (1-15) separated by 50 bp or less, corresponds to Figure 2C. (B) Number of putative AHR target genes within 10 kb of AHR ChIP-seq peaks containing 5 repeats of XRE/NC-XRE motifs separated by different numbers of spacers, corresponds to Figure 2B. (C) Enlarged view of Figure 2D, for detailed comparison of chromatin state distributions. 5 repeats of XRE or NC-XRE motifs with 50 bp spacers are mostly present near transcription start sites and enhancers. (D) Values for the data shown in (C). (E) Overview of histone modifications, indicating their functions (activation or repression) and typical locations (e.g., transcription start sites).

Supplementary Figure 3. Enrichment analysis of putative AHR-NC-XRE target genes. (A) Table corresponding to the ORA data shown in Figure 3B, including the enriched Gene Ontology Biological Processes (GOBP) gene sets and overlapping genes that have 5 NC-XRE repeats with a maximum of 50 bp spacers within 10 kb of the gene body. **(B)** GSEA enrichment plots of 6 upregulated pathways identified in the ORA analysis. These panels are ranked based on the number of hit genes in the ORA: Embryo development (8), tissue development (8), regulation of multicellular organismal development (7), animal organ morphogenesis (7), regulation of cell differentiation (7), and enzyme-linked receptor protein signaling pathway (6). **(C)** Normalized enrichment scores (NES) of the GOBP gene sets shown in Figure 3D.

Supplementary Figure 4. Putative transcription factors targeting 5 repeats of XRE or NC-XRE separated by 50 bp or less. (A) Corresponds to the ORA data shown in Figure 4B, including the upregulated CollecTRI transcription factors and overlapping genes that have 5 NC-XRE repeats with a maximum of 50 bp spacers within 10 kb of the gene body. (B) Corresponds to the CistromeDB giggle scores shown in Figure 4C, indicating the overlap of AHR peaks containing 5 or more XRE or NC-XRE with 50 bp spacers in our query dataset compared mouse liver AHR ChIP-seq datasets on the CistromeDB database. Higher giggle score means higher similarity. (C) Giggle scores of AHR peaks with 5 or more NC-XRE repeats with up to 50 bp spacers compared to some of the hit transcription factor (Fig 4A, 4B) peaks from mouse non-liver ChIP-seq experiments. Each point represents how strongly NC-XRE containing AHR binding sites overlap with the transcription factor binding sites. A higher giggle score indicates a higher likelihood of these transcription factors co-occupying the AHR peaks.

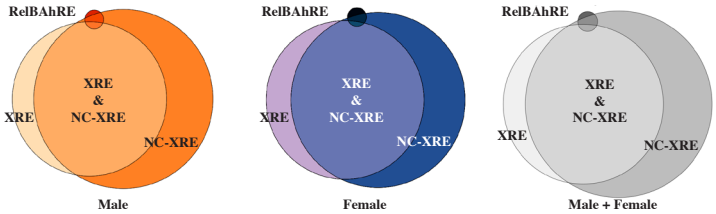
Supplementary Table 1: AHR Target Genes with Different Combinations of XRE or NC-XRE Near Gene Body. Frequency of repeated runs of XRE or NC-XRE motifs within 10000 bp of a gene body. It shows the total number of AHR ChIP peaks containing 5 repeats of XRE or NC-XRE separated by spacers ranging from 1 to 100 basepairs (bp) along with the expression pattern (log2FoldChange) of a particular gene (Fold Change > 1.5x; FDR <0.05).

Supplementary Table 2. Similarity of AHR ChIP-Seq Peaks containing 5 NC-XRE with Known Transcription Factor Peaks. Giggle scores of AHR peaks with 5 or more NC-XRE repeats with up to 50 bp spacers compared to known transcription factor peaks from mouse liver or liver cell ChIP-seq datasets. A higher giggle score indicates a higher likelihood of these transcription factors co-occupying the AHR peaks.

A

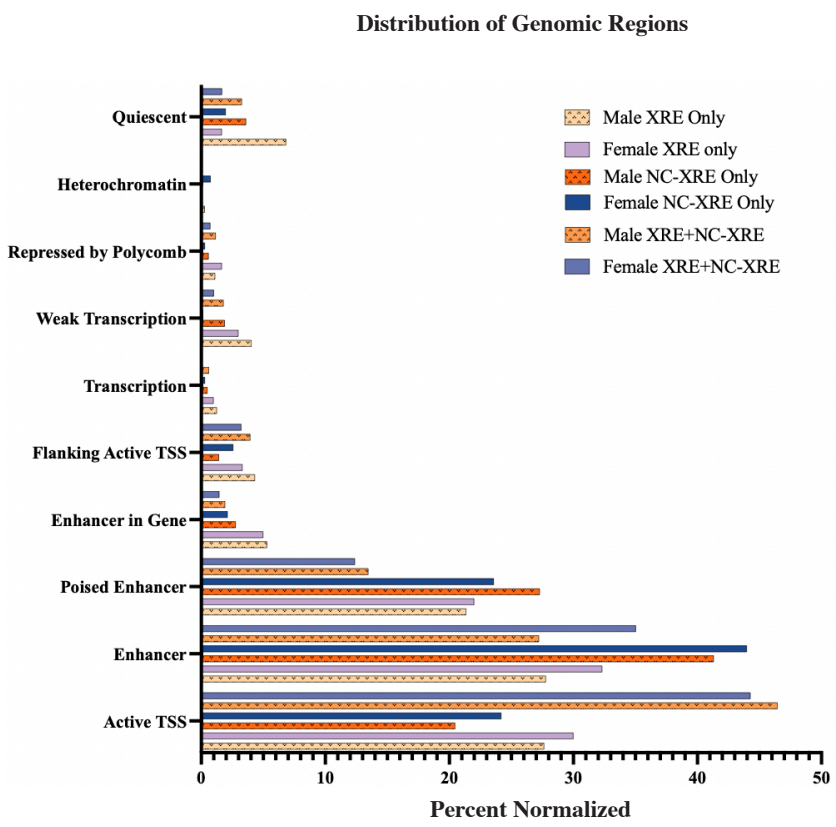
Feature	Sequence	Number in AHR Peaks	Number of Random Peaks	Odds Ratio	P value
XRE	GCGTC	4909	666	16.988	$<10^{-40}$
NC-XRE	GGGA	6664	6021	1.583073	9.53×10^{-34}
RelBAhRE	GGGTGCAT	39	32	1.219784	0.4757227

B



	Number of AHR Peaks					
Feature	Male	Percent Male	Female	Percent Female	Male + Female	Percent Male + Female
XRE Only	714	8.99	297	11.75	719	8.81
NC-XRE Only	2362	29.74	652	25.79	2462	30.18
RelBAHRE Only	14	0.18	9	0.36	16	0.20
XRE + NC-XRE	4108	51.72	1364	53.96	4188	51.34
XRE + RelBAHRE	25	0.31	8	0.32	25	0.31
NC-XRE + RelBAHRE	35	0.44	10	0.40	37	0.45
XRE + NC-XRE + RelBAHRE	23	0.29	5	0.20	23	0.28
Peaks without XRE, NC-XRE or RelBAHRE	746	9.39	207	8.19	773	9.47
Total AHR Peaks	7943	100	2528	100	8158	100

C



D

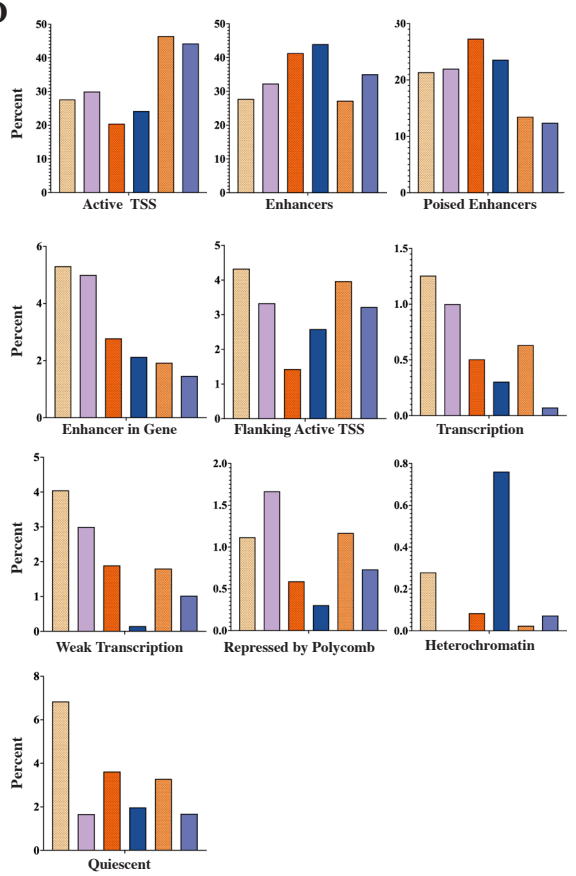
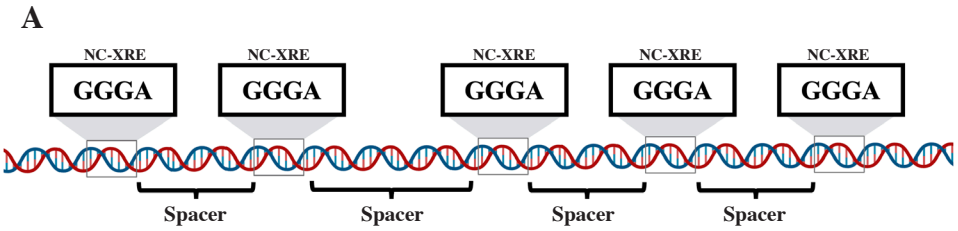
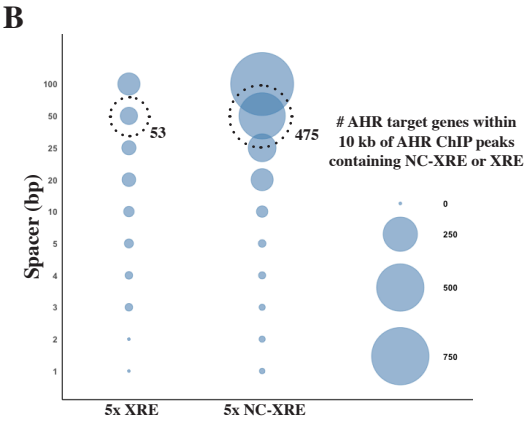
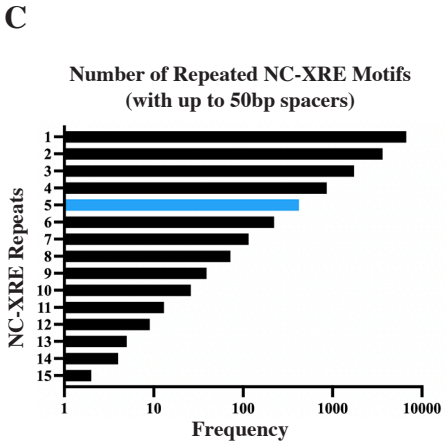
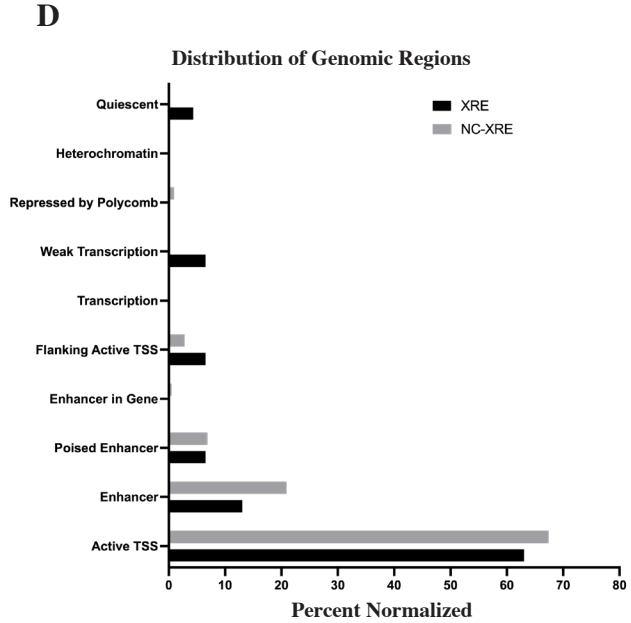


Figure 1. Enrichment and Distribution of NC-XRE, XRE, and RelBAhRE in AHR ChIP-seq Peaks. Following TCDD exposure, we analyzed mouse liver AHR ChIP-seq peaks from 5 male and 5 female mice for the presence of known AHR motifs: canonical xenobiotic response element (XRE), non-canonical xenobiotic response element (NC-XRE), and RelB AHR response element (RelBAHRE). **(A)** Likelihood of finding XRE, NC-XRE, or RelBAHRE motifs in AHR ChIP peaks versus in random genomic regions. The table shows the number of AHR ChIP peaks containing at least one XRE, NC-XRE, or RelBAHRE motif. Odds ratios and p-values derived from Fisher's exact test indicate significant enrichment of XRE and NC-XRE motifs in AHR peaks. **(B)** Proportion of AHR ChIP peaks containing XRE DNA sequence, NC-XRE, XRE + NC-XRE, or none in male and female mice. NC-XRE motifs are more prevalent than XRE motifs in both sexes. **(C)** Genomic distribution of AHR ChIP-seq peaks mapped to 10 chromatin states defined by ChromHMM using 8 histone modification states. The highest percentages of AHR peaks containing XRE and NC-XRE motifs are found in promoter and enhancer regions. **(D)** Enlarged view of panel C for detailed comparison of chromatin state distributions. This panel highlights the differences in the distribution of AHR peaks in various chromatin states. XRE and NC-XRE motifs are slightly more prominent in promoters and enhancers in female mice, whereas in male mice, these motifs are found at higher percentages in gene bodies and repressed DNA regions.









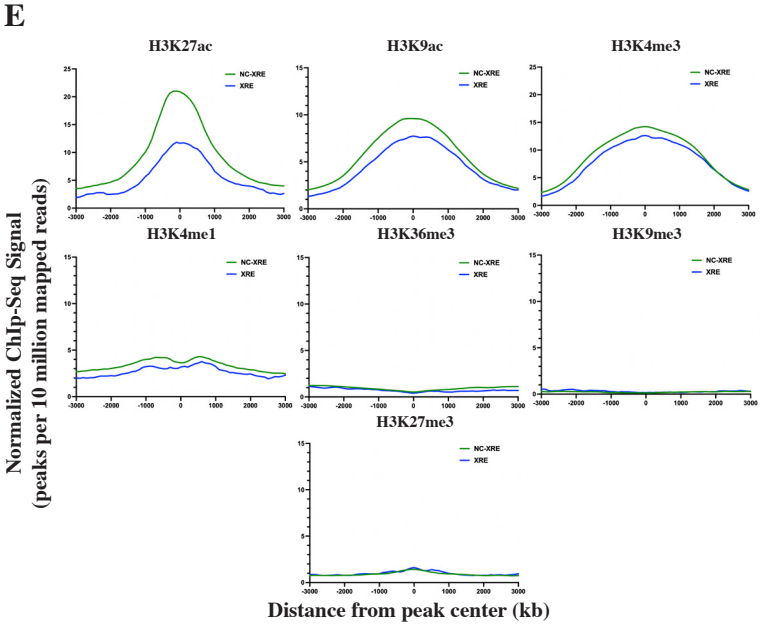
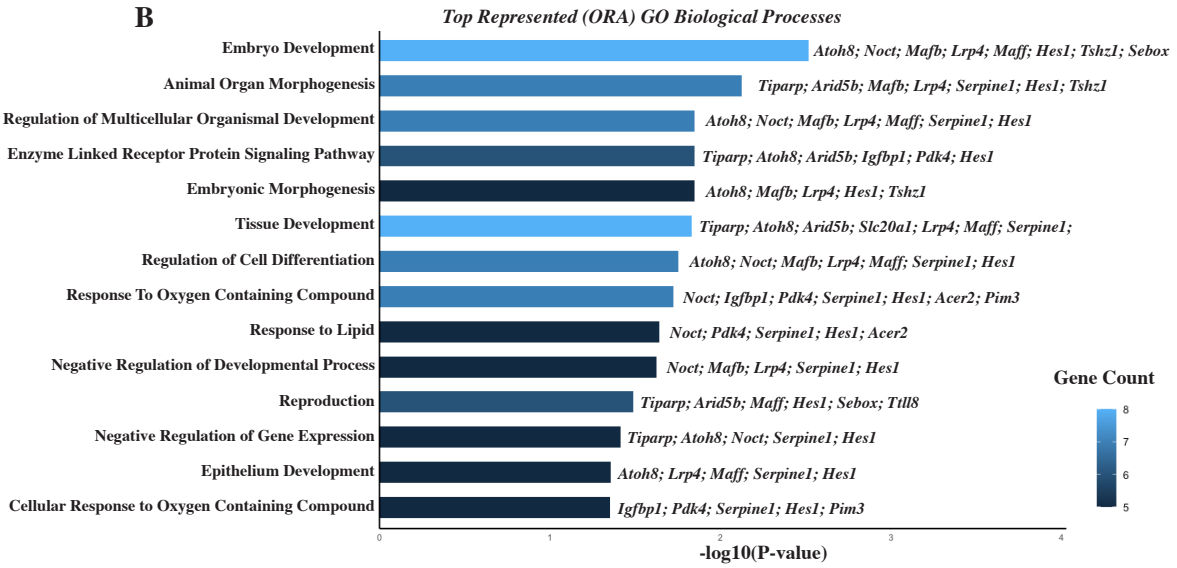


Figure 2. Analysis of NC-XRE Repeats and Their Genomic Context. (A) We categorized NC-XRE DNA motifs based on the number of repeated NC-XRE motifs and the distance separating those motifs (spacers) from each other. For example, 5 motifs separated by 50 basepairs (bp) would indicate 5 NC-XRE motifs where each spacer is up to 50 bp long. **(B)** Number of putative AHR target genes (from RNA-seq data) within 10 kb of AHR ChIP-seq peaks containing either 5 repeats of NC-XRE or XRE motifs separated by spacers ranging from 1 to 100 basepairs (bp). The size of the circles represents the number of AHR target genes. Highlighted circles denote that there are 9-fold more AHR target genes associated with AHR ChIP peaks containing 5 NC-XRE motifs versus 5 XRE motifs separated by 50 bp or less. **(C)** Frequency of repeated runs of NC-XRE motifs in AHR ChIP peaks. It shows the total number of AHR ChIP peaks containing different numbers of NC-XRE repeats (1-15) with a maximum of 50 bp spacer between any two motifs. **(D)** Genomic distribution of AHR ChIP-seq peaks containing 5 XRE or NC-XRE motifs separated by 50 bp or less. These are mapped to 10 chromatin states defined by ChromHMM using 8 histone modification states. The highest percentages of AHR peaks containing XRE and NC-XRE motifs are found in promoter and enhancer regions. NC-XRE motifs are prevalent in enhancers and promoters compared to XRE motifs but are absent in gene bodies (transcription, weak transcription, enhancer in gene). **(E)** Binding strength of indicated histone modifiers at AHR ChIP peaks containing 5 or more NC-XRE or XRE motifs separated by 50 bp or less. AHR-NC-XRE peaks are more associated with activating histone marks (H3K27ac, H3K9ac, H3K4me3, H3K4me1).

A

Gene Name	Fold Change	NC-XRE Location	XRE within 300 bp on an NC-XRE
Serpine1	92.62	promoter	No
Tiparp	22.70	promoter and after exon 1	Yes
Maff	4.92	promoter and between alternatively spliced exon 1	Yes
Ttll8	2.53	near exon 9	No
Sebox	2.39	promoter	No
Igfbp1	2.31	promoter	Yes
Pim3	2.17	promoter, exon1, intron between exon 4 & 5	Yes
Mafb	2.05	promoter	Yes
Lrp4	1.90	between exon 1 & 2	Yes
Gcc1	1.89	between exon 1 & 2	Yes

Gene Name	Fold Change	NC-XRE location	XRE within 300 bp on an NC-XRE
Slc20a1	1.81	promoter	Yes
Fam222a	1.79	promoter, after exon1, after exon 2	Yes (promoter and after exon 1) No (after exon 2)
Acer2	1.73	between exons 3 & 4 and exons 4 & 5	Yes
Pdk4	1.71	promoter	Yes
Iffo2	1.69	after exon 3	Yes
Arid5b	1.66	between exons 3 & 4 and after exon 7	Yes
Atoh8	1.60	between exons 1 & 2 and 2 & 3	No (exons 1 &2), Yes (exon 2 & 3)
Noct	1.60	promoter and between exons 1 & 2	Yes
Rfx7	1.60	promoter	Yes
Hjurp	1.58	promoter and near exon 7	Yes (promoter), No (near exon 7)
Tshz1	1.52	promoter	Yes



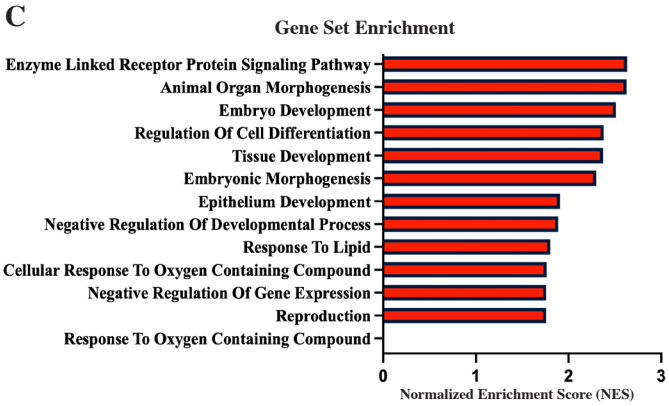
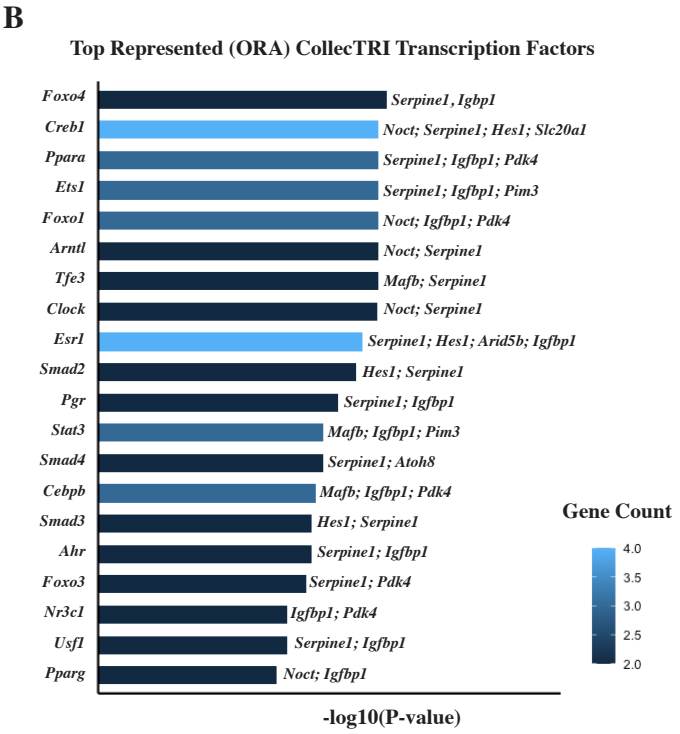
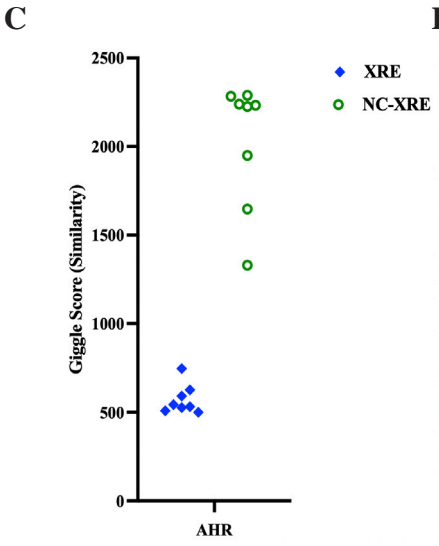


Figure 3. Analysis of AHR Target Genes. (A) Upregulated AHR target genes (fold change TCDD vs vehicle, 2-hour exposure, mouse liver) containing AHR ChIP peaks with 5 or more NC-XRE motifs separated by 50 base pairs or less. AHR target genes are defined as differentially expressed genes (RNA-seq, fold difference in TCDD vs vehicle > 1.5x, FDR < 5%) containing AHR ChIP peaks within 10 kb of a gene body. **(B)** Over-Representation Analysis (ORA) of Gene Ontology Biological Processes (GOBP) using the AHR target genes from panel A. Enriched pathways were ranked by their statistical significance (P-value) along with the number of hit genes and their names. For example, the embryo development GOBP pathway contains 8 upregulated genes (*Atoh8, Notch1, Mafb, Lrp4, Maff, Hes1, Tsgz1, Sebox*) that have 5 repeats of NC-XRE motifs with a maximum of 50 bp spacers between them. **(C)** Gene Set Enrichment Analysis (GSEA) displaying normalized enrichment scores (NES) for mouse liver global RNA-seq (TCDD vs DMSO, 2-hour exposure) ranked genes. Only the pathways identified in the ORA (panel B) are shown here. All except the 'response to oxygen-containing compound' pathway are enriched in the TCDD-exposed mouse liver.

A

Motif name	Motif sequence	Percent of AHR peaks with flanking motif & NC-XRE	Percent of AHR peaks with flanking motif & XRE	NC-XRE -log10(p value)	XRE -log10(p value)
Nr1a2	TRAGGTCA	50.62	46.52	215	71
AR halfsite	CCAGGAACAG	44.61	0	27	not significant
RARa	TTGAMCTTTG	40.3	36.23	418	158
Arnt:Ahr	TBGCACGCAA	36.39	67.08	3704	5093
Tgif2	TGTCANYT	33.88	0	36	not significant
ERRa	CAAAGGTCAG	33.42	30.63	307	112
Smad3	TWGTCTGV	31.28	0	17	not significant
Hif1b	RTACGTGC	30.78	50.69	1036	1566
Tgif1	YTGWCADY	30.33	0	15	not significant
Klf14	RGKGGGCGK-GGC	29.25	32.36	25	39
Hic1	TGCCAGCB	27.12	0	26	not significant
COUP-TFII	AGRGGTCA	26.32	23.42	365	119





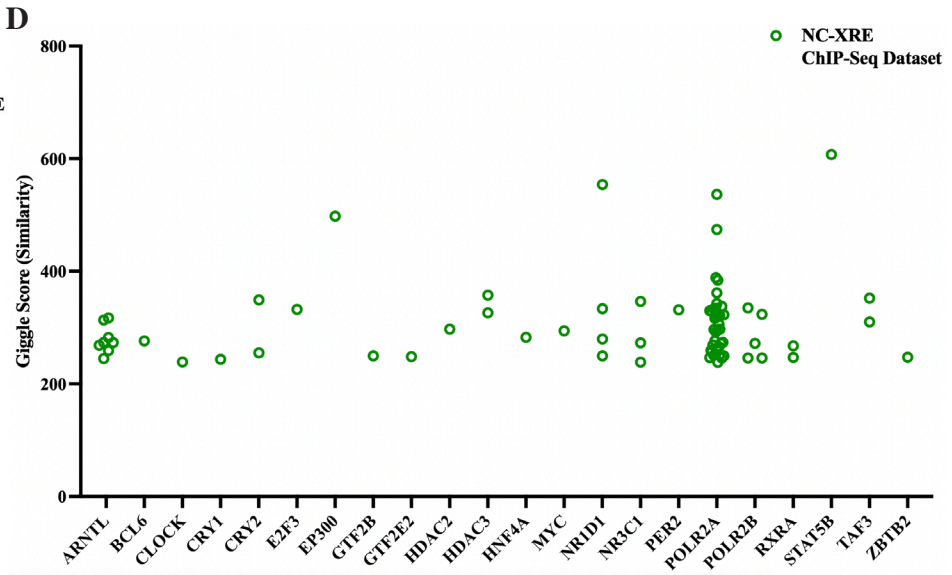


Figure 4. Putative Transcription Factors Acting with AHR at NC-XRE DNA. (A) Top 12 most frequently occurring known transcription factor motifs in the 50 basepairs flanking repeated NC-XRE and XRE motifs (2 or more motifs within 25 basepairs of each other) in genomic regions where AHR binds. Highlighted motifs are enriched flanking NC-XRE but not XRE. Significance defined as $p<10^{-11}$. **(B)** Over-Representation Analysis (ORA) using the CollecTRI comprehensive gene regulatory network to identify transcription factors known to upregulate AHR targets (described in Figure 3A), where the AHR target genes are associated with 5 NC-XRE repeats separated by 50 bp or less. Top represented transcription factors ranked by their statistical significance (-log10(P-value)). For example, FOXO4 has two hits (*Serpine1, Igfbp1*), indicating that the *Serpine1* and *Igfbp1* genes are (1) targets of the FOXO4 transcription factor protein, (2) are targets of AHR (RNAseq from TCDD in mouse liver), and (3) contain 5 NC-XRE repeats with 50 bp spacers in AHR ChIP peaks within 10 kb of the gene body (ChIP-seq from TCDD in mouse liver). **(C, D)** Giggle scores of AHR peaks with 5 or more XRE or NC-XRE repeats with up to 50 bp spacers compared to known transcription factor peaks from mouse liver or liver cell ChIP-seq. Each point represents how strongly XRE or NC-XRE containing AHR binding sites overlap with the known transcription factor binding sites. A higher giggle score indicates a higher likelihood of these transcription factors co-occupying the AHR peaks. For example, DNA sequence to which STAT5B binds is enriched in AHR ChIP peaks containing NC-XRE (5 or more repeats with 50 bp spacers). Positive control (C) shows high enrichment of AHR DNA binding motifs within AHR ChIP peaks.

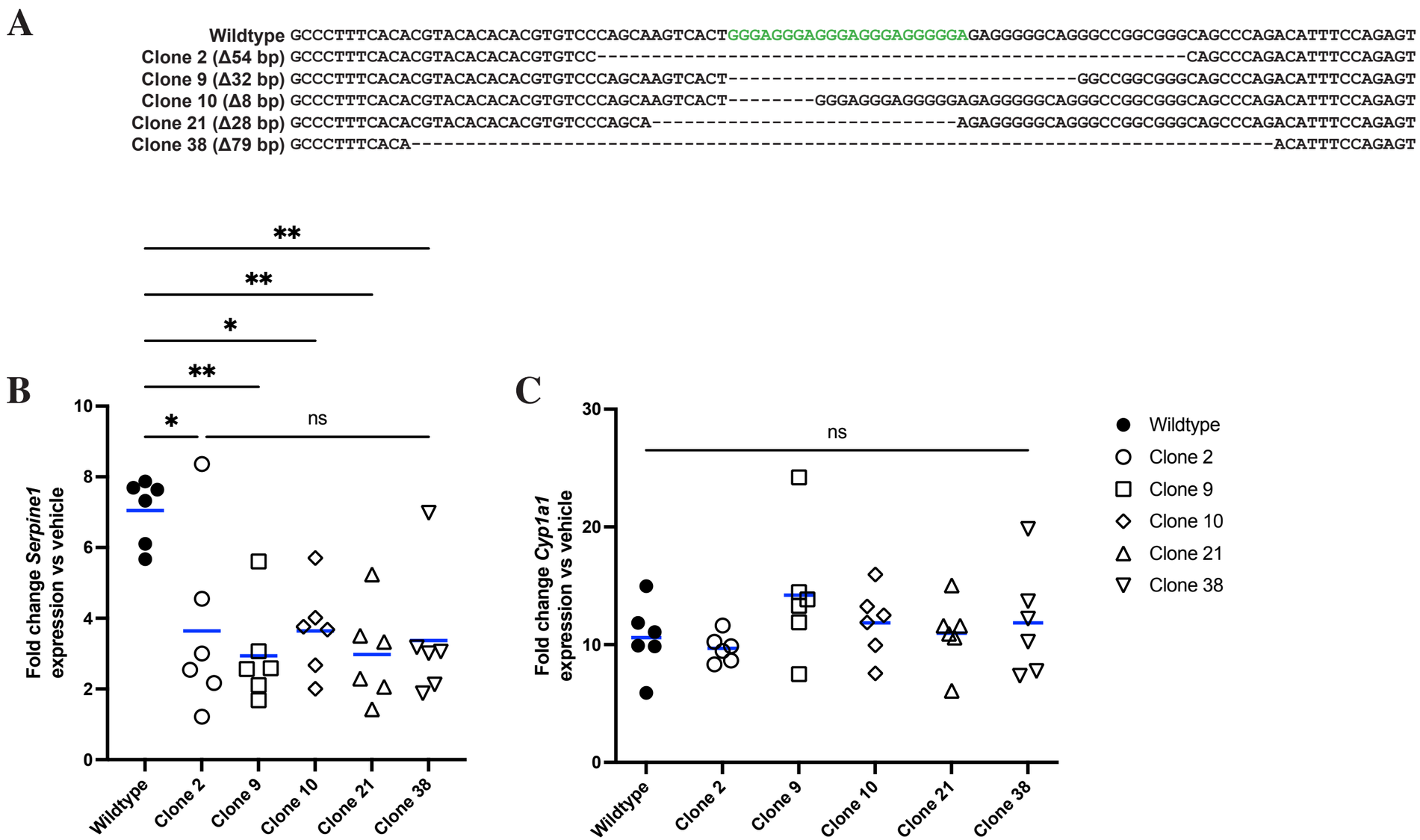
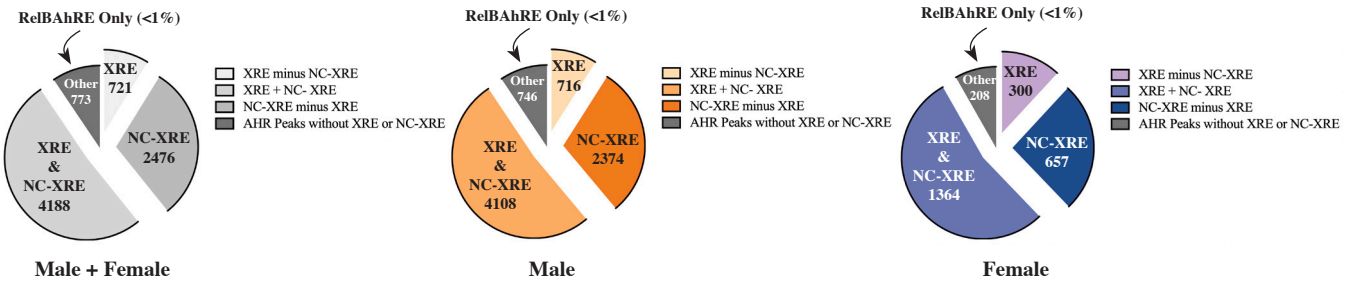


Figure 5. Mutations in NC-XRE DNA in *Serpine1* Promoter Reduced TCDD-Dependent Expression of *Serpine1*. (A) Schematic of mouse *Serpine1* gene showing AHR ChIP peak ~150 bp upstream of transcription start site, which contains 5x repeated NC-XRE (nucleotides in red) and no XRE sequences. Hepa1-6 cell clones containing deletions in the NC-XRE region (dashed lines) are shown below the wild-type sequence. (B, C) Hepa1-6 wildtype or mutant cells were exposed to 10 nM TCDD or vehicle for 2 hours, followed by RNA extraction and qPCR for *Serpine1* (A) or *Cyp1a1* (B). Mutant cell lines with full or partial deletion of NC-XRE exhibit reduced increase in *Serpine1* following TCDD exposure compared to wild-type cells. Mutant cells showed similar increase in *Cyp1a1* following TCDD exposure as wild-type cells. Gene expression normalized to vehicle, 18S rRNA used as reference gene. One-way ANOVA with Tukey's multiple comparisons test, * $p < 0.05$, ** $p < 0.01$, ns, not significant ($p > 0.05$).

A



B

Feature	No. of AHR Peaks					
	XRE (male)	XRE (Female)	NC-XRE (male)	NC-XRE (Female)	XRE + NC-XRE (male)	XRE + NC-XRE (Female)
Active TSS	198	90	486	159	1909	604
Enhancer	199	97	981	286	1119	478
Poised Enhancer	153	66	648	155	554	169
Enhancer in Gene	38	15	66	14	79	20
Flanking Active TSS	31	10	34	17	163	44
Transcription	9	3	12	2	26	1
Weak Transcription	29	9	45	1	74	14
Repressed by Polycomb	8	5	14	2	48	10
Heterochromatin	2	0	2	5	1	1
Quiescent	49	5	86	13	135	23
Total AHR Peaks	716	300	2374	657	4108	1364

Supplementary Figure 1. Distribution of AHR ChIP-seq Peaks and Chromatin States. (A) Pie chart showing the distribution of AHR ChIP-seq peaks from mouse liver containing XRE, NC-XRE, and RelBAhRE motifs, corresponding to data in ***Figure 1B***. The chart highlights that the proportion of XRE and NC-XRE motifs are similar between the sexes. The number of RelBAhRE motifs is miniscule compared to XRE and NC-XRE motifs. **(B)** Table detailing the genomic distribution of AHR ChIP-seq peaks across various chromatin states, corresponding to ***Figure 1(C-D)***. It includes the number of AHR peaks associated with each chromatin state for different motif variations (XRE/ NC-XRE for male/ female/ combined datasets). Chromatin states include Active TSS, Enhancer, Poised Enhancer, Weak Transcription, Repressed by Polycomb, Heterochromatin, and Quiescent regions.

A

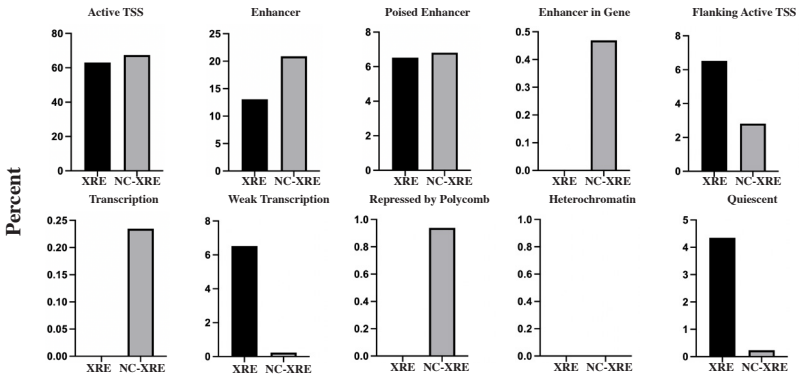
No. of NC-XRE Repeats	No. of AHR Peaks	Percent of Total AHR Peaks
1	6664	81.69
2	3629	44.48
3	1741	21.34
4	864	10.59
5	426	5.22
6	222	2.72
7	115	1.41
8	72	0.88
9	39	0.48
10	26	0.32
11	13	0.16
12	9	0.11
13	5	0.06
14	4	0.05
15	2	0.02

B

Motif Feature	No. of AHR Target Genes	
	XRE	NC-XRE
Repeats >= 5, Spacer <=1bp	0	2
Repeats >= 5, Spacer <=2bp	0	3
Repeats >= 5, Spacer <=3bp	6	3
Repeats >= 5, Spacer <=4bp	6	5
Repeats >= 5, Spacer <=5bp	10	6
Repeats >= 5, Spacer <=10bp	15	19
Repeats >= 5, Spacer <=20bp	30	94
Repeats >= 5, Spacer <=25bp	33	158
Repeats >= 5, Spacer <=50bp	53	475
Repeats >= 5, Spacer <=100bp	95	910

C

Distribution of Genomic Regions (Repeats >= 5, Spacer <=50bp)



D

Feature	No. of AHR Peaks	
	XRE (Repeats >= 5, Spacer <=50bp)	NC-XRE (Repeats >= 5, Spacer <=50bp)
Active TSS	29	287
Enhancer	6	89
Poised Enhancer	3	29
Enhancer in Gene	0	2
Flanking Active TSS	3	12
Transcription	0	1
Weak Transcription	3	1
Repressed by Polycomb	0	4
Heterochromatin	0	0
Quiescent	2	1
Total	46	426

E

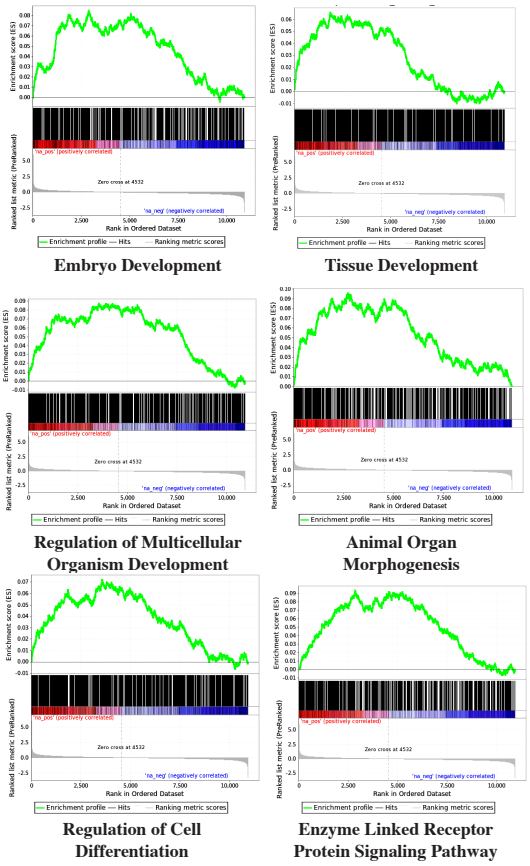
Histone Modification	Function	Location
H3K27ac	Activation	Enhancers, Transcription Start Sites
H3K9ac	Activation	Transcription Start Sites (high), Enhancers (low)
H3K4me3	Activation	Transcription Start Sites, Bivalent Transcription Start Sites
H3K4me1	Activation	Enhancers (high), Transcription Start Sites (low)
H3K36me3	Activation	Gene Bodies (Transcribed Regions/ Enhancer in Gene)
H3K9me3	Repression	Heterochromatin, Satellite Repeats, Telomeres, Pericentromeres
H3K27me3	Repression	Repressed by Polycomb, Bivalent Transcription Start Sites

Supplementary Figure 2. Detailed Analysis of NC-XRE Repeats and Their Genomic Context. **(A)** Number of AHR ChIP peaks containing NC-XRE repeats (1-15) separated by 50 bp or less, corresponds to Figure 2C. **(B)** Number of putative AHR target genes within 10 kb of AHR ChIP-seq peaks containing 5 repeats of XRE or NC-XRE motifs separated by different numbers of spacers, corresponds to Figure 2B. **(C)** Enlarged view of Figure 2D, for detailed comparison of chromatin state distributions. 5 repeats of XRE or NC-XRE motifs with 50 bp spacers are mostly present near transcription start sites and enhancers. **(D)** Values for the data shown in (C). **(E)** Overview of histone modifications, indicating their functions (activation or repression) and typical locations (e.g., transcription start sites).

A

Gene Set	-log10(P-value)	Genes in Gene Set	Total Number of Overlapping Genes	Overlapping Genes
Embryo Development	2.518	1062	8	*Atoh8; Noct; Mafb; Lrp4; Maff; Hes1; Tshz1; Sebox*
Animal Organ Morphogenesis	2.124	1012	7	*Tiparp; Arid5b; Mafb; Lrp4; Serpine1; Hes1; Tshz1*
Regulation Of Multicellular Organismal Development	1.847	1286	7	*Atoh8; Noct; Mafb; Lrp4; Maff; Serpine1; Hes1*
Enzyme Linked Receptor Protein Signaling Pathway	1.847	933	6	*Tiparp; Atoh8; Arid5b; Igfbp1; Pdk4; Hes1*
Embryonic Morphogenesis	1.847	590	5	*Atoh8; Mafb; Lrp4; Hes1; Tshz1*
Tissue Development	1.831	1798	8	*Tiparp; Atoh8; Arid5b; Slc20a1; Lrp4; Maff; Serpine1; Hes1*
Regulation Of Cell Differentiation	1.752	1445	7	*Atoh8; Noct; Mafb; Lrp4; Maff; Serpine1; Hes1*
Response To Oxygen Containing Compound	1.724	1536	7	*Noct; Igfbp1; Pdk4; Serpine1; Hes1; Acer2; Pim3*
Response To Lipid	1.640	831	5	*Noct; Pdk4; Serpine1; Hes1; Acer2*
Negative Regulation Of Developmental Process	1.624	844	5	*Noct; Mafb; Lrp4; Serpine1; Hes1*
Reproduction	1.487	1377	6	*Tiparp; Arid5b; Maff; Hes1; Sebox; Ttll8*
Negative Regulation Of Gene Expression	1.414	1017	5	*Tiparp; Atoh8; Noct; Serpine1; Hes1*
Epithelium Development	1.357	1077	5	*Atoh8; Lrp4; Maff; Serpine1; Hes1*
Cellular Response To Oxygen Containing Compound	1.351	1096	5	*Igfbp1; Pdk4; Serpine1; Hes1; Pim3*

B



C

Gene Set	NES Score
Embryo Development	2.5135262
Animal Organ Morphogenesis	2.6263876
Regulation Of Multicellular Organismal Development	2.7157314
Enzyme Linked Receptor Protein Signaling Pathway	2.633406
Embryonic Morphogenesis	2.298786
Tissue Development	2.3724976
Regulation Of Cell Differentiation	2.3799562
Response To Oxygen Containing Compound	0
Response To Lipid	1.8044732
Negative Regulation Of Developmental Process	1.8892956
Reproduction	1.7579603
Negative Regulation Of Gene Expression	1.758068
Epithelium Development	1.9074922
Cellular Response To Oxygen Containing Compound	1.7649908

Supplementary Figure 3. Enrichment Analysis of Putative AHR-NC-XRE Target Genes. (A) Table corresponding to the ORA data shown in Figure 3B, including the enriched Gene Ontology Biological Processes (GOBP) gene sets and overlapping genes that have 5 NC-XRE repeats with a maximum of 50 bp spacers within 10 kb of the gene body. **(B) GSEA enrichment plots of 6 upregulated pathways identified in the ORA analysis. These panels are ranked based on the number of hit genes in the ORA: Embryo development (8), tissue development (8), regulation of multicellular organismal development (7), animal organ morphogenesis (7), regulation of cell differentiation (7), and enzyme-linked receptor protein signaling pathway (6). (C)** Normalized enrichment scores (NES) of the GOBP gene sets shown in Figure 3D.

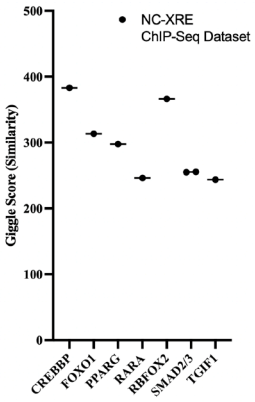
A

CollecTRI Transcription Factor	-log10(P-value)	Genes in Gene Set	Genes in Comparison (Fig 3A)	Total Number of Overlapping Genes	Overlapping Genes
FOXO4	2.315	25	22	2	*Serpine1; Igfbp1*
CREB1	2.243	372	22	4	*Noct; Serpine1; Hes1; Slc20a1*
PPARA	2.243	188	22	3	*Serpine1; Igfbp1; Pdk4*
ETS1	2.243	198	22	3	*Serpine1; Igfbp1; Pim3*
FOXO1	2.243	188	22	3	*Noct; Igfbp1; Pdk4*
ARNTL	2.243	52	22	2	*Noct; Serpine1*
TFE3	2.243	35	22	2	*Mafb; Serpine1*
CLOCK	2.237	57	22	2	*Noct; Serpine1*
ESR1	2.116	505	22	4	*Serpine1; Hes1; Arid5b; Igfbp1*
SMAD2	2.064	75	22	2	*Hes1; Serpine1*
PGR	1.920	92	22	2	*Serpine1; Igfbp1*
STAT3	1.802	350	22	2	*Mafb; Igfbp1; Pim3*
SMAD4	1.802	113	22	2	*Serpine1; Atoh8*
CEBPB	1.740	376	22	3	*Mafb; Igfbp1; Pdk4*
SMAD3	1.707	140	22	2	*Hes1; Serpine1*
AHR	1.707	139	22	2	*Serpine1; Igfbp1*
FOXO3	1.664	151	22	2	*Serpine1; Pdk4*
NR3C1	1.513	188	22	2	*Igfbp1; Pdk4*
USF1	1.513	190	22	2	*Serpine1; Igfbp1*
PPARG	1.431	215	22	2	*Noct; Igfbp1*

B

GSM ID	Gene	Source	Giggle Score (XRE)	Giggle Score (NC-XRE)
GSM2574037	AHR	Liver	745.94	2288.92
GSM2573809	AHR	Liver	625.85	2283.21
GSM2573805	AHR	Liver	591.48	2225.91
GSM2573807	AHR	Liver	543.091	2232.81
GSM2574040	AHR	Liver	531.76	1949.66
GSM2573806	AHR	Liver	526.41	1647.26
GSM2573808	AHR	Liver	508.52	2239.41
GSM2574038	AHR	Liver	499.79	1330.16

C



GSM ID	Gene	Source	Giggle Score (NC-XRE)
GSM1315601	CREBBP	Embryonic Fibroblast Cells	383.01
GSM998924	FOXO1	Treg Cells from Blood	313.38
GSM1199130	PPARG	3T3-L1 Adipocytes	297.72
GSM1047939	RARA	CMTi-1 Embryonic Stem Cell	246.23
GSM1397521	RBFOX2	Heart	366.21
GSM1782929	SMAD2/3	Embryonic Stem Cells	255.51
GSM1782932	SMAD2/3	Embryonic Stem Cells	254.89
GSM1333558	TGIF1	Bone Marrow Cells	243.65

Supplementary Figure 4. Putative Transcription Factors Targeting 5 Repeats of XRE or NC-XRE Separated by 50 bp or Less. (A) Corresponds to the ORA data shown in Figure 4B, including the upregulated CollecTRI transcription factors and overlapping genes that have 5 NC-XRE repeats with a maximum of 50 bp spacers within 10 kb of the gene body. **(B)** Corresponds to the CistromeDB giggle scores shown in Figure 4C, indicating the overlap of AHR peaks containing 5 or more XRE or NC-XRE with 50 bp spacers in our query dataset compared mouse liver AHR ChIP-seq datasets on the CistromeDB database. Higher giggle score means higher similarity. **(C)** Giggle scores of AHR peaks with 5 or more NC-XRE repeats with up to 50 bp spacers compared to some of the hit transcription factor (Fig 4A, 4B) peaks from mouse non-liver ChIP-seq experiments. Each point represents how strongly NC-XRE containing AHR binding sites overlap with the transcription factor binding sites. A higher giggle score indicates a higher likelihood of these transcription factors co-occupying the AHR peaks.

Table 1: ChIP-Seq Datasets Analyzed in This Study

GEO Series	PMID	GEO Samples	Target	Sex	Strain	Tissue	Age at Assay	Conditions
GSE97634	28213091 31015483	GSM2573805 GSM2573806 GSM2573807 GSM2573808 GSM2573809	AHR	Male	C57BL/6	Liver	Adult 28 Days	Exposed to 30 ug/kg TCDD via oral gavage for 2 hours, n = 5 mice
GSE97636	26582802	GSM2574037 GSM2574038 GSM2574039 GSM2574040 GSM2574041	AHR	Female	C57BL/6	Liver	Adult 28 Days	Exposed to 30 ug/kg TCDD via oral gavage for 2 hours, n = 5 mice
GSE31039	22889292	GSM1000140	H3K27ac	Male	C57BL/6	Liver	Adult 8 Weeks	No treatment
		GSM1000150	H3K27me3	Male	C57BL/6	Liver	Adult 8 Weeks	
		GSM1000151	H3K36me3	Male	C57BL/6	Liver	Adult 8 Weeks	
		GSM1000152	H3K79me2	Male	C57BL/6	Liver	Adult 8 Weeks	
		GSM1000153	H3K9ac	Male	C57BL/6	Liver	Adult 8 Weeks	
		GSM769014	H3K4me3	Male	C57BL/6	Liver	Adult 8 Weeks	
		GSM769015	H3K4me1	Male	C57BL/6	Liver	Adult 8 Weeks	
GSE36027	22889292	GSM918718	Control	Male	C57BL/6	Liver	Adult 8 Weeks	No treatment
GSE30298	23238790	GSM751035	H3K9me3	N/A	C57BL/6	Liver	Adult 8-10 Weeks	No treatment

Table 2: RNA-Seq Datasets Analyzed in This Study

GEO Series	PMID	GEO Samples	Treatment Duration (hr)	Sex	Strain	Tissue	Age at Assay	Conditions
GSE109863	29752288	GSM2971788 GSM2971789 GSM2971790 GSM2971791 GSM2971792 GSM2971793	2	Male	C57BL/6	Liver	Adult 28 Days	On postnatal day 28 mice were orally gavaged with sesame oil vehicle or 30 μg/kg TCDD
		GSM2971794 GSM2971795 GSM2971796 GSM2971797 GSM2971798 GSM2971799	4					
		GSM2971800 GSM2971801 GSM2971802 GSM2971803 GSM2971804 GSM2971805	8					
		GSM2971806 GSM2971807 GSM2971808 GSM2971809 GSM2971810 GSM2971811	12					
		GSM2971812 GSM2971813 GSM2971814 GSM2971815 GSM2971816 GSM2971817	24					
		GSM2971818 GSM2971819 GSM2971820 GSM2971821 GSM2971822 GSM2971823	72					
		GSM2971824 GSM2971825 GSM2971826 GSM2971827 GSM2971828 GSM2971829	168					