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

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

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

ABSTRACT

The aryl hydrocarbon receptor (AHR) is a ligand-dependent transcription factor that binds DNA and regulates genes in response to halogenated and polycyclic aromatic hydrocarbons. AHR also regulates the development and function of the liver and the immune system. In the canonical pathway, AHR binds a consensus DNA sequence, 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 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-NCXRE 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 also performed functional genomics at a single locus, the mouse *Serpine1* gene. 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.

NC-XRE motifs are prevalent throughout regions of the genome where AHR binds. Taken together, our results suggest that AHR regulates genes via NC-XRE motifs. Our results will also improve our ability to identify AHR target genes and their physiologic 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), ligand-dependent transcription factors that recruit protein coregulators to DNA and directly regulate gene expression (Denison et al., 2011). AHR forms a dimer with aryl hydrocarbon receptor nuclear translocator protein (ARNT1, also known as HIF1B) to regulate transcription (Hankinson, 2005; Reyes et al., 1992). The AHR-ARNT1 complex binds 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. A chromatin immunoprecipitation study found that 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, gel shift and reporter assays support the hypothesis that AHR binds a non-consensus XRE DNA sequence (nucleotides GGGA, referred to as NC-XRE motif) (Huang and Elferink, 2012; Jackson et al., 2014; Joshi et al., 2015; Wilson et al., 2013). Such studies were limited because they did not analyze AHR-NCXRE interactions on a genome-wide scale. Additionally, chromatin immunoprecipitation identifies where AHR binds to DNA 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-NCXRE 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 gene, directly test whether NC-XRE motifs are required for AHR-dependent target gene expression.

RESULTS

Frequency of NC-XRE motifs

To determine the 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 #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]. We found that, on average, 30% of peaks (AHR binding to DNA) contained a NC-XRE motif without a nearby XRE. 9% of peaks contained an XRE motif without a nearby NC-XRE, while 51% of peaks contained an XRE and NC-XRE (Fig 1A, C). Fewer than 1% of peaks contained RelBAHRE.

We wondered whether XRE and NC-XRE peaks were distributed equally among promoter and enhancer DNA. We observed more peaks at promoter regions (within 3 kb of transcription start sites) containing XRE motifs alone or XRE+NC-XRE motifs together compared to promoter regions containing NC-XRE alone (Fig 1B). This suggests that AHR binding to XRE is stronger in promoter regions, whereas AHR binding to NC-XRE is stronger in enhancer regions.

XRE, NC-XRE and RelBAHRE motifs are different lengths, which affects the frequency with which they are found in 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. NC-XRE and XRE motifs are statistically significantly enriched in AHR ChIP peaks, whereas RelBAHRE motifs are not (Fig 1D). Because of the low number of RelBAHRE in AHR peaks, we focus on the XRE and NC-XRE motifs in this study.

Repeated NC-XRE motifs associated with AHR target genes

The preceding analysis examined presence of NC-XRE motifs in all regions where AHR was bound to DNA. However, transcription factors like AHR may bind DNA without affecting gene expression. Therefore, we sought to identify AHR ChIP-seq peaks associated with AHR target genes. To identify AHR target genes, 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 treated with a variety of doses of TCDD for various durations of exposure (Table 1). We define 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. Of 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.

The preceding analysis is for any single instance of XRE or NC-XRE DNA sequence. For some transcription factors, repeated DNA binding sites produce a more robust response (Klein-Hitpass et al., 1988). We sought to examine the association between repeated runs of XRE or NCXRE motifs and target gene expression. Fig 2A shows the relative frequency of AHR target genes with AHR peaks containing between 3 and 10

124 NC-XRE motifs separated by up to 50 basepairs (bp), compared to target genes found
125 near similar patterns of XRE peaks. 24% of AHR peaks contain 2 or more NC-XRE
126 motifs separated by 25 bp or less (Fig 2B). AHR peaks containing 5 NC-XRE motifs
127 separated by 25 bp or less were found proximal to 82 AHR target genes. Compared to
128 repeated XRE motifs, runs of 5 or more NC-XREs are enriched at AHR binding sites
129 proximal to transcription start sites (H3K4me, H3K27ac, Fig 2C). This suggests that
130 NCXREs may be more potent when they occur in tandem repeats. The top differentially
131 regulated genes with 5 or more NCXRE motifs are shown in Fig 2D. As validation, the
132 known AHR target gene *Serpine1* (aka *Pai1*) contains 5 NCXRE motifs in the putative
133 promoter, consistent with results that AHR upregulates *Serpine1* via NC-XRE DNA
134 (Huang and Elferink, 2012). AHR binding was also detected at NC-XRE DNA near
135 genes like *Nqo1*, a canonical AHR target gene thought to be regulated by XRE motifs
136 (Yeager et al., 2009). It is possible that NC-XRE DNA also contributes to the regulation
137 of NQO1 expression.
138
139 **Motifs flanking XRE or NC-XRE sequence**
140
141 Transcription factors bind regulatory DNA in a complex composed of multiple different
142 transcription factors. For example, AHR is associated with the transcription factor
143 estrogen receptor alpha at regulatory DNA (Beischlag and Perdew, 2005; Klinge et al.,
144 2000; Madak-Erdogan and Katzenellenbogen, 2012). Another transcription factor, ARNT,
145 is thought to be an obligate co-regulator of AHR (Lo and Matthews, 2012; Reyes et al.,
146 1992). Previous studies have focused on transcription factor binding at XRE motifs. We
147 identified known transcription factor motifs flanking NC-XRE motifs. We focused on AHR
148 ChIP peaks containing at least 2 NC-XRE motifs within 25 basepairs of each other. The
149 top 12 most frequently occurring known transcription factor motifs are shown in Figure
150 3. We found several motifs enriched in peaks containing both XRE and NC-XRE motifs.
151 As expected, the ARNT motif was enriched in both, supporting the previously known
152 association with XRE motifs and suggesting that the AHR-ARNT complex could be
153 associated with NC-XRE motifs. Several motifs were enriched flanking NC-XRE but not
154 XRE, such as androgen receptor binding site and SMAD3 site. This suggests that there

155 may be crosstalk between AHR and other signaling pathways, such as TGF-beta, at
156 NC-XRE motifs.
157
158 **Testing NC-XRE function in a natural genomic context**
159
160 We mined published RNA-seq and ChIP-seq data and identified candidate genes
161 upregulated by TCDD in mouse liver, where such genes contain an AHR ChIP peak
162 containing XRE or NCXRE (or both) within 10 kb of the gene body. We exposed mouse
163 Hepa1-6 liver cells to 10 nM TCDD for 2 hours and assayed expression of 8 candidate
164 genes identified from our analysis of mouse liver. Three genes were upregulated more
165 than 2-fold compared to vehicle in Hepa1-6 cells (Figure 4A). For one of these genes,
166 *Serpine1*, ChIP studies in mouse liver identified an AHR binding site ~150 bp upstream
167 of the transcription start site, which contains a run of 5 NC-XRE motifs (Fig 4B) (Huang
168 and Elferink, 2012). Indirect evidence suggests that AHR upregulates *Serpine1*
169 expression via NC-XRE motifs, but this has never been tested directly.
170
171 We deleted the 5x NC-XRE motif in the *Serpine1* promoter, generated clonal mutant cell
172 lines, and directly tested whether the 5x NC-XRE motif contributes to TCDD-dependent
173 *Serpine1* expression. We transfected Hepa1-6 cells with Cas9 protein together with a
174 guide RNA targeting this NCXRE sequence. We picked single cells and established a
175 clonal cell line (clone 14) containing mutations in the target NCXRE (Fig 4B). Bulk
176 sequencing of clone 14 cells revealed a heterogenous population with 57% of the DNA
177 sequence containing a 24bp deletion of the 5x NC-XRE, 14% had a 2 bp deletion that
178 does not disrupt the NC-XRE motifs, and 28% of the DNA was wild-type. We then
179 derived additional single cell colonies from the original clone 14, denoted as clone 14-10
180 and clone 14-28. Bulk DNA sequencing of each of these clones revealed no wild-type
181 DNA. 95% of the DNA contained the expected 24bp deletion encompassing the 5x NC-
182 XRE, while 5% had the same 2bp deletion that does not disrupt the NC-XRE motifs.
183
184 We exposed mutant and wild-type cells to TCDD or vehicle for 2 hours and assayed
185 *Serpine1* expression by RT-qPCR. In all clones, deletion of the NCXRE reduced

upregulation of *Serpine1* in response to TCDD (Fig 4C). 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. Our results suggest that AHR transcription factor function is normal in NC-XRE clone 14 mutant cell lines. We conclude that AHR upregulates *Serpine1* via this NCXRE 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.

There are few examples where a transcription factor binding site DNA was mutated in an endogenous genomic context, allowing investigators to test whether a specific DNA sequence is required for transcription factor activity. Carleton and colleagues used catalytically inactive Cas9 to identify and block enhancer DNA bound by estrogen receptor alpha (Carleton et al., 2017). Other labs have deleted hundreds or thousands of basepairs in enhancer DNA to interrogate enhancer function (reviewed in (Lopes et al., 2016)). These studies are fruitful but lack the resolution to identify specific DNA sequence to which transcription factors bind. Here we used targeted, precise mutation to demonstrate the necessity of repeated NC-XRE sequence to regulate an AHR target gene.

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.

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). These results suggest that AHR may bind NC-XRE in the absence of ARNT1.

A limitation of the previous studies is that none considered ARNT2. ARNT2 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. One possibility is that ARNT1 preferentially associates with AHR-XRE complexes, while ARNT2 preferentially associates with AHR-NC-XRE complexes.

We appreciate that DNA regulatory elements may be incompletely conserved between species. Showing that AHR binds NCXRE DNA to regulate gene expression in any species is important foundational knowledge. While the specific number or configuration of NCXRE 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 NCXRE in mice but not in humans. Considering that AHR binds identical XRE DNA elements in zebrafish, mice and humans (Andreasen et al., 2002; Denison et

248 al., 1988; Tanguay et al., 1999; Yao and Denison, 1992), we argue that the same is likely
249 true for NCXRE.
250
251
252 **METHODS**
253
254 **Re-analysis of published murine liver ChIP-Seq datasets**
255 ChIP-Seq datasets of AHR in murine livers after 2hrs of TCDD treatment in male
256 (GSE97634) and female (GSE97636) mice were downloaded from NCBI Gene
257 Expression Omnibus (GEO) (Fader et al., 2017; Fader et al., 2019; Nault et al., 2016).
258 Data were trimmed using trim_galore, mapped to the mouse genome build UCSC
259 mm10 using bowtie2, and peaks were called using MACS2 using the provided controls
260 (Langmead and Salzberg, 2012; Zhang et al., 2008). Peaks were merged using
261 bedtoools across both sexes (Quinlan and Hall, 2010). Each peak was searched for
262 AHR motifs using an in-house script. To test the likelihood of finding AHR motifs
263 anywhere in the genome, random DNA peaks were generated matching the
264 chromosome and size distribution as previously described (Coarfa et al., 2020), then
265 searched for AHR motifs. Enrichment of AHR motifs compared to the random controls
266 was assessed using a Fisher's exact test, with significance achieved at $p\text{-value}<0.05$.
267 Venn diagrams of AHR motifs in the merged AHR peaks were derived using bedtools.
268 Genomic distribution of AHR peaks over genomic elements was inferred using bedtools
269 and visualized using GraphPad Prism.
270
271 Runs of AHR motifs, eg sequences of successive motifs within a maximum distance
272 from each other, were determined using an in-house Python script. Flanking sequences
273 were determined using the bedtools software. Enriched motifs within the flanking
274 sequences were determined using the HOMER software (Heinz et al., 2010).
275
276 Additional ChIP-Seq datasets of histone modifications in the mouse liver were
277 downloaded from NCBI GEO using Encode mouse liver from 8 week old mice:
278 GSM1000140, GSM769014 (Yue et al., 2014). Data were trimmed using trim_galore,

279 mapped to the mouse genome build UCSC mm10 using bowtie2 (Langmead and
280 Salzberg, 2012), and ChIP-Seq signal distribution over genomic regions was computed
281 using the HOMER software.
282
283 **Re-analysis of published murine liver AhR RNA-Seq datasets and integration with**
284 **ChIP-seq datasets**
285 Several RNA-seq datasets of mouse liver after treatment with TCDD at different doses
286 and for multiple timepoints were downloaded from NCBI GEO: GSE109863, GSE62902,
287 and GSE87519 (Fader et al., 2017; Nault et al., 2015; Nault et al., 2018)(Table 1). Data
288 was trimmed using trim_galore, then mapped onto the mouse genome build UCSC
289 mm10 using STAR (Dobin et al., 2013), and gene expression was quantified using
290 feature_counts (Liao et al., 2014). Differentially expressed genes between TCDD and
291 vehicle treatment were inferred using the EdgeR and RUVr R packages (Risso et al.,
292 2014; Robinson et al., 2010), with significance achieved at fold change exceeding 1.5x
293 and FDR-adjusted p-value<0.05. Genomic locations (represented as bed files) and
294 gene signatures were integrated using bedtools. To capture both promoter and proximal
295 enhancer effects, peaks were considered as associated with genes if the gene body
296 was within 10,000 basepairs of a peak.
297
298 **Cell culture**
299 The mouse hepatoma cell line Hepa 1–6 was obtained from ATCC (Manassas, VA)
300 (Darlington et al., 1980). Hepa 1-6 cells cells were cultured in 10 cm plates in
301 Dulbecco's Modified Eagle's medium (DMEM) supplemented with 1% (v/v)
302 Penicillin/Streptomycin and 10% (v/v) FBS in a humidified incubator (5% CO_2, 37°C).
303
304 **CRISPR mutagenesis of Hepa 1-6 cells**
305 We used CHOPCHOP (Labun et al., 2019) to design guide RNAs targeting the 5x NC-
306 XRE motif region 150bp upstream of transcription start in the mouse Serpine1 gene
307 (target sequence CAGCAAGTCACTGGGAGGGA<u>GGG</u>, PAM motif underlined).
308 Synthetic single guide RNA—modified with 2'-O-methyl at 3 first and last bases, and 3'-
309 phosphorothioate bonds between first 3 and last 2 bases—and purified SpCas9-2NLS

310 protein was purchased from Synthego (Redwood City, CA). We mixed guide RNA and
311 Cas9 (1.3:1 ratio) to form ribonucleoproteins according to Synthego's recommended
312 protocol. Guide RNA and Cas9 ribonucleoprotein complexes were transfected into
313 Hepa1-6 cells using Lipofectamine™ CRISPRMAX™ Cas9 Transfection reagent
314 (Thermo Fisher Scientific, Waltham, MA) according to the manufacturer's protocol.
315 100,000 Hepa-1-6 cells were mixed with the ribonucleoprotein-transfection solution and
316 split into two wells for genomic analysis and clonal expansion. Cells were incubated in a
317 humidified incubator (5% CO_2, 37 °C) for 24 hours. We then changed the media and
318 allowed the cells to incubate for another 3 days before genotyping or deriving single-cell
319 colonies for clonal expansion.
320
321 To derive colonies from single cells, we took 10 cells / mL suspension and aliquoted 100
322 µl into each well of the 96-well plate. The plates were checked daily and wells with one
323 cell were marked. Cells were monitored and allowed to grow for 1-2 weeks before
324 transfer to a 6 well plate for genotyping and further use.
325
326 **Genotyping cell lines**
327 Genomic DNA was extracted from cells using phenol-chloroform extraction. Cells were
328 lysed with DNA lysis buffer (10 mM Tris, 200 mM NaCl, 5 mM EDTA, 1% SDS, 0.4
329 mg/ml proteinase K). Equal volume of phenol:choloroform:isoamyl alcohol was added
330 for DNA extraction into upper phase followed by ethanol-sodium acetate precipitation of
331 DNA. We used PCR to amplify the region flanking the NC-XRE motifs using primers
332 5'- AAGCCAGGCCAACTTTTCCT and 5'- CGGTCCTCCTTCACAAAGCT. Amplicons
333 were then ligated into pCR4-TOPO vector using a TOPO TA cloning kit (catalogue
334 450030, ThermoFisherScientific, Waltham, MA) according to the manufacturer's
335 protocol. Plasmids were transfected into competent cells, plated and grown overnight.
336 10-20 bacterial colonies per plate were picked for miniprep DNA extraction and Sanger
337 sequencing using M13 forward primer. DNA sequence was aligned to wild-type using
338 MacVector software (MacVector Inc., Apex, NC). If 19 out of 20 colonies contained the
339 same mutant DNA sequence, and 1 colony contained wild-type DNA, we would
340 conclude that 95% of that cell line contained mutant DNA, and 5% wild-type. Clone 14

341 was selected based on the presence of a 24 bp deletion that encompassed the entire 5x
342 NC-XRE motif (see Figure 4).
343
344 **Quantitative reverse-transcription PCR**
345 70-80% confluent Hepa-1-6 wild-type and mutant cells were exposed to media
346 containing 10 nM TCDD or DMSO (vehicle) for 2 hours. To extract RNA, cells were
347 lysed in TRIzol followed by RNA extraction according to the manufacturer's protocol
348 using Direct-zol RNA Miniprep Kit, including the on-column DNase digestion (catalogue
349 11-331, Zymo Research, Irvine, CA). RNA concentration was measured using a
350 NanoDrop 2000 Spectrophotometer (ThermoFisherScientific, Waltham, MA). Purity of
351 the RNA (A260/A280) was ≥ 2 with the yield of 60 ng/µl or higher. We converted 1000
352 ng of the RNA into cDNA using iScript Reverse Transcription Supermix for RT-qPCR
353 (Biorad, Hercules, CA; cat no 1708841) according to the manufacturer's protocol (total
354 volume 20 µl, incubated for 5 min at 25°C, followed by incubation at 46°C for 20 min,
355 then 95°C for 1 min). For the qPCR reaction, 4 µl of 1:10 dilution of cDNA was mixed
356 with SsoAdvanced Universal SYBR Green Supermix (cat no 1725271, Bio-Rad,
357 Hercules, CA) and 500 nM each primer up to a total volume of 10 µl and amplified in a
358 CFX96 Real-Time System (Bio-Rad) with the following program: 95C for 30 sec, then
359 95C for 10 sec and 60C for 30 sec for 40 cycles. Primer pairs were either designed
360 using MacVector software or were previously published. Ct values were calculated
361 using CFX Maestro software version 4.1.2433.1219 (Bio-Rad). Target gene expression
362 was compared to reference gene expression using the $2^{-\Delta\Delta Ct}$ method. The results shown
363 were the average of three or more independent experiments. Statistical analysis was
364 performed using Prism version 9.4.2 (GraphPad Software, Boston, MA).
365
366 Primer sequences:
367 CYP1B1 gene (Liu et al., 2015)
368 Cyp1b1 FP: 5'-CCAGATCCCGCTGCTCTACA-3'
369 Cyp1b1 RP: 5'-TGGACTGTCTGCACTAAGGCTG-3'
370
371 AHRR (Aryl Hydrocarbon Receptor Repressor) gene (Bernshausen et al., 2006)

372 AHRR FP: 5’-GTTGGATCCTGTAGGGAGCA-3’
373 AHRR RP: 5’-AGTCCAGAGGCTCACGCTTA-3’
374
375 Mmp24os1 (BC029722) gene
376 BC029722 FP: 5’-CGCTTTCTAATCGCCTGCAC-3’
377 BC029722 RP: 5’-TGAGGAGATAAAAGCCAGGCC-3’
378
379 Cd36 gene (Niu et al., 2018)
380 CD36 FP: 5’-TTGTGGAGCTCAAAGACCTG-3’
381 CD36 RP: 5’-TGCAAGAAGCGGGATGTAGTC-3’
382
383 P21Cip1 gene (Jackson et al., 2014)
384 Cdkn1a FP: 5’-TGTCTTGCACTCTGGTGTCTGAG-3’
385 Cdkn1a RP: 5’-CAATCTGCGCTTGGAGTGATAG-3’
386
387 HROB gene
388 HROB FP: 5’-AAGAGGAGCTCTCAGAGGCA-3’
389 HROB RP: 5’-ATGGTGGATGCCCTGTCTTG-3’
390
391 IL1RN Gene (Isoda et al., 2005)
392 IL1RN FP: 5′-CTTTACCTTCATCCGCTCTGAGA-3′
393 IL1RN RP: 5′-TCTAGTGTTGTGCAGAGGAACCA-3′
394
395 MAF BZIP Transcription Factor F gene
396 MAFF FP: 5’ ATGGCTGTGGATCCCTTATCT-3’
397 MAFF RP: 5’ CATCAGCGCTTCATCCGACA-3’
398
399 UGT1A7C gene
400 UGT1A7C FP: 5’-GTCATCCAAAGACTCGGGCA-3’
401 UGT1A7C RP: 5’-GGGCATCATCACCATCGGAA-3’
402

403 18s rRNA (Turner et al., 2018)

404 m18s-FP: 5’-GTAACCCGTTGAACCCCATT-3’

405 m18s-RP: 5’-CCATCCAATCGGTAGTAGCG-3’

406

407 CYP1A1 gene (Jackson et al., 2014)

408 CYP1A1-FP: 5’-GCCTAACTCTTCCCTGGATGC-3’

409 CYP1A1-RP: 5’- GACATCACAGACAGCCTCATTGA -3’

410

411 Serpine1 gene (aka PAI-1) (Eren et al., 2014)

412 Serpine1-FP: 5′-ACGCCTGGTGCTGGTGAATGC-3′

413 Serpine1-RP: 5′-ACGGTGCTGCCATCAGACTTGTG-3′

414

415 **ACKNOWLEDGMENTS**

416 TDP, SLG and CC were partially supported by The Cancer Prevention Institute of Texas

417 (CPRIT) [RP210227, RP200504], NIH/NCI P30 shared resource grant [CA125123],

418 NIH/NIEHS center grants [P30 ES030285] and [P42 ES027725], and NIH/NIMHD

419 [P50MD015496]. DAG supported by NIEHS R01 ES026337.

420

421

FIGURE CAPTIONS

Figure 1. Frequency of NC-XRE motifs in DNA bound by AHR. (**A**) Following TCDD exposure, we analyzed mouse liver AHR ChIP-seq peaks for the presence of known AHR motifs: canonical xenobiotic response element (XRE), non-canonical xenobiotic response element (NC-XRE) or RelB AHR response element (RelBAHRE). Pie-chart shows the percent of AHR binding sites (AHR ChIP peaks) containing one or more of the motifs indicated. A majority of AHR peaks contain canonical and non-canonical xenobiotic response element sequences (XRE+NC-XRE). About a quarter of AHR peaks contain NC-XRE motifs alone. (**B**) Percent of AHR peaks containing XRE or NC-XRE sequences that are associated with promoters, enhancers or indicated genomic feature. A higher ratio of XRE+NC-XRE sites overlap with promoters compared to NC-XRE sites alone. (**C**) Venn diagram showing number of AHR peaks containing XRE, NCXRE, RelBAHRE alone or in combination. No AHR peaks contain RelBAHRE alone, two AHR peaks contain RelBAHRE + XRE. (**D**) Likelihood of finding XRE, NC-XRE or RelBAHRE motifs in AHR ChIP peaks versus in a random region of DNA. Table shows number of AHR peaks containing at least one XRE, NC-XRE or RelBAHRE motif. P value derived from Fisher's exact test.

Figure 2. Repeated NC-XRE motifs associated with AHR target genes. (**A**) Relative frequency of putative AHR target genes within 10kb of AHR ChIP peaks containing 3-10 NC-XRE motifs with 1-50 basepair (bp) spacers, compared to AHR peaks containing 3-10 XRE motifs with similar spacing. AHR ChIP peaks containing 5 NC-XRE motifs separated by 25 basepairs or less were found within 10 kb of 82 target genes. There are 4.9 fold more AHR target genes associated with AHR ChIP peaks containing 5 NCXRE motifs vs 5 XRE motifs separated by 25 bp or less. (**B**) Frequency of repeated runs of NCXRE motifs in AHR ChIP peaks. In 24% of AHR ChIP-seq peaks there are 2 or more NCXRE motifs separated by 25 basepairs or less. (**C**) Binding strength of H3K4me3 and H3K27ac at AHR peaks containing 5 or more NC-XRE or XRE motifs separated by 25 basepairs or less. Runs of NCXRE are enriched at AHR binding sites proximal to transcription start sites, as marked by H3K4me or H3K27ac ChIP peaks. (**D**) Examples of differentially expressed AHR target genes (fold change TCDD vs vehicle, mouse liver) containing AHR ChIP peaks with 5 or more NC-XRE motifs separated by 25 basepairs or less. We define AHR target genes as differentially expressed genes (RNA-seq, fold difference in TCDD vs vehicle > 1.5x, FDR<5%) containing AHR ChIP peak within 10 kb of a gene body.

460 **Figure 3. Transcription factor motifs flanking XRE or NC-XRE motifs in AHR peaks.**
461 Top 12 most frequently occuring known transcription factor motifs in the 50 basepairs
462 flanking repeated NCXRE and XRE motifs (2 or more motifs within 25 basepairs of each
463 other) in genomic regions where AHR binds. Highlighted motifs are enriched flanking
464 NCXRE but not XRE. Significance defined as $p<10^{-11}$
465

466 **Figure 4. Mutations in NCXRE DNA in *Serpine1* promoter reduced TCDD-dependent**
467 **expression of *Serpine1***. (**A**) Hepa1-6 cells were exposed to 10 nM TCDD or vehicle for
468 2 hours, followed by RNA extraction and qPCR for the indicated genes. Gene expression
469 normalized to vehicle, *18S rRNA* used as reference gene. Each circle represents a
470 different biological replicate, horizontal black lines indicate average expression. Dotted
471 line indicates 1x fold change. *Cyp1a1*, *Cyp1b1* and *Ahrr* were used as positive controls
472 since they are known targets of TCDD-AHR. *Cdkn1a*, *Maff* and *Serpine1* exhibit more
473 than 2x increase in expression following TCDD exposure. (**B**) Schematic of mouse
474 *Serpine1* gene showing AHR ChIP peak upstream of transcription start site (TSS). An
475 AHR peak 150 basepairs (bp) upstream of TSS contains NCXRE and no XRE sequences.
476 We transfected Hepa1-6 cells with Cas9 plus guide RNA to generate mutations in these
477 NCXRE motifs, picked single cells and derived a clonal population (clone 14) that
478 completely lacks the NCXRE motifs. We then picked single cells from clone 14 and
479 derived additional clonal populations, designated clone 14-10 and 14-28, with indentical
480 mutations (deletion of the entire NCXRE motifs). (**C**) Hepa1-6 wild-type or mutant cells
481 were exposed to 10 nM TCDD or vehicle for 2 hours, followed by RNA extraction and
482 qPCR for *Serpine1* or *Cyp1a1*. Mutant cell lines lacking the entire NCXRE exhibit reduced
483 increase in *Serpine1* following TCDD exposure compared to wild-type cells. Mutant cells
484 showed similar increase in *Cyp1a1* following TCDD exposure as wild-type cells. One-
485 tailed paired t test, * $p < 0.05$, ** $p < 0.01$, ns not significant ($p > 0.05$).
486
487
488

489 **REFERENCES**

490 **Andreasen, E. A., Hahn, M. E., Heideman, W., Peterson, R. E. and Tanguay, R. L.**
491 (2002). The zebrafish (Danio rerio) aryl hydrocarbon receptor type 1 is a novel
492 vertebrate receptor. *Mol. Pharmacol.* **62**, 234–249.

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

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

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

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

505 **Darlington, G. J., Bernhard, H. P., Miller, R. A. and Ruddle, F. H.** (1980). Expression
506 of liver phenotypes in cultured mouse hepatoma cells. *J Natl Cancer Inst* **64**, 809–
507 819.

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

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

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

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

520 **Eren, M., Boe, A. E., Murphy, S. B., Place, A. T., Nagpal, V., Morales-Nebreda, L.,**
521 **Urich, D., Quaggin, S. E., Budinger, G. R. S., Mutlu, G. M., et al.** (2014). PAI-1-
522 regulated extracellular proteolysis governs senescence and survival in Klotho mice.
523 *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.

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

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

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

567 **Liao, Y., Smyth, G. K. and Shi, W.** (2014). featureCounts: an efficient general purpose
568 program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–
569 930.

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

573 **Lopes, R., Korkmaz, G. and Agami, R.** (2016). Applying CRISPR-Cas9 tools to
574 identify and characterize transcriptional enhancers. *Nat. Rev. Mol. Cell Biol.* **17**, 597–
575 604.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

623 **Yao, E. F. and Denison, M. S.** (1992). DNA sequence determinants for binding of
624 transformed Ah receptor to a dioxin-responsive enhancer. *Biochemistry* **31**, 5060–
625 5067.

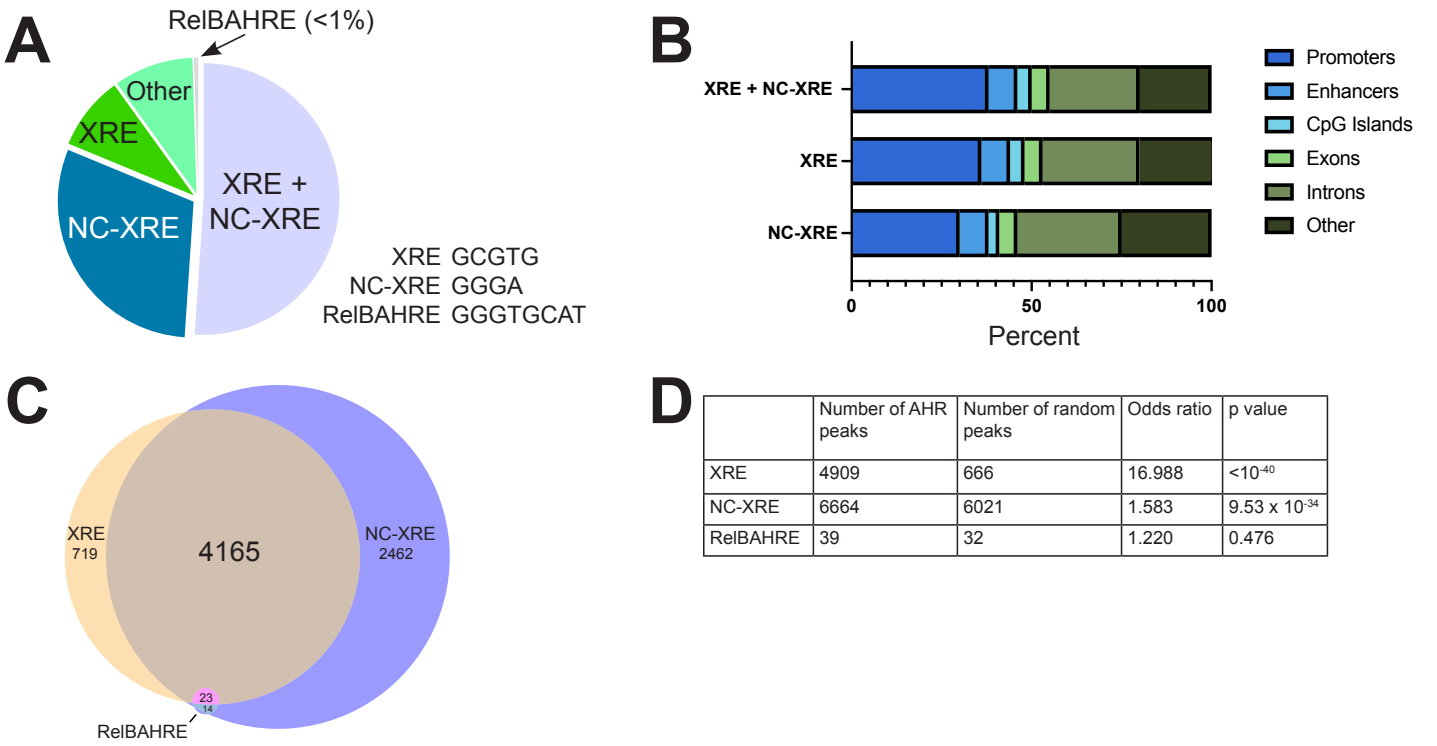
626 **Yeager, R. L., Reisman, S. A., Aleksunes, L. M. and Klaassen, C. D.** (2009).
627 Introducing the "TCDD-inducible AhR-Nrf2 gene battery". *Toxicol. Sci.* **111**, 238–246.

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

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

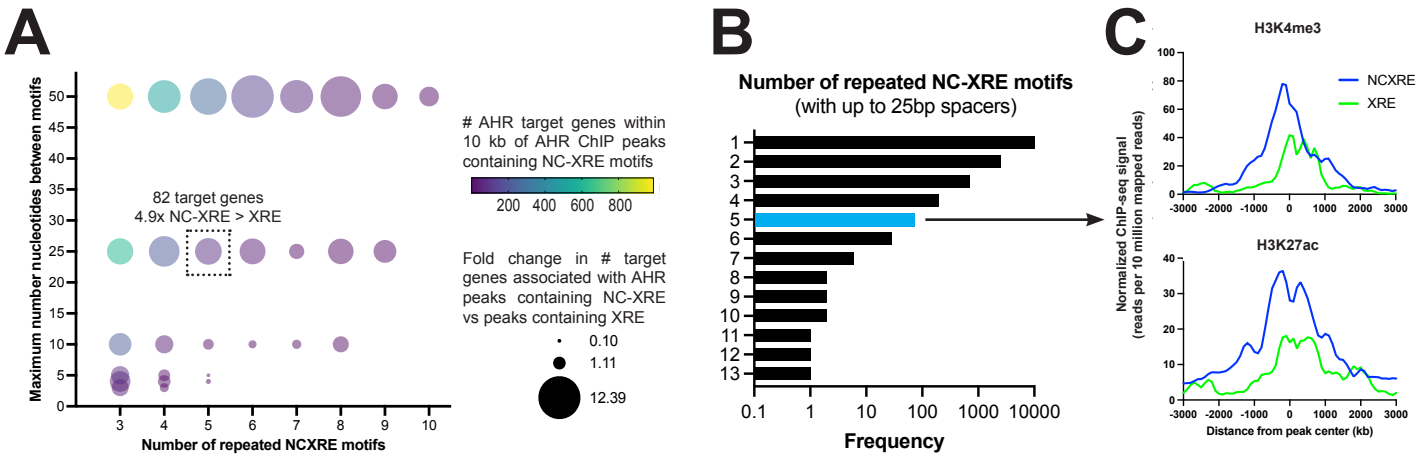
GEO accession	PMID	Assay	Sex	Conditions
GSE97634	28213091, 31015483	ChIP-seq	male C57BL/6 liver	exposed to 30 ug/kg TCDD via oral gavage for 2 hours, n=5 mice
GSE97636	26582802	ChIP-seq	female C57BL/6 liver	exposed to 30 ug/kg TCDD via oral gavage for 2 hours, n=5 mice
GSE62902	25958198	RNA-seq	female C57BL/6 liver	exposed to 0.01, 0.03, 0.1, 0.3, 1, 3, 10, or 30 ug/kg TCDD or sesame oil vehicle via oral gavage, every 4 days for 28 days
GSE87519	28213091	RNA-seq	male C57BL/6 liver	exposed to 0.01, 0.03, 0.1, 0.3, 1, 3, 10, or 30 ug/kg TCDD or sesame oil vehicle via oral gavage, every 4 days for 28 days. At 7, 15, and 23d after the first dose, mice (vehicle and 30 µg/kg TCDD groups) were fasted for 6 hours (access to water but not food). At 22d after the first dose, oral glucose tolerance tests (OGTT) were performed (vehicle and 30 µg/kg TCDD groups; fasted for 6h). Briefly, at time 0 minutes (min) animals were orally gavaged with 2 g/kg glucose in a 25% solution and tail blood glucose was measured at 0, 5, 15, 25, 30, 60, and 120 min. At 26d after the first dose, mice (vehicle and 30 µg/kg TCDD groups) were transferred to Innocages lacking bedding for 2h, with access to water but not food.
GSE109863	29752288	RNA-seq	male C57BL/6 liver	On postnatal day 28 mice were orally gavaged with sesame oil vehicle or 30 µg/kg TCDD for 2, 4, 8, 12, 24, 72, and 168 hours

Table 1. Datasets analyzed in this manuscript



	Number of AHR peaks	Number of random peaks	Odds ratio	p value
XRE	4909	666	16.988	$<10^{-40}$
NC-XRE	6664	6021	1.583	9.53×10^{-34}
RelBAHRE	39	32	1.220	0.476

Figure 1. Frequency of NC-XRE motifs in DNA bound by AHR. (**A**) Following TCDD exposure, we analyzed mouse liver AHR ChIP-seq peaks for the presence of known AHR motifs: canonical xenobiotic response element (XRE), non-canonical xenobiotic response element (NC-XRE) or RelB AHR response element (RelBAHRE). Pie-chart shows the percent of AHR binding sites (AHR ChIP peaks) containing one or more of the motifs indicated. A majority of AHR peaks contain canonical and non-canonical xenobiotic response element sequences (XRE+NC-XRE). About a quarter of AHR peaks contain NC-XRE motifs alone. (**B**) Percent of AHR peaks containing XRE or NC-XRE sequences that are associated with promoters, enhancers or indicated genomic feature. A higher ratio of XRE+NC-XRE sites overlap with promoters compared to NC-XRE sites alone. (**C**) Venn diagram showing number of AHR peaks containing XRE, NCXRE, RelBAHRE alone or in combination. No AHR peaks contain RelBAHRE alone, two AHR peaks contain RelBAHRE + XRE. (**D**) Likelihood of finding XRE, NC-XRE or RelBAHRE motifs in AHR ChIP peaks versus in a random region of DNA. Table shows number of AHR peaks containing at least one XRE, NC-XRE or RelBAHRE motif. P value derived from Fisher's exact test.



D

Gene name	Fold change	NCXRE location	XRE within 300 bp	Gene function	Relation to AHR
Serpine1	93	promoter	No	serine protease inhibitor that inhibits plasminogen activator, involved in blood clotting	AHR target gene
Crlf1	66	after last exon	Yes	cytokine receptor-like factor-1, a soluble protein that is a member of the ciliary neuro-trophic factor receptor pathway. CRLF1 forms a complex with CLCF1 and CNTFR to induce downstream signaling events	unknown
Maff	30	intron	Yes	coregulator of Nrf1 and Nrf2 transcription factors	AHR-Nrf2 crosstalk
Pdk4	26	promoter	Yes	PDK4 reversibly inactivates the mitochondrial PDH complex	In HepaRG cells, AHR ligands reduced expression of *PDK4*
Hrk	23	promoter	Yes	Harakiri is a member of the BCL2 gene family, which encode apoptosis regulatory pro-teins.	AHR ligands upregulate/activate HRK to cause cell death in ovaries
Lin7a	22	intron	No	intracellular vesicle trafficking and exocytosis	unknown
Nqo1	16	promoter	Yes	2-electron reductase that detoxifies quinones derived from the oxidation of phenolic metabolites of benzene	Upregulated by TCDD, considered a canonical AHR target gene
Klf6	8	promoter	Yes	transcription factor and tumor suppressor	An AHR binding partner at NC-XRE

Figure 2. Repeated NC-XRE motifs associated with AHR target genes. (**A**) Relative frequency of putative AHR target genes within 10kb of AHR ChIP peaks containing 3-10 NC-XRE motifs with 1-50 basepair (bp) spacers, compared to AHR peaks containing 3-10 XRE motifs with similar spacing. AHR ChIP peaks containing 5 NC-XRE motifs separated by 25 basepairs or less were found within 10 kb of 82 target genes. There are 4.9 fold more AHR target genes associated with AHR ChIP peaks containing 5 NCXRE motifs vs 5 XRE motifs separated by 25 bp or less. (**B**) Frequency of repeated runs of NCXRE motifs in AHR ChIP peaks. In 24% of AHR ChIP-seq peaks there are 2 or more NCXRE motifs separated by 25 basepairs or less. (**C**) Binding strength of H3K4me3 and H3K27ac at AHR peaks containing 5 or more NC-XRE or XRE motifs separated by 25 basepairs or less. Runs of NCXRE are enriched at AHR binding sites proximal to transcription start sites, as marked by H3K4me or H3K27ac ChIP peaks. (**D**) Examples of differentially expressed AHR target genes (fold change TCDD vs vehicle, mouse liver) containing AHR ChIP peaks with 5 or more NC-XRE motifs separated by 25 basepairs or less. We define AHR target genes as differentially expressed genes (RNA-seq, fold difference in TCDD vs vehicle > 1.5x, FDR<5%) containing AHR ChIP peak within 10 kb of a gene body.

Motif name	Motif sequence	Percent of AHR peaks with flanking motif & NCXRE	Percent of AHR peaks with flanking motif & XRE	NCXRE -log10(p value)	XRE -log10(p value)
Nr1a2	TRAGGTCA	50.62	46.52	215	71
AR half-site	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 3. Transcription factor motifs flanking XRE or NC-XRE motifs in AHR peaks. Top 12 most frequently occuring known transcription factor motifs in the 50 basepairs flanking repeated NCXRE and XRE motifs (2 or more motifs within 25 basepairs of each other) in genomic regions where AHR binds. Highlighted motifs are enriched flanking NCXRE but not XRE. Significance defined as $p<10^{-11}$

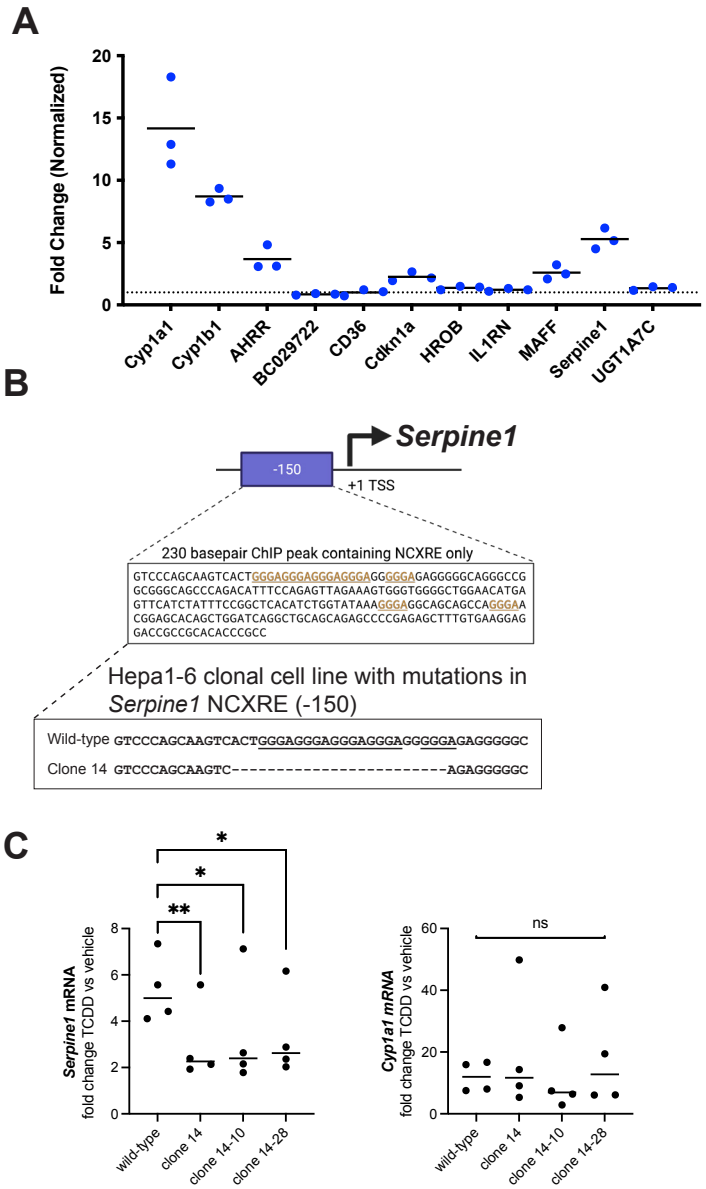


Figure 4. Mutations in NCXRE DNA in *Serpine1* promoter reduced TCDD-dependent expression of *Serpine1*. (**A**) Hepa1-6 cells were exposed to 10 nM TCDD or vehicle for 2 hours, followed by RNA extraction and qPCR for the indicated genes. Gene expression normalized to vehicle, *18S rRNA* used as reference gene. Each circle represents a different biological replicate, horizontal black lines indicate average expression. Dotted line indicates 1x fold change. *Cyp1a1*, *Cyp1b1* and *Ahrr* were used as positive controls since they are known targets of TCDD-AHR. *Cdkn1a*, *Maff* and *Serpine1* exhibit more than 2x increase in expression following TCDD exposure. (**B**) Schematic of mouse *Serpine1* gene showing AHR ChIP peak upstream of transcription start site (TSS). An AHR peak 150 basepairs (bp) upstream of TSS contains NCXRE and no XRE sequences. We transfected Hepa1-6 cells with Cas9 plus guide RNA to generate mutations in these NCXRE motifs, picked single cells and derived a clonal population (clone 14) that completely lacks the NCXRE motifs. We then picked single cells from clone 14 and derived additional clonal populations, designated clone 14-10 and 14-28, with indentical mutations (deletion of the entire NCXRE motifs). (**C**) Hepa1-6 wild-type or mutant cells were exposed to 10 nM TCDD or vehicle for 2 hours, followed by RNA extraction and qPCR for *Serpine1* or *Cyp1a1*. Mutant cell lines lacking the entire NCXRE 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. One-tailed paired t test, * $p < 0.05$, ** $p < 0.01$, ns not significant ($p > 0.05$).