Title Page:

Single Nucleotide Polymorphisms (SNPs) Distant from Xenobiotic Response Elements Can Modulate Aryl Hydrocarbon Receptor Function: SNP-Dependent *CYP1A1* Induction

Duan Liu, Sisi Qin, Bamiki Ray, Krishna R. Kalari, Liewei Wang, Richard M. Weinshilboum

Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN 55905, USA (D.L., S.Q., B.R., L.W. and R.M.W)

Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA (K.R.K)

Current affiliation: Assurex Health Inc., Mason, OH (B.R.)

Running title: SNPs distant from XREs modulate AHR Function

Corresponding author:

Richard M. Weinshilboum, M.D.,

200 First Street SW, Mayo Clinic

Rochester, MN 55905,

Phone: 507-284-2790, FAX: 507-284-4455, Email: weinshilboum.richard@mayo.edu

Manuscript information:

Number of text pages: 35

Number of tables: 1

Number of figures: 5

Number of references: 50

Number of words in Abstract: 249/250

Number of words in Introduction: 750/750

Number of words in Discussion: 1181/1500

Abbreviations:

3MC, 3-methylcholanthrene; AHR, aryl hydrocarbon receptor; ChIP, chromatin immunoprecipitation; Co-IP, co-immunoprecipitation; CYP1A1, cytochrome P450 1A1; DMSO, dimethyl sulfoxide; EMSA, electrophoretic mobility shift assay; KD, knockdown; LCLs, lymphoblastoid cell lines; MAF, minor allele frequency; MS, mass spectrometry; SNPs, single nucleotide polymorphisms; V, variant; WT, wildtype; XREs, xenobiotic response elements

Abstract:

CYP1A1 expression can be up-regulated by the ligand-activated aryl hydrocarbon receptor (AHR). Based on prior observations with estrogen receptors and estrogen response elements, we tested the hypothesis that single-nucleotide polymorphisms (SNPs) mapping hundreds of base pairs (bp) from xenobiotic response elements (XREs) might influence AHR binding and subsequent gene expression. Specifically, we analyzed DNA sequences 5 Kb up- and down-stream of the *CYP1A1* gene for putative XREs. SNPs located ±500 bp of these putative XREs were studied using a genomic data-rich human lymphoblastoid cell line (LCL) model system. CYP1A1 mRNA levels were determined after treatment with varying concentrations of 3-methylcholanthrene (3MC). The rs2470893 (-1694G>A) SNP, located 196 bp from an XRE in the *CYP1A1* promoter was associated with two-fold variation in AHR-XRE binding in a SNP-dependent fashion. LCLs with the AA genotype displayed significantly higher AHR-XRE binding and CYP1A1 mRNA expression after 3MC treatment than did those with the GG genotype. Electrophoretic mobility shift assay (EMSA) showed that oligonucleotides with the AA genotype displayed higher LCL nuclear extract binding after 3MC treatment than did those with the GG genotype, and mass spectrometric analysis of EMSA protein-DNA complex bands identified three candidate proteins, two of which were co-immunoprecipitated with AHR. In conclusion, we have demonstrated that the rs2470893 SNP, which maps 196 bp from a *CYP1A1* promoter XRE, is associated with variation in 3MC-dependent AHR binding and *CYP1A1* expression. Similar “distant SNP effects” on AHR binding to an XRE motif and subsequent gene expression might occur for additional AHR-regulated genes.

INTRODUCTION:

The aryl hydrocarbon receptor (AHR) is a ligand-activated transcription factor that is best known for its role in response to environmental toxins (Denison and Nagy, 2003; Murray et al., 2014). Once activated, AHR forms a heterodimer with the AHR nuclear transporter (ARNT) that binds to specific DNA sequence motifs, xenobiotic response elements (XREs), resulting in the regulation of gene transcription (Mimura and Fujii-Kuriyama, 2003). Although a structurally diverse group of chemical compounds are AHR ligands, the best characterized high affinity ligands are hydrophobic agents such as halogenated aromatic hydrocarbons and non-halogenated polycyclic aromatic hydrocarbons (Denison and Nagy, 2003). A prototypic gene that is induced by AHR activation is *CYP1A1*, which encodes cytochrome P450 1A1, an enzyme that metabolizes xenobiotics and environmental toxins, many of which are AHR ligands (Denison and Nagy, 2003). Several XRE motifs have been identified across the *CYP1A1* gene (Denison et al., 1988; Lo and Matthews, 2012).

CYP1A1 expression can vary widely, which theoretically might contribute to variation in cancer risk after carcinogen exposure, e.g., lung cancer risk in smokers (Rannug et al., 1995; Stucker et al., 2000). As a result, many studies have been performed searching for possible associations of *CYP1A1* SNPs with the induction of CYP1A1 expression and carcinogenesis risk. However, the results have often been contradictory — with some studies, especially those performed in East Asia, reporting a correlation of *CYP1A1* genetic polymorphisms with CYP1A1 induction and/or cancer risk (Kawajiri et al., 1990; Kiyohara et al., 1998), while others have failed to replicate those observations (Inoue et al., 2000; Anttila et al., 2001; Smith et al., 2001). Most of those studies focused on two common *CYP1A1* SNPs — rs4646903, also known as *CYP1A1*2A*, that

maps 3' of *CYP1A1* (Spurr et al., 1987) and rs1048943, a non-synonymous SNP that results in an Ile462Val substitution without change in enzyme activity (Zhang et al., 1996; Persson et al., 1997). Both of those SNPs have higher minor allele frequencies in East Asian populations than in Europeans, which might explain the contradictory results of attempts to associate *CYP1A1* genetic polymorphisms with CYP1A1 induction and/or cancer risk. However, exactly how those two SNPs might alter CYP1A1 activity remains unclear. With the exception of one *AHR* SNP (Smart and Daly, 2000) that was reported to influence CYP1A1 induction, no SNPs have been found to be associated with the induction of expression of this gene. As a result, the molecular basis for individual variation in the inducibility of CYP1A1 is not completely understood.

Recently, during studies of estrogen receptor alpha (ERα), we observed that SNPs at a distance (ie hundreds of bp) from estrogen response elements (EREs) could strongly influence both ER binding to those EREs and subsequent gene expression (Ingle et al., 2013; Ho et al., 2016; Ingle et al., 2016; Liu et al., 2017). To determine whether this type of phenomenon might also occur with other nuclear receptors, we set out in the present study to test the hypothesis that SNPs at a significant distance from XREs might modulate AHR-XRE binding and influence the transcription of AHR-regulated genes such as *CYP1A1*, a process that might, in part, explain individual variation in CYP1A1 inducibility. ERα acts in a fashion very similar to AHR. Once activated by estrogen binding, ERα forms a homodimer and binds to specific DNA sequence motifs in estrogen response elements (EREs), thus regulating gene expression. For example, a previous genome-wide association study (GWAS) for response to tamoxifen breast cancer prevention therapy identified SNP signals in intron 2 of the *ZNF423* gene (Ingle et al., 2013). Functional genomic studies demonstrated that the *ZNF423* rs9940645 SNP, which maps 200 bp

away from an ERE cluster, could influence ERα-ERE binding and *ZNF423* transcription (Ingle et al., 2013). Furthermore, subsequent studies identified calmodulin-like protein 3 as a co-regulator of ERα that "sensed" genotypes for the *ZNF423* intron 2 SNP (Qin et al., 2017).

In present study, we aim to determine whether SNPs at distance from XREs might modulate AHR-XRE binding and influence down-stream gene transcription in a fashion similar to the effect of the rs9940645 SNP in *ZNF423* on ERα binding to EREs by studying SNPs and XREs in and near *CYP1A1*. As described subsequently, we identified a SNP, rs2470893, that maps 5'- of *CYP1A1* that can influence both AHR binding to an XRE located 196 bp distant from the SNP and subsequent *CYP1A1* transcription induced by the AHR agonist, 3-methylcholanthrene (3MC) (Riddick et al., 1994). We also identified possible AHR interacting proteins that might help to explain the mechanism responsible for this process.

MATERIALS AND METHODS:

Identification of Putative XREs

A motif scanning tool, the Find Individual Motif Occurrences (FIMO) (http://meme-suite.org/tools/fimo), was applied to identify sequences matched to known XRE motifs (Denison et al., 1988; Sogawa et al., 2004). DNA sequences of the two known XRE motifs were used as templates for putative XRE scanning, the XRE-I motif with a sequence of "CACGCNA" (Denison et al., 1988) and the XRE-II motif, with a sequence of "$CATGN_6C[T/A]TG$" (Sogawa et al., 2004). Based on known transcriptional activity sites for the *CYP1A1* gene included in ENCODE (https://genome.ucsc.edu/ENCODE), DNA sequences located 5 Kb up- and down-stream of the *CYP1A1* gene as well as across the gene, were analyzed using FIMO to identify putative XREs.

Cell Culture

The "Human Variation Panel" of lymphoblastoid cell lines (LCLs) was obtained from the Coriell Institute (Camden, NJ). The DNA from these 300 LCLs had been genotyped in the Coriell Institute using the Affymetrix Human SNP Array 6.0 (Affymetrix, Santa Clara, CA), and in our laboratory using Illumina HumanHap 550K and HumanExon 510S-Duo BeadChips (Illumina, San Diego, CA) to obtain genome-wide SNP data. Imputation was then performed using 1000 Genomes data. In addition, we generated gene expression data for all of the LCLs by using Affymetrix U133 2.0 Plus GeneChip expression arrays. These genotype and gene expression data were deposited in the National Center for Biotechnology Information Gene Expression Omnibus (GEO accession: GSE23120) (Niu et al., 2010). This LCL model system has previously been used successfully to preform many studies of SNP-dependent gene transcription (Ingle et al.,

2013; Ho et al., 2016; Liu et al., 2017). The LCLs were cultured in RPMI 1640 media containing 15% fetal bovine serum (FBS).

Undifferentiated HepaRG human hepatic cells (HPR101) were obtained from Biopredict (Rennes, France) and were cultured and differentiated into functional hepatocyte-like cells according to the manufacture's protocol. HPR101 cells were grown in William's E media (Gibco, Grand Island, NY) supplemented with 1× GlutaMAX (Gibco, Grand Island, NY) and 10% HepaRG Growth Medium Supplement (Biopredict, Rennes, France) for 10-14 days until confluent. The cells were then switched into HepaRG Differentiation media for 2 weeks. Differentiated HepaRG cells were used to preform transfection studies.

Human SK-N-BE(2) neuroblastoma and U-87 MG glioblastoma cell lines were obtained from the American Type Culture Collection (ATCC). Original stocks of the cells were stored in liquid nitrogen and had not been passaged. Both of these cell lines were cultured in Eagle's Minimum Essential Medium with 10% FBS.

3-Methylcholanthrene (3MC) Treatment

A "prototypic" AHR ligand, 3MC (Sigma, St. Louis, MI), was used to treat LCLs, HepaRG and U87 MG cells to activate AHR signaling pathways. Control cells were treated with dimethyl sulfoxide (DMSO), which had been used to dissolve the 3MC. The DMSO final concentration in the cell culture media was 0.1% (v:v) for all of the control treatments. For CYP1A1 mRNA quantification, LCLs were treated with 3MC at a series of concentrations ranging from 1E-06 to 10 μM for 24 hours and were then harvested for total RNA extraction. For chromatin

immunoprecipitation (ChIP) assays, cells were incubated with 0.1 µM 3MC for 2 hours and were then fixed by exposure to 1% formaldehyde. For reporter gene assays, cells were treated with 0.1 µM 3MC for 24 hours before harvest for use in subsequent experiments.

mRNA Quantification

mRNA levels were determined by quantitative reverse transcription PCR (qRT-PCR) using the Power SYBR™ Green RNA-to-CT™ 1-Step Kit (Applied Biosystems Inc., Foster City, CA). Total RNA was extracted from cells with the Quick-RNA™ MiniPrep Plus kit (Zymo Research, Irvine, CA). A total of 100 ng of total RNA was used in each reaction. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) mRNA expression was measured for use as an internal control. Specific primers for CYP1A1, AHR, GAPDH mRNA as well as other genes were purchased from Integrated DNA Technologies (Coralville, IA). Primer sequences for those genes are listed in **Supplemental Table 1**.

Chromatin Immunoprecipitation (ChIP) Assay

ChIP assays were conducted for LCLs using the EpiTect® ChIP OneDay Kit (QIAGEN, Germantown, MD). In brief, approximately 5E-06 cells were treated with 0.1 µM 3MC for 2 hours. The same number of cells that had been treated with DMSO was used as a control. After the cells were cross-linked and harvested, 400 µL of lysis buffer was added to lyse the cells. DNA was sheared by sonication into 200~1000 base pair (bp) fragments, with an average size of 500 bp. Aliquots of 100 µL of sheared chromatin/DNA was used for immunoprecipitation with an AHR antibody (Cat#: 83200S; Cell Signaling, Danvers, MA) or IgG (Cat#: 2729S; Cell Signaling). Two µL of purified ChIP DNA was added to a final 10 µL qPCR reaction mixture

for the quantification of putative XRE sequences. SYBR® Green ROX™ qPCR Mastermix (QIAGEN, Germantown, MA) was used to amplify ChIP DNA. The qPCR primers used to amplify specific DNA fragments including putative XREs and SNPs are listed in **Supplemental Table 2.** Functional XREs were defined on the basis of greater than 2-fold enrichment when compared with DMSO treated cells after exposure to AHR antibody in 3MC treated cells.

Transient Transfection

LCLs were transfected with DNA plasmids for reporter gene assay or with siRNA for gene knockdown (KD) by electroporation using the Amaxa® Cell Line Nucleofector® Solution V Kit (Lonza, Cologne, Germany). Specifically, LCLs were maintained at 1E-06 cells/mL for 2 days before use. For each transfection, 1E-06 cells were mixed with 2µg of plasmid DNA, or 100 pmol of siRNA, in 100 uL of Nucleofector® Solution V and were incubated in cuvettes for 5 min without air bubbles. Sample suspensions in cuvettes were then "pulsed" on a Nucleofector® II Device (Lonza, Cologne, Germany) with program X-001. After the addition to the cuvette of 500 µL of pre-equilibrated culture media, samples were gently transferred to culture plates and were incubated at 37 °C overnight. Cells were then treated with 3MC as described above before use. SK-N-BE(2) cells were transfected in the same fashion as LCLs using the Nucleofector® II Device, but with the X-005 program. HepaRG and U87 MG cells were transfected with reporter gene constructs and/or siRNA by using the Lipofectamine™ 3000 Transfection Reagent (ThermoFisher, Grand Island, NY). Twenty four hours after transfection, cell culture media was replaced with fresh media containing 0.1 µM 3MC for an additional 24 hours. The cells were then harvested for the extraction of total RNA and for luciferase activity.

Reporter Gene Assay

To determine SNP effects on *CYP1A1* transcription, an approximately 2.7 Kb DNA segment that included sequence upstream (5') and exon 1 of *CYP1A1* were sub-cloned upstream of the *luc2* gene (encoding firefly luciferase) in the pGL4-Basic vector (Promega, Madison, WI). DNA sequences of the primers used for sub-cloning were 5'-ttgcCTCGAGTAATTGGACGGAGCAAAAGG-3' (forward primer with the *Xho*I restriction enzyme site underlined) and 5'-gcgAAGCTTAGAAAGGGCAAGCCAGAAGT-3' (reverse primer with the *Hind*III restriction enzyme site underlined). Genomic DNA from LCLs with known genotypes for the SNP of interest was used as template for PCR reactions designed to amplify 2.7 Kb DNA fragments containing wildtype (WT) or variant (V) SNP alleles. After confirmation of the sequence by DNA sequencing, reporter gene constructs containing WT or V SNP genotypes were transfected into cells as described above. A pRL-CMV vector (Promega, Madison, WI) which expresses *renilla* luciferase was co-transfected as an internal control. After 3MC treatment, the cells were lysed, and luciferase activity was determined using the Dual-Luciferase® Reporter Assay System (Promega, Madison, WI).

Electrophoretic Mobility Shift Assays (EMSA) and Mass Spectrometry

To determine whether the DNA sequence containing the SNP of interest might bind to nuclear proteins, EMSA was performed with nuclear extracts from LCLs after treatment with 3MC or DMSO. DNA oligonucleotides (Oligoes) 21 nucleotides in length with the SNP in the middle were synthesized by Integrated DNA Technologies (Coralville, IA). The DNA sequences of oligonucleotides for the SNP of interest are listed in **Supplemental Table 3**. A similar set of oligonucleotides was synthesized with a biotin label at the 5'-end. Double-stranded (ds) DNA fragments were obtained by annealing "sense" and "anti-sense" oligonucleotides, and those

DMD Fast Forward. Published on July 6, 2018 as DOI: 10.1124/dmd.118.082164
This article has not been copyedited and formatted. The final version may differ from this version.

dsDNA fragments were incubated with LCL nuclear extracts by using the LightShift™ EMSA Optimization & Control Kit (Thermo Scientific, Rockford, IL). Nuclear extracts from LCLs treated with 3MC were prepared by using the Nuclear Extract Kit (Active Motif, Carlsbad, CA). After incubation, the dsDNA-protein complexes were separated in parallel using two 5% Mini-PROTEAN® TBE Precast Gels (Bio-Rad, Hercules, CA). After separation, one gel was used to transfer dsDNA fragments to Biodyne® B Pre-Cut Modified Nylon Membranes, 0.45µm (Thermo Scientific, Rockford, IL). Biotin labeled dsDNA on the membranes was then visualized using the Chemiluminescent Nucleic Acid Detection Module (Thermo Scientific, Rockford, IL). The other gel was stained using the Pierce™ Silver Stain Kit (Thermo Scientific, Rockford, IL). EMSA bands were then excised from the gel and sent for mass spectrometry (MS) analysis. Sample processing and MS assays were performed by the Taplin Biological Mass Spectrometry Facility at Harvard Medical School.

Co-Immunoprecipitation (Co-IP)

LCLs with homozygous variant genotypes (*V/V*) for the rs2470893 SNP (n=3) were treated with 3MC for 24 hours. Similar amount of each LCL, approximately 10^7 cells, were mixed for total protein extraction using NETN buffer (100 mM NaCl, 20 mM Tris–Cl (pH 8.0), 0.5 mM EDTA, and 0.5% Nonidet P-40) with protease and phosphatase inhibitor cocktails (Roche). Co-IP was then performed as described previously (Yu et al., 2018). Antibodies against AHR, RCC1, NAP1L4 and DDX39B, or control IgG were used to perform Co-IP and Western blot analysis. Reverse Co-IP was performed using antibodies directed against RCC1, NAP1L4 and UAP56, respectively. AHR antibody was used for the Western blot analysis. Information with regard to the antibodies used in the Co-IP and Western blot analyses is listed in **Supplemental Table 4**.

Statistical Analysis

Data analyses were performed and graphs were plotted using GraphPad Prism (GraphPad Software, La Jolla, CA). Statistical comparisons were made using Student's *t*-test or One-way ANOVA.

RESULTS:

Identification of Putative XREs and SNP Selection

To test the hypothesis that SNPs at distance from XREs in the *CYP1A1* gene might influence its induction, we first identified putative XREs near and across the gene by "scanning" DNA sequences across and 5 Kb up- and down-stream of the *CYP1A1* gene for sequences that match known XRE sequence motifs. Fifteen "putative" XREs were identified, of which 7 matched XRE-I and 8 matched XRE-II sequences (see **Supplemental Table 5**). We failed to identify any SNP that either disrupted or created XRE sequences. Since our purpose was to determine the possible function of SNPs at a distance from XREs, SNPs located ±500 bp of these putative XREs were included in the studies described subsequently (**Supplemental Table 5**). We chose to study SNPs that mapped ±500 bp of an XRE because those SNPs could be studied with the greatest accuracy with regard to AHR-XRE binding by the application of ChIP-qPCR assay with the amplification of both the XRE and SNP sequences (Ingle et al., 2013; Liu et al., 2017). Specifically, during ChIP assays, the chromatin is often sheared into fragments 200~1000 bp in length, with an average size of 500 bp (Nelson et al., 2006; Dahl and Collas, 2008; Saleh et al., 2008; Carey et al., 2009). Chromatin fragments 400–500 bp in length have proven to be most suitable for PCR-based ChIP assays since they cover two to three nucleosomes (Dahl and Collas, 2008). This length of the chromatin fragments limited our SNP selection since only fragments with AHR bound to the XRE would be precipitated by the AHR antibody and only SNP(s) within the same chromatin fragment would be detected. SNPs located further away from the XRE being studied would not be precipitated. To determine possible associations of SNPs with CYP1A1 induction, we used a human lymphoblastoid cell line (LCL) model system consisting of cells from 300 individuals of three different ethnic groups, for which genome-wide SNP

DMD Fast Forward. Published on July 6, 2018 as DOI: 10.1124/dmd.118.082164
This article has not been copyedited and formatted. The final version may differ from this version.

genotype and mRNA expression data had been generated in our laboratory (Niu et al., 2010). Specifically, LCLs with WT or homozygous variant genotypes for the SNPs of interest were treated with an AHR agonist (3MC), followed by determination of CYP1A1 mRNA levels. We excluded SNPs having a minor allele frequency (MAF) < 0.10 because it would be practically difficult to study those SNPs as a result of the limited number of LCLs homozygous for variant SNP genotypes. There were four SNPs within the area studied that met these criteria, including the *CYP1A1 MspI* variant, and three other SNPs that mapped 5' of the *CYP1A1* gene (**Table 1**).

SNP-dependent CYP1A1 **Induction**

To determine whether any of these four SNPs (**Table 1**) might be associated with variation in CYP1A1 induction, we chose 14 LCLs based on their genotypes for these four SNPs (**Supplemental Table 6**). Since a previous GWAS that we had performed for symptom severity of major depressive disorder had shown that the *AHR* rs17137566 SNP was an eQTL in European populations (Liu et al., 2018), and a previous study had reported that the *AHR* rs2066853 SNP appeared to influence CYP1A1 induction (Smart and Daly, 2000), we excluded LCLs that were homozygous variant genotype for those two *AHR* SNPs to avoid significant differences in baseline AHR activities in the 14 LCLs that we chose to study. These 14 LCLs were then treated for 24 hours with 3MC at concentrations that ranged from 1E-06 to 10 μM. CYP1A1 mRNA levels were then determined by qRT-PCR. We found that the induction of CYP1A1 mRNA differed by genotype for the rs2470893 SNP (**Fig. 1**). This SNP was located 1,570 bp from the 5'-end of the *CYP1A1* transcription-start site, 196 bp away from a putative XRE (**Fig. 1A**). CYP1A1 induction in LCLs with homozygous variant genotypes (*V/V*) for the rs2470893 SNP was significantly higher than in LCLs homozygous for the WT genotype (*W/W*)

(**Fig. 1B**). Genotypes for the three other SNPs listed in **Table 1**, either individually or in combination, were not associated with differences in CYP1A1 induction after 3MC treatment (data not shown). To determine whether this rs2470893 SNP-dependent difference in CYP1A1 induction might be caused by differences in AHR binding to the nearby putative XRE sequence, ChIP assays performed with anti-AHR antibody were performed using LCLs with *W/W* and *V/V* genotypes after 2 hours of 3MC treatment. A significant increase in enrichment of SNP/XRE DNA fragments was observed in LCLs with the *V/V* genotype but not those with the *W/W* genotype (**Fig. 1C**). Except for the rs2470893 SNP, we did not observe a significant SNP-dependent difference in enrichment for the other three SNP/XRE fragments in ChIP assays (data not shown).

We selected LCLs for these initial studies of possible SNP-dependent CYP1A1 induction because of the availability of genome-wide SNP genotype information for these LCLs, which made it possible for us to select an adequate number of LCLs with the SNPs of interest for study. To determine whether the SNP-dependent CYP1A1 induction that we observed in LCLs might also exist in other cell lines, we created reporter gene constructs that included approximately 2.7 Kb of the *CYP1A1* promoter with WT and variant genotypes for the rs2470893 SNP (**Fig. 2A**). Sequences of the WT and variant reporter gene constructs were confirmed by DNA sequencing, which showed that they differed only in the single nucleotide that defined the rs2470893 SNP. These reporter gene constructs were then transfected into different cell lines to compare CYP1A1 transcriptional activities for WT and variant SNP genotypes in those cells.

The reporter gene constructs were first transfected into LCLs and those cell lines were then treated with 3MC for 24 hours. LCLs transfected with the variant construct showed 1.5-fold higher CYP1A1 transcription activity than did cells transfected with the WT construct (**Fig. 2B**). The same constructs were then transfected into HepaRG cells, a cell line of hepatocyte origin that retains many of the characteristics of human hepatocytes, including expression of most of the CYP enzymes (Guillouzo et al., 2007; Kanebratt and Andersson, 2008). The variant genotype construct once again showed significantly greater CYP1A1 transcriptional activity than did the WT construct (**Fig. 2C**). Since we had previously reported that AHR has a function in the regulation of kynurenine pathway genes in cells of central nervous system (CNS) origin (Liu et al., 2018), the reporter gene constructs were also transfected into a neuroblastoma cell line, SK-N-BE(2), and a glioblastoma cell line, U-87 MG. A significant increase in transcriptional activity was observed in SK-N-BE(2) cells transfected with the variant construct when compared with that for the WT construct (**Fig. 2D**) while no significant differences were observed in U-87 MG (**Fig. 2E**). These reporter gene assay results suggested that rs2470893 SNP-dependent CYP1A1 induction can be cell line-dependent.

Identification of Possible rs2470893 SNP-binding Proteins

To test potential mechanisms for the rs2470893 SNP-dependent CYP1A1 induction, EMSA was performed to identify potential AHR "co-regulator(s)" that might also bind to the SNP region (**Fig. 3A**). Specifically, nuclear extract from LCLs treated with 3MC was incubated with 21 bp DNA fragments with the SNP at middle of the sequence. DNA fragments with variant SNP genotype showed stronger binding to nuclear proteins than did those with the WT genotype (**Fig. 3B**). We then analyzed the EMSA bands by mass spectrometry (MS) in an attempt to identify

proteins bound to the DNA fragments. “Blank” samples excised from the margins of gel were also submitted for MS analysis as controls. The proteins identified in each sample by MS are listed in **Supplemental Table 7**. Proteins present in “blank” samples, and proteins that are commonly seen and, as a result might represent nonspecific binding due to their abundance in cells (Malovannaya et al., 2010; Hodge et al., 2013) were not pursued. Since MS identifies proteins based on the identity of peptide fragments, proteins homologous to those present in “blank”-samples were also not included in subsequent studies. In addition, proteins encoded by pseudogenes and proteins that are not known to be present in the cell nucleus were also excluded from further study. As a result, a total of 28 candidate proteins, including 19 proteins in common in both WT and variant samples and 9 proteins identified only in the variant sample, were selected for further study of their possible involvement in SNP-specific CYP1A1 induction (**Fig. 3C**).

To identify which of the rs2470893 SNP-binding proteins might influence *CYP1A1* induction in a SNP-dependent fashion, we knocked down (KD) each of the 28 candidate genes encoding these proteins in an LCL with homozygous variant rs2470893 SNP genotype, followed by the quantification of CYP1A1 mRNA expression after treatment with vehicle (DMSO) or 3MC. The effects of the KD of AHR were also determined as a positive control. CYP1A1 mRNA levels after the KD of AHR and the 28 candidate genes were then normalized to the negative control KD (Neg. siRNA). After treatment with either DMSO or 3MC, CYP1A1 mRNA levels were dramatically decreased after the KD of AHR when compared with the control KD (**Fig. 3D**). Meanwhile, KD of several of the candidate genes also resulted in a decrease in CYP1A1 mRNA levels (**Fig. 3D**). To determine which gene KD might have an effect similar to that of AHR KD,

DMD Fast Forward. Published on July 6, 2018 as DOI: 10.1124/dmd.118.082164
This article has not been copyedited and formatted. The final version may differ from this version.

CYP1A1 mRNA expression levels after the KD of each gene by siRNA transfection were plotted with results for treatment with the DMSO vehicle on the x-axis and results for treatment with 3MC on the y-axis (**Fig. 3E**). Each dot in the plot shown graphically in **Fig. 3E** represents results for gene KD by siRNA transfection, with its position in this figure dependent on CYP1A1 mRNA level after treatment with either DMSO and 3MC. The dot for AHR KD (siAHR) was closest to the origin (**Fig. 3E**), as anticipated, because CYP1A1 mRNA expression was dramatically decreased after the KD of AHR after both treatment with the vehicle, DMSO, and 3MC. CYP1A1 mRNA levels after the KD of Nucleosome Assembly Protein 1 Like 4 (NAP1L4), Regulator of Chromosome Condensation 1 (RCC1) or DExD-Box Helicase 39B (DDX39B) were close to the position of results seen after AHR KD (**Fig. 3E**), raising the possibility that those three proteins might be involved in the AHR/ARNT transcriptional complex which binds near the rs2470893 SNP.

NAP1L4, RCC1 and DDX39B Contribute to rs2470893 SNP-Dependent CYP1A1 **Induction**

To determine whether NAP1L4, RCC1 and DDX39B proteins might contribute to SNP-dependent CYP1A1 induction, KD of those three genes was repeated in LCLs with homozygous variant (*V/V*) and WT (*W/W*) rs2470893 SNP genotypes, followed by CYP1A1 mRNA quantification. Although CYP1A1 mRNA expression was significantly increased after treatment with 3MC when compared with DMSO vehicle treatment, KD of those three genes, as well as the KD of AHR, significantly decreased CYP1A1 induction when compared with control KD in LCLs with a *V/V* genotype (**Fig. 4A,** left). After plotting CYP1A1 mRNA levels after KD of those three candidate genes in LCLs homozygous for *V/V* SNP genotypes with treatments that consisted of DMSO vehicle shown on the x-axis and 3MC treatment results shown on the y-axis,

DMD Fast Forward. Published on July 6, 2018 as DOI: 10.1124/dmd.118.082164
This article has not been copyedited and formatted. The final version may differ from this version.

dots for KD of those three genes were once again located near the dot showing results for AHR KD (**Fig. 4A,** right). However, in LCLs homozygous for the *W/W* rs2470893 SNP genotype, KD of those three proteins did not change CYP1A1 mRNA levels induced by 3MC treatment (**Fig. 4B,** left). In that case, dots depicting results for KD of those three genes were close to that for the non-targeting control KD (**Fig. 4B,** right), indicating that the NAP1L4, RCC1 and DDX39B proteins did not appear to significantly affect CYP1A1 induction in LCLs with *W/W* genotypes for the rs2470893 SNP.

Using reporter gene constructs for WT and variant SNP genotypes for the rs2470893 SNP, we had demonstrated SNP-dependent CYP1A1 transcriptional activity in HepaRG cells (**Fig. 2B**). To determine whether NAP1L4, RCC1 and DDX39B might influence CYP1A1 induction in HepaRG cells, these reporter gene assays were performed once again after separate KD of each of those three genes. In the control KD group (Neg. siRNA), CYP1A1 transcriptional activity was induced by 3MC treatment and the activity was once again significantly increased in cells transfected with the variant reporter gene construct when compared with results for the WT construct (**Supplemental Fig. 1A**). As expected, KD of AHR (siAHR) repressed *CYP1A1* transcriptional activity for both WT and variant reporter gene constructs and no significant difference in 3MC-induced CYP1A1 transcriptional activity was found between WT and V reporter gene constructs (**Supplemental Fig. 1A**). Furthermore, compared to control KD, KD of NAP1L4, RCC1 or DDX39B in HepaRG cells also significantly repressed CYP1A1 transcriptional activity and, as in the case of AHR KD, no differences in CYP1A1 transcriptional activity were observed between WT and variant reporter gene constructs (**Supplemental Fig. 1A**). Finally, AHR, RCC1, NAP1L4 and DDX39B mRNA levels were decreased significantly

DMD Fast Forward. Published on July 6, 2018 as DOI: 10.1124/dmd.118.082164
This article has not been copyedited and formatted. The final version may differ from this version.

after specific siRNA transfections, as measured by qRT-PCR (**Supplemental Fig. 1B**). These results also indicated that NAP1L4, RCC1 and DDX39B might play a role in the SNP-dependent induction of CYP1A1 in HepaRG cells.

RCC1 and NAP1L4 Interact Physically with AHR

To determine whether NAP1L4, RCC1 or DDX39B might physically interact with the AHR transcriptional complex, co-immunoprecipitation (Co-IP) with anti-AHR antibody was performed using LCLs treated with 3MC. After precipitation with anti-AHR antibody, the samples were analyzed by Western blot analysis to determine whether NAP1L4, RCC1 or DDX39B might precipitate with AHR. As shown in **Fig. 5A**, RCC1 and NAP1L4 proteins were precipitated by the anti-AHR antibody, but DDX39B protein was not. To further confirm protein-protein interactions with AHR, "reverse" co-IP was performed using antibodies against NAP1L4, RCC1 and DDX39B, respectively, followed by detection performed by AHR precipitation. AHR was detected after precipitation with anti-RCC1 (**Fig. 5B**) and NAP1L4 (**Fig. 5C**) antibodies but not after precipitation with anti-DDX39B antibody (data not shown). These Co-IP studies confirmed that RCC1 and NAP1L4 could physically interact with AHR. When joined with the EMSA assay results, CYP1A1 mRNA expression results and the luciferase reporter gene assay results, these observations suggest that RCC1 and NAP1L4 proteins might interact with the rs2470893 SNP and enhance the binding of AHR to the nearby XRE, thus contributing to the regulation of CYP1A1 expression.

DISCUSSION:

Based on our previous observations with respect to the modulation of ERα-ERE binding and downstream gene transcription (Ingle et al., 2013; Ho et al., 2016; Ingle et al., 2016; Liu et al., 2017; Qin et al., 2017), in the present study we set out to test the hypothesis that SNPs hundreds of bp distant from XREs might influence AHR-XRE binding and downstream gene expression. We chose to study *CYP1A1*, a prototypical AHR regulated gene, by analyzing DNA sequences across and both up and downstream of the open reading frame for putative XREs and common SNPs that mapped ±500 bp of those XREs. We identified one SNP, rs2470893, that was located 5' of *CYP1A1*, was associated with significant variation in AHR binding to a nearby (196 bp distant) XRE and with CYP1A1 transcription in LCLs (**Fig. 1**). This rs2470893 SNP not only influenced CYP1A1 transcriptional activity in LCLs but also appeared to be functionally significant in cell lines of different tissue origins, including the hepatocyte-like HepaRG cells (**Fig. 2**). We then performed a series of studies to investigate possible molecular mechanisms for distant-SNP modulated AHR-XRE binding and gene transcription. Three proteins, NAP1L4, RCC1, and DDX39B were identified by mass spectrometry following EMSA assay that could interact with DNA sequences that included the rs2470893 SNP and which were associated with AHR-dependent and SNP-dependent CYP1A1 induction. Two of those proteins also interacted physically with AHR, which indicates that they may be involved in the mechanism underlying distant-SNP modulated CYP1A1 transcription. Those findings provide, to our knowledge, the first example of SNPs that are near but not in an XRE and which can regulate AHR-XRE binding and downstream gene transcription.

Because of its important role in the metabolism of environmental toxins and procarcinogens, *CYP1A1* genetic polymorphisms have been studied for their possible association with risk of the

occurrence of cancer, but the results have often been contradictory. Those studies mainly focused on two common *CYP1A1* SNPs, rs4646903 (also known as *MspI* mutant or *CYP1A1*2A*) and the rs1048943 (Ile462Val) SNP. However, the molecular mechanism by which those two SNPs might function in the regulation of CYP1A1 activity remains unclear, which has hampered our understanding of the possible relationship between *CYP1A1* genetic polymorphisms and clinical phenotypes. By systematically testing a novel mechanism by which SNPs at a distance from XREs might modulate AHR function and CYP1A1 induction, we have demonstrated that the *CYP1A1* rs2470893 SNP variant genotype results in enhanced AHR binding to a nearby XRE, with increased CYP1A1 transcription after 3MC treatment. Our results fit with an observation reported previously that the rs2470893 SNP was significantly associated with circulating polychlorinated biphenyl (PCB) 118 levels in a general elderly population (n=1016), with the variant genotype being associated with a decrease in PCB118 concentrations (Lind et al., 2014). PCBs are AHR agonists and they are metabolized by CYP1A1. The decrease in PCB118 plasma concentrations in subjects who carry the rs2470893 variant genotype might be a consequence of enhanced induction of CYP1A1 expression.

The *CYP1A1* and *CYP1A2* genes are located “head to head” and, as a result, they share the same 5’ region on chromosome 15. The rs2470893 SNP is located between the two genes, approximately 1.5 Kb upstream of *CYP1A1* and 29.5 Kb upstream of the *CYP1A2* transcription start site. CYP1A2 is the major enzyme that metabolizes caffeine (Butler et al., 1989; Berthou et al., 1991). Several large-cohort GWAS studies for coffee consumption have identified rs2470893 as one of the “top hits” for association with that phenotype, with the variant genotype being associated with less coffee consumption (Cornelis et al., 2011; Amin et al., 2012) and decreased caffeine metabolism (Cornelis et al., 2016). Caffeine was not found to induce CYP1A2 activity

in human hepatocytes (Vaynshteyn and Jeong, 2012) but it did repress CYP1A1 expression in LCLs (Amin et al., 2012). It would be interesting to determine whether the rs2470893 variant genotype might enhance the caffeine effect on CYP1A2 repression and decrease caffeine metabolism. However, since CYP1A2 is not expressed in LCLs (data not shown), we were not able to study possible effects of the SNP on *CYP1A2* transcription in our cell line-based model system. Future studies using hepatocytes from individuals with different SNP genotypes might help to clarify the possible role of this SNP, or other genetically linked SNPs, in caffeine effect on CYP1A2.

Functional genomic studies have been most highly focused on genetic variation that maps to exons that alters the encoded protein amino acid sequence or within canonical splice sequences that alter RNA splicing. However, the majority of GWAS top hit SNP signals map to introns or intergenic regions, which indicates that non-coding SNPs play an important functional role, a role that often involves variation in transcription regulation (Battle et al., 2017; Weinshilboum and Wang, 2017; Wang et al., 2018). Gene transcription is often initiated by transcription factor binding to specific DNA sequence motifs, with the AHR/ARNT heterodimer binding to XREs, following by the recruitment of co-regulators and RNA polymerase to form a transcriptional complex (Malovannaya et al., 2011). We have previously reported that proteins encoded by members of the testis specific Y-like (*TSPYL*) gene family were co-regulators for the transcription of cytochrome P450s, including CYP3A4, CYP2C9, CYP2C19, CYP17A1 and CYP19A1 (Liu et al., 2013; Qin et al., 2018). Co-regulators like members of the TSPYL gene family may, among other mechanism, bind to specific DNA sequences, alter the nucleosome structure or recruit RNA polymerase, thus facilitating gene transcription (Qin et al., 2017). Therefore, changes in the binding of co-regulators to DNA sequences as a result of SNPs that

map hundreds of bp from response elements might well affect downstream gene transcription. In the present study, we identified NAP1L4 and RCC1 as possible AHR co-regulators that interact with rs2470893 SNP sequences and influence CYP1A1 transcription in LCLs and in HepaRG cells. NAP1L4, also known as nucleosome assembly protein-2, is a histone chaperone protein that can transfer histones onto naked DNA templates (Rodriguez et al., 1997) and it functions within complexes with other proteins that are distinct from histones (Rodriguez et al., 2004). RCC1 is a chromatin-bound protein that binds directly with the nucleosome (Makde et al., 2010). Whether those two proteins regulate AHR transcriptional activity though interaction with nucleosomes needs to be determined in the course of future studies. Finally, the DDX39B protein, which was shown to bind the rs2470893 SNP sequence and affect CYP1A1 transcriptional activity (**Fig. 3D & E**), did not physically interact with AHR (**Fig. 5A**). Whether DDX39B might interact with other proteins recruited to the AHR-ARNT transcription complex needs to be determined.

In summary, by testing the hypothesis that SNPs distant from XRE motifs can modulate AHR transcriptional activity, we have demonstrated that the rs2470893 SNP modulates AHR-dependent CYP1A1 transcription. This “distant SNP” modulation of AHR transcriptional activity might be a mechanism that, in part, explains variation in the induction of CYP1A1 and, perhaps other SNPs that influence the expression of other AHR-regulated genes. RNA sequencing and ChIP sequencing using LCLs and other genome-data rich cell lines might result in the discovery of additional SNPs distant from XREs that modulate AHR-dependent gene transcription.

Authorship Contributions:

Participated in research design: Liu, Qin, Ray, Weinshilboum

Conducted experiments: Liu, Qin, Kalari

Contributed new reagents and analytical procedures: Liu, Qin, Wang, Weinshilboum

Performed data analysis: Liu, Qin, Kalari, Weinshilboum

Wrote or contributed to the writing of the manuscript: Liu, Weinshilboum

REFERENCES:

Amin N, Byrne E, Johnson J, Chenevix-Trench G, Walter S, Nolte IM, Vink JM, Rawal R, Mangino M, Teumer A, Keers JC, Verwoert G, Baumeister S, Biffar R, Petersmann A, Dahmen N, Doering A, Isaacs A, Broer L, Wray NR, Montgomery GW, Levy D, Psaty BM, Gudnason V, Chakravarti A, Sulem P, Gudbjartsson DF, Kiemeney LA, Thorsteinsdottir U, Stefansson K, van Rooij FJ, Aulchenko YS, Hottenga JJ, Rivadeneira FR, Hofman A, Uitterlinden AG, Hammond CJ, Shin SY, Ikram A, Witteman JC, Janssens AC, Snieder H, Tiemeier H, Wolfenbuttel BH, Oostra BA, Heath AC, Wichmann E, Spector TD, Grabe HJ, Boomsma DI, Martin NG, and van Duijn CM (2012) Genome-wide association analysis of coffee drinking suggests association with CYP1A1/CYP1A2 and NRCAM. *Mol Psychiatry* **17:**1116-1129.

Anttila S, Tuominen P, Hirvonen A, Nurminen M, Karjalainen A, Hankinson O, and Elovaara E (2001) CYP1A1 levels in lung tissue of tobacco smokers and polymorphisms of CYP1A1 and aromatic hydrocarbon receptor. *Pharmacogenetics* **11:**501-509.

Battle A, Brown CD, Engelhardt BE, and Montgomery SB (2017) Genetic effects on gene expression across human tissues. *Nature* **550:**204-213.

Berthou F, Flinois JP, Ratanasavanh D, Beaune P, Riche C, and Guillouzo A (1991) Evidence for the involvement of several cytochromes P-450 in the first steps of caffeine metabolism by human liver microsomes. *Drug Metab Dispos* **19:**561-567.

Butler MA, Iwasaki M, Guengerich FP, and Kadlubar FF (1989) Human cytochrome P-450PA (P-450IA2), the phenacetin O-deethylase, is primarily responsible for the hepatic 3-demethylation of caffeine and N-oxidation of carcinogenic arylamines. *Proc Natl Acad Sci U S A* **86:**7696-7700.

Carey MF, Peterson CL, and Smale ST (2009) Chromatin immunoprecipitation (ChIP). *Cold Spring Harb Protoc* **2009:**pdb prot5279.

Cornelis MC, Kacprowski T, Menni C, Gustafsson S, Pivin E, Adamski J, Artati A, Eap CB, Ehret G, Friedrich N, Ganna A, Guessous I, Homuth G, Lind L, Magnusson PK, Mangino M, Pedersen NL, Pietzner M, Suhre K, Volzke H, Bochud M, Spector TD, Grabe HJ, and Ingelsson E (2016) Genome-wide association study of caffeine metabolites provides new insights to caffeine metabolism and dietary caffeine-consumption behavior. *Hum Mol Genet* **25:**5472-5482.

Cornelis MC, Monda KL, Yu K, Paynter N, Azzato EM, Bennett SN, Berndt SI, Boerwinkle E, Chanock S, Chatterjee N, Couper D, Curhan G, Heiss G, Hu FB, Hunter DJ, Jacobs K, Jensen MK, Kraft P, Landi MT, Nettleton JA, Purdue MP, Rajaraman P, Rimm EB, Rose LM, Rothman N, Silverman D, Stolzenberg-Solomon R, Subar A, Yeager M, Chasman DI, van Dam RM, and Caporaso NE (2011) Genome-wide meta-analysis identifies regions on 7p21 (AHR) and 15q24 (CYP1A2) as determinants of habitual caffeine consumption. *PLoS Genet* **7:**e1002033.

Dahl JA and Collas P (2008) A rapid micro chromatin immunoprecipitation assay (microChIP). *Nat Protoc* **3:**1032-1045.

Denison MS, Fisher JM, and Whitlock JP, Jr. (1988) The DNA recognition site for the dioxin-Ah receptor complex. Nucleotide sequence and functional analysis. *J Biol Chem* **263:**17221-17224.

Denison MS and Nagy SR (2003) Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annu Rev Pharmacol Toxicol* **43:**309-334.

Guillouzo A, Corlu A, Aninat C, Glaise D, Morel F, and Guguen-Guillouzo C (2007) The human hepatoma HepaRG cells: a highly differentiated model for studies of liver metabolism and toxicity of xenobiotics. *Chem Biol Interact* **168:**66-73.

Ho MF, Bongartz T, Liu M, Kalari KR, Goss PE, Shepherd LE, Goetz MP, Kubo M, Ingle JN, Wang L, and Weinshilboum RM (2016) Estrogen, SNP-Dependent Chemokine Expression and Selective Estrogen Receptor Modulator Regulation. *Mol Endocrinol* **30:**382-398.

Hodge K, Have ST, Hutton L, and Lamond AI (2013) Cleaning up the masses: exclusion lists to reduce contamination with HPLC-MS/MS. *J Proteomics* **88:**92-103.

Ingle JN, Liu M, Wickerham DL, Schaid DJ, Wang L, Mushiroda T, Kubo M, Costantino JP, Vogel VG, Paik S, Goetz MP, Ames MM, Jenkins GD, Batzler A, Carlson EE, Flockhart DA, Wolmark N, Nakamura Y, and Weinshilboum RM (2013) Selective estrogen receptor modulators and pharmacogenomic variation in ZNF423 regulation of BRCA1 expression: individualized breast cancer prevention. *Cancer Discov* **3:**812-825.

Ingle JN, Xie F, Ellis MJ, Goss PE, Shepherd LE, Chapman JW, Chen BE, Kubo M, Furukawa Y, Momozawa Y, Stearns V, Pritchard KI, Barman P, Carlson EE, Goetz MP, Weinshilboum RM, Kalari KR, and Wang L (2016) Genetic Polymorphisms in the Long Noncoding RNA MIR2052HG Offer a Pharmacogenomic Basis for the Response of Breast Cancer Patients to Aromatase Inhibitor Therapy. *Cancer Res* **76:**7012-7023.

Inoue H, Kiyohara C, Marugame T, Shinomiya S, Tsuji E, Handa K, Hayabuchi H, Onuma K, Hamada H, Koga H, and Kono S (2000) Cigarette smoking, CYP1A1 MspI and GSTM1 genotypes, and colorectal adenomas. *Cancer Res* **60:**3749-3752.

Kanebratt KP and Andersson TB (2008) Evaluation of HepaRG cells as an in vitro model for human drug metabolism studies. *Drug Metab Dispos* **36:**1444-1452.

Kawajiri K, Nakachi K, Imai K, Yoshii A, Shinoda N, and Watanabe J (1990) Identification of genetically high risk individuals to lung cancer by DNA polymorphisms of the cytochrome P450IA1 gene. *FEBS Lett* **263:**131-133.

Kiyohara C, Nakanishi Y, Inutsuka S, Takayama K, Hara N, Motohiro A, Tanaka K, Kono S, and Hirohata T (1998) The relationship between CYP1A1 aryl hydrocarbon hydroxylase activity and lung cancer in a Japanese population. *Pharmacogenetics* **8:**315-323.

Lind L, Penell J, Syvanen AC, Axelsson T, Ingelsson E, Morris AP, Lindgren C, Salihovic S, van Bavel B, and Lind PM (2014) Genetic variation in the CYP1A1 gene is related to circulating PCB118 levels in a population-based sample. *Environ Res* **133:**135-140.

Liu D, Ho MF, Schaid DJ, Scherer SE, Kalari K, Liu M, Biernacka J, Yee V, Evans J, Carlson E, Goetz MP, Kubo M, Wickerham DL, Wang L, Ingle JN, and Weinshilboum RM (2017) Breast cancer chemoprevention pharmacogenomics: Deep sequencing and functional genomics of the ZNF423 and CTSO genes. *npj Breast Cancer* **3:**30.

Liu D, Ray B, Neavin DR, Zhang J, Athreya AP, Biernacka JM, Bobo WV, Hall-Flavin DK, Skime MK, Zhu H, Jenkins GD, Batzler A, Kalari KR, Boakye-Agyeman F, Matson WR, Bhasin SS, Mushiroda T, Nakamura Y, Kubo M, Iyer RK, Wang L, Frye MA, Kaddurah-Daouk R, and Weinshilboum RM (2018) Beta-defensin 1, aryl hydrocarbon receptor and plasma kynurenine in major depressive disorder: metabolomics-informed genomics. *Transl Psychiatry* **8:**10.

Liu M, Ingle JN, Fridley BL, Buzdar AU, Robson ME, Kubo M, Wang L, Batzler A, Jenkins GD, Pietrzak TL, Carlson EE, Goetz MP, Northfelt DW, Perez EA, Williard CV, Schaid DJ,

Nakamura Y, and Weinshilboum RM (2013) TSPYL5 SNPs: association with plasma estradiol concentrations and aromatase expression. *Mol Endocrinol* **27:**657-670.

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.

Makde RD, England JR, Yennawar HP, and Tan S (2010) Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **467:**562-566.

Malovannaya A, Lanz RB, Jung SY, Bulynko Y, Le NT, Chan DW, Ding C, Shi Y, Yucer N, Krenciute G, Kim BJ, Li C, Chen R, Li W, Wang Y, O'Malley BW, and Qin J (2011) Analysis of the human endogenous coregulator complexome. *Cell* **145:**787-799.

Malovannaya A, Li Y, Bulynko Y, Jung SY, Wang Y, Lanz RB, O'Malley BW, and Qin J (2010) Streamlined analysis schema for high-throughput identification of endogenous protein complexes. *Proc Natl Acad Sci U S A* **107:**2431-2436.

Mimura J and Fujii-Kuriyama Y (2003) Functional role of AhR in the expression of toxic effects by TCDD. *Biochimica et biophysica acta* **1619:**263-268.

Murray IA, Patterson AD, and Perdew GH (2014) Aryl hydrocarbon receptor ligands in cancer: friend and foe. *Nat Rev Cancer* **14:**801-814.

Nelson JD, Denisenko O, and Bomsztyk K (2006) Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat Protoc* **1:**179-185.

Niu N, Qin Y, Fridley BL, Hou J, Kalari KR, Zhu M, Wu TY, Jenkins GD, Batzler A, and Wang L (2010) Radiation pharmacogenomics: a genome-wide association approach to identify radiation response biomarkers using human lymphoblastoid cell lines. *Genome Res* **20:**1482-1492.

Persson I, Johansson I, and Ingelman-Sundberg M (1997) In vitro kinetics of two human CYP1A1 variant enzymes suggested to be associated with interindividual differences in cancer susceptibility. *Biochem Biophys Res Commun* **231:**227-230.

Qin S, Ingle JN, Liu M, Yu J, Wickerham DL, Kubo M, Weinshilboum RM, and Wang L (2017) Calmodulin-like protein 3 is an estrogen receptor alpha coregulator for gene expression and drug response in a SNP, estrogen, and SERM-dependent fashion. *Breast Cancer Res* **19:**95.

Qin S, Liu D, Kohli M, Wang L, Vedell PT, Hillman DW, Niu N, Yu J, and Weinshilboum RM (2018) TSPYL Family Regulates CYP17A1 and CYP3A4 Expression: Potential Mechanism Contributing to Abiraterone Response in Metastatic Castration-Resistant Prostate Cancer. *Clin Pharmacol Ther* **104:**201-210.

Rannug A, Alexandrie AK, Persson I, and Ingelman-Sundberg M (1995) Genetic polymorphism of cytochromes P450 1A1, 2D6 and 2E1: regulation and toxicological significance. *J Occup Environ Med* **37:**25-36.

Riddick DS, Huang Y, Harper PA, and Okey AB (1994) 2,3,7,8-Tetrachlorodibenzo-p-dioxin versus 3-methylcholanthrene: comparative studies of Ah receptor binding, transformation, and induction of CYP1A1. *J Biol Chem* **269:**12118-12128.

Rodriguez P, Munroe D, Prawitt D, Chu LL, Bric E, Kim J, Reid LH, Davies C, Nakagama H, Loebbert R, Winterpacht A, Petruzzi MJ, Higgins MJ, Nowak N, Evans G, Shows T, Weissman BE, Zabel B, Housman DE, and Pelletier J (1997) Functional characterization of human nucleosome assembly protein-2 (NAP1L4) suggests a role as a histone chaperone. *Genomics* **44:**253-265.

Rodriguez P, Ruiz MT, Price GB, and Zannis-Hadjopoulos M (2004) NAP-2 is part of multi-protein complexes in HeLa cells. *J Cell Biochem* **93:**398-408.

Saleh A, Alvarez-Venegas R, and Avramova Z (2008) An efficient chromatin immunoprecipitation (ChIP) protocol for studying histone modifications in Arabidopsis plants. *Nat Protoc* **3:**1018-1025.

Smart J and Daly AK (2000) Variation in induced CYP1A1 levels: relationship to CYP1A1, Ah receptor and GSTM1 polymorphisms. *Pharmacogenetics* **10:**11-24.

Smith GB, Harper PA, Wong JM, Lam MS, Reid KR, Petsikas D, and Massey TE (2001) Human lung microsomal cytochrome P4501A1 (CYP1A1) activities: impact of smoking status and CYP1A1, aryl hydrocarbon receptor, and glutathione S-transferase M1 genetic polymorphisms. *Cancer Epidemiol Biomarkers Prev* **10:**839-853.

Sogawa K, Numayama-Tsuruta K, Takahashi T, Matsushita N, Miura C, Nikawa J, Gotoh O, Kikuchi Y, and Fujii-Kuriyama Y (2004) A novel induction mechanism of the rat CYP1A2 gene mediated by Ah receptor-Arnt heterodimer. *Biochem Biophys Res Commun* **318:**746-755.

Spurr NK, Gough AC, Stevenson K, and Wolf CR (1987) Msp-1 polymorphism detected with a cDNA probe for the P-450 I family on chromosome 15. *Nucleic Acids Res* **15:**5901.

Stucker I, Jacquet M, de Waziers I, Cenee S, Beaune P, Kremers P, and Hemon D (2000) Relation between inducibility of CYP1A1, GSTM1 and lung cancer in a French population. *Pharmacogenetics* **10:**617-627.

Vaynshteyn D and Jeong H (2012) Caffeine induces CYP1A2 expression in rat hepatocytes but not in human hepatocytes. *Drug Metab Lett* **6:**116-119.

Wang L, Ingle J, and Weinshilboum R (2018) Pharmacogenomic Discovery to Function and Mechanism: Breast Cancer as a Case Study. *Clin Pharmacol Ther* **103:**243-252.

Weinshilboum RM and Wang L (2017) Pharmacogenomics: Precision Medicine and Drug Response. *Mayo Clin Proc* **92:**1711-1722.

Yu J, Qin B, Moyer AM, Nowsheen S, Liu T, Qin S, Zhuang Y, Liu D, Lu SW, Kalari KR, Visscher DW, Copland JA, McLaughlin SA, Moreno-Aspitia A, Northfelt DW, Gray RJ, Lou Z, Suman VJ, Weinshilboum R, Boughey JC, Goetz MP, and Wang L (2018) DNA methyltransferase expression in triple-negative breast cancer predicts sensitivity to decitabine. *J Clin Invest* **128:**2376-2388.

Zhang ZY, Fasco MJ, Huang L, Guengerich FP, and Kaminsky LS (1996) Characterization of purified human recombinant cytochrome P4501A1-Ile462 and -Val462: assessment of a role for the rare allele in carcinogenesis. *Cancer Res* **56:**3926-3933.

FOOTNOTES

This work was supported by the National Institutes of Health (NIH) the National Institute of General Medical Sciences (NIGMS) [Grants U19 GM61388, RO1 GM28157 and RO1 GM125633].

FIGURE LEGENDS

Fig.1. A) Schematic depiction of the location of the rs2470893 SNP and the nearby XRE at the 5'-end of the *CYP1A1* gene. Annealing sites for the qPCR primers that were used in the ChIP assay are indicated. **B**) After 24-hours of treatment with 3MC at a series of concentrations, LCLs homozygous for the rs2470893 *V/V* genotype (n = 6) showed significantly higher CYP1A1 mRNA expression than did LCLs homozygous for the *W/W* genotype (n = 8). **C**) ChIP assays performed with anti-AHR antibody showed a 3-fold increase in AHR binding to the XRE 196 bp away from the *V/V* genotype rs2470893 SNP when compared with those homozygous for the *W/W* genotype (n = 3). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Values shown are mean ± SEM.

Fig. 2. Reporter Gene Assays for the *CYP1A1* rs2470893 SNP. **A**) Structure of the *CYP1A1* promotor reporter gene construct within the basic pGL4 plasmid (Promega). The rs2470893 SNP was included in the *CYP1A1* promotor sequence that was sub-cloned upstream of the luciferase reporter gene, *Luc2*. The plasmids contained either WT and variant sequences for the SNP within the *CYP1A1* promoter and were transfected into: **B**) LCLs, **C**) HepaRG, **D**) SK-NB-E(2) and **E**) U87-MG cells. Values shown for **B)**, **C)**, **D)** and **E)** are ratios of relative light units (RLU) compared to the internal reference after 3MC or DMSO vehicle treatment. ** $p < 0.01$, ns = not significant. Values shown are mean ± SEM for 3 determinations. Data for individual assays are shown in **B), C), D)** and **E).**

Fig. 3. A) Schematic depiction of the hypothesis that co-regulators might bind to the SNP and the AHR/ARNT complex to regulate CYP1A1 transcription. **B)** Oligonucleotides 21 bp in length

DMD Fast Forward. Published on July 6, 2018 as DOI: 10.1124/dmd.118.082164
This article has not been copyedited and formatted. The final version may differ from this version.

with the SNP in the middle were synthesized for EMSA assays. EMSA assays showed that oligonucleotides with the variant genotype displayed greater LCL nuclear extract binding after 3MC treatment than did those with the WT genotype. The EMSA bands were sent for mass spectrometric analysis to identify proteins that bound to the SNP DNA sequences. **C**) Venn diagram for the mass spectrometric data summary. A total of 28 proteins, including 19 present in both WT and variant bands and 9 specific for the variant band, were selected for subsequent study. **D**) Comparison of CYP1A1 mRNA expression levels after KD of 28 candidate proteins that had been identified by mass spectrometry together with data for the KD of AHR, followed by DMSO or 3MC treatment; * $p < .05$, ** $p < .01$, *** $p < .001$. Values shown are mean ± SEM for 3 determinations. **E**) Correlations of ratios of CYP1A1 mRNA expression after DMSO or 3MC treatment indicated that KD of the *RCC1, NAP1L4* and *DDX39B* genes (see red arrows) resulted in *CYP1A1* ratios similar to that found after the KD of AHR. Dots represent mean values for 3 determinations.

Fig. 4. A, left) CYP1A1 induction by 3MC was decreased in LCLs with the rs2470893 *V/V* genotype after the KD of RCC1, NAP1L4 or DDX39B**. A, right**) The correlation of CYP1A1 mRNA levels after DMSO or 3MC treatment showing that CYP1A1 mRNA expression after the KD of RCC1, NAP1L4 or DDX39B was similar to results obtained after the KD of AHR in rs2470893 *V/V* LCLs. **B, left**) CYP1A1 induction by 3MC was unchanged in LCLs with the rs2470893 *W/W* genotype after the KD of RCC1, NAP1L4 or DDX39B, and **B**, **right)** in contrast, CYP1A1 mRNA expression after the KD of RCC1, NAP1L4 or DDX39B was similar to data for the KD control in LCLs with the rs2470893 *W/W* genotype. * $p < 0.05$, ** $p < 0.01$,

*** $p < 0.001$. Values shown on the left are mean ± SEM for 4 determinations. Dots and squares shown on the right represent mean values for 4 determinations.

Fig. 5. A) Co-immunoprecipitation (Co-IP) with anti-AHR antibody identified interactions between AHR and RCC1 and NAP1L4 in LCLs that were treated with 3MC, but not for DDX39B. "Reverse" Co-IP with **B**) anti-RCC1 and **C**) anti-NAP1L4 antibodies confirmed interactions between AHR and RCC1 and between AHR and NAP1L4. The data shown is representative of the results of two individual experiments.

Table 1. Common SNPs that mapped ±500 bp of XREs across or 5 Kb up- and down-stream of the *CYP1A1* gene.

SNP rsID	Common Allele	Minor Allele	MAF (1000 Genomes Data)					Putative XRE Sequences	Motif Type[a]	*p* Value[b]	Distance (bp) XRE to Gene[c]	Distance (bp) SNP to XRE
			EUR	EAS	AFR	AMR	SAS					
rs17861109	C	A	0.107	0.427	0.371	0.419	0.338	CACGCCA	XRE-I	8.78E-05	-2106	259
rs2470893	**G**	**A**	**0.285**	**0.000**	**0.016**	**0.138**	**0.050**	**CAGGGAA GGCCTTG**	**XRE-II**	**8.38E-05**	**-1766**	**196**
rs3826041	A	C	0.154	0.538	0.570	0.470	0.363	CACGCGA	XRE-I	4.39E-05	-1045	33
rs4646903 (*MspI*)	T	C	0.107	0.430	0.235	0.411	0.340	CACGCCA	XRE-I	8.78E-05	294	53

EUR: European, n= 1006; EAS: East Asian, n= 1008; AFR: African, n =1322; AMR: Ad Mixed American, n=694; SAS: South Asian, n=978. Base on 1000 Genomes data, the rs2470893 SNP was not in LD with the other three SNPs in European populations (r^2 <0.090). The other three SNPs were in LD with each other in European populations ($r^2 \geq 0.630$). The rs17861109 and rs4646903 SNPs were tightly linked (r^2 =1.000).

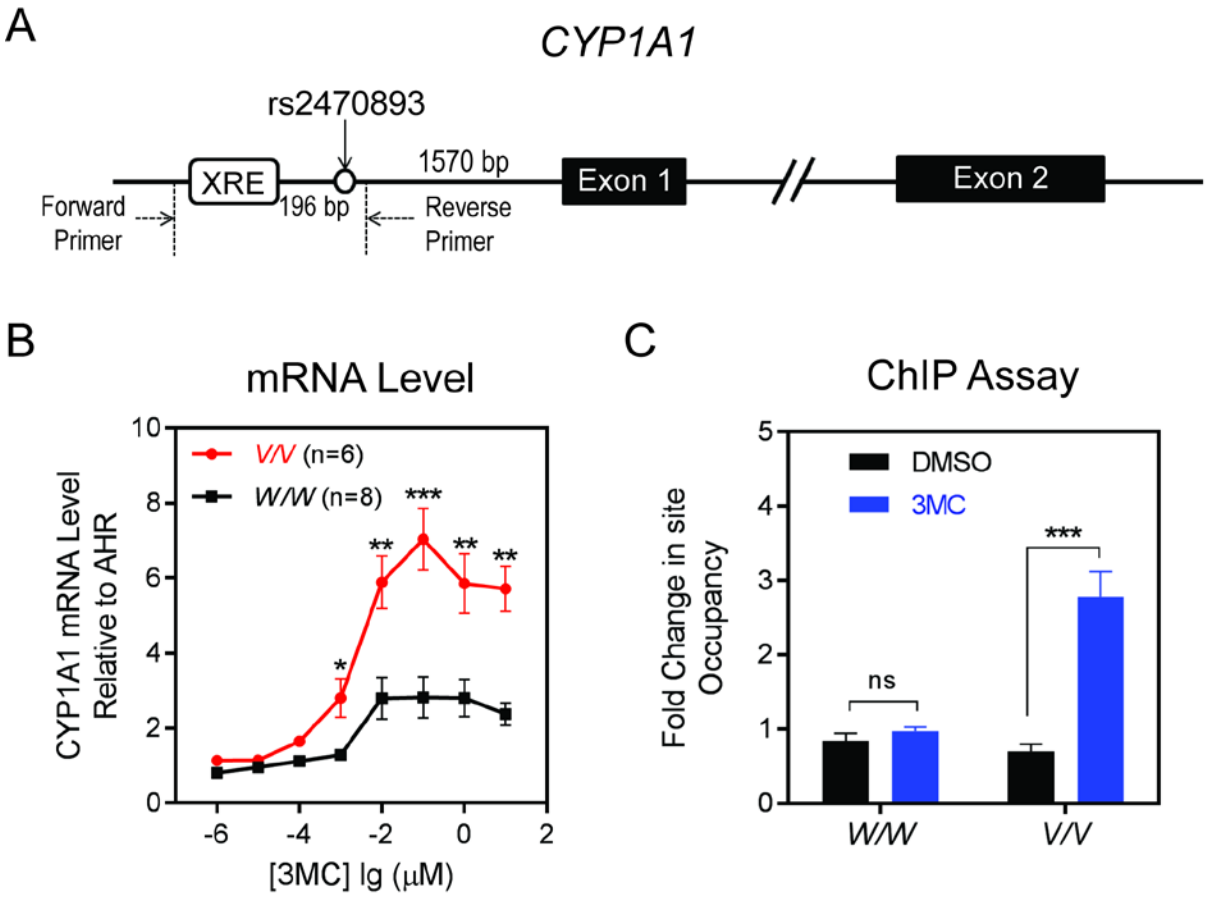
[a] The XRE-I sequence motif is "CACGCNA" and the XRE-II motif is "$CATGN_6C$[T/A]TG" as described in the *Materials and Methods* section;

[b] *p* value is given by the program FIMO to show the significance of the matching between putative XRE and the XRE motifs.

[c] Negative and positive values refer to the distance from the 5'- or 3'- ends of gene, respectively. Data for the rs2470893 SNP have been bolded.

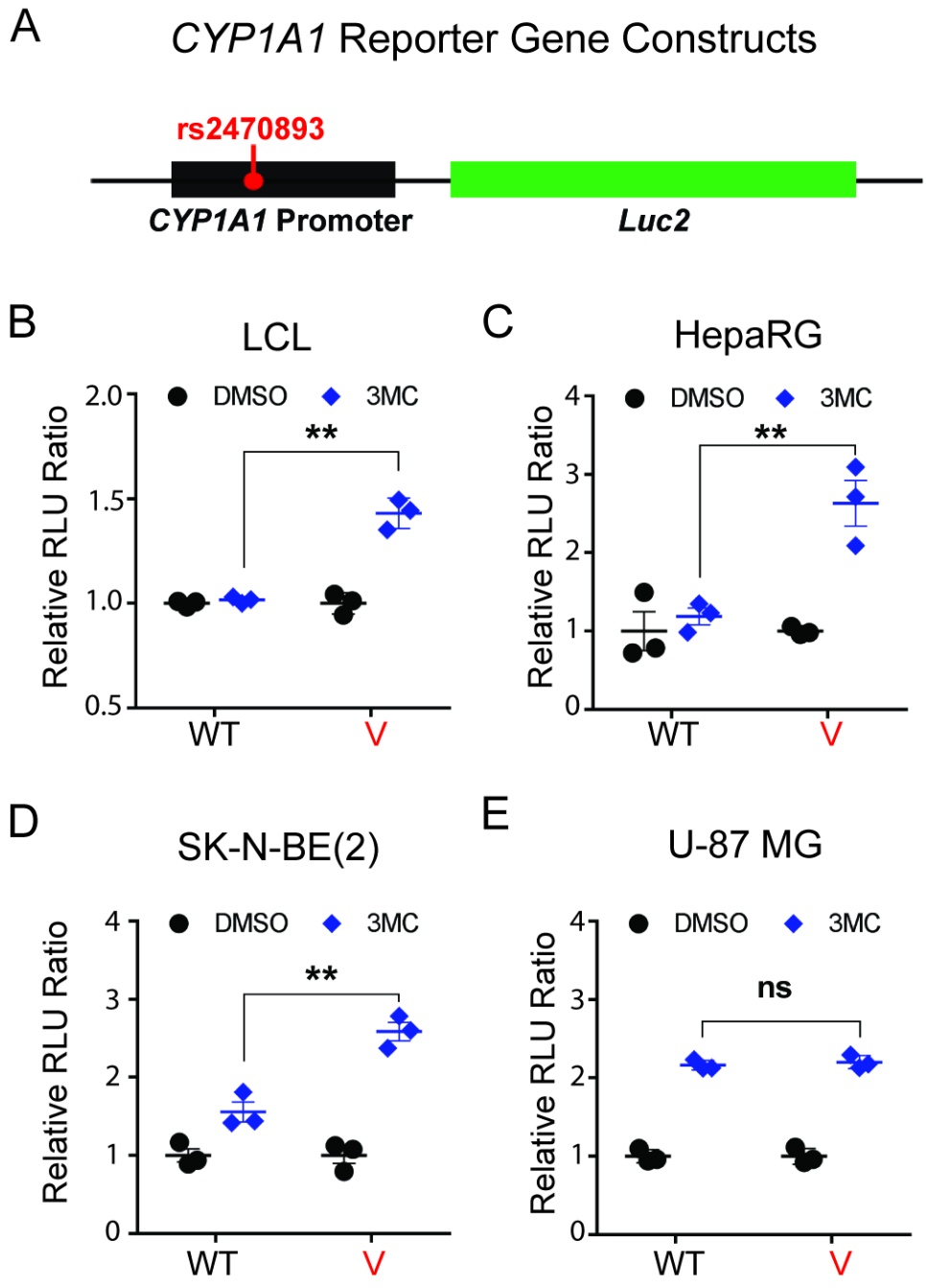
DMD Fast Forward. Published on July 6, 2018 as DOI: 10.1124/dmd.118.082164
This article has not been copyedited and formatted. The final version may differ from this version.

Figure 1.



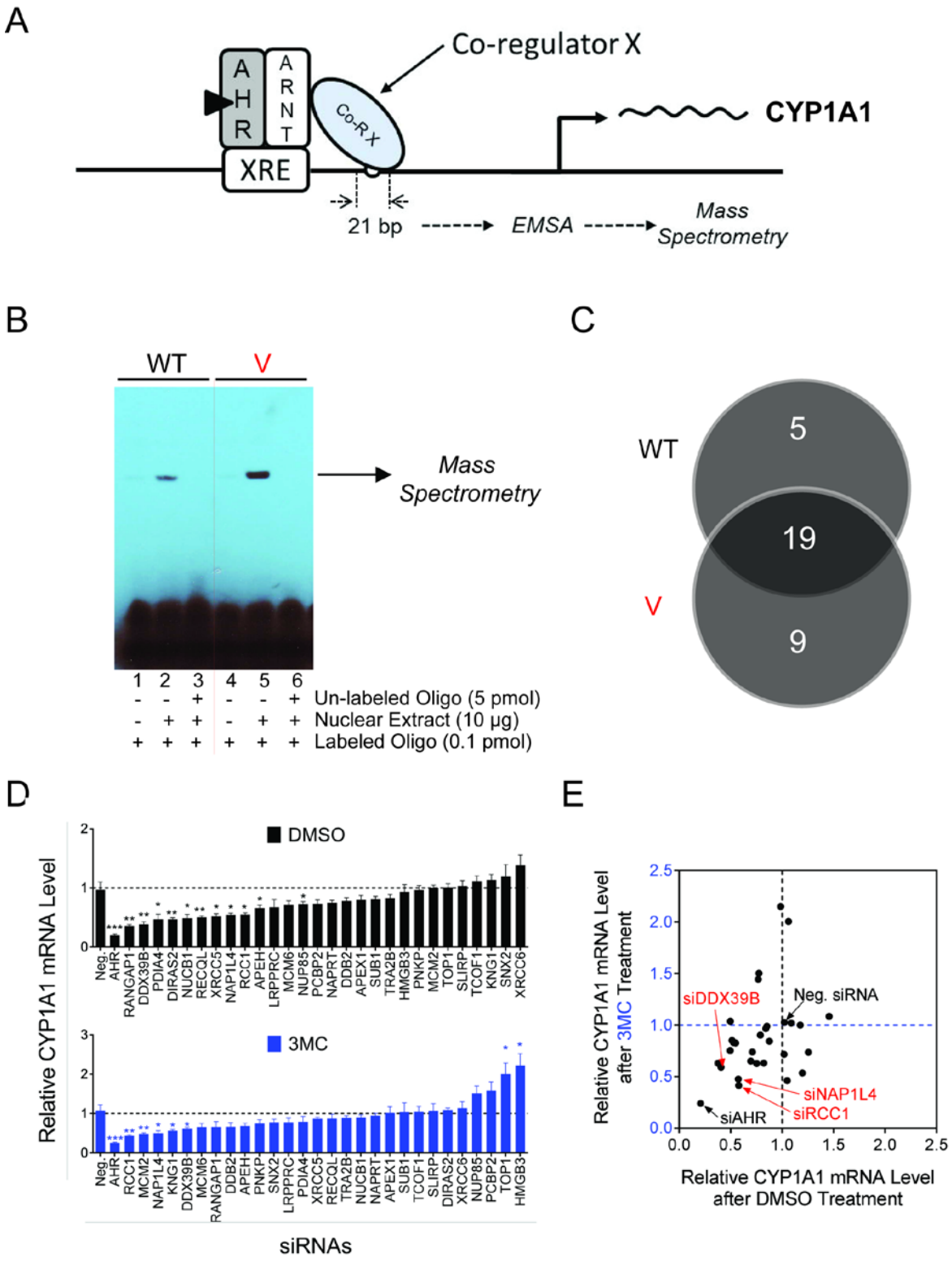
DMD Fast Forward. Published on July 6, 2018 as DOI: 10.1124/dmd.118.082164
This article has not been copyedited and formatted. The final version may differ from this version.

Figure 2.



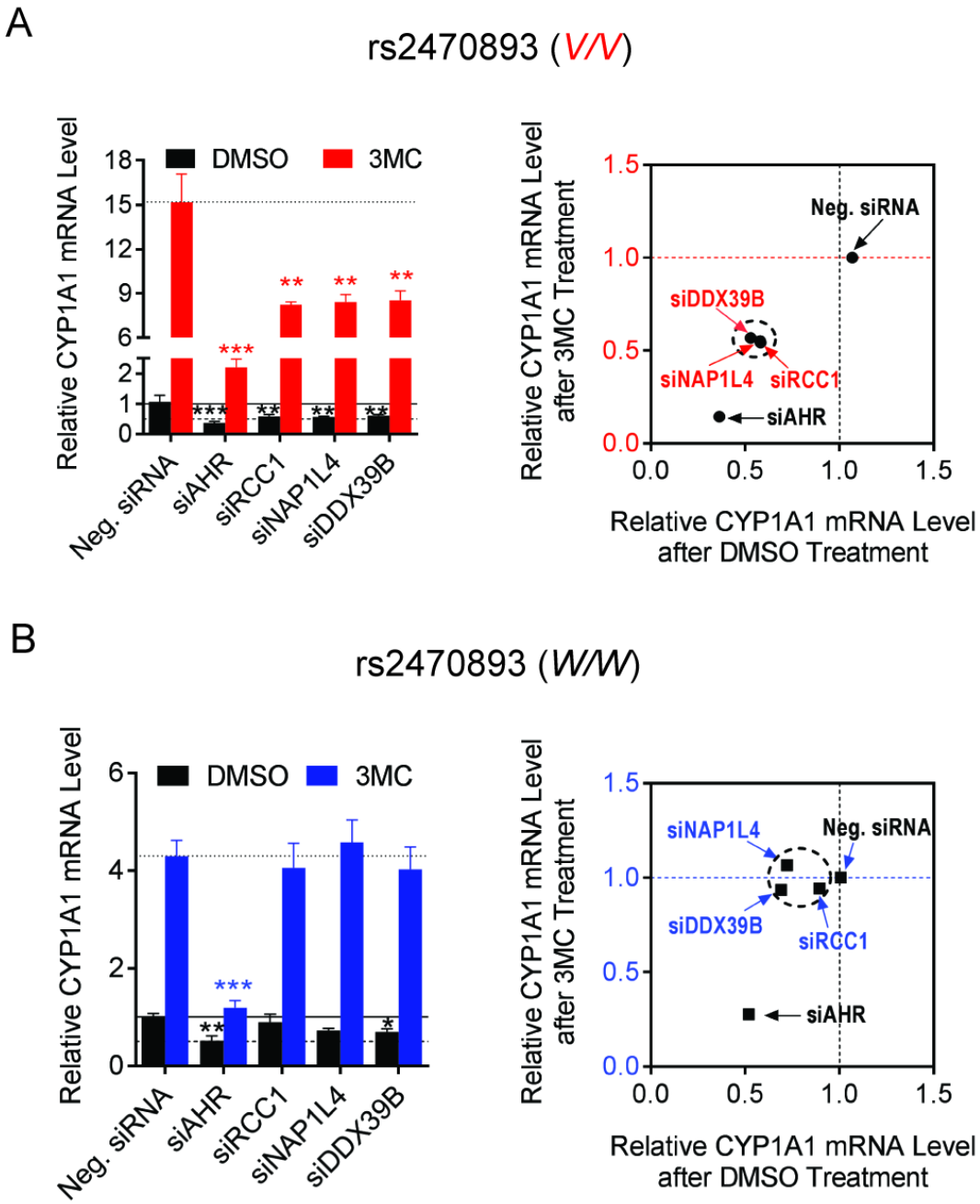
DMD Fast Forward. Published on July 6, 2018 as DOI: 10.1124/dmd.118.082164
This article has not been copyedited and formatted. The final version may differ from this version.

Figure 3.



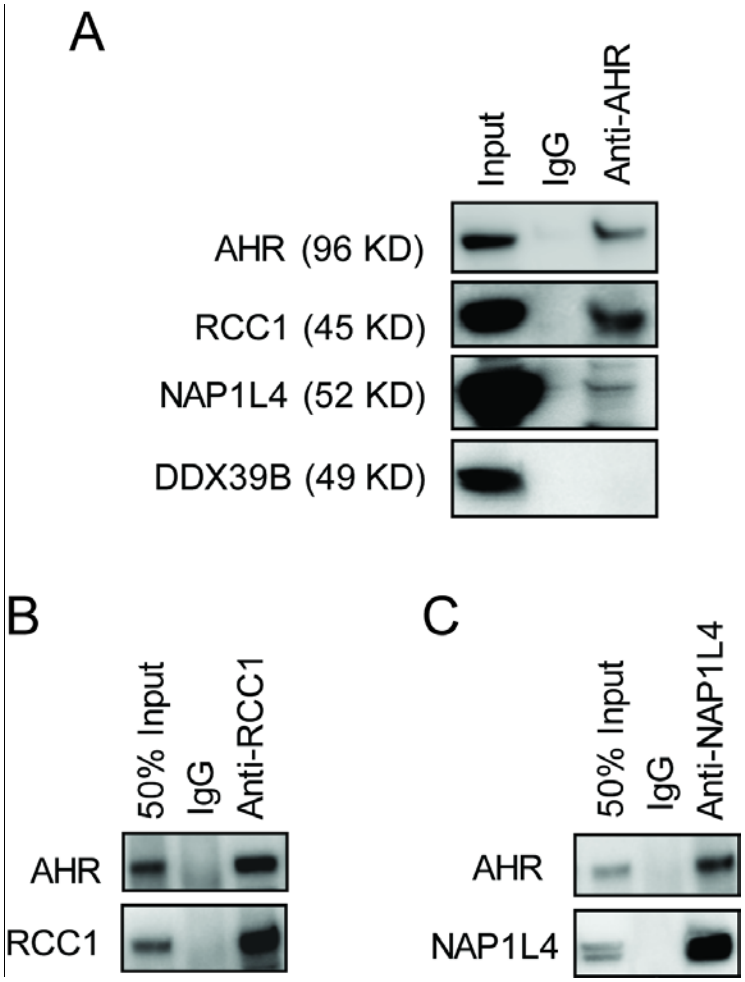
DMD Fast Forward. Published on July 6, 2018 as DOI: 10.1124/dmd.118.082164
This article has not been copyedited and formatted. The final version may differ from this version.

Figure 4.



DMD Fast Forward. Published on July 6, 2018 as DOI: 10.1124/dmd.118.082164
This article has not been copyedited and formatted. The final version may differ from this version.

Figure 5



DMD/2018/082164

Supplemental Data:

Single Nucleotide Polymorphisms (SNPs) Distant from Xenobiotic Response Elements Can Modulate Aryl Hydrocarbon Receptor Function: SNP-Dependent *CYP1A1* Induction

Duan Liu, Sisi Qin, Bamiki Ray, Krishna R. Kalari, Liewei Wang, Richard M. Weinshilboum

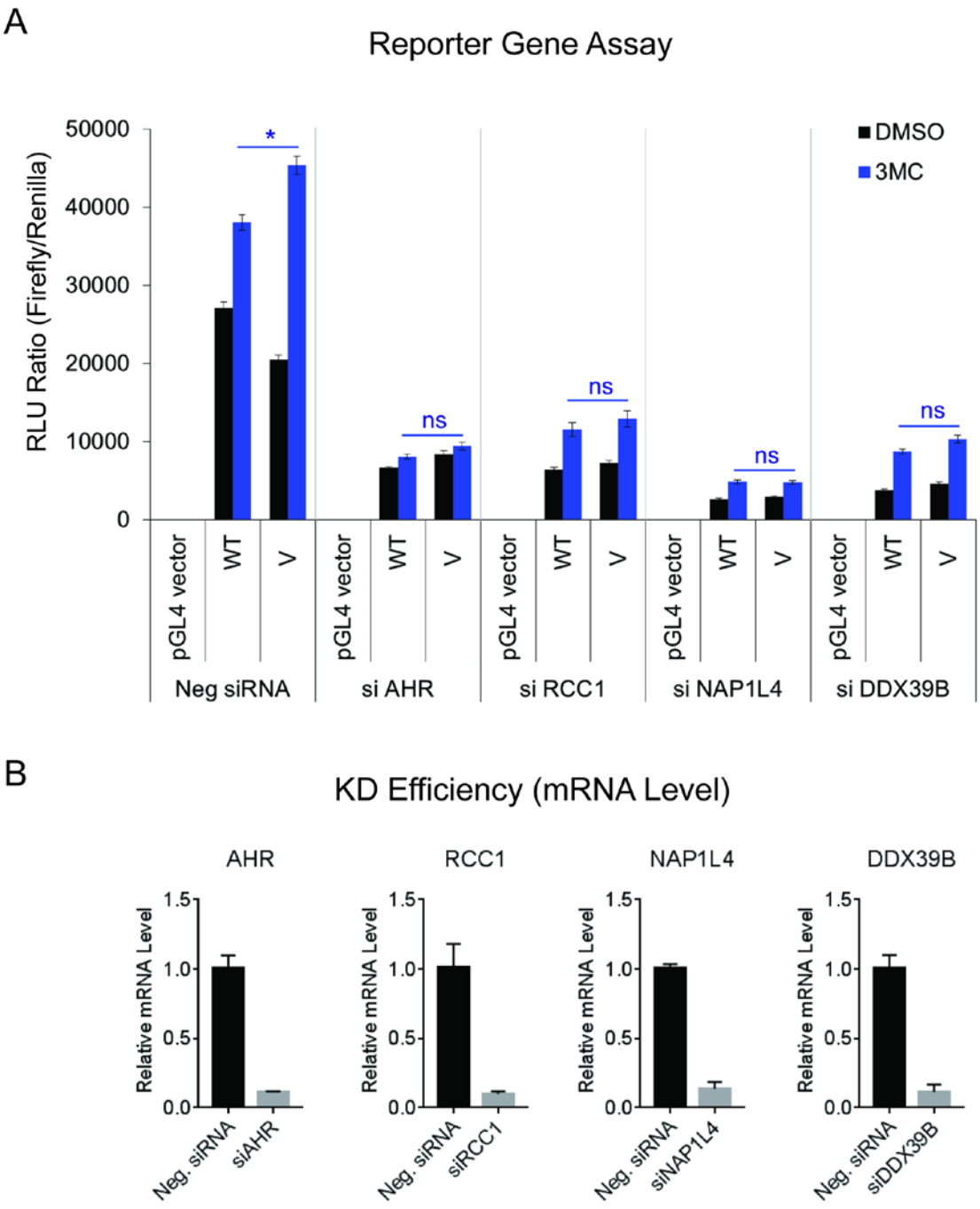
Division of Clinical Pharmacology, Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN 55905, USA (D.L., S.Q., B.R., L.W. and R.M.W)

Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA (K.R.K)

Current affiliation: Assurex Health Inc., Mason, OH (B.R.)

Supplemental Figure:

Supplemental Fig. 1. *CYP1A1* Promoter Reporter Gene Assay Results in HepaRG Cells. A) Relative light unit (RLU) ratio of firefly to renilla luciferase in HepaRG cells transfected with pGL4 empty vector and with *CYP1A1* promoter reporter gene constructs that contained either wildtype (WT) or variant (V) rs2470893 SNP genotypes after knock-down (KD) of AHR, RCC1, NAP1L4 or DDX39B by siRNA transfection. **B**) KD efficiencies for AHR, RCC1, NAP1L4 and DDX39B in HepaRG cells were measured by qRT-PCT.



Supplemental Tables 1-7:

Supplemental Table 1. Primers for mRNA quantification using qRT-PCR assays.

Gene	Forward Primer	Reverse Primer
CYP1A1	GAACCTTCCCTGATCCTTGTG	TGGAGATTGGGAAAAGCATGA
AHR	GCATTAAATCCTTCTCAGTGTACAG	CGACATATGAAGCACCTCTCC
GAPDH	ACATCGCTCAGACACCATG	TGTAGTTGAGGTCAATGAAGGG
RCC1	TCTCCAACTACCATCAGCTTG	GGACTTGGTGGAATTCTTGAAG
NAP1L4	GCTGATCCCTTTTCCTTTGAAG	CGACCCTTATGCTTCTGCTT
DDX39B	CTTCGCCACACTCGTTCC	CCCCCACTGTTATCTTTATTCTCT

Supplemental Table 2. Primers for ChIP assays using qPCR.

SNP rsID	Forward Primer	Reverse Primer
rs17861109	GGGAGTTCTCCTGGCATGTA	CTCTTACACCCCAGCCATGT
rs2470893	TCCACGAGATTAGGAGTGCTG	CTATGTGCCAGGAGCTGTTC
rs3826041	GCGTTGCGTGAGAAGGAC	CGAGGCTGGCCCTTTAAGA
rs4646903	AGGCTGGAGTGCACTGGTA	CATGGTGAAACCCCATCTCT

Supplemental Table 3. Oligonucleotides for EMSA assays. The locations of the SNPs have been bolded.

SNP rsID	SNP Genotype	Sense	Anti-sense
rs17861109	C	CATGAGACTT **C** CTCTGTGCGA	TCGCACAGAG **G** AAGTCTCATG
	A	CATGAGACTT **A** CTCTGTGCGA	TCGCACAGAG **T** AAGTCTCATG
rs2470893	G	AATCGCCAGC **G** CCTCCGAACA	TGTTCGGAGG **C** GCTGGCGATT
	A	AATCGCCAGC **A** CCTCCGAACA	TGTTCGGAGG **T** GCTGGCGATT
rs3826041	A	CACCCAGCCG **A** CCCATTCCCC	GGGGAATGGG **T** CGGCTGGGTG
	C	CACCCAGCCG **C** CCCATTCCCC	GGGGAATGGG **G** CGGCTGGGTG
rs4646903	T	CTCCACCTCC **T** GGGCTCACAC	GTGTGAGCCC **A** GGAGGTGGAG
	C	CTCCACCTCC **C** GGGCTCACAC	GTGTGAGCCC **G** GGAGGTGGAG

Supplemental Table 4. Primary Antibodies Used to Perform Immunoprecipitation and Western Blot Analyses

Antibody	Company	Catalog Number
AHR	Cell Signaling Technology	83200S
RCC1	Abcam	ab109379
RCC1	Santa Cruz Biotechnology	sc-55559
NAP1L4	Abcam	ab21631
DDX39B (UAP56)	Abcam	ab47955
DDX39B (BAT1)	Santa Cruz Biotechnology	sc-271395
IgG	Cell Signaling Technology	2729S

Supplemental Table 5. Putative XREs across the CYP1A1 Gene and SNPs Located ±500 bp of Putative XREs

Putative XRE									SNP		
XRE Number	Motif Type[a]	Start Site (bp)[b]	End Site (bp)[b]	Sequence	FIMO Score	FIMO *p* Value	Strand[c]	Distance (bp) XRE to Gene	rs ID	Location (bp)[b]	Distance (bp) SNP to XRE
XRE-1	XRE-I	75020941	75020948	CACGCCA	11.8424	8.78E-05	-	-3063	N.A	N.A	N.A
XRE-2	XRE-I	75019984	75019991	CACGCCA	11.8424	8.78E-05	+	-2106	rs17861109	75020250	259
									rs17861108	75020247	256
									rs17861107	75020018	27
									rs7495708	75019843	141
XRE-3	XRE-II	75019644	75019658	CAGGGAAGGCCTTG	9.37576	8.38E-05	+	-1766	rs17861107	75020018	360
									rs7495708	75019843	185
									rs2470893	75019448	196
									rs3809585	75019289	355
									rs36090160	75019185	459
XRE-4	XRE-I	75018923	75018930	CACGCGA	11.8485	4.39E-05	+	-1045	rs3809585	75019289	359
									rs36090160	75019185	255
									rs17861104	75018932	2
									rs4646417	75018931	1
									rs3826042	75018905	18
									rs3826041	75018890	33
XRE-5	XRE-I	75018365	75018372	CACGCGA	11.8485	4.39E-05	-	-487	rs17861099	75018340	25
									rs35686934	75018330	35
									rs111578048	75017934	431
XRE-6	XRE-II	75018117	75018131	CTTGGGACGCCTTG	9.56364	6.71E-05	-	-239	rs17861099	75018340	209
									rs35686934	75018330	199
									rs111578048	75017934	183
XRE-7	XRE-II	75014369	75014383	CAGGGTCCTGCATG	9.36364	9.72E-05	+	0	N.A	N.A	N.A
XRE-8	XRE-II	75014369	75014383	CATGCAGGACCCTG	10.2909	3.41E-05	-	0	N.A	N.A	N.A
XRE-9	XRE-II	75013803	75013817	CATGAGGCTCCAGG	9.3697	9.03E-05	+	0	N.A	N.A	N.A
XRE-10	XRE-I	75011581	75011588	CACGCCA	11.8424	8.78E-05	-	294	rs17861084	75012019	431
									rs4646903	75011641	53
									rs17861082	75011211	370
									rs17861081	75011197	384
XRE-11	XRE-II	75010706	75010720	CACGCCAGGGCATG	9.59394	6.20E-05	+	1162	rs17861082	75011211	491
									rs17861081	75011197	477
									rs45586835	75011039	319
									rs45597535	75010895	175

									rs45527133	75010207	499
XRE-12	XRE-II	75010706	75010720	CATGCCCTGGCGTG	10.5818	3.02E-05	-	1162	rs17861082	75011211	491
									rs17861081	75011197	477
									rs45586835	75011039	319
									rs45597535	75010895	175
									rs45527133	75010207	499
XRE-13	XRE-I	75010706	75010713	CACGCCA	11.8424	8.78E-05	+	1169	rs17861082	75011211	498
									rs17861081	75011197	484
									rs45586835	75011039	326
									rs45597535	75010895	182
									rs45527133	75010207	499
XRE-14	XRE-I	75009415	75009422	CACGCCA	11.8424	8.78E-05	-	2460	rs145571222	75009499	77
									rs75705772	75009033	382
XRE-15	XRE-II	75008622	75008636	CCTGGTGGCCCTTG	9.57576	6.54E-05	+	3246	rs75705772	75009033	397
									rs74607204	75008882	246

[a] The XRE-I sequence motif is “CACGCNA” and the XRE-II motif is “CATGN_6C[T/A]TG”, as described in the *Materials and Methods* section; [b] Putative XRE start and end sites and the SNP locations are based on the human genome assembly GRCh37; [c] The positive strand is the DNA strand that encode CYP1A1 mRNA, and the negative strand is complementary to the positive strand. "N.A." indicate that no SNP was present within ±500 bp of that XRE.

Supplemental Table 6. SNP genotypes for LCLs used in the study.

LCL ID	Catalog ID[a]	Genotype					
		CYP1A1				*AHR*	
		rs17861109 (C>A)	rs2470893 (G>A)	rs3826041 (A>C)	rs4646903 (T>C)	rs17137566 (A>G)	rs2066853 (G>A)
CA01	GM17201	C C	A A	A A	T T	A A	G G
CA02	GM17202	C C	A A	A A	T T	A A	G G
CA12	GM17212	C C	A A	A A	T T	A A	G G
CA23	GM17223	C C	A A	A A	T T	A A	G G
CA29	GM17229	A A	G G	C C	C C	A A	G G
CA34	GM17234	A A	G G	C C	C C	A A	G G
CA36	GM17236	C C	G G	A A	T T	A A	G G
CA40	GM17240	C C	G G	A A	T T	A A	G G
CA43	GM17243	C C	G G	A C	T T	A G	G G
CA52	GM17252	A A	G G	C C	C C	A A	G G
CA64	GM17264	C C	G G	A A	T T	A A	G G
CA76	GM17276	A C	G G	A C	C T	A A	G G
CA81	GM17281	C C	A A	A A	T T	A A	G G
CA95	GM17295	C C	A A	A A	T T	A G	G G

[a] Original catalog ID from the Coriell Institute (Camden, NJ) from which the LCLs were obtained.

Supplemental Table 7. Proteins in the EMSA bands identified by mass spectromerty.

Blank	WT	VR
ACTA2	ACTA2	ACTA2
ACTB	ACTB	ACTB
ACTR5	ADCY5	ACTR5
AHNAK	ANXA2	ANXA2
ALOX12B	ANXA6	ANXA6
ANXA1	**APEH**	**APEH**
ANXA2	**APEX1**	**APEX1**
APOD	ARHGDIB	ARHGDIB
ASAH1	ATP6V1A	ATP5B
ATP5B	AZGP1	ATP6V1A
ATR	BCLAF1	CAT
AZGP1	CAT	CLIC1
BLMH	CHMP4B	CST6
BPIFA1	CLIC1	DARS
CAT	CPSF2	DDB1
CCDC153	DARS	**DDB2**
CDSN	**DDB2**	**DDX39B**
CHD3	DDRGK1	DHX9
COL1A1	**DDX39B**	**DIRAS2**
COL22A1	DHX9	DKFZp686J1372
CPA4	ECHS1	ECM1
CST6	EHD1	EHD4
DMBT1	EHD4	ELAVL1
ECM1	ELAVL1	ERO1L
EIF6	ERO1L	FDXR
FABP5	FDXR	GSS
FAM26F	FERMT3	HAL
GAPDH	GAPDH	HARS
GGCT	GDI1	HK1
GGH	GRHPR	HMGB2
HAL	GRSF1	**HMGB3**
HBB	GSTK1	HNRNPA2B1
HNRNPA2B1	HBA1	HNRNPAB
IAH1	HDGFRP2	HNRNPAB
IGHA1	HK1	HNRNPH1
IGHG1	HMGA1	HNRNPK
IGJ	HMGB1P1	HSP90AA1
IGKC	HMGB2	HSP90AB3P
INTS7	**HMGB3**	HSP90B1
KHSRP	HNRNPA0	HSP90B2P
LMNA	HNRNPA2B1	HSPA5
LTF	HNRNPAB	HSPD1
LYZ	HNRNPK	IDE
MAGEF1	HNRNPL	IRF4
NELFE	HNRNPM	**KNG1**
NUSAP1	HSP90AA1	KPNB1
PKM	HSP90AA2	LDHB
POF1B	HSP90AB3P	LONP1
PPP1R3F	HSP90B1	**LRPPRC**
PRDX2	HSP90B2P	LSP1

PRSS3
PSMA1
PSMA3
PSMA5
PSMA6
PSMA7
PSMA8
PSMB1
PSMB3
PSMB5
PSMB6
RABGEF1
RNF180
RPS3
S100A11
S100A14
SBSN
SDR9C7
SERPINA12
SERPINB8
SUMO2
TFAP2D
TNNI3K
TUBB4A
TXN
VCL
XP32
YWHAB
YWHAZ
ZDHHC15
ZG16B
ZMYM3
ZNF12

HSPA5
HSPD1
IGJ
ILF3
KPNB1
LDHB
LONP1
LRPPRC
LSP1
MCM2
MCM3
MCM6
MX1
NAGK
NAP1L4
NCL
NOLC1
PBX2
PCBP2
PDCD4
PDIA4
PKM
PMPCB
PNKP
PPIB
PRKCSH
RANGAP1
RBBP5
RECQL
RPA1
RPSA
RRP9
SLAIN2
SRRT
SRSF3
SRSF4
SRSF7
SSR4
SUB1
SUMO2
SYNCRIP
TCOF1
TMX1
TOP1
TPM4
TUBB
UBA1
UBA2
UQCRC1
XRCC5
XRCC6
YARS
ZPR1

LTF
MAGOHB
MCM2
MCM3
MCM6
MDC1
MTHFD2
MX1
NAGK
NAP1L4
NAPRT
NCL
NHP2L1
NOLC1
NPEPPSL1
NUCB1
NUP85
PCBP1
PCBP2
PDIA4
PKM
PNKP
PPP1R2P3
PPP2R1A
PSMA6
PSMA8
PSMB5
PSMB6
RANGAP1
RAP1A
RCC1
RECQL
RPA1
RPSA
RPSAP58
SARNP
SBSN
SERPINB6
SHMT2
SLAIN2
SLIRP
SNX2
SRRT
SRSF2
SRSF3
SRSF4
SRSF6
SRSF7
SSBP1
SUB1
SUMO2
SYNCRIP
TCOF1
TMX1
TMX4

TOP1
TPM3
TPM4
TRA2B
TSN
TUBA4A
TUBB
TUBB4A
TUBG1
UBA1
UBA2
UBA3
UQCRC1
XRCC5
XRCC6
ZPR1

The 28 candidate proteins that were selected for further study are **bolded**, and the 9 proteins specifically identified in the variant (V) sample are highlighted in **red**.