ChIPing the cistrome of PXR in mouse liver

Julia Yue Cui[1], Sumedha S. Gunewardena[2], Cheryl E. Rockwell[1] and
Curtis D. Klaassen[1,*]

[1]Department of Pharmacology, Toxicology, and Therapeutics and [2]Department of Molecular and Integrative
Physiology, University of Kansas Medical Center, Kansas City, KS 66160, USA

ABSTRACT

The pregnane X receptor (PXR) is a key regulator of xenobiotic metabolism and disposition in liver. However, little is known about the PXR DNA-binding signatures *in vivo,* or how PXR regulates novel direct targets on a genome-wide scale. Therefore, we generated a roadmap of hepatic PXR bindings in the entire mouse genome [chromatin immunoprecipitation (ChIP)-Seq]. The most frequent PXR DNA-binding motif is the AGTTCA -like direct repeat with a 4bp spacer [direct repeat (DR)-4]. Surprisingly, there are also high motif occurrences with spacers of a periodicity of 5 bp, forming a novel DR-$(5n+4)$ pattern for PXR binding. PXR-binding overlaps with the epigenetic mark for gene activation (histone-H3K4-di-methylation), but not with epigenetic marks for gene suppression (DNA methylation or histone-H3K27-tri-methylation) (ChIP-on-chip). After administering a PXR agonist, changes in mRNA of most PXR-direct target genes correlate with increased PXR binding. Specifically, increased PXR binding triggers the *trans*-activation of critical drug-metabolizing enzymes and transporters. The mRNA induction of these genes is absent in PXR-null mice. The current work provides the first *in vivo* evidence of PXR DNA-binding signatures in the mouse genome, paving the path for predicting and further understanding the multifaceted roles of PXR in liver.

INTRODUCTION

Liver is an essential 'buffer' for drugs and nutrient homeostasis in the body. The pregnane X receptor (PXR; NR1I2) is a key regulator in mediating xeno- and endobiotic metabolism and disposition in liver, serving as a critical component of the liver's defense mechanisms against toxic substances (1). As a member of the nuclear receptor family, PXR is a ligand-activated transcription factor (TF) that functions with its binding partner, the retinoid X receptor (RXR). Once activated, the PXR/RXR complex binds to DNA in the nucleus and regulates gene transcription (2). PXR is highly expressed in mammalian liver, and its DNA-binding domain is highly conserved across species (1). However, the ligand-binding domain displays more variability, allowing PXR to be activated by a wide spectrum of chemicals, including various drugs like the antibiotic rifampicin, the anti-inflammatory drug dexamethasone, and the anticonvulsant phenobarbital (1), environmental polybrominated diphenyl ethers (PBDE) (3) and endogenous chemicals, such as bile acids (4,5). Although PXR ligands can be species specific, the target gene profiles share great similarities between rodents and humans, including genes encoding drug metabolizing enzymes and transporters (together termed 'drug-processing genes' in this study). The synthetic compound, pregnenolone-16α-carbonitrile (PCN), is a potent activator of mouse PXR, which is widely used by researchers to recapitulate human PXR activation (2).

Data from this and other laboratories have shown that numerous drug-processing genes in mouse liver are up-regulated following PCN administration (6–10) (e.g. our laboratory has shown that 18 drug-processing genes are induced by PXR in mouse liver), but only a few genes have been shown to be direct PXR targets (1,11,12). It remains to be determined whether the induction of critical drug-processing genes is due to direct trans-activation by PXR or due to secondary effects. More importantly, as it is becoming increasingly evident that PXR has novel physiological functions, such as trans-activating genes involved in lipid metabolism (11,12) and cell cycle (13), novel PXR-target genes need to be characterized, which will fill a critical knowledge gap

in predicting and further understanding the multifaceted roles of PXR in liver.

Numerous studies have characterized PXR response elements, which have shown that PXR binds to AGGTCA-like direct repeats separated by 3 or 4 bp [direct repeat (DR)-3 and DR-4] and everted repeats separated by 6 or 8 bp (ER-6 and ER-8) (1). Unfortunately, most of these findings are based on naked DNAs, or cell cultures of non-hepatic origin with certain proteins artificially over-expressed, or cryopreserved hepatocytes that may have lost many features of *in vivo* cells. In addition, most studies have limited the detection range for PXR binding to the gene promoter regions. Such designs are inherently biased, in that they do not seek to detect novel genomic PXR-binding sites that may be equally important for gene regulation. It is therefore necessary to determine the most preferred DNA-binding signatures for PXR *in vivo*. Recent technological advancements, including chromatin immunoprecipitation (ChIP)-on-chip and ChIP-sequencing (ChIP-Seq), have made such unbiased genome-wide investigations possible (14,15).

Next-generation sequencing technology has offered tremendous opportunities for genome scale analysis of regulatory elements. However, the large amount of data generated by this technology pose new challenges to *in silico* tools designed for identifying the exact locations of binding sites in DNA for TFs. For example, ChIP-Seq is limited by the resolution of the output DNA fragments that localize TF binding sites to ∼500 bp. The precise location of the binding sites within the extracted DNA fragments needs to be identified *in silico*. Many different algorithms have been proposed to address the problem of identifying TF binding sites in promoter sequences but with limited success (16). Traditional algorithms for *de-novo* prediction of TF binding sites such as AlignAce (5), MEME (6), MonoDi (7), etc. were not designed to analyze large volumes of sequence data generated by modern high-throughput technologies. As a result, applying these algorithms to these data sets is cumbersome, time consuming and unpractical in many situations with solutions circumventing the problem being sub-optimal. Therefore, for the present study, using PXR as an example, we describe a fast, sensitive and efficient method for identifying the TF binding sites in large input data sets generated from genome-scale TF binding assays such as ChIP-chip and ChIP-Seq. The efficiency of the method is derived from transforming the sequence space to a k-mer space that has the advantage of drastically reducing the sampling space for the sampling algorithm to efficiently operate. The k-mer transformed sequence space makes explicit the implicit role played by k-mers in characterizing TF binding sites. This is observed in the fact that every TF binding site is an element of some set of k-mers of appropriate k. The algorithm takes advantage of the fact that the total number of k-mers in a set of input nucleotide sequences is upper bounded by the minimum of 4^k or $n - k+1$, where n is the total length of the input sequences, and, k, the length of a k-mer. For many biological sequences, this number is far less due to k-mer duplication. The task of identifying k-mers of

putative TF binding sites is made simpler, faster and more efficient once the k-mer distribution of the input sequences is determined. For example, if we are interested in finding putative TF binding sites of a length of 6 in 500 input sequences with an average length of 500 bp, instead of sampling ∼250 000 sites in the sequence space, one only needs to look at the frequency distribution of 4096 sites in the k-mer space. The upper bound of the k-mer space ensures that its size does not exceed that of the sequence space, and, due to repeated k-mers, will almost always be much smaller than the sequence space. The k-mer distribution of a given set of input sequences can be determined in linear time. This enables the algorithm to be repeated efficiently with different lengths of k, if k is not known *a priori*, in order to determine the optimum k for the TF binding sites of interest. The motif detection algorithm that we use is a deterministic maximally discriminative algorithm that operates on the whole k-mer space.

A critical question arises from the common *in vitro* approaches used to assess the interactions between 'naked' DNA and protein: how does PXR manage to access its target genes in living cells where most DNA is tightly packed in nucleosomes? Recently, a large body of evidence has indicated that distinct chromatin epigenetic signatures are essential in determining whether there is a permissive chromatin environment for TFs to access DNA and initiate gene transcription (17,18). DNA methylation (DNAMe) occurs on 5′-cytosine residues, which in general prevents the recruitment of TFs to DNA, or attracts transcription co-repressors resulting in gene silencing. Among various types of histone modifications (acetylation, methylation, phosphorylation, ADP-ribosylation, ubiquitination and sumoylation), a pattern has emerged for histone H3K4 di-methylation (H3K4Me2) of actively transcribed genes, and H3K27 tri-methylation (H3K27Me3) of silenced genes, as evident by high-resolution profiling of histone methylations of the entire human genome (18). These epigenetic signatures may be critical in predefining the transcription potential of PXR-target genes.

Briefly, the purpose of the present study is to provide the first *in vivo* knowledge of the PXR-nuclear-binding signatures and direct-target gene profiles in liver both under basal and PXR-activated conditions, to perform an unbiased investigation of novel PXR DNA-binding motifs and their genomic locations, to characterize the epigenetic signatures which predefine the transcription potential of PXR-target genes, and to determine whether critical drug-processing genes are bona fide PXR targets using PXR-null mice.

MATERIALS AND METHODS

Materials and reagents

PCN was purchased from Sigma-Aldrich (St Louis, MO). PXR antibody (sc-25381) was from Santa Cruz (Santa Cruz, CA). Polyclonal antibody against 5-methyl-cytosine (ab51552) was from Abcam (Cambridge, MA). Primers for ChIP-quantitative PCR (qPCR) were purchased

from Integrated DNA Technologies (Coralville, IA). Polyclonal antibody against methylated histone H3 proteins at lysine 4 and 27 were from Millipore Upstate (Billerica, MA).

Animals and treatments

Wild-type (WT) C57BL/6 male mice were purchased from Charles River Laboratories (Wilmington, MA). Breeding pairs of PXR-null mice (2) in the C57BL/6 background were kindly provided by Dr. Frank Gonzalez (National Cancer Institute, Bethesda, MD). Mice were housed according to the American Animal Association Laboratory Animal Care guidelines, and were bred under standard conditions at the University of Kansas Medical Center. The animal treatments are described in detail in the Supplementary Experimental Procedures. All animal procedures were reviewed and approved by the Institutional Animal Care and Use Committee at KUMC. For ChIP-qPCR and ChIP-Seq, 8-week-old male C57BL/6 mice were administered a single dose of PCN (200 mg/kg, i.p. in corn oil) ($n = 2$ per time point) or vehicle ($n = 1$ per time point), and livers were removed 2, 8 and 12 h thereafter. For mRNA assays in WT and PXR-null mice, 8-week-old male mice (WT and PXR-null, $n = 5$ per group) were administered PCN or vehicle once daily (200 mg/kg, i.p. in corn oil) for 4 days, and livers were removed on the 5th day. Livers were frozen immediately in liquid nitrogen, and stored at $-80°C$.

RNA isolation

Total RNA was isolated using RNA Bee reagent (Tel-Test Inc., Friendswood, TX) as per the manufacturer's protocol. RNA concentrations were quantified using a NanoDrop Spectrophotometer (NanoDrop Technologies, Wilmington, DE) at a wavelength of 260 nm. The integrity of the total RNA samples was evaluated by formaldehyde-agarose gel electrophoresis, and confirmed by visualization of 18S and 28S rRNA bands.

Branched DNA Amplification (bDNA) technology

The mRNAs of drug-processing genes Cyp3a11, Gstm3 and Oatp1a4 as direct PXR-target genes were determined by bDNA assays (QuantiGene bDNA Signal Amplification Kit, Panomics, Fremont, CA). Multiple oligonucleotide probe sets (including capture, label and blocker probes) were designed using ProbeDesigner Software v1.0 (Bayer Corp., Diagnostics Div.) as described previously (19). Probe sequences for Cyp3a11, Oatp1a4 and Gstm3 have been reported previously (7,9,19). Equal amounts of total RNA were added to each well of a 96-well plate. The mRNA was captured by specific probe sets and attached to a branched DNA amplifier. Enzymatic reactions occur upon substrate addition and the luminescence for each well is reported as Relative Light Units (RLU).

ChIP-qPCR assay

Livers were fixed with 1% formaldehyde for 15 min and quenched with 0.125 M glycine. The tissues were disaggregated with a Tissue Tearor (BioSpec Products, Bartlesville, OK) and chromatin was isolated by adding lysis buffer, followed by disruption with a Dounce homogenizer. Lysates were sonicated and the DNA sheared to an average length of 300–500 bp. Genomic DNA (input) was prepared by treating aliquots of chromatin with RNase, proteinase K and heat for de-crosslinking, followed by ethanol precipitation. Pellets were re-suspended and the resulting DNA was quantified with a Nanodrop spectrophotometer. Extrapolation to the original chromatin volume allowed quantification of the total chromatin yield. An aliquot of chromatin (30 μg) was precleared with protein-A agarose beads (Invitrogen, Carlsbad, CA). Genomic DNA regions of interest were isolated using an antibody against PXR. After incubation at 4°C overnight, protein-A beads were used to isolate the immune complexes. Complexes were washed, eluted from the beads with SDS buffer, and subjected to RNase and proteinase-K treatment. Crosslinks were reversed by incubation overnight at 65°C, and ChIP DNA was purified by phenol-chloroform extraction and ethanol precipitation. qPCR reactions were carried out in triplicate on specific genomic regions using SYBR Green Supermix (Bio-Rad, Hercules, CA). The resulting signals were normalized for primer efficiency by carrying out QPCR for each primer pair using Input DNA.

ChIP-Seq analysis

Livers of control and PCN-treated mice at 8 h were used for ChIP-Seq experiments. Fragments of DNA were tagged by 35-nt identifiers and subjected to sequencing by the Illumina Genome Analyzer Sequencer based on Solexa Technology (Illumina, San Diego, CA). Preprocessing of the ChIP-Seq data was performed by Genpathway (San Diego, CA). Briefly, the tags identified were mapped to the genome using Eland Software, which resulted in a list of their chromosome coordinates. Only tags that mapped uniquely and that have no more than 1 mismatch were retained. Because the 5′-end of the sequence tags represented the end of ChIP-fragments, the tags were extended *in silico* using Genpathway software at their 3′-ends to a length of 110 bp, which was the average fragment length in the size selected library. To identify the density of fragments (extended tags) along the mouse genome, the genome was divided into 32-nt bins, and the number of fragments in each bin was determined and stored together in a BAR (Binary Analysis Results) file. The BAR files were viewed in the Affymetrix Integrated Genome Browser (IGB) for PXR binding in the mouse genome. The locations of fragment-density peaks, defined by chromosome number and a start and end coordinate, were termed as 'intervals'. For each BAR file, intervals were calculated using the Affymetrix Tiling Analysis Software (TAS) and compiled into a Browser Extensible Data (BED) file. Three parameters of intervals were identified: threshold, MaxGap and MinRun. Twenty-fold of input background

signal was used as the threshold value based on calculations of false discovery rate (FDR). MaxGap and MinRun were set at 100 bp. The exact locations of intervals along with their proximities to gene annotations and other genomic features were then determined. In addition, average and peak fragment densities within intervals were compiled. Control- and PCN-treated samples were normalized to the sequencing depth of uniquely aligned reads as described previously (20–24). Peaks for PXR bindings and significant differences between PCN and control conditions were identified using the MACS algorithm (24).

Motif analysis

In order to refine the targets of PXR binding, the top 500 ChIP-DNA sequences with highest enrichment peak values in both control and PCN-treated conditions were retrieved from the UCSC genome browser. ChIP-Seq data, which localizes PXR-binding regions to ~500 bp, were further analyzed *in silico* to identify the exact PXR-binding targets. PXR-binding profiles were generated using a novel *k*-mer-based approach designed for analyzing high-throughput sequence data. The method is highly sensitive compared to other tools for identifying putative TF binding sites (16). As a result, it facilitates a finer analysis of site profiles. It is a deterministic algorithm that searches for sites or combined half sites separated by a variable spacer distance. The detailed description of this algorithm is shown in a supplementary document entitled 'motif analysis'.

Nuclear extraction from control and PCN-treated mouse livers

Nuclear protein was extracted from liver samples using the Nuclear Extract Kit from Active Motif (Carlsbad, CA). Briefly, livers from vehicle or PCN-treated mice at 8 h after dosing were weighed and diced into small pieces using a clean razor blade, and collected in a pre-chilled Dounce homogenizer. Livers were homogenized on ice in ice-cold $1\times$ Hypotonic Buffer supplemented with DTT and Detergent (3 μl of the provided 1 M DTT and 3 μl of the provided Detergent) (ratio: 3 ml/ g tissue). The liver homogenates were incubated on ice for 15 min, centrifuged for 10 min at 850g at 4°C, and the supernatant discarded. The pellet was resuspended in $1\times$ Hypotonic Buffer and incubated for an additional 15 min on ice. Detergent was added to the supernatant (at a ratio of 1:20, detergent to supernatant) and vortexed for 10 s at the highest setting. The suspension was then centrifuged for 30 s at 14 000g in a microcentrifuge pre-cooled at 4°C. The supernatant (cytoplasmic fraction) was discarded and the pellet (nuclear fraction collection) was resuspended in $1\times$ Complete Lysis Buffer. The nuclear protein fraction was vortexed for 10 s and incubated for 30 min on ice on a rocking platform. The nuclear fractions were vortexed for 10 s at the highest setting and centrifuged for 10 min at $14\,000 \times g$. The supernatants (nuclear protein fractions) were transferred to new tubes and subjected to an enzyme-linked immunosorbent assay (ELISA)-based TF DNA-binding assay.

Oligonucleotide design for the ELISA-based TF DNA-binding assay

Briefly, for each DR-$(5n+4)$-like DNA sequence, two complementary oligonucleotides were designed based on genes that have such sequence patterns within ± 10 kb of the gene loci according to the ChIP-Seq motif analysis, to determine the biological significance of the assay. The forward sequence was labeled with biotin at the 3′-end. All probes have a minimum length of 50 bp, with 7-bp linker region at the 5′-end, and 40-bp background sequences after the consensus sequence and before the 3′-biotinylation. The background DNA was obtained from sequences adjacent to the known PXR response element DR-3 upstream of the *Cyp3a11* gene locus. Non-biotinylated oligonucleotides identical to the biotinylated sequences, as well as non-biotinyated mutated oligonucleotide that have two mutated nucleotides in the middle of the half site (e.g. AGTTCA→ AG AACA), were designed for competition assays. DR-3-like sequences were used as positive controls. All oligonucleotides were examined by NHR-scan (25) to make sure there are no known PXR-consensus sequences in the probe other than the binding site of interest, and were synthesized by Sigma Aldrich Inc. (Atlanta, GA). All the sequences are shown in Supplementary Table S4.

ELISA-based TF DNA-binding assay

The quantification of PXR binding to the DR-3 and DR-$(5n+4)$-binding sites was determined by a modified TransAM TF DNA-binding assay (Active Motif, Carlsbad, CA). Because a PXR-binding kit is not commercially available, the existing Active Motif TransAM Flexi NFκB kit was modified and developed for the detection of PXR binding to the following binding sites: DR-3 (positive control), DR-4 (when $n = 0$), DR-9 (when $n = 1$), DR-14 (when $n = 2$) and DR-19 (when $n = 3$). Briefly, 10 μg of nuclear protein extract and 1 pmol of biotinylated oligonucleotides, corresponding to a DR-3, DR-4, DR-9, DR-14 or DR-19 binding sites were added to binding buffer, 1 mM DTT + 10 ng/ml Herring sperm DNA + Protease inhibitor cocktail (supplied as $100\times$ solution) in DDH$_2$O, mixed and incubated for 30 min at room temperature (RT). Each reaction was then transferred to an individual well of a streptavidin-coated plate and incubated for 1 h at room temperature with mild agitation. The plate was washed three times with phosphate-buffered saline (PBS) and primary antibody was added (Santa Cruz sc-25381 diluted 1:1000 in $1\times$ TransAm Ab binding buffer for most assays, but for comparison of all the binding motifs (Figure 9F), the primary antibody was diluted 1:1500 in order to better compare the different sequences). The plate was then incubated for 1 h at room temperature without agitation. The plate was then washed three times with PBS and secondary antibody was added, HRP-conjugated anti-rabbit IgG (from Active Motif) diluted 1:1000 in $1\times$ TransAM-binding buffer AM1. The plate was then

incubated for 1 h at room temperature without agitation. The plate was washed four times with TransAM wash buffer AM2 (from Active Motif), after which Developing Solution (from Active Motif) was added to each well. The plate was allowed to develop for 2–10 min after which time the reaction was stopped by adding Stop Solution (from Active Motif). The plate was then read at a wavelength of 450 nm by a Uquant microplate reader (BioTek, Winooski, VT). A non-biotinylated WT oligonucleotide, identical to the biotinylated sequence, was used to compete off PXR binding in order to determine the specificity of each assay (used at 10 pmol/well). A non-biotinyated mutated (Mut) oligonucleotide was used to further confirm the specificity of each assay (used at 10 pmol/well).

Microarray

Gene expression in livers of control and PCN-treated WT mice at 12 h was determined by the Affymetrix Mouse 430.20 arrays in the KUMC Microarray Core Facility. Data were preprocessed by the Affy Expression Console and RMA-PARTEK, and gene annotations were obtained using GeneSpring (Agilent Technologies, CA). The detection P-value (Student's test) was <0.05. An MA plot was generated by R-programming and Bioconductor (26). Briefly, the probe-level microarray data (as stored in CEL files) from control and PCN-treated groups were pre-processed using the RMA method. The expression values were in $\log_2$ scale. The probe intensities higher than $\log_2 100$ in either control or PCN-treatment group were selected for graphing. An MA plot was generated by plotting fold-changes of probe intensities (PCN/control) against the average probe intensities between control and PCN using the smoothScatter function in the prada package. An interactive output of the top 5000 differentially expressed genes after PCN treatment was generated using the annaffy package. Gene symbols were obtained from the mouse 4302 package. All annotations, including gene descriptions, chromosome IDs, chromosome locations, GeneBank IDs, Gene IDs, Cytoband IDs, UniGene IDs, PubMed IDs, Gene Ontoglogies and Pathways were built in one table and hyperlinked to on-line databases. The gene expression data in $\log_2$ scale were constructed in the same table, with the cells colored with varying intensities of green to show relative expression values. The fold-changes of gene expression (PCN/control) were listed in the last column (descending order), with increase colored by purple, and decrease colored by red.

Pathway analysis

Pathway analysis for control and PCN-treated microarray data were performed using the Ingenuity Pathway Analysis Software (IPA) (Ingenuity® Systems, http://www.ingenuity.com).

ChIP-on-chip assay of DNA methylation and histone methylation

Genpathway's ChIP-on-chip assays (San Diego, CA) using Affymetrix GeneChip Mouse Tiling 2.0R E array, were used to determine the following epigenetic profiles on mouse chromosome 5, 12 and 15: DNAMe, H3K4Me2 and H3K27Me3 as described previously (27). The raw and processed data are stored in the Gene Expression Omnibus (GEO) database with the accession number GSE14620.

Multiplex suspension bead array

The mRNA expression of drug metabolizing enzymes and transporters in WT and PXR-null mouse livers were quantified by the multiplex suspension bead array. Briefly, individual bead-based oligonucleotide probe sets specific for each gene examined were developed by Panomics Inc. The probe sequences are available at http://www.panomics.com. Samples were analyzed using a Bio-Plex 200 System Array reader with Luminex 100 X-MAP technology, and data were acquired using a Bio-Plex Data Manager Software Version 5.0 (Bio-Rad, Hercules, CA). Assays were performed according to the manufacturer's protocol. Data are expressed as the ratio of relative light units (RLU) specific to the mRNA expression, and normalized to the internal control Gapdh. (The raw and processed ChIP-Seq data will be stored in the GEO database.)

RESULTS

Time-dependent mRNA induction and ChIP-qPCR of known PXR-target genes

As a first approach to determine whether PCN effectively activates PXR, mRNA expression of known PXR-target genes, namely cytochrome P450 3a11 (Cyp3a11), organic anion transporting polypeptide 1a4 (Oatp1a4) and glutathione *S*-transferase m3 (Gstm3) (2,7,19), was quantified in livers of control and PCN-treated WT mice at 2, 8 and 12 h after treatment (Figure 1A). Minimal to no mRNA induction by PCN was observed 2 h after treatment, followed by a prominent mRNA induction at 8 h (Cyp3a11: 1.84-fold; Oatp1a4: 3.80-fold; Gstm3, 1.44-fold), with even more induction of Cyp3a11 and Oatp1a4 mRNA at 12 h. With this dose of PCN (200 mg/kg), mice appeared normal with no liver injury, as evident by normal levels of serum biomarkers and histological examinations (data not shown), suggesting that mRNA induction was due to PXR activation rather than pathological responses.

To determine when maximal PXR-nuclear occupancy occurs in liver, ChIP-qPCR was performed to detect the presence of PXR-response elements in the promoter regions of the three PXR-target genes (Figure 1B and C). Prior to this work, DR-3 was considered the most preferred PXR-binding motif (28–30); therefore, DR-3s were selected *in silico* using NHRscan (25) within 10 kb upstream of the transcription start sites (TSS) of the three known PXR-target genes. Prominent PXR-binding events were observed at −62-bp upstream of *Cyp3a11* and −5203bp upstream of *Gstm3* gene loci under both control and PCN-treated conditions. For these two sites, both maximal PXR-binding events as well as PXR-binding fold-change (PCN/control) were observed
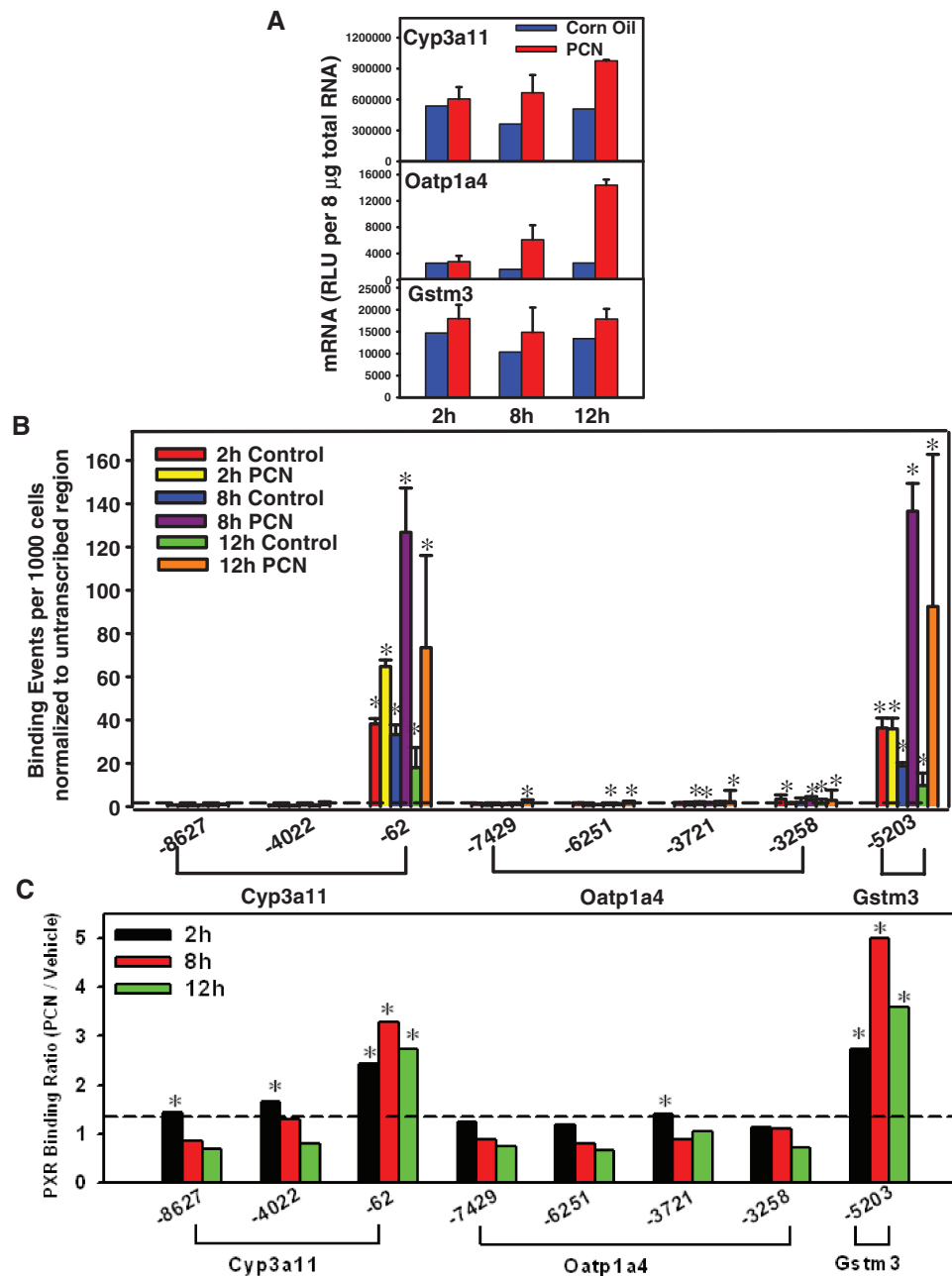
Figure 1. Time-dependent mRNA induction of known PXR-target genes (Cyp3a11, Oatp1a4, and Gstm3), and ChIP-qPCR for the optimal PXR-binding time. (**A**) Messenger RNA of Cyp3a11, Oatp1a4, Gstm3 in livers of control and PCN-treated WT mice at 2, 8 and 12 h after treatment. Data are expressed as relative light units (RLU) per 8 μg of total RNA. (**B**) PXR-binding events per 1000 liver cells, normalized to untranscribed region (background), in livers of control and PCN groups at 2, 8 and 12 h after treatment. DR-3 was selected *in silico* as the candidate PXR-response elements based on reports in the literature. ChIP was performed as described in 'Materials and Methods' section. qPCR reactions were carried out in triplicate on selected genomic regions. Signals were normalized for primer efficiency by carrying out qPCR for each primer pair using input DNA. Dashed line indicates threshold (1.5-fold of background values). Asterisks indicate regions with PXR binding at least 50% higher than background. (**C**) PXR-binding ratio (PCN/control) around the target gene loci in livers of control and PCN groups at 2, 8 and 12 h after treatment. Dashed line indicates threshold (1.3-fold of control). Asterisks indicate regions with PXR-binding ratio higher than 1.3-fold.

8 h after PCN administration (Figure 1B and C). Surprisingly, although it has been reported that Oatp1a4 is a direct PXR target in rat liver (30), none of the selected motifs appeared to be targeted by PXR in mice. In addition, the two distal DR-3 motifs from *Cyp3a11* TSS did not appear to be PXR targets.

In summary, PCN administration effectively *trans*-activated the PXR-target genes, and although maximal mRNA induction was observed 12 h after PCN treatment, optimal PXR-nuclear occupancy was observed 4 h earlier. Therefore, 8 h was selected for subsequent PXR-binding assays.

Genome-wide mapping of PXR-nuclear-binding signatures

In order to characterize the PXR-binding signatures in the entire mouse genome, ChIP-Seq was performed using livers from control and PCN-treated WT mice 8 h after dosing. The sequencing reads had ∼11 million unique matches from both control and PCN-treated mouse livers, with a similar percentage of unique matches out of the total number of tags that passed QC (control: 81%; PCN: 82%) (Table 1). After normalization, in control mice there were 3812 active regions for positive PXR-binding (Figure 2A), indicating a need for PXR binding for basal functions. PCN treatment increased the total number of active regions to 6446. This increase in PXR binding occurred in all mouse chromosomes (Supplementary Figure S1). Interestingly, whereas there were 3026 common intervals between control and PCN-treated groups, there were unique PXR-binding sites in control only (786), and in PCN-treatment only (3420) (Figure 2B). There were 2591 genes targeted by PXR within ±10 kb of the gene loci in control, which increased to 3509 genes after PCN treatment (Figure 2C). Although >70% of the PXR bindings were located within ±10 kb of the NCBI-annotated genes, positive PXR binding was observed in the intergenic regions (∼30%) (Figure 2D), suggesting that PXR might have novel functions in regulating non-protein-coding transcripts.

Using the MACS algorithm (24), we identified 1450 peaks for PXR binding that were significantly

Table 1. General sequencing information for the PXR ChIP-Seq experiments in mouse livers

Sequencing information	Control	PCN
Total tags (lines in export file)	21 993 489	21 051 280
Unique match	14 389 148	13 985 170
No match	5 616 115	5 167 805
QC (excess Ns)	208 511	194 388
non-unique match	1 729 781	1 671 711
musribosomal	17 244	15 876
newcontam.fa	32 690	16 330
Total tags passed quality control filtering	**14 248 265**	**13 097 656**
Unique match	**11 510 955**	**10 803 383**
% Unique match/total tags passed	**81%**	**82%**
No match	1 262 350	931 692
QC (excess Ns)	14 838	10 756
non-unique match	1 419 985	1 326 822
musribosomal	14 059	12 574
newcontam.fa	26 078	12 429
Total tags failed quality control filtering	7 745 224	7 953 624
Unique match	2 878 193	3 181 787
No match	4 353 765	4 236 113
QC (excess Ns)	193 673	183 632
non-unique match	309 796	344 889
musribosomal	3185	3302
newcontam.fa	6612	3901

ChIP-Seq was performed using livers from control and PCN-treated wild-typeWT mice as described in 'Materials and Methods' section. Total tags and the total tags that passed quality control filtering ([including unique match, no match, QC [(excess Ns]), non-unique match, musribosomal, and newcontam.fa), as well as percentage of unique match/total tags that passed quanlity control filtering, are listed in this table.
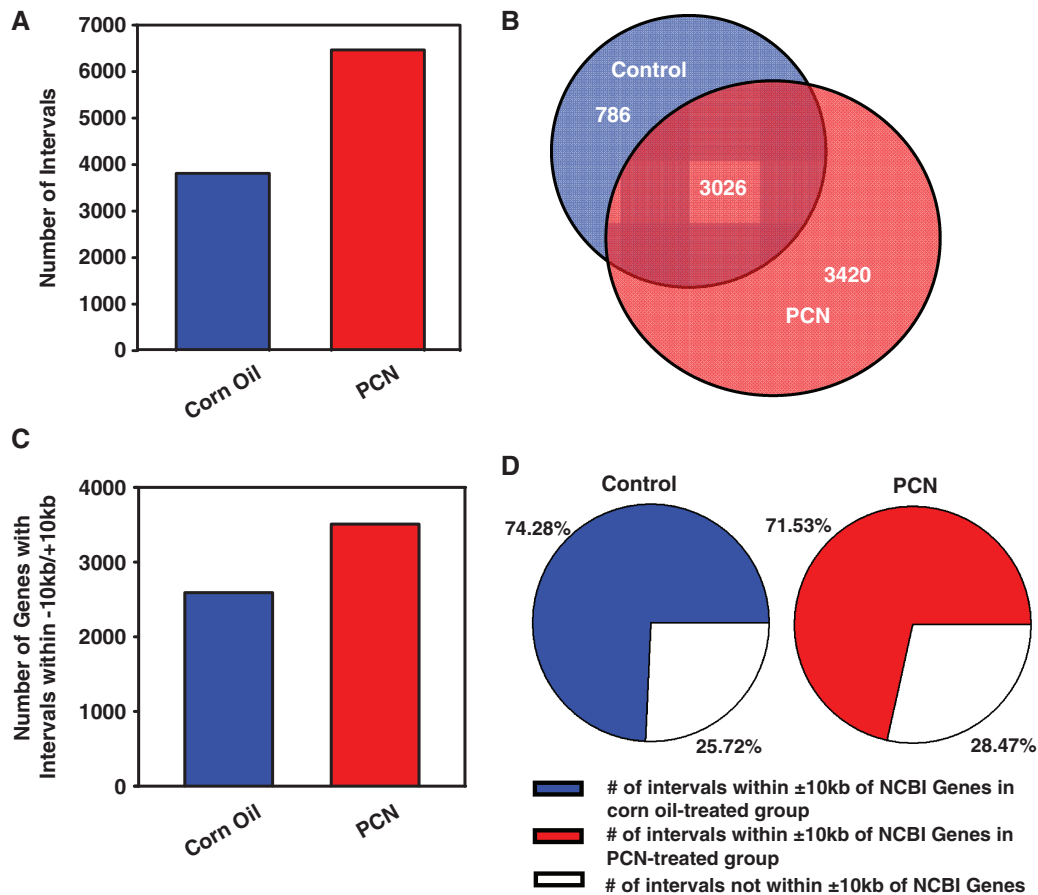
increased in the PCN-treated mouse liver ($P = 10e^{-10}$) (Supplementary Table S1). To determine the biological importance of the different PXR-binding signatures in control and PCN-treated conditions, the Ingenuity Pathway Analysis Software (IPA) (Ingenuity® Systems, http://www.ingenuity.com) revealed that the top networks for genes with significantly increased PXR-binding peaks were mainly involved in regulation of gene expression, drug metabolism, lipid metabolism and metabolic diseases, indicating an adaptive mechanism for the liver to combat chemical insult after PXR activation.

In summary, basal PXR binding markedly increased after ligand activation, resulting in more PXR binding to genes involved in both xeno- and endo-biotic metabolism.

Novel DNA-binding patterns of PXR in the mouse genome

PXR-binding profiles were generated from the top 500 ChIP-DNA sequences, with highest enrichment peak values, of both control and PCN-treated assays (control: Supplementary Table S2 and Supplementary Figure S2; PCN: Supplementary Table S3 and Supplementary Figure S3). The enrichment peak values of the top 500 ChIP-DNA sequences varied between 1575 and 47 (mean 94, and SD 134.8) in the control group, and 2099 and 79 (mean 154 and SD 170.5) in the PCN-treated group. The overlap between the two groups was quantified in terms of the number of common 150-mers (150 bp was the length of the analyzed sequence segments, set as a parameter, k_0, in the motif detection algorithm described in methods) between them. Of the 171 324 unique 150-mers in the control group and 225 746 unique 150-mers in the PCN-treated group, only 2287 were common to both groups, suggesting a significant difference between the sequences of the two groups. The reason for the difference can be attributed to the marked increase in the number of bound sequences, and the increased binding intensity of PXR after PCN treatment.

The current work offers new insights into the binding patterns of PXR, suggesting a distinct correlation in the spacer distance between the two nuclear receptor half sites of high-affinity binding sites that span a distance much further than the furthest spacer (8 bp) reported in the literature. Whereas the direct repeats, DR-3 and DR-4, of the well-established consensus motif AGTTCA featured prominently in our analysis, the everted repeats, ER-6 and ER-8, cited in the literature as PXR-binding motifs (1), was less evident (Figure 8). The current analysis suggests that by far the most favored spacer configuration for PXR binding in both control and PCN-treated mouse livers is DR-4 (Figure 3A–C: PCN treatment; Supplementary Figure S2 top motif A–C: control). Interestingly, the spacer distance distribution of half sites of the consensus direct repeat AGTTCA shows clear peaks at isochronal intervals of 5 bp, corresponding to spacer distances of 4, 9, 14, 19 and so on (i.e. spacer distances of the form $5n + 4$, $n = 0, 1, 2,...$), extending to over a hundred base pairs (spacer distance distribution,

Figure 2. A roadmap of enrichment in PXR binding (intervals) in the entire mouse genome by ChIP-Seq analysis. (**A**) Total number of active regions for PXR binding in livers from control and PCN-treated mice. (**B**) An overlay of PXR-binding active regions between control and PCN-treated groups for common and distinct binding sites. (**C**) Total number of genes with PXR bindings (intervals) within $\pm 10\,\text{kb}$ of the gene loci in control and PCN-treated groups. (**D**) Classification of PXR-binding intervals based on their genomic locations (within $\pm 10\,\text{kb}$ of NCBI genes or in intergenic regions) in control and PCN-treated groups.

Figure 3C: PCN treatment; Supplementary Figure S2 top motif C: control). This reveals a new structural configuration of half sites that favors a spacer distance that is periodically correlated with a period of 5 bp (sample partial correlations, Figure 3D: PCN treatment; Supplementary Figure S2 top motif D: control). It also indicates that the binding of PXR to the consensus direct repeat AGTTCA is partial to response elements with a spacer distance of $5n+4$ between half sites over the proximity of the half sites to each other. The observed tapering of peak densities of these sites can be, as expected, explained by the increasing motif length [a regression fit on spacer distance explains 80% of the variation between peak densities (Figure 3C: regression line)]; however, the significantly high density of DR-4 suggests that proximity between half sites is desired by PXR on a secondary level.

To confirm that DR-$(5n+4)$ is indeed the most preferred PXR-binding motifs, we analyzed the distribution of binding affinities of sequences with response elements of the form DR-$(5n+0)$, $(5n+1)$, $(5n+2)$, $(5n+3)$ and $(5n+4)$ by comparing their mean sequence ranks, ranked on binding affinity. The enrichment peak values from ChIP-Seq data were used as a surrogate for binding affinity. Ranks were assigned in descending order of binding affinity, such that sequences with high affinity received a low rank. This analysis did indeed reveal a strong association between sequences with DR-$(5n+4)$ response elements and high-binding affinity (multiple comparison, Figure 3E: PCN treatment; Supplementary Figure S2 top motif E: control). Figure 3E shows a multiple comparison plot of the mean sequence ranks with comparison intervals around them. Means in Figure 3E with comparison intervals that do not overlap are different at a significance level of $\alpha < 0.02$. Furthermore, the results of a pair-wise comparison of the mean sequence rank between sequences with the different forms of response elements is shown in Supplementary Table S5, after a Bonferroni multiple test correction at a 98% confidence level. It is evident from Figure 3E and Supplementary Table S5 that DR-$(5n+4)$, with the lowest rank, is the most common PXR-binding configuration. Similar results to the above were obtained when the multiple comparison tests were carried out directly on peak value instead of sequence rank (data not shown). Sequences with DR-$(5n+3)$ response elements had the second highest binding affinity, of which DR-3 is the first element. However, there is no
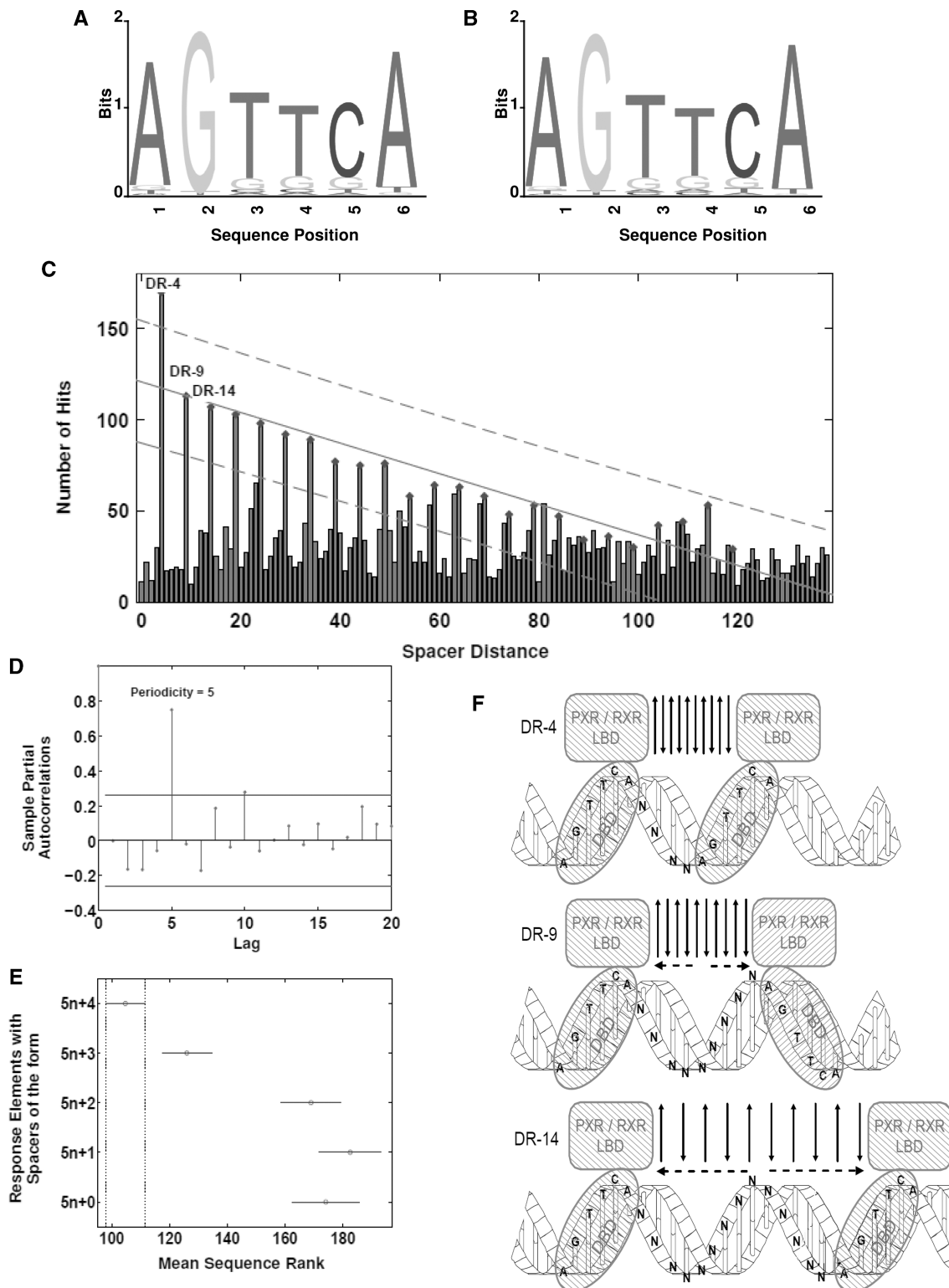
Figure 3. PXR DNA-binding profiles of the highest scoring motif (MAP score 6345.15, number of hits = 4744, number of seq. hit = 370) obtained from *in silico* analysis of ChIP-Seq data of PXR binding after PCN treatment. (**A**) Sequence logo representing the sequence conservation of the 3′ half site. (**B**) Sequence logo representing the sequence conservation of the 5′ half site. (**C**) Spacer distance distribution between the two half sites. Diamonds mark densities at spacer distances of $5n+4$. The linear regression line ($\beta_0 = -0.83$, $\beta_1 = 120.75$, norm of residuals = 70.14) of these points is shown with 95% confidence bounds (dash lines). (**D**) Sample partial autocorrelation function of the spacer distance distribution in (C), with 95% confidence bounds. The lag corresponds to distance in base pairs. A significantly high lag of 5 bp in the PACF suggests the presence of a recurring pattern of high site density occurring at 5-bp intervals. (**E**) Multiple comparison plot of mean sequence rank on binding affinity of the AGTTCA-like direct repeat with spacer distances of the form $5n+0$, $5n+1$, $5n+2$, $5n+3$ and $5n+4$, with comparison intervals around the mean estimates. Means with comparison intervals that do not overlap are significantly different at a significance level of $\alpha < 0.02$. (**F**) Illustration of the accordion model to explain the binding density of the PXR homodimer to AGTTCA like direct repeats with a $5n+4$ bp spacer distance. The PXR homodimer interface has a 10 β-strand structure that stretches to accommodate the spacer distances between the two DNA binding domains of the PXR/RXR complex as n increases.

evidence of any association with binding affinity for response elements with other spacer distances. The distribution of binding affinities did not vary significantly between the sequences with DR-$(5n+4)$ response elements (a *t*-test comparison between the top and bottom 50th percentile of sequences with DR-$(5n+4)$ response elements revealed no significant difference in their mean sequence ranks, $P = 0.59$). As can be seen, these results are observed in both PXR binding under PCN-treated and control conditions, however, as expected the observed patterns are much stronger in the PCN-treated assay than control.

As a negative control, we analyzed the background sequences of the same ChIP-Seq data for characteristics of the observed high-affinity response elements for both control and PCN-treated assays. The background sequences were comprised of all non-overlapping 500-bp upstream and 500-bp downstream sequence segments of the top 500 ChIP-DNA sequences. As expected, the relative abundance of high-affinity binding sites was substantially low in the background sequences. For example, the hit ratio (i.e. the ratio of the number of observed sites to the total number of possible sites) of the consensus direct repeat AGTTCA in the background sequences of the control and PCN-treated assays was 0.91 and 0.89%, respectively, compared to 3.70 and 4.83% in the ChIP DNA sequences. Less than 50% of the background sequences contained the consensus direct repeat AGTTCA as opposed to 74% of the PCN-treated ChIP DNA sequences and 52% of the control ChIP DNA sequences. Furthermore, unlike the ChIP DNA sequences, the background sequences displayed a uniform spacer distance distribution between half sites devoid of any periodic relationship in both control and PCN-treated groups (Supplementary Figure S4: control; Supplementary Figure S5: PCN treatment). This further corroborates the significance of the $(5n+4)$-bp spacer distance configuration between half sites in the binding of PXR to DNA. In order to ensure that no one sequence disproportionately influenced the observed PXR-binding patterns, we first augmented the total number of sequences by 10 to include the top 510 ChIP DNA sequences, and then randomly selected 500 of those sequences for a similar analysis. The procedure was repeated 100 times. The results from these experiments were consistent with our initial observations for ChIP DNA sequences from both PCN-treated and control groups (data not shown).

In summary, genome-wide motif analysis has revealed a novel structural configuration of DR-$(5n+4)$ as the major PXR-binding signature in the mouse genome, with the strongest preference towards DR-4. An 'accordion model' (Figure 3F), described in the 'Discussion' section, has been proposed to explain this observation.

ELISA-based TF binding assay confirmed novel DR-$(5n+4)$-like sequences are functional PXR-binding sites

In order to quantify binding of PXR to the different DNA-binding motifs, we developed a quantitative ELISA-based TF DNA-binding assay. The assay was developed from an existing commercially available kit for the detection of other TFs (Active Motif, Carlsbad, CA), and modified accordingly for the detection of PXR. A detailed description of the modifications to the assay can be found in the 'Materials and methods' section. As shown in Figure 9, this assay can detect basal binding (CON) and PCN-induced binding (PCN) of PXR to the following binding sites: DR-3, DR-4, DR-9, DR-14 and DR-19. For each binding site, nuclear protein from PCN-treated mouse liver resulted in increased PXR binding to the DNA sequence. In addition, binding of PXR to each DNA-binding site was effectively competed off with an unlabeled competitor oligonucleotide, whereas mutated unlabeled oligonucleotides were either ineffective or much less effective in competing with the labeled DNA sequences. Interestingly, under these binding conditions, PXR binding was strongest to DR-19, followed by DR-14, whereas DR-9, DR-4 and DR-3 were relatively comparable (Figure 9F). Collectively, these data demonstrate that PXR binds to the DR-$(5n+4)$-binding motif in PCN-treated mouse liver.

To further justify the functional significance of DR-$(5n+4)$s in gene regulation, we performed a global analysis of the genes that carry DR-$(5n+4)$-like sequences within the PXR-ChIP DNA fragments. For these genes, we compared the fold changes of PXR bindings and mRNA differential expression after PCN treatment. Among all the PXR-direct target genes that carry DR-$(5n+4)$, ∼36% of genes fall into the top 5000 differentially regulated genes (Supplementary Table S6) after PCN treatment. Among these differentially regulated genes, the majority of the altered mRNA expression (both induction and suppression) corresponds to increased PXR-binding to the DR-$(5n+4)$ (60% for induced genes, and 75% for suppressed genes), suggesting that DR-$(5n+4)$ indeed has functional significance in PXR-mediated gene regulation.

Microarray analysis reveals both PCN-induced and suppressed gene batteries correlate with increased PXR binding

To determine whether alterations in PXR binding have biological effects on gene regulation, mRNA microarray experiments were performed in control and PCN-treated WT mice 12 h after administration. PCN-treatment revealed both induced (Figure 4A, above red line) and suppressed (Figure 4A, below blue line) gene batteries. Approximately 7.4% of mouse liver genes were induced and 4.5% suppressed by PCN treatment (Figure 4B). Because PXR is a well-known transcription activator for numerous drug-processing genes, we hypothesized that more PXR binding after PCN treatment would result in mRNA induction. To test this hypothesis, an overlay was performed between microarray and ChIP-seq data (Figure 4C and D). For the direct PXR targets, there were 401 genes in the induction battery by PCN (26.1%), and the majority of these genes had increased PXR binding (18.0%) (Figure 4C). For the PCN-suppressed gene battery, 300 genes (32.1%) were direct PXR targets, and surprisingly, the majority of
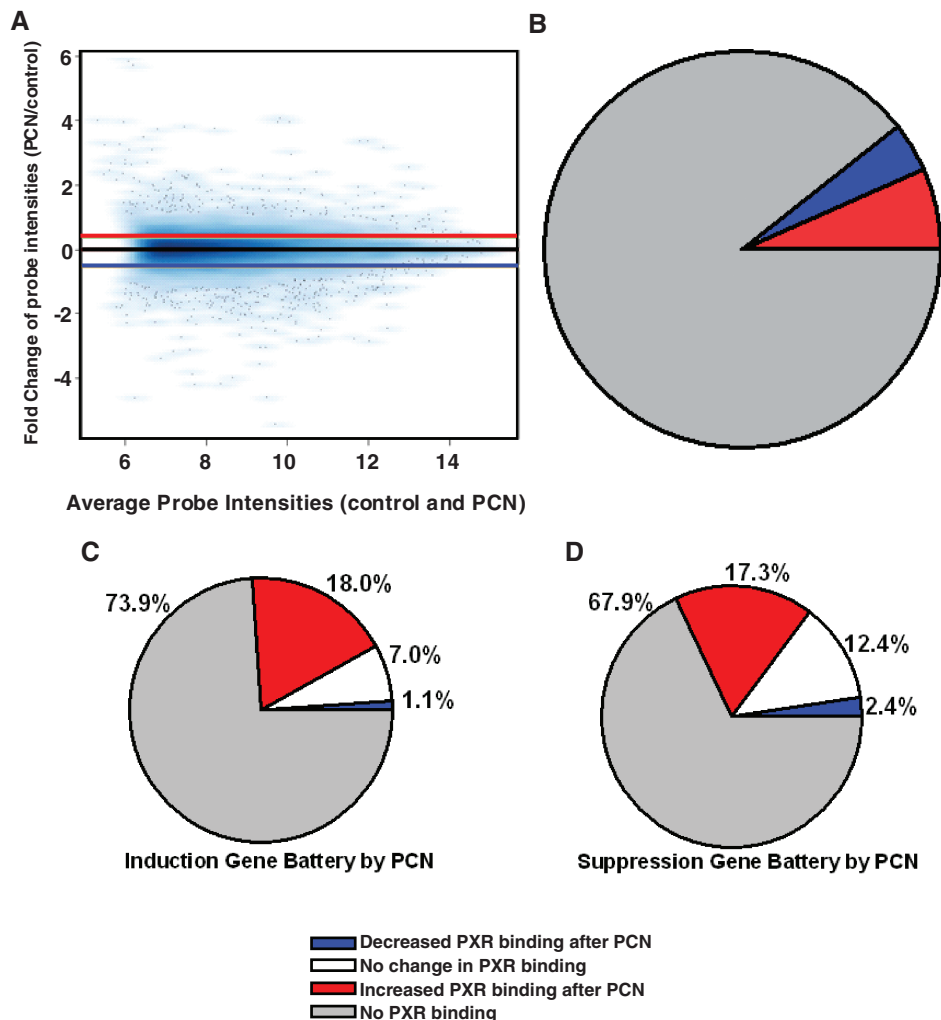
Figure 4. Differential gene expression between control and PCN-treated groups and the corresponding changes in PXR binding. (**A**) An *M–A* plot (*M* is the intensity ratio, and *A* is the average intensity for a dot in the plot) of mRNA fold-change (PCN/control) against the average of mRNAs between control and PCN-treated groups. The mRNAs were determined by microarray. Midline (black): mRNA fold change = 0. Red line: 30% increase in mRNA expression after PCN treatment. Blue line: 30% decrease in mRNA expression after PCN treatment. (**B**) Percentage of genes with mRNA induced, suppressed, or not altered by PCN treatment in the entire mouse genome. (**C**) An overlay between the genes induced by PCN (microarray) and presence of PXR binding ±10kb around the gene loci (ChIP-Seq). (**D**) An overlay between the genes suppressed by PCN (microarray) and presence of PXR binding ±kb around the gene loci (ChIP-Seq).

these down-regulated genes (17.3%) also had increased PXR binding, indicating that PXR might have both in-duction and suppression functions in gene regulation (Figure 4D). Ingenuity Pathway Analysis demonstrated that induced direct PXR-target gene batteries were mainly involved in adaptive responses like cell prolifer-ation, cell signaling, and drug metabolism. In contrast, the suppressed direct PXR-target gene batteries were mainly involved in basal maintenance functions like amino acid and carbohydrate metabolism (Supplementary Table S7). Fold changes of gene expres-sion corresponding to alterations in PXR binding are listed in Supplementary Tables S8 (induced gene battery) and S9 (suppressed gene battery). An interactive output of the top 5000 differentially regulated genes by PCN linked to their online databases (probe ID, symbol, description, chromosome location, GeneBank ID, PubMed ID, Gene Ontology, Pathways, etc.) was constructed and available in Supplementary Table S10. Approximately 70% of the

genes in both PCN induced and suppressed gene batteries were indirect PXR targets, likely due to a 'butterfly effect' triggered by PXR binding to genes encoding upstream regulatory factors, or due to the off-target effects of PCN. In summary, PXR appears to have dual functions in gene regulation. The majority of both induction and suppression of direct PXR-target genes appear to be due to increased PXR binding. Whereas the suppressed PXR direct target gene battery was mainly involved in nutrient homeostasis and cell maintenance, the induced gene battery was important for drug metabolism and cell proliferation.

Co-existence of distinct epigenetic signatures favors PXR binding to DNA

The PXR-nuclear-binding signatures have raised the question of whether a permissive chromatin environment pre-exists at PXR-binding sites favoring the interaction

between PXR protein and DNA. To explore the pre-existence of chromatin epigenetic signatures in predefining the binding signatures of PXR, three epigenetic marks, namely DNAMe, H3K4Me2 and H3K27Me3, were examined by ChIP-on-chip.

Because PXR is a key regulator of drug metabolism and transport in liver, the present study first aims to understand how epigenetic marks regulate the recruitment of PXR to its prototypical target genes encoding drug metabolizing enzymes and transporters. Therefore, as a first attempt to discover the roles of epigenetic signatures in regulating PXR binding to drug-processing genes, we selected chromosomes 5, 12 and 15, because these chromosomes are enriched with drug-processing genes (e.g. these chromosomes contain all the Cyp3a gene subfamily, Ugt2 and Ugt3 gene clusters, and the Mdr gene cluster). An overlay was performed between these epigenetic signatures and PXR binding (ChIP-Seq) in both control and PCN-treated conditions. As expected, there were more nucleotides covered by PXR after PCN treatment (Figure 5A). Minimal overlapping was observed between PXR binding and the non-permissive epigenetic mark DNAMe in control mice (0, 491 and 62 bp on chr 5, 12 and 15, respectively), and even less in PCN-treated mice (0, 337 and 30 bp on chr 5, 12 and chr15, respectively) (Figure 5B). In contrast, prominent portions of PXR-bound nucleotides overlapped with the permissive mark H3K4Me2 in control (824, 596 and 1130 bp on chr 5, 12 and 15, respectively), and even more after PCN treatment (1379, 607 and 1964 bp on chr 5, 12 and 15, respectively) (Figure 5C). There was no overlapping between PXR binding and the non-permissive mark H3K27Me3 in either control or PCN-treated conditions (0 bp overlap) (Figure 5D). The total number of nucleotides covered by each epigenetic mark, as well as other overlapping information among the three epigenetic marks and PXR binding (control and PCN) are shown in Supplementary Table S11 (A: numbers of overlapping nucleotides; B: percentage of overlapping nucleotides).

Specifically, on chromosome 5, as we reported previously (31), the prototypical PXR-target gene *Cyp3a11* is marked with strong enrichment of H3K4Me2, without any enrichment of DNAMe or H3K27Me3. Corresponding to such a permissive chromatin environment, there is enriched PXR binding in control mice with two positive binding sites, with even more PXR bindings after PCN-treatment (11 positive binding sites, highest peak value: 108-fold) (Figure 6A), and the Cyp3a11 mRNA was up-regulated by PCN in a PXR-dependent manner (Figure 1A and Table 2).

In addition, we analyzed the chromatin epigenetic signatures, the PXR-binding signatures, and target gene expression of a few other known PXR direct target genes defined by in vitro assays, including *CD36,* which encodes a fatty acid transporter for lipid homeostasis (12), as well as *Mdr1a* and *Mdr1b,* which encode the P-glycoprotein transporters for drug-excretion (29,32). Interestingly, there was no enrichment of H3K4Me2 around the *CD36* gene locus (data not shown), and corresponding to the absence of the H3K4Me2, there was no PXR bindings observed within ±10 kb of the entire *CD36* gene locus
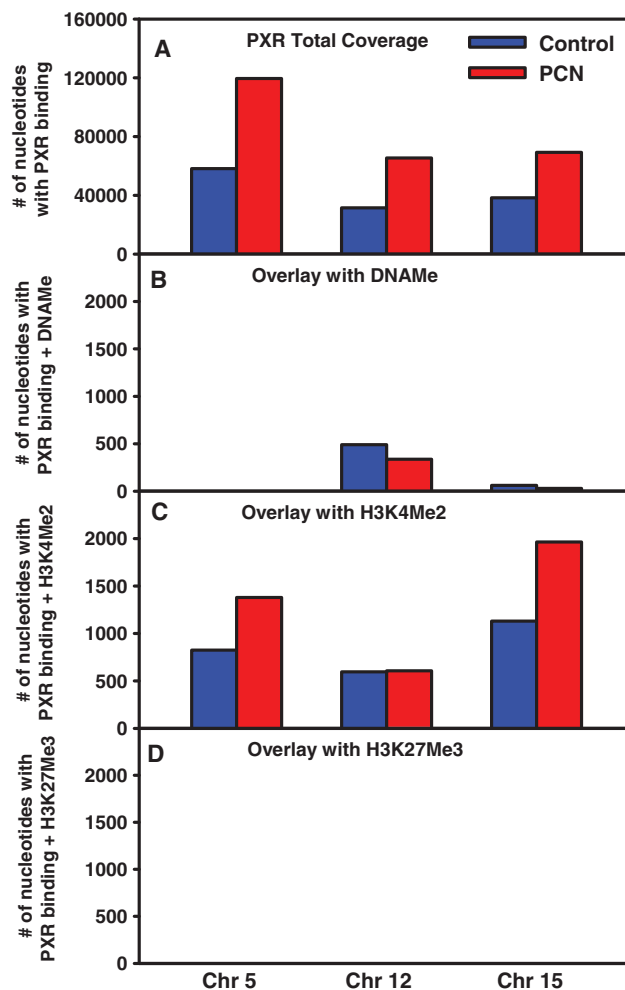

Figure 5. An overlay between PXR binding and presence of three epigenetic marks: DNAMe, H3K4Me2 and H3K27Me3 on mouse chromosome 5, 12 and 15. Presence of the three epigenetic marks was determined by ChIP-on-chip as described in 'Materials and methods' section. (**A**) Total number of nucleotides bound by PXR in control and PCN-treated groups on mouse chromosome 5, 12 and 15, based on ChIP-Seq analysis. (**B**) An overlay between PXR binding and presence of DNAMe. (**C**) An overlay between PXR binding and presence of H3K4Me2. (**D**) An overlay between PXR binding and presence of H3K27Me3.

(Supplementary Figure S6A), indicating that PXR-binding sites are likely masked in chromatin when a permissive epigenetic mark is absent. In addition, the CD36 mRNA was not induced after PCN-treatment *in vivo* (data not shown). For the Mdr1s, there was a strong enrichment in the non-permissive mark DNAMe in the 3′-UTR of both Mdr1a and 1b (Mdr1a, 5.47-fold; Mdr1b: 5.06-fold; threshold: 3.0-fold of input background signal) (Supplementary Figure S6B), and corresponding to such a non-permissive chromatin environment marked by DNAMe, there were no observed PXR bindings in control mouse liver, and after PCN-treatment, only one site with very low enrichment of PXR bindings was observed in *Mdr1a* (40-fold) and *Mdr1b* (31-fold) gene loci (Supplementary Figure S6C). As we previously reported (33), Mdr1a and 1b mRNAs are low in liver under basal
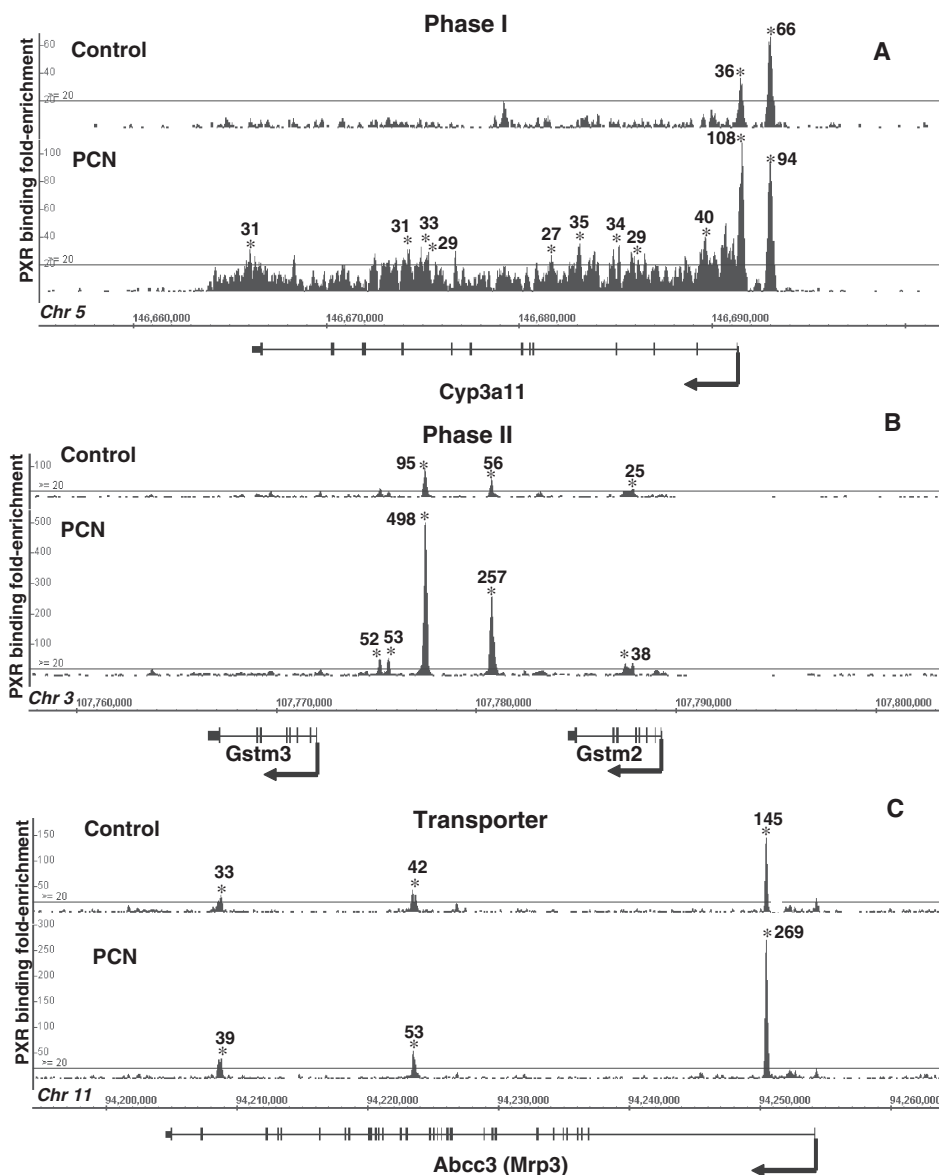
Figure 6. Location and fold enrichment of PXR binding to the representative drug-processing gene loci in control and PCN-treated groups. Image was generated by Integrated Genome Browser. (**A**) Cyp3a11; (**B**) Gstm2 and 3; (**C**) Abcc3 (Mrp3).

conditions, and neither was inducible by PXR ligands. These data serve as a negative control that further corroborates the importance of a pre-existing chromatin epigenetic environment in defining the transcription potential of PXR-target genes. The known PXR-direct targets, evident by *in vitro* assays in the literature, are not necessarily targeted by PXR *in vivo*, unless a permissive epigenetic signal pre-exists around the chromatin.

Taken together, these observations indicate the importance of a permissive chromatin environment for PXR binding and *trans*-activation of target genes. Genomic regions marked with non-permissive epigenetic signatures, namely DNAMe and H3K27Me3, did not appear to interact with PXR protein. Conversely, the presence of the permissive epigenetic mark H3K4Me2 appears to at least partially contribute in recruiting PXR to target DNA sequences.

ChIP-Seq reveals novel PXR-binding sites within critical drug-processing gene loci

Most PXR-binding studies in the literature have only examined the promoter regions of a few genes. To determine whether novel PXR-binding sites exist in other regions of a particular gene, PXR binding within ±10 kb of an entire gene locus was examined in the Integrated Genome Browser (IGB). Because PXR is a critical xenobiotic sensor, we focused on genes encoding Phase I and II drug-metabolizing enzymes and transporters. *Cyp3a11* has two positive PXR-binding sites in controls (−58 and −1578 bp from the TSS), with peak values 36- and 46-fold higher than background signals, respectively (threshold: average peak value >20-fold) (Figure 6A). After PCN treatment, PXR bindings at −1594 bp further increased to 94-fold above background,

Table 2. PXR-DNA-binding enrichments and mRNA expression of selected drug-processing genes in mouse liver in control and PCN-treated conditions

Genes	PXR binding average peak	PXR binding average peak	PXR binding fold change	mRNA fold change	mRNA fold change
Phase I	WT-Corn Oil	WT-PCN	WT PCN/Corn Oil	WT PCN/Corn Oil	PXR-null PCN/Corn Oil
Cyp3a11	14.73	**44.64[a]**	**3.03 (increase)**	**1.74[b]**	1.072
Cyp2b10	47.33[a]	221.33[a]	**4.68 (increase)**	**3.63[b]**	1.30
Aldh1a1	**38.00[a]**	**63.00[a]**	**1.66 (increase)**	**1.26[b]**	0.98
Aldh3a2	NA	NA	NA	0.89	1.07
Aldh6a1	NA	NA	NA	0.89	1.06
Aldh7a1	NA	NA	NA	0.98	1.14
Phase II	WT-Corn Oil	WT-PCN	PXR binding fold change	WT PCN/Corn Oil	PXR-null PCN/Corn Oil
Gsta1	**40.00[a]**	**66.00[a]**	**1.65 (increase)**	**15.65[b]**	0.79
Gstm1	**42.00[a]**	**104.00[a]**	**2.48 (increase)**	**1.99[b]**	1.08
Gstm2	**44.60[a]**	**179.60[a]**	**4.03 (increase)**	**1.98[b]**	0.85
Gstm3	**49.50[a]**	**215.00[a]**	**4.34 (increase)**	**16.02[b]**	0.81
Gstm4	NA	NA	NA	1.25	1.00
Gstm5	NA	NA	NA	0.82	1.00
PAPss2	**69.50[a]**	**119.00[a]**	**1.71 (increase)**	**1.29[b]**	1.03
Transporters	WT-corn oil	WT-PCN	PCN/corn oil	WT-PCN/corn oil	PXR-null PCN/corn oil
Oatp1a4	**277.00[a]**	**549.00[a]**	**1.98 (increase)**	4.24[b]	1.05
Mrp3	**73.33[a]**	**120.33[a]**	**1.64 (increase)**	**2.52[b]**	1.05
Mrp5	NA	NA	NA	0.63	0.91

The average peaks for PXR–DNA binding, fold-changes of PXR–DNA binding, as well as the mRNA fold-changes in WT and PXR-null mouse livers are shown in the table. The drug-processing genes are classified into three groups: phase-I, phase-II, and transporters.
NA: no PXR binding.
[a]PXR binding above threshold (20).
[b]Statistically significant.

and in addition, 10 new binding sites emerged, 9 of which were intronic (108-fold at 374-bp downstream of TSS, 40-fold at 1622 bp, 29-fold at 5558 bp, 34-fold at 6246 bp, 35-fold at 8326 bp, 27-fold at 9750 bp, 29-fold at 16 198 bp, 33-fold at 16 519 bp, 31-fold at 17 270 bp downstream of TSS), and one downstream of the gene locus (31-fold at 25 398-bp downstream of TSS). Similarly, *Cyp2b10*, which encodes another cytochrome P450 enzyme for Phase I metabolism, showed both enhanced PXR-binding intensity (64- to 255-fold at −2299 bp, and 44- to 174-fold at −4267 bp) and an increase in the number of binding sites (two sites in control, three sites after PCN treatment; new site: 235-fold at −9611 bp) after PCN treatment, but all the binding locations were restricted to the upstream region from the TSS (Supplementary Figure S7A). *Aldh1a1*, which encodes a critical aldehyde dehydrogenase for the oxidation of aldehyde products from lipid peroxidation, had only one PXR-binding site localized to an intron region (38-fold at 29 256-bp downstream of TSS), with further enrichment in PXR binding at approximately the same site after PCN administration (63-fold at 29 240-bp downstream of TSS) (Supplementary Figure S7B).

For Phase II metabolism, a few Gst genes, which encode glutathione *S*-transferases, essential for detoxifying electrophiles and free radicals, were selected. Gstm2 and Gstm3 form a polycistron cluster on the chromosome (Figure 6B). In control conditions, there were three PXR-binding sites observed in this cluster. After PCN treatment, there was a marked increase in PXR binding in all three sites, and in addition, two new PXR-binding sites emerged within the cluster. In contrast, *Gstm1*

had only one PXR-binding site that was located upstream of the TSS (42-fold at −7525 bp), with higher binding intensity after PCN-treatment (104-fold at −7573 bp). *Gsta1* had two PXR-binding sites upstream of the TSS in control (−797 and −7277 bp), and it gained one more binding site located inside the gene after PCN treatment (75-fold at 1538-bp downstream of TSS), in addition to increased binding intensities at the other two positions (Supplementary Figure S8A and B). The 3′-phosphoadenosine 5′-phosphosulfate synthase 2 (Papss2), a protein which synthesizes the active sulfate (PAPS) for the sulfation of chemicals, had PXR bindings exclusively within the *Papss2* gene in control (62-fold at 2305 bp, 210-fold at 6913 bp, 30-fold at 7249 bp, 37-fold at 8241 bp and 80-fold at 17 521 bp downstream of TSS), and both the binding intensity and number of binding sites increased after PCN-treatment (112-fold at 2321 bp, 364-fold at 7089 bp, 133-fold at 17 537 bp, 57-fold at 31 233 bp and 30-fold at 33 137 bp) (Supplementary Figure S8C).

For transporters, the uptake transporter Slco1a4 (Oatp1a4) and the efflux transporter Abcc3 (multidrug resistance related protein 3, Mrp3) are important for moving drugs in and out of hepatocytes. A strong enrichment in PXR binding was observed as far as 10-kb upstream from the *Oatp1a4* TSS (277-fold), which was further enriched at this position after PCN treatment (549-fold) (Supplementary Figure S9). There were no PXR-binding sites within the gene or further downstream. In contrast, *Mrp3* had three positive PXR-binding sites only inside the gene (45 538, 30 722 and 3810 bp downstream of TSS), with further enriched PXR binding at approximately the same regions after PCN treatment

(45 586, 30 722 and 3778 bp downstream of TSS) (Figure 6C).

In summary, contrary to the biased detection of PXR binding in the literature, which restricted the detection only to upstream promoter regions, ChIP-seq revealed novel PXR-binding sites in drug-processing gene loci, highlighting several interesting patterns: (i) PXR proteins can bind to upstream, downstream, or the intronic regions of a gene locus, (ii) polycistron clusters of genes can share the same PXR-binding sites and (iii) after administering the PXR-ligand, the target genes either gain more PXR-binding sites, and/or more enriched binding intensity at the original binding position.

To determine the global significance of PXR bindings in gene regulation, all the PXR DNA-binding intervals that are specifically located upstream, downstream and within a particular gene were selected as shown in Supplementary Table S12, and the mRNA expression of the genes with PXR bindings in these regions was examined and shown in Supplementary Table S13. Overall, there are 1346 intervals located specifically to the upstream of the TSS of genes, 550 in control, and 796 in PCN-treated group; 664 intervals downstream (242 in control and 422 in PCN-treated group); 3871 intervals located in genes (1389 in control and 2482 in PCN-treated group) (Supplementary Table S12). A confidence score (*P*-value) was calculated between altered PXR binding and changes in gene expression. Interestingly, although the *P*-values for the upstream and downstream genes are not significant, it is significant for PXR binding within the genes, highlighting the importance of intronic PXR bindings in regulating target gene expression.

To strengthen the evidence that critical drug-processing genes in liver are bona fide PXR targets, we performed mRNA assays of these genes in control and PCN-treated WT and PXR-null mice (Table 2). The Phase I metabolism genes, *Cyp3a11*, *Cyp2b10* and *Aldh1a1,* all had increased PXR binding after PCN treatment, corresponding to a significant increase in mRNA expression of these genes in livers of WT mice. Such mRNA induction was not observed in PXR-null mice (Table 2). As a negative control, *Aldh3a2, 6a1 and 7a1*, which have no PXR binding regardless of PCN treatment, were selected. As expected, the mRNAs of these genes were not altered by PCN in either WT or PXR-null mice.

In Phase II metabolism, the mRNAs of Gsta1, m1, m2, m3, as well as Papss2 were induced by PCN in WT mice corresponding to more PXR binding. Such mRNA induction was blocked in PXR-null mice. For Gstm4 and m5, PXR did not bind in either control mice or PCN-treated mice, and there was no change in mRNA expression in livers of either WT or PXR-null mice.

Similarly, the mRNAs of the transporters, Oatp1a4 and Mrp3 were induced by PCN in WT mice, corresponding to increased PXR binding, but no mRNA induction was observed in PXR-null mice. As a negative control, Mrp5, which had no PXR binding regardless of

PCN-treatment, had no changes in mRNA expression in either WT or PXR-null mice.

DR-4 is the motif of choice for PXR-targeted drug-processing genes

For all these drug-processing genes directly targeted by PXR, local motif analysis of their ChIP DNA fragments further justified DR-4 as the most prevalent DNA-binding motif for PXR after PCN treatment (Figure 7), suggesting that *trans*-activation of these critical drug-processing genes is mainly due to a DR-4 effect.

Taken together, the present data have established positive associations between PXR-binding and mRNA induction of critical drug-processing genes, and using PXR-null mice, we have confirmed that PXR is necessary in trans-activating these genes.

DISCUSSION

The present study has provided the first *in vivo* evidence of PXR-nuclear binding signatures in mouse liver on a genome-wide scale, both under basal conditions and
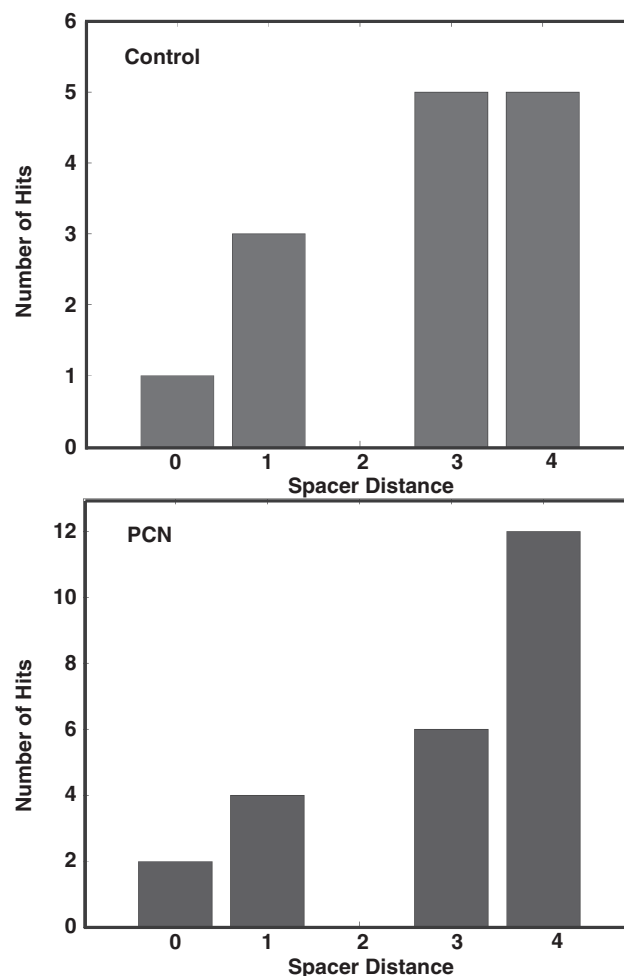
Figure 7. The densities of the PXR DNA-binding motifs (DR-0 to DR-4) among all the PXR-bound drug-processing genes in control (**A**) and PCN-treated mouse livers (**B**).

after chemical activation. Unbiased ChIP-Seq has demonstrated that DR-4 is the most preferred DNA-binding motif of the PXR protein, challenging the existing paradigm that implies DR-3 and ER-6 as the common PXR-binding motifs. Furthermore, *in silico* motif analysis has shed light to a novel DNA-binding pattern for PXR binding, with a strict spacer requirement of $5n + 4$ nucleotides between the two direct-repeat half sites. PXR appears to have dual functions in gene regulation, as the majority of both induced and suppressed PXR-direct target genes correspond to more PXR binding. In addition, using ChIP-on-chip, we have investigated the importance of three distinct epigenetic signatures in predefining the PXR-binding patterns and transcription potential. Finally, we have characterized novel PXR-binding locations within the drug-processing gene loci, and used PXR-null mice to justify that these drug-processing genes are bona fide PXR targets.

The arrival of a new generation of high-throughput technologies for querying TF binding sites in their native environment in a whole genome scale has brought about a new set of challenges to computational tools that perform *ab initio* prediction of DNA-binding targets. Foremost among them is the need to deal with very large datasets. Large datasets create a strain on traditional algorithms designed for this task; as a result, many of them place strict limitations on the size of their input datasets. We addressed this problem by transforming the sequence space, on which traditional algorithms operate, to a k-mer space. The transformation from sequence space to k-mer space can be achieved in linear time for all reasonable length TF binding sites. Working in the k-mer space is efficient, not only due to its substantially smaller size, but also because of its indexing properties that provide rapid access to individual k-mers. This not only paves the way for analyzing large data sets but also facilitates the analysis of a wider contrast of the target binding sites. This is very advantageous when working with response elements that are nuclear receptor half sites with variable spacer distances. A testament to this is the comprehensive analysis of PXR-binding targets in the mouse genome described in this article.

More interestingly, we have established a new binding paradigm that emphasizes on a isochronously extending spacer distance between half sites. The PXR protein shows a clear binding preference to the AGTTCA consensus direct repeat with a spacer distance of $(5n + 4)$ bps between the two half sites, suggesting that PXR binding depends more on a distinct periodically related configuration, rather than the simple proximity between the two half sites. There is no observable difference in binding affinity of response elements with half sites on the same face (half sites with spacer distances of the form, $10n + 4$), or on opposite faces (half sites with spacer distances of the form, $10n + 9$), (*t*-test, $P = 0.76$). This novel and interesting observation raises the question of whether there are any intrinsic features in the DNA or the PXR–protein complex enforcing such a periodic pattern. As demonstrated by Watson and Crick in 1953, the double helical structure of DNA has a 3.4 Å interval between the two adjacent bases, with 10 such intervals per turn

(34 Å) (34). If one assumes that at the very beginning of PXR binding, the double helical structure of DNA is still preserved with only partial destruction of hydrogen bonds, then the length of one AGTTCA-like half site exactly matches the length of half a helical turn from trough to peak [$3.4 Å \times (6–1) = 17 Å$]. Therefore, DR-4, which is most preferred by the PXR–protein complex (Figure 3C, Supplementary Figure S2 top motif C and S3 top motif C), will form a structure of three half turns in tandem, with the 4bp spacers occupying the middle half turn (Figure 3F). Similarly, DR-9 incorporates one more half turn between the two half sites, as five new intervals are introduced in the spacer region (Figure 3F). Following this rule, as n half turns are incorporated between the two half sites, a DR $(5n + 4)$ pattern emerges ($n = 0,1,2,...$), which has a tandem structure of two half sites on the outermost half turns, and $(n + 1)$ half turns in between, forming the spacer region. Based on the observations in the present study (Figure 3C), it seems that DNA interacts with PXR protein in a functional unit of half a helical turn, which means DNA is more likely targeted by PXR if there are two AGTTCA-like half sites still preserved in the helical structure, separated by at least one half turn in between.

If the proposed structure of DR $(5n + 4)$ is really the preferred DNA configuration for PXR binding, it suggests that some symmetry exists in the PXR–protein complex. Based on amino acid sequence alignment, we predicted the 3D structure of the mouse PXR protein. The mouse PXR ligand binding domain shows highest similarity to the tethered human PXR-LBDSRC-1p apoprotein (35) (PDB ID: 3CTB; E value: $8e-139$), which forms a PXR–PXR homodimer interfaced with 10-β strands. Very recently, computational biologists have demonstrated that PXR functions as a heterotetramer with the unique β-stranded interface of the PXR homodimer in the middle, and two RXR proteins on the outside (36). In addition, the key amino acids forming the homodimer interface are highly conserved across species and the disruption of these amino acids result in decreased transcription activity of PXR (37).

Based on the above evidence, we propose an 'accordion model' to explain the interesting phenomenon of the DR-$(5n + 4)$ periodic PXR-DNA binding patterns (Figure 3F). The basic assumption of this model is that the PXR homodimer interface, which has a β-sheet structure with 10 β-strands, is stretchable like the bellows of an accordion that fine-tunes the distances between the two DNA-binding domains of the PXR/RXR complex. The lowest energy configuration of the PXR/RXR complex corresponds to the binding to DR-4, and as the bellows of the accordion stretch out with less favorable configuration, the PXR/RXR–protein complex allows the incorporation of an integral number of half helical turns between the two DNA-binding domains. As more half turns of DNA are incorporated in the spacer region, the interface is further stretched out, which still allows the interactions between the PXR–protein complex and its response element, but with decreased predilection.

Prior to this work, there were a few studies showing that the distance between two TF binding sites can be changed

by an integral multiple of 10.5 bp (the approximate number of base pairs in one turn of the DNA helix in solution without impairing function (38–40), but changing the distance by a number of base pairs very different from multiples of 10.5 bp causes severe impairment (41). The present study also revealed that the PXR–protein complex seems to interact with DNA with a periodic pattern. It has provided novel aspects of the periodicity of DNA related to transcriptional binding, by demonstrating that PXR interacts with the smallest unit of a half turn of DNA instead of a full turn. Indeed, our ELISA-based TF-binding assays provided direct evidence to this fact by showing that PXR indeed binds to DR-4 and DR-14 (which has one-half turn and 1.5-half turns between the two half sites, respectively).

Contrary to reports in the literature, many motifs such as ER-6 and ER-8, which are known PXR-response elements *in vitro* (1), do not appear to be the prominent PXR targets *in vivo* (Figure 8). Many well established PXR-direct target genes as evident *in vitro*, including MDR1/Mdr1 and CD36 (12,29), do not appear to have significant enrichment of PXR bindings *in vivo*. In addition, although the CYP4F12 gene was recently identified as a novel PXR-direct target in cryopreserved human hepatocytes (42), its mouse homolog Cyp4f14 is not targeted by PXR in either control or PCN-treated conditions *in vivo* liver (data not shown). Such discrepancies in observations are likely due to the distinct *in vivo* chromatin structures that mask numerous PXR response elements by packing them into nucleosomes, so that only regions marked by a permissive chromatin signal can be accessed by the PXR protein. Genomic regions enriched in H3K4me2 have been shown to associate with DNase I hypersensitivity, a marker of active-regulatory regions (43), suggesting that the genomic regions marked with H3K4Me2 are more accessible to TFs. Furthermore, removal of H3K4Me2 prevents the recruitment of the TF FoxA1 to its response elements (44), providing direct evidence that H3K4Me2 predefines the binding signature of some TFs. In contrast, minimal PXR bindings occurred at genomic regions marked with DNAMe or H3K27Me3, suggesting that these are non-permissive regions for PXR binding, that likely explains why the Mdr1s, which are PXR-direct-target genes according to *in vitro* evidence (29), do not have PXR binding sites *in vivo* (as both of them are marked by DNAMe).

For CD36, it has been shown that PXR binds to and induces the mRNA expression of the CD36 gene in mouse liver following PXR agonist administration, suggesting that CD36 is a direct PXR-target gene (12). However, in the present study, corresponding to the absence of a permissive chromatin environment, PXR binding was not observed in the CD36 gene locus or ±10kb. The absence of PXR binding to the CD36 gene locus is not likely due to false negative detection, because as a positive control, we have identified positive PXR binding (Figure 6A) and presence of the permissive epigenetic mark H3K4Me2 upstream of the prototypical target gene Cyp3a11 (31), and the detection threshold was set with sufficient statistical justification depending on the

number of tags sequenced, information on positive and negative test sites, as well as estimation of FDR, as recommended by Genpathway (San Diego, CA). Such differences between our observations and the previous findings (12) could be due to differences in detection methods. For example, Zhou *et al.* used virus protein under the control of the albumin promoter to over-express human PXR. Although humanized transgenic mouse is a powerful tool in terms of clinical significance, the Alb-VP-hPXR does not necessarily represent the normal amount of mouse PXR protein under physiological conditions, and the human PXR does not necessarily have the same DNA-binding signatures compared to the mouse PXR. In the previous manuscript (12), ChIP assays identified the recruitment of mPXR onto the mCD36 promoter. However, the PXR detected was a HA-tagged protein by a hydrodynamic gene delivery method, and therefore does not represent the physiological levels of mouse PXR. The advantages of using ChIP-sequencing technology to detect PXR DNA-binding signatures are that it provides an unbiased genome scale information of the cistrome at the chromatin level, instead of naked DNA (as commonly used for gel shift assays), and reflects the PXR bindings under physiological conditions, rather than artificially over-expressed conditions.

It is interesting to note that the basolateral uptake transporter Slco1a4 (Oatp1a4) gene has been reported to be a direct PXR-target gene in rat liver, because PCN significantly enhanced the rat Oatp1a4 mRNA expression, and electrophoretic mobility shift assays showed that PXR binds to a DR-3 consensus sequence with the highest affinity at around 8-kb upstream of the TSS, and it binds to two other DR-3 motifs (−5 and −8 kb, respectively) with a lower affinity (45). However, in mouse liver, none of the four PXR-consensus sequences within 10-kb upstream of the TSS of *Oatp1a4*, appeared to bind PXR protein in either control or PCN-treated conditions (Figure 1B), suggesting that PXR-DNA binding signatures are species specific. Interestingly, unbiased
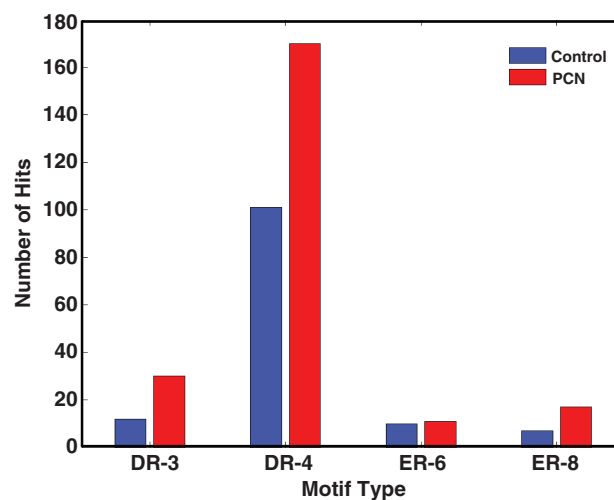

Figure 8. The densities of the known PXR DNA-binding motifs (DR-3, DR-4, ER-6 and ER-8) among all the top 500 PXR-ChIP DNA sequences.

ChIP-sequencing has identified one site at 10-kb upstream of the TSS of *Oatp1a4* with strong enrichment of PXR binding in control liver, which further increased after PCN treatment (Supplementary Figure S9). Although there were no known PXR-consensus sequences (DR-3, DR-4, ER-6 and ER-8) observed in this region using conventional methods, which have limited settings for the spacer distance (NHR-scan and NUBI-scan, data not shown), our novel motif detection algorithm has identified one DR-9-like consensus sequence present in the PXR-binding site for Oatp1a4, which is shown for the first time in the present study to be a PXR DNA-binding motif (DR-$(5n + 4)$ periodic pattern, $n = 1$). TF-binding assays validated that PXR protein indeed binds to the DR-9 sequence (Figure 9), and PXR-dependent mRNA induction of Oatp1a4 was observed in mouse liver (Table 2). Taken together, our results from ChIP-sequencing, motif analysis, TF-binding assays, as well as mRNA assays in WT and PXR-null mice have lead to the conclusion that Oatp1a4 is a direct PXR-target gene in mouse liver, with a novel PXR DNA-binding signature *in vivo*.

The current study has revealed both PXR-induced and suppressed direct-target gene batteries, raising the question of how the same TF exerts different gene regulatory functions. Previous studies have shown that the C-terminal helix of PXR (termed AF-2) is responsible for recruiting either co-activators, like the steroid receptor co-activator 1 (SRC1), or co-repressors, like the nuclear receptor co-repressor 2 (NCoR2/SMRT) (42,46). In addition, PXR can interact with other TFs and alter their target-gene expression (47). This shows the
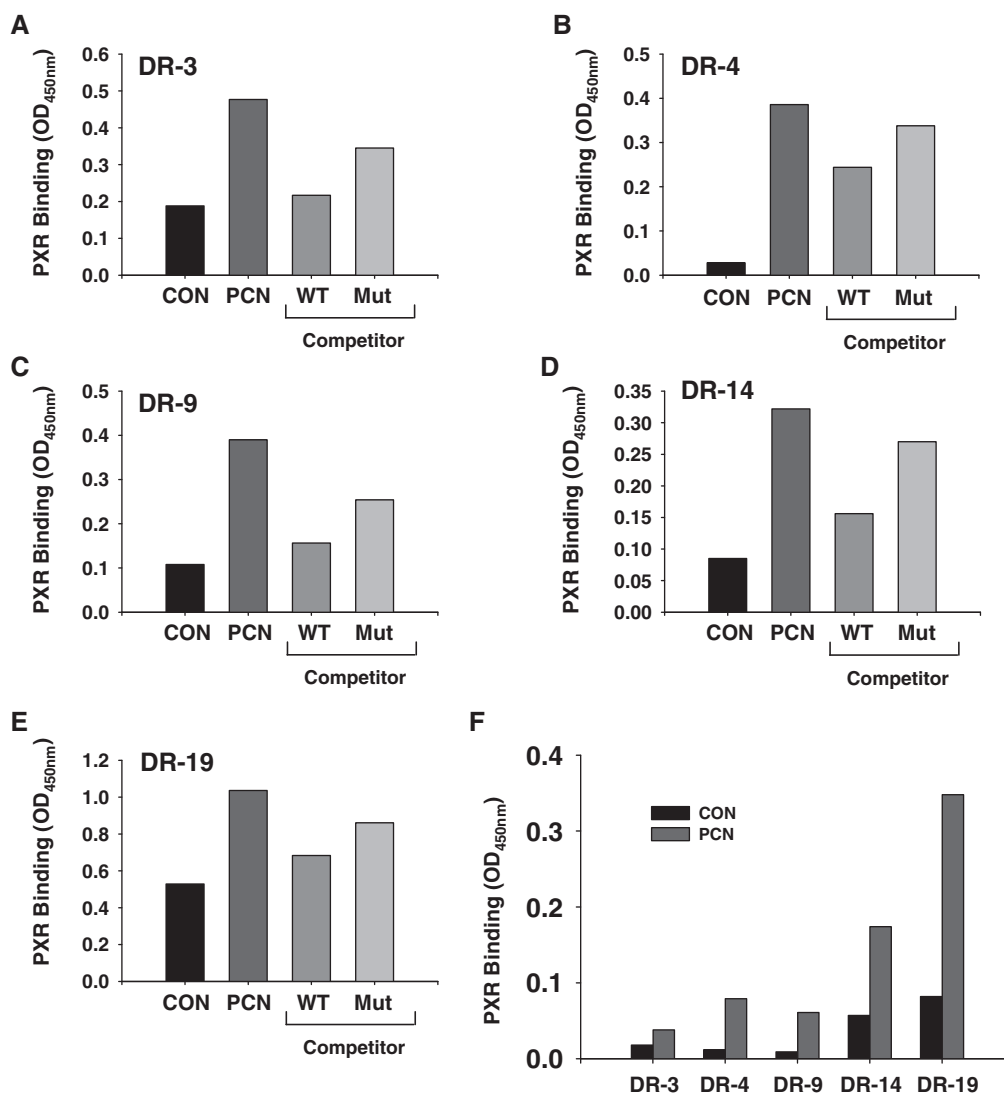

Figure 9. Quantification of PXR binding to the DR-3, DR-4, DR-9, DR-14 and DR-19 DNA-binding motifs. Nuclear protein extracts from the livers of mice treated with corn oil vehicle control (CON) or the PXR agonist, PCN, were incubated with oligonucleotides corresponding to various DNA-binding motifs. The binding of PXR to the various DNA-binding motifs was quantified using an ELISA-based TF binding assay, as detailed in the Methods section. Binding of PXR to (**A**) DR-3, (**B**) DR-4, (**C**) DR-9, (**D**) DR-14 and (**E**) DR-19. An unlabeled oligonucleotide competitor was included for each DR DNA-binding motif to confirm the specificity of the assay (WT Comp) as well as a mutated oligonucleotide competitor which should not compete effectively with the positive control (PCN treatment). (**F**) PXR binding to all the DNA-binding motifs was analyzed simultaneously in order to directly compare the intensity of binding among the different DR sequences.

complexity of the role of PXR in regulating gene expression *in vivo*, where numerous other transcription regulators are present.

Unbiased ChIP-Seq analysis has identified numerous PXR-binding sites within introns and downstream of the target gene loci, in addition to binding sites found in promoters (Figure 6 and Supplementary Figures S3–S5), which are in agreement with genome-wide ChIP-based studies for PXR binding, using cryopreserved human hepatocytes (42). Numerous studies in the literature have shown that regulatory elements in introns, exons and further downstream are critical in regulating transcription (48–51). Therefore, it appears that PXR-binding sites are likely important wherever they are positioned around the gene loci, and do not necessarily need to be located in the 5′ flanking regions. It has been shown that widely spaced genomic response elements (e.g. in promoters and terminators) can communicate in the third dimension by forming 'chromatin loops', which facilitates the re-initiation of transcription (52). It is possible that PXR might use the same strategy to enhance the transcription efficiency of target genes by drifting among multiple binding sites within a gene locus.

It is generally considered that PXR is a key regulator of drug metabolism and transport in liver. Therefore, it is critical to understand how PXR is recruited to the response elements of the target genes encoding drug-metabolizing enzymes and transporters. As a first attempt to discover the roles of epigenetic signatures in regulating PXR bindings to drug-processing genes, we selected chromosomes 5, 12 and 15, because these chromosomes are enriched with drug-processing genes. For example, for Phase I drug metabolism, the entire Cyp3a family members are transcribed from chromosome 5, and it is well known that cytochrome P450 3A enzymes are major drug-metabolizing enzymes, responsible for metabolizing >50% of prescribed drugs and some endogenous compounds (53). For Phase II drug metabolism, the entire Ugt2 and Ugt3 gene clusters are localized to chromosome 5 and 15, respectively. Ugts encode UDP-glucuronosyltransferases that conjugate lipophilic substrates with glucuronic acid, thereby increasing hydrophilicity and enhancing excretion through bile and urine, serving as one of the major mechanisms for detoxification of chemicals. For transporters, the entire Mdr gene cluster (Mdr1a, 1b and 2), which encodes critical multidrug-resistance transporters for the excretion of drugs (Mdr1a and 1b) and phospolipids (Mdr2), are localized on chromosome 5. In addition, these chromosomes contain genes encoding several critical TFs and cofactors for xenobiotic metabolism and disposition, including the aryl hydrocarbon receptor (Ahr) on chromosome 12, peroxisome proliferator activated receptor alpha (PPARα) on chromosome 15, and peroxisome proliferative activated receptor, gamma, coactivator 1 alpha (PGC-1α) on chromosome 5. Therefore, these three chromosomes were selected to determine the roles of epigenetic signatures in PXR–DNA bindings and the target gene expression. The roles of various epigenetic marks, including DNA and histone methylation and various other histone modifications (acetylation, phosphorylation,

etc.) in PXR–DNA binding in the entire genome will be examined in future studies.

PXR-mediated gene regulation is apparently tissue and gender specific. For example, the PXR agonist PCN induces Mdr1a, 1b and Mdr2 gene expression in mouse brain, induces Mdr1a in intestine, but induces none of the Mdrs in liver (33). For Gsts, it has been shown by Gong *et al.* that PXR regulates glutathione *S*-transferase expression in an isozyme-, tissue- and gender-specific manner (54). It still remains a mystery how PXR-target genes are regulated in tissue and gender specific manner. Accumulating evidence in the literature suggests that the tissue specificity of drug-processing gene expression may be due to distinct epigenetic mechanisms. For example, it has been shown by Dr Sugiyama's group that epigenetic regulation underlies the tissue-specific expression of mouse liver-specific transporters (55). They examined the DNA methylation and histone modification profiles near the TSS of various transporters in liver, kidney and cerebrum. Genome-wide DNA methylation profiling demonstrated that the CpG dinucleotides around the TSS of liver-enriched transporters were hypomethylated in liver, but hypermethylated in kidney and cerebrum where these transporters are expressed lowly, whereas the opposite pattern was observed for kidney and cerebrum-enriched transporters; whereas the histone H3 associated with liver-enriched transporter gene promoters was hyperacetylated in the liver but was acetylated very little in kidney and cerebrum, and the opposite pattern holds true for kidney and cerebrum-enriched genes. These results suggest the epigenetic mechanisms are critical in determining the tissue specificity of genes. In addition, tissue-specific cofactors may contribute to tissue specificity of gene expression. For example, very recently, the DNA-binding signatures of FXR in mouse liver and ileum have been characterized using ChIP-Seq, and a large degree of tissue-specific FXR binding has been identified, with only 11% of total sites shared between liver and intestine. Motif analysis revealed a half nuclear receptor binding site, normally bound by a few orphan nuclear receptors, adjacent to the FXR-response elements, indicating possible involvement of some orphan nuclear receptors in modulating FXR function (56). In summary, the tissue specificity of nuclear-receptor-mediated gene expression may be determined by epigenetic mechanisms and tissue-specific cofactors. Presence of PXR–DNA bindings, co-existence of various DNA and histone epigenetic marks, as well as the interactions between PXR and critical cofactors, will be determined by ChIP-Seq in future studies.

In conclusion, the current work offers new insights in understanding the nuclear binding signatures, direct targets and potential epigenetic cofactors of the hepatic nuclear receptor PXR *in vivo*, paving the path for further understanding the multifaceted roles of hepatic PXR in mediating the physiological and pharmacological responses in humans. In spite of extensive research in the area, we are still learning the true nature of the nuclear receptor–DNA interactions that play a key part of gene regulation during physiological and pharmacological responses. By looking beyond the generally expected short

nuclear receptor binding motif (no more than eight spacers as reported in the literature) using a novel computational method, we were able to identify not only very long regulatory elements, but also an underlying periodic pattern in the spacer distance. In addition, with the advent of next generation DNA sequencing and other genomic-scale technologies, it is now possible to obtain genome-scale *in vivo* protein–DNA interaction data, which not only allows us to determine the DNA-binding signatures of a particular nuclear receptor, but also its interactions between co-localizing regulatory factors, including epigenetic marks. By combining the genome-wide location analysis of the nuclear receptor PXR, extensive computational analysis on its DNA-binding motifs, its co-localization with epigenetic signatures, microarray profiling of PXR-target gene expression and pathway analysis, and validation in PXR-null mice, the present study has set an example of exploring novel functions of nuclear receptors *in vivo*.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Kliewer,S.A., Goodwin,B. and Willson,T.M. (2002) The nuclear pregnane X receptor: a key regulator of xenobiotic metabolism. *Endocr. Rev.*, **23**, 687–702.
2. Kliewer,S.A., Moore,J.T., Wade,L., Staudinger,J.L., Watson,M.A., Jones,S.A., McKee,D.D., Oliver,B.B., Willson,T.M., Zetterstrom,R.H. *et al.* (1998) An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway. *Cell*, **92**, 73–82.
3. Pacyniak,E.K., Cheng,X., Cunningham,M.L., Crofton,K., Klaassen,C.D. and Guo,G.L. (2007) The flame retardants, polybrominated diphenyl ethers, are pregnane X receptor activators. *Toxicol. Sci.*, **97**, 94–102.
4. Staudinger,J.L., Goodwin,B., Jones,S.A., Hawkins-Brown,D., MacKenzie,K.I., LaTour,A., Liu,Y., Klaassen,C.D., Brown,K.K.,

5. Reinhard,J. *et al.* (2001) The nuclear receptor PXR is a lithocholic acid sensor that protects against liver toxicity. *Proc. Natl Acad. Sci. USA*, **98**, 3369–3374.
5. Hughes,J.D., Estep,P.W., Tavazoie,S. and Church,G.M. (2000) Computational identification of cis-regulatory elements associated with functionally coherent groups of genes in Saccharomyces cerevisiae. *J. Mol. Biol.*, **296**, 1205–1214.
6. Bailey,T.L. and Elkan,C. (1995) The value of prior knowledge in discovering motifs with MEME. In *Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology*, 21–29.
7. Gunewardena,S. and Zhang,Z. (2008) A hybrid model for robust detection of transcription factor binding sites. *Bioinformatics*, **24**, 484–491.
8. Maher,J.M., Cheng,X., Slitt,A.L., Dieter,M.Z. and Klaassen,C.D. (2005) Induction of the multidrug resistance-associated protein family of transporters by chemical activators of receptor-mediated pathways in mouse liver. *Drug Metab. Dispos.*, **33**, 956–962.
9. Cheng,X., Maher,J., Dieter,M.Z. and Klaassen,C.D. (2005) Regulation of mouse organic anion-transporting polypeptides (Oatps) in liver by prototypical microsomal enzyme inducers that activate distinct transcription factor pathways. *Drug Metab. Dispos.*, **33**, 1276–1282.
10. Alnouti,Y. and Klaassen,C.D. (2008) Tissue distribution, ontogeny, and regulation of aldehyde dehydrogenase (Aldh) enzymes mRNA by prototypical microsomal enzyme inducers in mice. *Toxicol. Sci.*, **101**, 51–64.
11. Zhou,J., Febbraio,M., Wada,T., Zhai,Y., Kuruba,R., He,J., Lee,J.H., Khadem,S., Ren,S., Li,S. *et al.* (2008) Hepatic fatty acid transporter Cd36 is a common target of LXR, PXR, and PPARgamma in promoting steatosis. *Gastroenterology*, **134**, 556–567.
12. Zhou,J., Zhai,Y., Mu,Y., Gong,H., Uppal,H., Toma,D., Ren,S., Evans,R.M. and Xie,W. (2006) A novel pregnane X receptor-mediated and sterol regulatory element-binding protein-independent lipogenic pathway. *J. Biol. Chem.*, **281**, 15013–15020.
13. Guzelian,J., Barwick,J.L., Hunter,L., Phang,T.L., Quattrochi,L.C. and Guzelian,P.S. (2006) Identification of genes controlled by the pregnane X receptor by microarray analysis of mRNAs from pregnenolone 16alpha-carbonitrile-treated rats. *Toxicol. Sci.*, **94**, 379–387.
14. Margolin,A.A., Palomero,T., Sumazin,P., Califano,A., Ferrando,A.A. and Stolovitzky,G. (2009) ChIP-on-chip significance analysis reveals large-scale binding and regulation by human transcription factor oncogenes. *Proc. Natl Acad. Sci. USA*, **106**, 244–249.
15. Schmidt,D., Wilson,M.D., Spyrou,C., Brown,G.D., Hadfield,J. and Odom,D.T. (2009) ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods*, **48**, 240–248.
16. Tompa,M., Li,N., Bailey,T.L., Church,G.M., De Moor,B., Eskin,E., Favorov,A.V., Frith,M.C., Fu,Y., Kent,W.J. *et al.* (2005) Assessing computational tools for the discovery of transcription factor binding sites. *Nat. Biotechnol.*, **23**, 137–144.
17. Bird,A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
18. Barski,A., Cuddapah,S., Cui,K., Roh,T.Y., Schones,D.E., Wang,Z., Wei,G., Chepelev,I. and Zhao,K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823–837.
19. Petrick,J.S. and Klaassen,C.D. (2007) Importance of hepatic induction of constitutive androstane receptor and other transcription factors that regulate xenobiotic metabolism and transport. *Drug Metab. Dispos.*, **35**, 1806–1815.
20. Feng,W., Liu,Y., Wu,J., Nephew,K.P., Huang,T.H. and Li,L. (2008) A Poisson mixture model to identify changes in RNA polymerase II binding quantity using high-throughput sequencing technology. *BMC Genomics*, **9(Suppl. 2)**, S23.
21. Ji,H., Jiang,H., Ma,W., Johnson,D.S., Myers,R.M. and Wong,W.H. (2008) An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat. Biotechnol.*, **26**, 1293–1300.

22. Kharchenko,P.V., Tolstorukov,M.Y. and Park,P.J. (2008) Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.*, **26**, 1351–1359.

23. Xu,H., Wei,C.L., Lin,F. and Sung,W.K. (2008) An HMM approach to genome-wide identification of differential histone modification sites from ChIP-seq data. *Bioinformatics*, **24**, 2344–2349.

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

25. Sandelin,A. and Wasserman,W.W. (2005) Prediction of nuclear hormone receptor response elements. *Mol. Endocrinol.*, **19**, 595–606.

26. Gentleman,R.C., Carey,V.J. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.

27. Cui,Y.J., Yeager,R.L., Zhong,X.B. and Klaassen,C.D. (2009) Ontogenic expression of hepatic Ahr mRNA is associated with histone H3K4 di-methylation during mouse liver development. *Toxicol. Lett.*, **189**, 184–190.

28. Huss,J.M. and Kasper,C.B. (2000) Two-stage glucocorticoid induction of CYP3A23 through both the glucocorticoid and pregnane X receptors. *Mol. Pharmacol.*, **58**, 48–57.

29. Geick,A., Eichelbaum,M. and Burk,O. (2001) Nuclear receptor response elements mediate induction of intestinal MDR1 by rifampin. *J. Biol. Chem.*, **276**, 14581–14587.

30. Guo,G.L., Choudhuri,S. and Klaassen,C.D. (2002) Induction profile of rat organic anion transporting polypeptide 2 (oatp2) by prototypical drug-metabolizing enzyme inducers that activate gene expression through ligand-activated transcription factor pathways. *J. Pharmacol. Exp. Ther.*, **300**, 206–212.

31. Li,Y., Cui,Y., Hart,S.N., Klaassen,C.D. and Zhong,X.B. (2009) Dynamic patterns of histone methylation are associated with ontogenic expression of the Cyp3a genes during mouse liver maturation. *Mol. Pharmacol.*, **75**, 1171–1179.

32. Synold,T.W., Dussault,I. and Forman,B.M. (2001) The orphan nuclear receptor SXR coordinately regulates drug metabolism and efflux. *Nat. Med.*, **7**, 584–590.

33. Cui,Y.J., Cheng,X., Weaver,Y.M. and Klaassen,C.D. (2009) Tissue distribution, gender-divergent expression, ontogeny, and chemical induction of multidrug resistance transporter genes (Mdr1a, Mdr1b, Mdr2) in mice. *Drug Metab. Dispos.*, **37**, 203–210.

34. Watson,J.D. and Crick,F.H. (1953) The structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.*, **18**, 123–131.

35. Wang,W., Prosise,W.W., Chen,J., Taremi,S.S., Le,H.V., Madison,V., Cui,X., Thomas,A., Cheng,K.C. and Lesburg,C.A. (2008) Construction and characterization of a fully active PXR/SRC-1 tethered protein with increased stability. *Protein Eng. Des. Sel.*, **21**, 425–433.

36. Teotico,D.G., Frazier,M.L., Ding,F., Dokholyan,N.V., Temple,B.R. and Redinbo,M.R. (2008) Active nuclear receptors exhibit highly correlated AF-2 domain motions. *PLoS Comput. Biol.*, **4**, e1000111.

37. Noble,S.M., Carnahan,V.E., Moore,L.B., Luntz,T., Wang,H., Ittoop,O.R., Stimmel,J.B., Davis-Searles,P.R., Watkins,R.E., Wisely,G.B. *et al.* (2006) Human PXR forms a tryptophan zipper-mediated homodimer. *Biochemistry*, **45**, 8579–8589.

38. Fickett,J.W. (1996) Coordinate positioning of MEF2 and myogenin binding sites. *Gene*, **172**, GC19–GC32.

39. Ioshikhes,I., Trifonov,E.N. and Zhang,M.Q. (1999) Periodical distribution of transcription factor sites in promoter regions and connection with chromatin structure. *Proc. Natl Acad. Sci. USA*, **96**, 2891–2895.

40. Rhodes,D. and Klug,A. (1980) Helical periodicity of DNA determined by enzyme digestion. *Nature*, **286**, 573–578.

41. Li,H. and Capetanaki,Y. (1994) An E box in the desmin promoter cooperates with the E box and MEF-2 sites of a distal enhancer to direct muscle-specific transcription. *EMBO J.*, **13**, 3580–3589.

42. Hariparsad,N., Chu,X., Yabut,J., Labhart,P., Hartley,D.P., Dai,X. and Evers,R. (2009) Identification of pregnane-X receptor target genes and coactivator and corepressor binding to promoter elements in human hepatocytes. *Nucleic Acids Res.*, **37**, 1160–1173.

43. Xi,H., Shulha,H.P., Lin,J.M., Vales,T.R., Fu,Y., Bodine,D.M., McKay,R.D., Chenoweth,J.G., Tesar,P.J., Furey,T.S. *et al.* (2007) Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet.*, **3**, e136.

44. Lupien,M. and Brown,M. (2009) Cistromics of hormone-dependent cancer. *Endocr. Relat. Cancer*, **16**, 381–389.

45. Guo,G.L., Staudinger,J., Ogura,K. and Klaassen,C.D. (2002) Induction of rat organic anion transporting polypeptide 2 by pregnenolone-16alpha-carbonitrile is via interaction with pregnane X receptor. *Mol. Pharmacol.*, **61**, 832–839.

46. Moore,D.D., Kato,S., Xie,W., Mangelsdorf,D.J., Schmidt,D.R., Xiao,R. and Kliewer,S.A. (2006) International Union of Pharmacology. LXII. The NR1H and NR1I receptors: constitutive androstane receptor, pregnene X receptor, farnesoid X receptor alpha, farnesoid X receptor beta, liver X receptor alpha, liver X receptor beta, and vitamin D receptor. *Pharmacol. Rev.*, **58**, 742–759.

47. Li,T. and Chiang,J.Y. (2005) Mechanism of rifampicin and pregnane X receptor inhibition of human cholesterol 7 alpha-hydroxylase gene transcription. *Am. J. Physiol. Gastrointest. Liver Physiol.*, **288**, G74–G84.

48. Tumpel,S., Cambronero,F., Sims,C., Krumlauf,R. and Wiedemann,L.M. (2008) A regulatory module embedded in the coding region of Hoxa2 controls expression in rhombomere 2. *Proc. Natl Acad. Sci. USA*, **105**, 20077–20082.

49. Nobuyoshi,M., Lin,X.H., Takimoto,Y., Deuel,T.F. and Wang,Z.Y. (1997) Transcription regulation of the PDGF A-chain gene by first intron elements. *Biochem. Biophys. Res. Commun.*, **230**, 569–572.

50. Yu,H., Di Nicolantonio,R., Lan,L. and Wilks,A. (1995) Mutations in the first intron of the SHR renin gene disrupt putative regulatory elements. *Clin. Exp. Pharmacol. Physiol.*, **22**, 450–451.

51. Kikuchi,M., Miki,T., Kumagai,T., Fukuda,T., Kamiyama,R., Miyasaka,N. and Hirosawa,S. (2000) Identification of negative regulatory regions within the first exon and intron of the BCL6 gene. *Oncogene*, **19**, 4941–4945.

52. Dekker,J. (2008) Gene regulation in the third dimension. *Science*, **319**, 1793–1794.

53. Wilkinson,G.R. (2005) Drug metabolism and variability among patients in drug response. *N. Engl. J. Med.*, **352**, 2211–2221.

54. Gong,H., Singh,S.V., Singh,S.P., Mu,Y., Lee,J.H., Saini,S.P., Toma,D., Ren,S., Kagan,V.E., Day,B.W. *et al.* (2006) Orphan nuclear receptor pregnane X receptor sensitizes oxidative stress responses in transgenic mice and cancerous cells. *Mol. Endocrinol.*, **20**, 279–290.

55. Imai,S., Kikuchi,R., Kusuhara,H., Yagi,S., Shiota,K. and Sugiyama,Y. (2009) Analysis of DNA methylation and histone modification profiles of liver-specific transporters. *Mol. Pharmacol.*, **75**, 568–576.

56. Thomas,A.M., Hart,S.N., Kong,B., Fang,J., Zhong,X.B. and Guo,G.L. Genome-wide tissue-specific farnesoid X receptor binding in mouse liver and intestine. *Hepatology*, **51**, 1410–1419.