



Gene co-regulation and co-expression in the aryl hydrocarbon receptor-mediated transcriptional regulatory network in the mouse liver

Navya Josyula¹ · Melvin E. Andersen² · Norbert E. Kaminski^{3,4,7} · Edward Dere^{5,9} · Timothy R. Zacharewski^{5,4} · Sudin Bhattacharya^{3,4,6,7,8}

Received: 15 September 2019 / Accepted: 6 November 2019 / Published online: 14 November 2019
© The Author(s) 2019

Abstract

Four decades after its discovery, the aryl hydrocarbon receptor (AHR), a ligand-inducible transcription factor (TF) activated by the persistent environmental contaminant 2,3,7,8-Tetrachlorodibenzo-*p*-dioxin (TCDD), remains an enigmatic molecule with a controversial endogenous role. Here, we have assembled a global map of the AHR gene regulatory network in female C57BL/6 mice orally gavaged with 30 µg/kg of TCDD from a combination of previously published gene expression and genome-wide TF-binding data sets. Using Kohonen self-organizing maps and subspace clustering, we show that genes co-regulated by common upstream TFs in the AHR network exhibit a pattern of co-expression. Directly bound, indirectly bound, and non-genomic AHR target genes exhibit distinct expression patterns, with the directly bound targets associated with highest median expression. Interestingly, among the directly bound AHR target genes, the expression level increases with the number of AHR-binding sites in the proximal promoter regions. Finally, we show that co-regulated genes in the AHR network activate distinct groups of downstream biological processes. Although the specific findings described here are restricted to hepatic effects under short-term TCDD exposure, this work describes a generalizable approach to the reconstruction and analysis of transcriptional regulatory cascades underlying cellular stress response, revealing network hierarchy and the nature of information flow from the initial signaling events to phenotypic outcomes. Such reconstructed networks can form the basis of a new generation of quantitative adverse outcome pathways.

Keywords Ligand-activated transcription factors · DNA binding · Dioxin response element · Signaling · Co-regulation · Co-expression · Phenotypic outcomes

Abbreviations

AHR	Aryl hydrocarbon receptor
ARNT	Aryl hydrocarbon nuclear translocator
bHLH	Basic helix-loop-helix
ChEA2	ChIP-X enrichment analysis

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00204-019-02620-5>) contains supplementary material, which is available to authorized users.

✉ Sudin Bhattacharya
sbhattach@msu.edu

¹ Biomedical and Translational Informatics Program, Geisinger Health System, Rockville, MD 20850, USA

² ScitoVation LLC, Durham, NC 27713, USA

³ Department of Pharmacology and Toxicology, Michigan State University, East Lansing, MI 48824, USA

⁴ Institute for Integrative Toxicology, Michigan State University, East Lansing, MI 48824, USA

⁵ Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA

⁶ Department of Biomedical Engineering, Michigan State University, East Lansing, MI 48824-1226, USA

⁷ Center for Research on Ingredient Safety, Michigan State University, East Lansing, MI 48824, USA

⁸ Institute for Quantitative Health Science and Engineering, Michigan State University, East Lansing, MI 48824, USA

⁹ Present Address: Genentech, South San Francisco, CA 94080, USA

ChIP	Chromatin immunoprecipitation
CYP	Cytochrome P450
DE	Differentially expressed
DRE	Dioxin response element
GO	Gene ontology
MS	Matrix similarity
NF- κ B	Nuclear factor- κ B
NRF2	Nuclear factor erythroid 2-related factor
PAS	PER-ARNT-SIM
PPAR α	Peroxisome proliferator-activated receptor alpha
SOM	Self-organizing map
STAT	Signal transducers and activators of transcription
TCDD	2,3,7,8-Tetrachlorodibenzo- <i>p</i> -dioxin
TF	Transcription factor
TSS	Transcription start site

Introduction

Intracellular signaling pathways, when sufficiently perturbed by exogenous chemicals, can lead to an adverse outcome at the cellular level, and potentially at the level of tissues and the whole organism. These perturbed pathways have been described as “toxicity pathways” (NRC 2007; Whelan and Andersen 2013). Signaling, transcriptional, and post-transcriptional regulatory networks underlie toxicity pathways and their dynamic behavior in response to endogenous and exogenous perturbation. It is crucial to understand the organization, structure, and dynamics of these networks through mapping and modeling approaches for a quantitative assessment of the risks of chemical exposure to biological systems. Tissue-specific network models of chemical-induced perturbation can improve our understanding of the intracellular events leading to adverse effects and eventual injury from chemical exposure.

The major cellular response pathways are governed both transcriptionally and post-translationally. A core set of master regulatory transcription factors (TFs) are central actors in most molecular pathways leading to altered expression of suites of genes in response to exposure to a variety of chemical compounds (Jennings et al. 2013). These TFs, including the nuclear receptors, p53, nuclear factor erythroid 2-related factor (NRF2), nuclear factor- κ B (NF- κ B), the STAT (signal transducers and activators of transcription) family, and the aryl hydrocarbon receptor (AHR), typically coordinate a broad range of physiological processes like metabolism, oxidative stress response, differentiation, tumor suppression, reproduction, development, and homeostasis (Audet-Walsh and Giguère 2015; Evans and Mangelsdorf 2014; Ma 2013; Tyagi et al. 2011; Wright et al. 2017). They thus act as sentinels of normal biological activity, but their inappropriate activation or inhibition can lead to adverse outcomes at the cellular or tissue level (Andersen et al. 2013).

Here, we describe a network model of the AHR pathway in the mouse liver, assembled from previously published genomic data sets, and newly analyzed using various computational methods. The AHR is a ligand-activated TF that belongs to the basic helix–loop–helix (bHLH)–PER–ARNT–SIM (PAS) family of proteins, which serve as sensors of developmental and environmental signals (Gu et al. 2000). The prototypical AHR ligand is TCDD (Poland et al. 1976), a persistent environmental toxicant that produces a variety of adverse effects in laboratory animals, including immune suppression, reproductive and endocrine effects, neurochemical alterations, developmental toxicity, chloracne, and tumor promotion (Birnbaum 1994; Pohjanvirta and Tuomisto 1994). These effects are mediated by the transcriptional activity of the AHR, as shown by their absence or amelioration in AHR-null mice and mice with low-affinity AHR alleles (Gonzalez and Fernandez-Salguero 1998; Okey et al. 1989; Peters et al. 1999), as well as in mice with mutations in the DNA-binding domain or nuclear localization sequence of the AHR (Bunger et al. 2003, 2008). Ligand binding causes the AHR in the cytosol to undergo a conformational change, resulting in dissociation from its chaperone protein complex and translocation to the nucleus, where it forms a heterodimer with the related nuclear protein aryl hydrocarbon nuclear translocator (ARNT) (Hoffman et al. 1991; Whitelaw et al. 1993). The AHR–ARNT complex then binds to specific DNA sequences on target genes called dioxin response elements (DRE) containing the core sequence 5'-GCGTG-3' (Denison et al. 1988), leading to the regulation of a diverse battery of genes (Hankinson 1995; Poland and Knutson 1982). While the 5'-GCGTG-3' nucleotide core is substitution-intolerant, the flanking 5' and 3' nucleotides adjacent to the core sequence also contribute to a functional AHR-binding site (Denison et al. 1988; Gillesby et al. 1997; Lusska et al. 1993; Shen and Whitlock Jr 1992). DRE-independent mechanisms of AHR binding have also been reported (Dere et al. 2011b; Huang and Elferink 2012).

While the density of AHR-bound regions in the genome of hepatic tissue from TCDD-treated mice is greatest in proximal promoter regions close to the transcription start site (TSS) of annotated genes, AHR also binds to sites distal from a TSS, e.g., in intergenic regions and 3' UTRs (Dere et al. 2011b). Moreover, only a third of the differentially expressed genes identified by microarray analysis showed AHR binding at a DRE in their proximal promoter regions, suggesting additional mechanisms of gene regulation by AHR beyond the canonical model described above (Dere et al. 2011b). These mechanisms may include target gene regulation from distal AHR-bound regions through DNA looping, or indirect regulation by AHR through tethering with a secondary TF (Farnham 2009). Such an indirect mechanism has been demonstrated in the regulation of the rat CYP1A2 gene by AHR (Sogawa et al. 2004).

Here, we have mapped the TCDD-induced AHR regulatory network from a combination of previously published gene expression and ChIP-on-chip data from the liver of female C57BL/6 mice orally gavaged with 30 µg/kg of TCDD (Dere et al. 2011b), which provides us a system-wide view of AHR-mediated gene regulation under short-term TCDD exposure. Specifically, statistical and visualization tools were used to establish a relationship between gene co-regulation by multiple TFs and gene co-expression, and link groups of co-regulated genes to distinct downstream functional outcomes. Such reconstructed networks can form the basis of a new generation of quantitative adverse outcome pathways (Conolly et al. 2017; Perkins et al. 2019). Our focus here is on the early stages of hepatic response to TCDD exposure—longer term exposure may lead to a different suite of adaptive responses at the cellular and tissue level.

Materials and methods

Microarray data

Our network analysis was based on results from a previous study of gene expression profiling using whole-genome oligonucleotide arrays (Agilent Technologies, Santa Clara, CA) of hepatic tissues from female C57BL/6 mice orally gavaged with 30 µg/kg of TCDD (Boverhof et al. 2005; Dere et al. 2011b). The gene expression analysis was performed in hepatic tissue from mice exposed to TCDD for 2, 4, 8, 12, 18, 24, 72, and 168 h. Differentially responsive genes were identified using previously described cutoffs for fold change and statistical significance (lfold change ≥ 1.5 and posterior probabilities $P_1(t) \geq 0.999$) (Dere et al. 2011b; Eckel et al. 2004).

ChIP-on-chip data

Genome-wide AHR location data were taken from the previously described ChIP-on-chip experiments (Dere et al. 2011b), where ChIP assays were performed with hepatic tissue from female C57BL/6 mice exposed to TCDD for 2 and 24 h. Genes were associated with AHR-enriched regions if the position of maximum fold enrichment was within 10 kb upstream of a transcriptional start site (TSS) through to the end of the 3' UTR. For the present analysis, the ChIP data for 2 and 24 h were combined to obtain a unique list of ChIP-enriched regions associated with annotated genes (Supplementary Methods; Supplementary Code 1). The choice of data sets for our analysis was constrained by the requirement of matched mouse liver gene expression and ChIP data sets under similar conditions.

DRE analysis in ChIP-enriched regions

The ChIP-enriched regions for the differentially expressed (DE) genes were computationally searched for the presence of 5'-GCGTG-3' DRE core sequences to infer the nature of AHR binding to the target genes. The putative DRE search algorithm, written in R (R Core Team 2016) (Supplementary Methods; Supplementary Code 2), was based on a previously described approach (Sun et al. 2004). Briefly, the genomic sequences of the enriched regions were obtained from UCSC Genome Browser (<https://genome.ucsc.edu>) and scanned for exact matches to the DRE core sequences on both positive and negative strands. For each matched region, the 5-bp core sequence was extended 7 bp upstream and downstream of the core. The matrix similarity (MS) scores (Quandt et al. 1995) for the 19-bp DRE sequences were calculated and compared to an MS score threshold of 0.8473 based on the lowest MS score of 13 bona fide AHR-binding sequences (Dere et al. 2011a) (i.e., sites from the literature confirmed to bind AHR). The DRE sequences with high MS scores (MS score ≥ 0.8473) were defined as putative DREs capable of binding AHR. The DE genes that were AHR-enriched and had a putative DRE in the enriched region were described as “directly bound” by AHR, while AHR-enriched genes without a putative DRE were described as “indirectly bound”. The remaining DE genes that were not AHR-enriched were regarded as “unbound”/“non-genomic” targets.

Construction and visualization of the AHR transcriptional regulatory network

The DE genes from the Agilent oligonucleotide array data were searched against online databases to obtain a list of TFs that regulate these genes. The ChIP-X Enrichment Analysis (ChEA2) database (Kou et al. 2013) was used to obtain the list of regulatory TFs. To obtain the mouse-liver specific list of transcription factors, the mouse-specific TFs from ChEA2 were screened for expression in the liver using the TRANSFAC[®] database (Matys et al. 2003). The ensemble of DE genes including the directly and indirectly AHR-bound genes, together with their inferred transcriptional regulators, form a comprehensive network for TF–gene interactions under AHR-mediated TCDD induction. The landscape of this regulatory network was rendered using the open-source network visualization tool Cytoscape (Shannon et al. 2003). The gene expression values at each time point of TCDD exposure were superposed on this network to visualize the temporal changes associated with each gene. A lfold ratio threshold of 1.20 was used to identify the key target genes that are themselves TFs regulating other genes in the data set (a less stringent fold change threshold was used for TFs than other genes as TFs tend to be more tightly regulated). To generate and annotate the network in Cytoscape, three

input files describing the network topology and gene expression values were used: an AHR–gene interaction file and a TF–gene interaction file (“network files”), and a gene expression file (“attributes file”). \log_2 scaling of the fold ratios was used for visualizing gene expression. The network files were merged together to form the complete layout.

Gene expression analysis based on transcriptional groupings

A binary TF–gene interaction matrix with 43 TFs in addition to AHR was created indicating which TFs interact with which target genes. If a gene is regulated by a particular TF, then the corresponding interaction is represented as ‘1’; otherwise, it is represented as ‘0’. We used this TF–gene interaction matrix to classify target genes into co-regulated groups in a transcriptional cascade, to examine any possible relation between co-regulation and co-expression. To generate this grouping, AHR and other key TFs that were also target genes were considered in all possible combinations to identify the expression trends for target genes in each group. The total number of genes in each co-regulated group was counted by referring to the TF–gene interaction matrix, and all groups with at least five genes were considered for examination of the expression patterns. A graphical analysis was performed in *R* to identify the expression patterns of target genes for each combination of regulatory TFs (Supplementary Methods; Supplementary Code 3).

Kohonen self-organizing maps to visualize gene co-expression

To further examine the relationship between the transcriptional groups and target gene expression patterns, a self-organizing map (SOM) for the AHR network was generated using the Kohonen SOM package in *R* (Wehrens and Buydens 2007). The same TF–gene interaction matrix described above was used as input for this analysis. The SOM algorithm follows a clustering technique to group the target genes according to their TF-binding patterns. Target genes with similar TF-binding patterns are grouped into the same cluster or adjacent clusters, referred to as ‘units’ (Supplementary Methods; Supplementary Code 4).

Subspace clustering

The ORCLUS subspace clustering algorithm (Aggarwal and Yu 2000) and corresponding *R* package (Szepannek 2013) were used to cluster the differentially expressed genes into 16 non-overlapping groups. The number of clusters $k=16$ and the dimensionality of each cluster $l=4$ were chosen so as to minimize the cluster sparsity coefficient (Aggarwal and Yu 2000) (Supplementary Code 5).

Functional categorization of genes in each cluster

Gene ontology (GO) functional analysis was performed for the DE genes present in each ORCLUS cluster. Enriched GO “process” categories were identified for genes in each cluster using the *GOrilla* tool (Eden et al. 2009) with a p -value threshold of 10^{-3} and the list of all DE genes as background. *REViGO* (Supek et al. 2011) was used to arrange the enriched processes into a “treemap”, which was then rendered as an image using the downloadable R script generated by the program (Supplementary Code 6, Supplementary Code 7).

Results

Differential gene expression

The raw array data set (Dere et al. 2011b) consisted of 41,267 records with annotated genes, fold ratio and significance [$P1(t)$ values] at 2, 4, 8, 12, 18, 24, 72, and 168 h post-TCDD exposure. For genes with multiple occurrences in the dataset, the fold ratios and $P1(t)$ values were averaged, resulting in a total of 21,307 unique gene records. After applying the statistical cutoff values for fold change and $P1(t)$ at each expression time point, the resulting number of unique differentially expressed (DE) genes was 1407. All 1,407 DE genes were used to generate the AHR regulatory network map.

Analysis of AHR-enriched genomic regions associated with DE genes

The ChIP-on-chip data sets for 2 and 24 h time points consisted of 14,446 and 974 AHR-enriched regions respectively, with associated genes (Dere et al. 2011b). The two data sets were combined to yield a unique list of genes associated with at least one enriched region. This list of enriched genes was compared against the list of 1407 DE genes, yielding 632 genes associated with one or more AHR-enriched regions. The AHR-enriched regions around these 632 genes were searched for putative DREs, producing three kinds of regions depending on presence and location of DREs:

- Regions with one or more 5-bp DRE cores centrally located, such that a 7-bp upstream and downstream extension was possible for MS score calculations.
- Regions with DRE cores present only at the edge of the region, so that the 7-bp extension in both directions was not possible.
- Regions with no DRE core.

A total of 144 genes were associated with AHR-enriched regions where MS score calculations were possible, and that had putative DREs, i.e., 19-bp DRE sequences with an MS score ≥ 0.8473 (see “Materials and methods”). These genes were considered to be “directly bound” by AHR. For the AHR-enriched regions with (1) non-putative DRE core (i.e., MS score < 0.8473), (2) DRE core located at edges, or (3) DRE core not present in the enriched region, the associated genes were considered to be “indirectly bound” by AHR. In total, among the 1407 differentially expressed genes, 632 were bound by AHR with 144 genes directly bound and 488 indirectly bound, and the remaining 775 genes unbound by AHR.

Other transcriptional regulators of the DE genes

The ChEA2 database (Kou et al. 2013) provides a comprehensive record of transcription factor/target gene interactions from genome-wide ChIP studies for both mouse and human. The list of 1407 DE genes from our analysis was uploaded to the ChEA2 server, which identified 104 unique mouse-specific transcriptional regulators for these genes. These 104 TFs were searched against the TRANSFAC[®] database to filter for expression in liver tissue, which after accounting for discrepancies in naming between ChEA2 and TRANSFAC[®] resulted in a list of 43 unique mouse liver-expressed TFs. Out of our 1407 DE genes, 1198 had interactions with at least one of these 43 transcription factors. Among the 43 TFs, seven were themselves target genes of other identified TFs differentially expressed at $\text{lfold ratio} > 1.2$ in the microarray dataset. In addition to AHR, these seven TFs: NRF2, FLI1, KLF4, SOX17, CCND1, PPARG, and GATA1, form the regulatory “hubs” of the mouse liver AHR network. They are both: (a) direct or indirect targets of AHR; and (b) themselves in turn regulate clusters of other downstream genes (Fig. 1a).

The AHR regulatory network

All interactions of the DE genes with AHR and the other 43 identified TFs together form the mouse liver AHR regulatory network (Fig. 1a), which consists of 44 “source” nodes interacting with 1,241 “target” nodes.

AHR and the other seven hub TFs act as both source and target nodes (AHR regulates itself). Two of these hub TFs are regulated by AHR: NRF2 is a direct target and Fli1 an indirect target (network schematic in Fig. 1a). The expression levels for up- and down-regulated genes were superposed on this network layout for each of the eight time points in the gene array study (Supplementary Fig. 1a–h), illustrating that the gene expression levels were not monotonic in time.

To examine the transcriptional regulatory hierarchy in the network, AHR and four of the seven hub TFs (FLI1, NRF2, KLF4, and SOX17), which were all expressed at $\text{lfold ratio} > 1.5$, were grouped in all possible combinations (Supplementary Table 1) to assess the expression of their target genes. Genes differentially expressed at least at one time point at $\text{lfold ratio} > 1.5$ were chosen for this analysis, yielding 1,191 target genes regulated individually or in combination by the above five TFs (Supplementary Table 1). We then examined the expression pattern of groups of co-regulated genes with a count of 5 or more (Table 1). The time courses of genes that were up-regulated at the 168 h time point (Fig. 1b) suggest that genes with the same upstream regulators have similar expression patterns.

Co-regulation and co-expression in the AHR network

To take a closer look at whether co-regulation in the AHR network is associated with co-expression, we clustered the 1191 target genes into self-organizing maps (SOMs) based on the factors that regulate them (Fig. 2a). Each circular unit in the SOM represents a grouping of genes (individual dots within a unit). The SOM algorithm groups genes into units, such that genes in a single unit or adjacent units have a similar combination of TFs regulating them (no gene expression values were used for clustering), while genes in distant units have more dissimilar regulators. The median expression level (\log_2 fold change) of the genes in each unit was then superposed on the SOM as a continuous color scale with blue indicating suppression and red activation (panels in Fig. 2b). A distinct pattern emerges over the time course, with the units with high median expression at 168 h localized at the lower right corner of the SOM (Fig. 2b). The median time courses of the genes in adjacent units are also quite similar (Supplementary Fig. 2). This analysis shows a strong association between gene co-regulation and co-expression in the AHR network.

Localized clustering of co-regulated genes

We further attempted to cluster the 1,191 target genes considered in the SOM analysis above based on regulation by the 44 TFs. Fundamentally, the clustering problem may be stated as: “Given a set of data points, partition them into a set of groups which are as similar as possible” (Aggarwal 2014). If we consider the binary TF–gene connectivity matrix, with genes in rows (observations), TFs in columns (features), and each matrix element equaling 1 or 0 depending on whether a TF binds a gene, we have a high-dimensional clustering problem with feature localization, i.e., different groups of genes are regulated by different subsets of TFs. Global clustering methods like k-means or

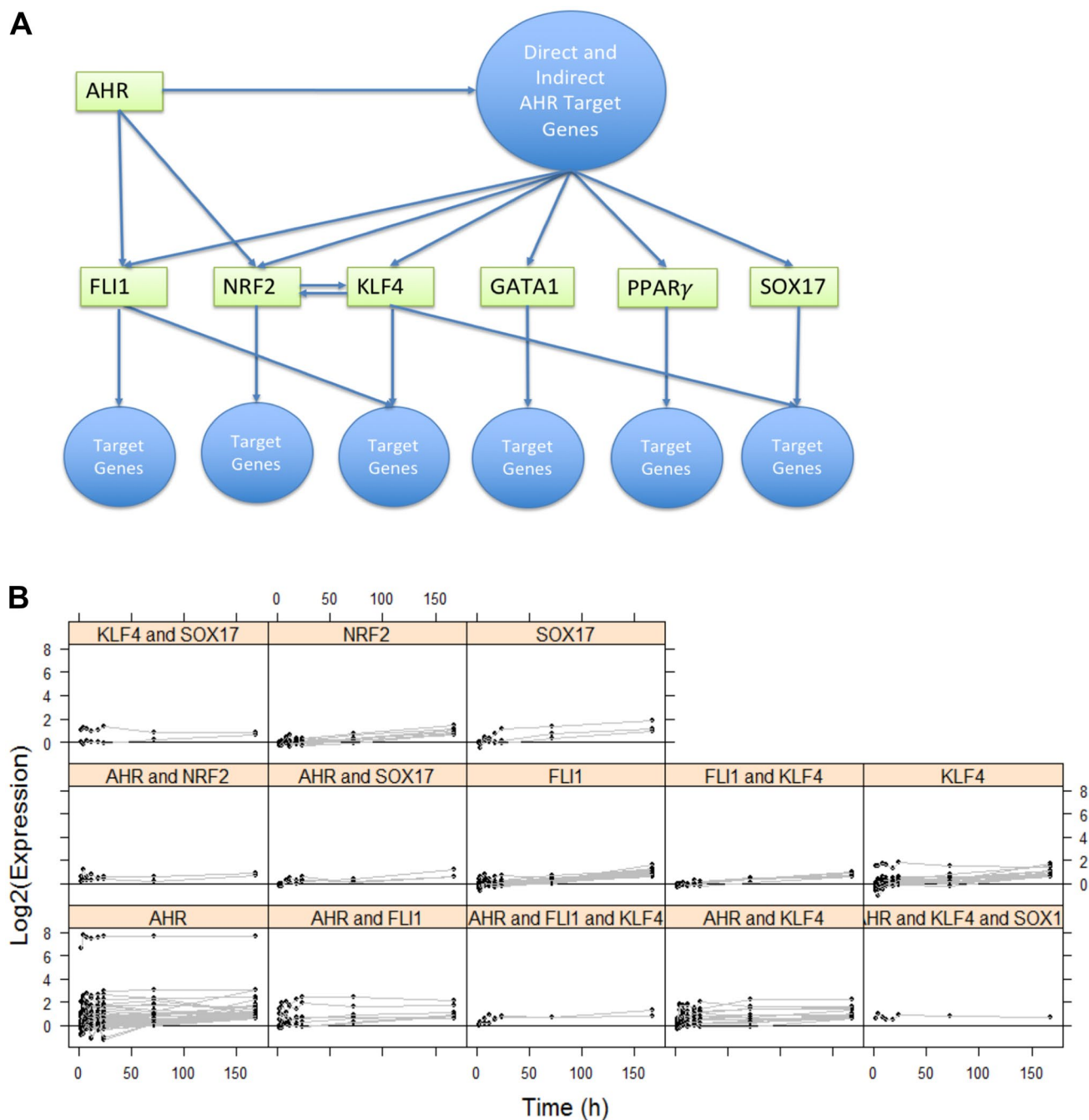


Fig. 1 AHR transcriptional regulatory landscape in the mouse liver. **a** A schematic map of the AHR transcriptional regulatory network in the mouse liver, showing some of the key transcription factors (green boxes) and clustered groups of co-regulated genes (blue circles), with the edges indicating gene regulatory interactions. The map reveals a hierarchical structure in the AHR network. The full detailed net-

work map is shown in Supplementary Fig. 1a–h. **b** Time courses of genes grouped by transcriptional regulators (only genes up-regulated at 168 h shown). Genes grouped by transcriptional regulators show similar expression patterns. The vertical axis denotes \log_2 fold change (color figure online)

dimensionality reduction approaches like principal components analysis do not perform well in this situation, which motivated the development of high-dimensional subspace clustering methods (Aggarwal 2014). These methods include “projected clustering” or “subspace clustering”

approaches like PROCLUS (Aggarwal et al. 1999), CLIQUE (Agrawal et al. 2005), and ORCLUS (Aggarwal and Yu 2000), where feature selection or transformation is performed specific to different localities of the data (Aggarwal 2014). ORCLUS in particular is suited for data

Table 1 Count of genes in each transcriptional grouping

	Combinations	TF.Groups	Gene.Count
1	AHR	G1	210
2	FLI1	G2	36
3	NFE2L2	G3	13
4	KLF4	G4	55
5	SOX17	G5	10
6	AHRIFLI1	G6	15
7	AHRINFE2L2	G7	12
8	AHRIKLF4	G8	54
9	AHRISOX17	G9	7
10	FLI1KLF4	G11	9
11	KLF4SOX17	G15	6
12	AHRIFLI1KLF4	G17	7
13	AHRIKLF4SOX17	G21	5

All possible groupings among the “hub” TFs in the network, AHR, FLI1, NRF2 (NFE2L2), KLF4, and SOX17 were considered. Groups with gene counts > 5 are shown in the table

sets like ours where relevant subspaces may be arbitrarily oriented due to inter-feature correlations (Aggarwal and Yu 2000), i.e., many TFs are correlated in term of which genes they regulate.

We used the ORCLUS algorithm to group the DE genes by TF connectivity into 16 clusters (Fig. 3a), illustrating both the sparsity of the TF–gene connectivity matrix and the fact that different clusters of genes are regulated by different subsets of TFs. In particular, there is a marked contrast between Cluster 2, where none of 157 genes is bound by AHR, and Cluster 6, where all 123 genes are. Cluster 2 genes can thus be said to comprise a “non-genomic pathway” and Cluster 6 genes a “genomic pathway” with respect to regulation by AHR. The most frequent regulators in Cluster 2 are PPARG, regulating 46 genes, STAT3 (33 genes), CEBPB, NANOG (30 genes each), CREB1, GATA2 and SUZ12 (27 genes each). In contrast, the most frequent regulators in Cluster 6 are AHR (all 123 genes), followed by SUZ12 (55 genes), PPARG (42 genes), CREB1 (29 genes), MYC (25 genes), NANOG (24 genes), and E2F1 and TAL1 (22 genes each).

Distinct gene clusters activate distinct biological processes

We carried out gene ontology (GO) analysis on the six major clusters of genes labeled in Fig. 3a. The genes in the six clusters enrich for different groups of biological processes (Fig. 3b). In particular, the genomic cluster (Cluster 6) is enriched for genes associated with metabolic processes and ribosome biogenesis, whereas the major GO categories associated with the non-genomic cluster (Cluster 2) are immune

regulatory processes. Interestingly, Cluster 5 is enriched for cell migration and activation of cellular defense mechanisms. Presumably, this reflects immune cell infiltration into the mouse liver under exposure to TCDD (Fader et al. 2015). Cluster 16 is also enriched for immune system response. Thus, co-regulated genes in the AHR network in the mouse liver show patterns of co-expression and lead to differential downstream activation of biological processes.

AHR binding and gene expression

We plotted the time courses of \log_2 fold change values for all genes in Clusters 2 and 6 (Fig. 4a, b). Genes in the “non-genomic” Cluster 2 are down-regulated or moderately up-regulated at earlier times (Fig. 4a), with about two-thirds of the genes showing up-regulation at the later time points. On the contrary, a majority of genes in the “genomic” Cluster 6 (Fig. 4b) are moderately to strongly up-regulated at all time points, with a smaller subset showing consistent down-regulation.

This led us to suspect that there may be a link between binding of a gene by AHR and its expression level, and we separately plotted the time courses of genes that are (a) directly bound (b) indirectly bound, and (c) unbound by AHR (Fig. 5a–c). Nearly all genes in the directly bound group (Fig. 5a) are up-regulated moderately or strongly at multiple time points, whereas in the indirectly bound group (Fig. 5b), about half of the genes are consistently down-regulated. Finally, the unbound group shows an unusual pattern of gene expression (Fig. 5c), where most genes are down-regulated at the first two time points, but then, about two-thirds of the genes show progressive up-regulation up to the 168 h time point. This observation suggests a cascade structure in the AHR network, where genes not proximally bound by AHR are bound by other TFs activated at intermediate-to-later points in the time course, or are targets of long-range interaction with distally bound AHR, leading to their up-regulation at later times.

These differences between direct, indirect, and unbound AHR target genes are also highlighted in overlaid box and violin plots (Fig. 5d–g; Supplementary Fig. 3), showing the respective distributions of expression level of the three groups of genes at multiple time points. At each time point shown, the middle 50% (first to third quartile) of the directly bound genes are all up-regulated, while the indirectly bound group is symmetrically distributed with about half of the genes up-regulated. In the unbound group, most genes are down-regulated at earlier time points, but at 168 h, the distribution is considerably right-skewed with many genes up-regulated. Overall, the directly regulated group has the highest median expression (except at 168 h), and also has the most outliers on the high expression end, the furthest outlier being the CYP1A1 gene.

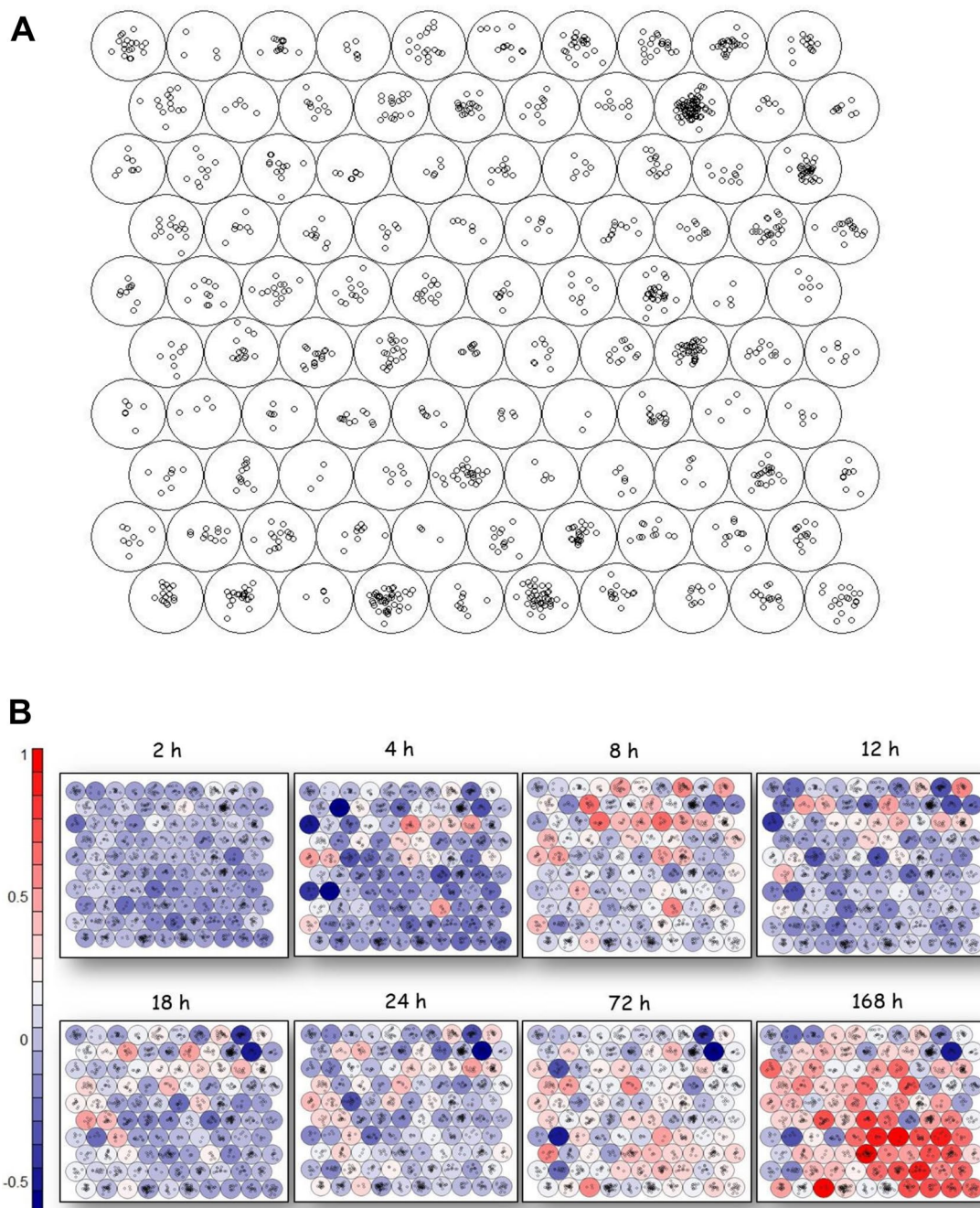


Fig. 2 Median expression of genes organized into Kohonen Self-Organizing Map (SOM). **a** The mapping of genes clustered in each unit according to TF-binding patterns. **b** The temporal gene expression patterns of the SOM units, confirming the co-regulation and co-

expression patterns of the genes. The continuous color scale shows the median \log_2 fold change expression values for the genes in each unit, with blue indicating suppression and red activation (color figure online)

Given that the CYP1A1 gene has a high number of DREs in its proximal promoter region (Li et al. 2014), we examined the relationship among expression level and number of proximal promoter DREs for the direct target genes at various time points (Fig. 6a–d). There is an increase in mean expression level with increasing number of DREs, a trend

that gets stronger at later time points. This is likely due to a larger number of AHR molecules binding to DREs in the promoter regions, leading to a higher degree of activation of proximal genes. This hypothesis is supported by the previous findings that in the human liver, genes with more

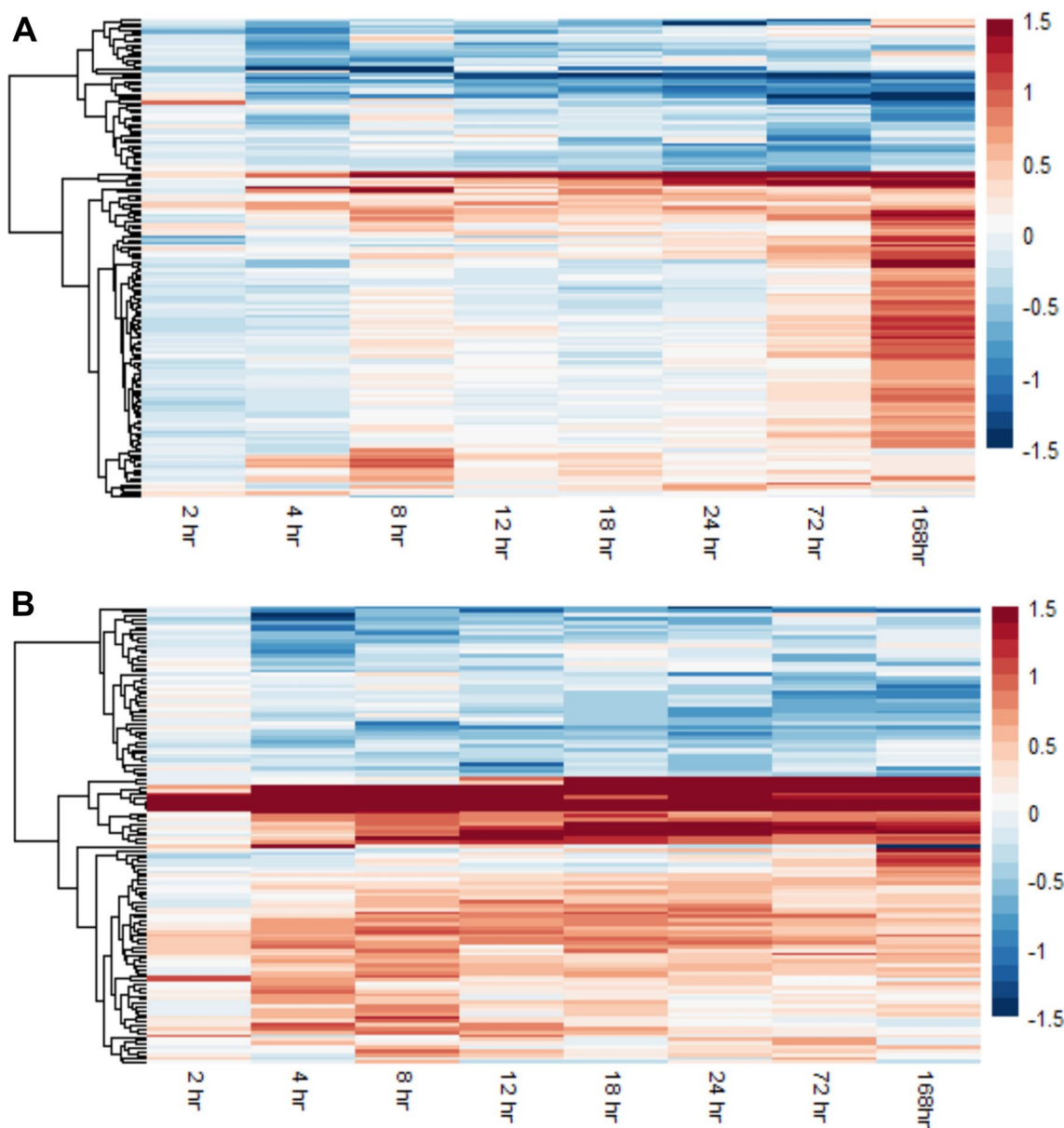


Fig. 4 Heatmaps showing time courses of \log_2 fold change for all 157 genes in Cluster 2 (a) and all 123 genes in Cluster 6 (b). For visualization of the heatmap, \log_2 fold change values > 1.5 were set to 1.5

and values < -1.5 to -1.5 . Blue indicates down-regulation and red up-regulation (color figure online)

a common feature of cellular pathways, since binding sites are often clustered in the genome, allowing multiple TFs to act in a coordinated fashion to induce or suppress groups of genes in specific cell types under particular conditions (George et al. 2011). In addition, a surprisingly large number of genes are activated or inhibited in a “non-genomic” manner, showing no evidence of binding by the master regulatory TF of the stimulated pathway in their promoter regions (Dere et al. 2011b; McMullen et al. 2014; Shen et al. 2011; van der Meer et al. 2010). These observations collectively suggest that combining gene expression data from

transcriptome profiling with high-throughput genome-wide analysis of TF binding can provide an integrated, systems-level view of the structure and function of transcription factor-governed molecular pathways (Blais and Dynlacht 2005; Dere et al. 2011b; Limonciel et al. 2015; Walhout 2006).

Accordingly, we have integrated TCDD-induced gene expression and multiple genome-wide TF-binding data sets for a global view of the AHR regulatory pathway in the mouse liver. Using a combination of self-organizing maps and subspace clustering, we show that there is a pattern of co-regulated genes in the AHR pathway being co-expressed,

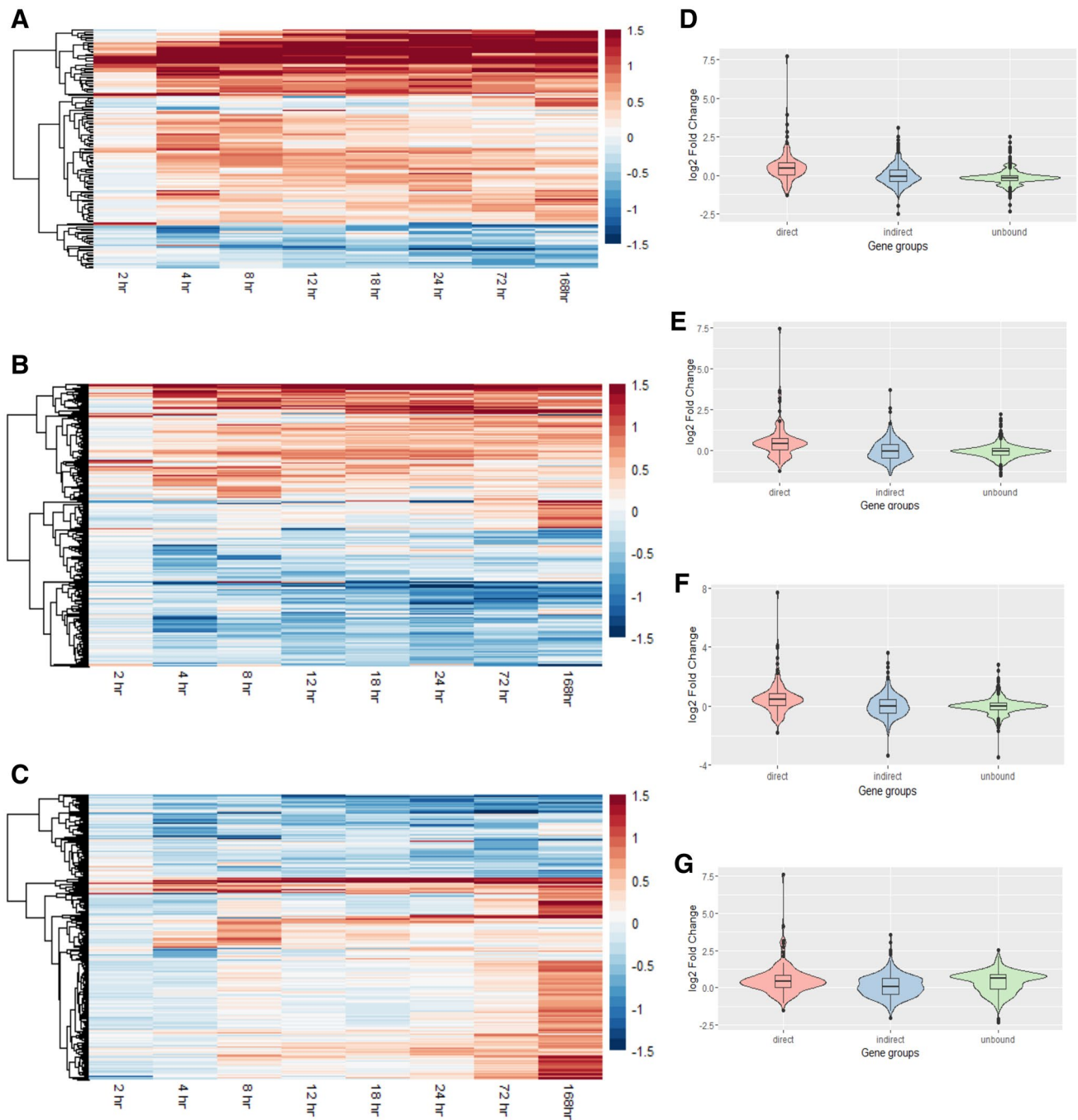


Fig. 5 Distinct expression patterns among the genes directly bound, indirectly bound, and unbound by AHR. The heatmaps show time courses of log₂ fold change for **a** 140 genes directly bound, **b** 477 genes indirectly bound, and **c** 574 genes unbound by AHR. For visualization of the heatmap, log₂ fold change values > 1.5 were set to 1.5 and values < -1.5 to -1.5. There are proportionately more up-regulated genes in **a** compared to **b**, and in **b** compared to **c**. Blue indicates down-regulation and red up-regulation. **d–g** Overlaid box

and violin plots showing the distribution in differential expression of direct ($n=140$), indirect ($n=477$) and unbound ($n=574$) AHR target genes at **d** 4 h, **e** 12 h, **f** 24 h, and **g** 168 h. These plots illustrate the respective distributions of expression level of the three groups of genes at multiple time points, with the box plots illustrating the median, first and third quartile, and outliers; and the overlaid violin plots showing a rotated histogram of the distribution of gene expression (color figure online)

as previously observed in *Saccharomyces cerevisiae* (Allocco et al. 2004; Yu et al. 2003). In particular, directly bound, indirectly bound, and unbound AHR target genes

have distinct patterns of gene expression, with the directly bound group showing higher median expression. Furthermore, among the direct AHR target genes, the expression

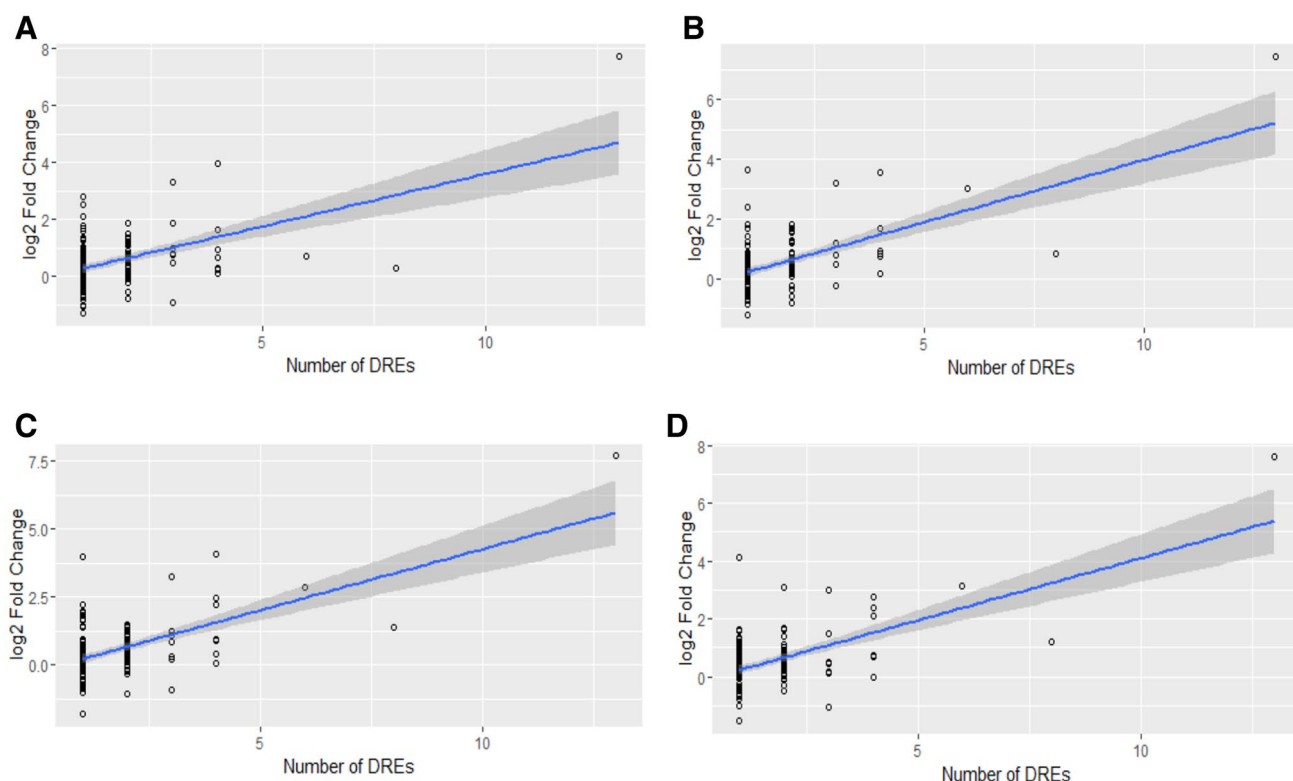


Fig. 6 Increase in expression level of direct AHR target genes with number of DREs in proximal promoter regions at 4 h (a), 12 h (b), 24 h (c) and 168 h (d). Circles denote individual genes; linear regression fit shown in blue line with shaded region showing 95% confidence interval

level increases with the number of AHR-binding DRE sites in the proximal promoter regions. Finally, we found that co-regulated gene clusters activated distinct groups of downstream biological processes, with the AHR-bound genomic cluster enriched for metabolic processes and the AHR-unbound non-genomic cluster primarily activating immune processes. This work, together with the other recent studies of the peroxisome proliferator-activated receptor alpha (PPAR α) and estrogen receptor pathways (McMullen et al. 2014; Pendse et al. 2016), illustrates the application of bioinformatic and statistical tools for reconstruction and analysis of the transcriptional regulatory cascades underlying cellular stress response. While these network reconstructions are species, tissue, and condition-specific, we anticipate that next-generation models that use machine learning to predict network structure and dynamics from genomic sequence and epigenomic features will soon be available. Such models will reduce our reliance on expensive assays for gene expression and genome-wide protein binding for different animal models, tissues, and exposures.

The work presented here describes a detailed map of the AHR transcriptional regulatory network activated by TCDD exposure in the mouse liver. This map can be used to derive a predictive model of TCDD dose-dependent

genomic response. The overall map (Supplementary Fig. 1) contains a large number of network edges that may be difficult to model quantitatively. However, our findings regarding gene co-regulation and co-expression imply that simplified representations of the network with groups of co-expressed genes represented as individual entities (as in Fig. 1a) may be sufficient to model the overall network response. Such a simplified representation would also be more easily quantifiable in terms of parameterizing the regulatory interactions. We foresee such quantitative dose–response models as a crucial step in the development of more rigorous mechanistic risk assessment protocols.

Acknowledgements The authors would like to thank Agnes Karmaus, Arindam Banerjee, Rory Conolly, and Qiang Zhang for helpful discussions. This work was supported by the US EPA STAR Program (EPA Grant number: R835000), the USDA National Institute of Food and Agriculture and Michigan AgBioResearch, and the Superfund Research Program of the National Institute of Environmental Health Sciences (Grant number: P42ES04911). No funding body played any role in the design of the study, collection, analysis, and interpretation of data, or in writing the manuscript.

Author contributions The study was conceived and coordinated by SB. NJ and SB analyzed and interpreted the data. The manuscript was written by NJ and SB with input from MEA, NEK, ED, and TRZ. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Aggarwal C, Yu P (2000) Finding generalized projected clusters in high dimensional spaces. In: Proceedings of ACM SIGMOD international conference on management of data, pp 70–81
- Aggarwal CC (2014) An introduction to cluster analysis. In: Aggarwal CC, Reddy CK (eds) Data clustering: algorithms and applications. Chapman and Hall Press, Boca Raton.
- Aggarwal CC, Wolf JL, Yu PS, Procopiuc C, Park JS (1999) Fast algorithms for projected clustering. ACM SIGMOD Rec 28(2):61
- Agrawal R, Gehrke J, Gunopulos D, Raghavan P (2005) Automatic subspace clustering of high dimensional data. Data Min Knowl Disc 11(1):5
- Allocco DJ, Kohane IS, Butte AJ (2004) Quantifying the relationship between co-expression, co-regulation and gene function. BMC Bioinform. <https://doi.org/10.1186/1471-2105-5-18>
- Andersen ME, McMullen PD, Bhattacharya S (2013) Toxicogenomics for transcription factor-governed molecular pathways: moving on to roles beyond classification and prediction. Arch Toxicol 87(1):7–11
- Audet-Walsh É, Giguère V (2015) The multiple universes of estrogen-related receptor α and γ in metabolic control and related diseases. Acta Pharmacol Sin 36(1):51–61. <https://doi.org/10.1038/aps.2014.121>
- Birnbaum LS (1994) The mechanism of dioxin toxicity: relationship to risk assessment. Environ Health Perspect 102(SUPPL. 9):157–167
- Blais A, Dynlacht BD (2005) Constructing transcriptional regulatory networks. Genes Dev 19(13):1499–1511. <https://doi.org/10.1101/gad.1325605>
- Boverhof DR, Burgoon LD, Tashiro C et al (2005) Temporal and dose-dependent hepatic gene expression patterns in mice provide new insights into TCDD-mediated hepatotoxicity. Toxicol Sci 85(2):1048–1063. <https://doi.org/10.1093/toxsci/kfi162>
- Bunger MK, Glover E, Moran SM et al (2008) Abnormal liver development and resistance to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin toxicity in mice carrying a mutation in the DNA-binding domain of the aryl hydrocarbon receptor. Toxicol Sci 106(1):83–92. <https://doi.org/10.1093/toxsci/kfn149>
- Bunger MK, Moran SM, Glover E et al (2003) Resistance to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin toxicity and abnormal liver development in mice carrying a mutation in the nuclear localization sequence of the aryl hydrocarbon receptor. J Biol Chem 278(20):17767–17774. <https://doi.org/10.1074/jbc.M209594200>
- Conolly RB, Ankley GT, Cheng W et al (2017) Quantitative adverse outcome pathways and their application to predictive toxicology. Environ Sci Technol 51:4661–4672. <https://doi.org/10.1021/acs.est.6b06230>
- Denison MS, Fisher JM, Whitlock JP Jr (1988) The DNA recognition site for the dioxin-Ah receptor complex Nucleotide sequence and functional analysis. J Biol Chem 263(33):17221–17224
- Dere E, Forgacs AL, Zacharewski TR, Burgoon LD (2011a) Genome-wide computational analysis of dioxin response element location and distribution in the human, mouse, and rat genomes. Chem Res Toxicol 24(4):494–504. <https://doi.org/10.1021/tx100328r>
- Dere E, Lo R, Celius T, Matthews J, Zacharewski TR (2011b) Integration of genome-wide computation DRE search, AhR ChIP-chip and gene expression analyses of TCDD-elicited responses in the mouse liver. BMC Genom. <https://doi.org/10.1186/1471-2164-12-365>
- Eckel JE, Gennings C, Chinchilli VM, Burgoon LD, Zacharewski TR (2004) Empirical bayes gene screening tool for time-course or dose-response microarray data. J Biopharm Stat 14(3):647–670. <https://doi.org/10.1081/bip-200025656>
- Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinform. <https://doi.org/10.1186/1471-2105-10-48>
- Evans RM, Mangelsdorf DJ (2014) Nuclear receptors, RXR, and the big bang. Cell 157(1):255–266. <https://doi.org/10.1016/j.cell.2014.03.012>
- Fader KA, Nault R, Ammendolia DA et al (2015) 2,3,7,8-tetrachlorodibenzo-*p*-dioxin alters lipid metabolism and depletes immune cell populations in the Jejunum of C57BL/6 mice. Toxicol Sci 148(2):567–580. <https://doi.org/10.1093/toxsci/kfv206>
- Farnham PJ (2009) Insights from genomic profiling of transcription factors. Nat Rev Genet 10(9):605–616. <https://doi.org/10.1038/nrg2636>
- George CL, Lightman SL, Biddie SC (2011) Transcription factor interactions in genomic nuclear receptor function. Epigenomics 3(4):471–485
- Gillesby BE, Stanostefano M, Porter W, Safe S, Wu ZF, Zacharewski TR (1997) Identification of a motif within the 5' regulatory region of pS2 which is responsible for AP-1 binding and TCDD-mediated suppression. Biochemistry 36(20):6080–6089. <https://doi.org/10.1021/bi962131b>
- Gonzalez FJ, Fernandez-Salguero P (1998) The aryl hydrocarbon receptor Studies using the AHR-null mice. Drug Metab Dispos 26(12):1194–1198
- Gu YZ, Hogenesch JB, Bradfield CA (2000) The PAS superfamily: sensors of environmental and developmental signals. Annu Rev Pharmacol Toxicol 40:519–561
- Hankinson O (1995) The aryl hydrocarbon receptor complex. Annu Rev Pharmacol Toxicol 35:307–340
- Hoffman EC, Reyes H, Chu FF et al (1991) Cloning of a factor required for activity of the Ah (dioxin) receptor. Science 252(5008):954–958
- Huang G, Elferink CJ (2012) A novel nonconsensus xenobiotic response element capable of mediating aryl hydrocarbon receptor-dependent gene expression. Mol Pharmacol 81(3):338–347. <https://doi.org/10.1124/mol.111.075952>
- Jennings P, Limonciel A, Felice L, Leonard MO (2013) An overview of transcriptional regulation in response to toxicological insult. Arch Toxicol 87(1):49–72. <https://doi.org/10.1007/s00204-012-0919-y>
- Kou Y, Chen EY, Clark NR, Duan Q, Tan CM, Ma'ayan A (2013) ChEA2: gene-set libraries from ChIP-X experiments to decode the transcription regulome. In: Cuzzocrea A, Kittl C, Simos DE, Weippl E, Xu L (eds) Availability, reliability, and security in information systems and HCI: IFIP WG 84, 89, TC 5 international cross-domain conference, CD-ARES 2013, Regensburg, Germany, September 2–6, 2013 proceedings. Springer, Berlin, pp 416–430
- Li S, Pei X, Zhang W, Xie H, Zhao B (2014) Functional analysis of the dioxin response elements (DREs) of the murine CYP1A1 gene promoter: beyond the core DRE sequence. Int J Mol Sci 15(4):6475

- Limonciel A, Moenks K, Stanzel S et al (2015) Transcriptomics hit the target: monitoring of ligand-activated and stress response pathways for chemical testing. *Toxicol In Vitro*. <https://doi.org/10.1016/j.tiv.2014.12.011>
- Lusska A, Shen E, Whitlock JP Jr (1993) Protein-DNA interactions at a dioxin-responsive enhancer: analysis of six bona fide DNA-binding sites for the liganded Ah receptor. *J Biol Chem* 268(9):6575–6580
- Ma Q (2013) Role of Nrf2 in oxidative stress and toxicity. *Annu Rev Pharmacol Toxicol* 53:401–426
- Matys V, Fricke E, Geffers R et al (2003) TRANSFAC®: Transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 31(1):374–378. <https://doi.org/10.1093/nar/gkg108>
- McMullen PD, Bhattacharya S, Woods CG et al (2014) A map of the PPAR α transcription regulatory network for primary human hepatocytes. *Chem Biol Interact* 209(1):14–24
- NRC (2007) Toxicity testing in the 21st century: a vision and a strategy. The National Academies Press, Washington, DC
- Odom DT, Dowell RD, Jacobsen ES et al (2006) Core transcriptional regulatory circuitry in human hepatocytes. *Mol Syst Biol* 2(2006):0017. <https://doi.org/10.1038/msb4100059>
- Okey AB, Vella LM, Harper PA (1989) Detection and characterization of a low affinity form of cytosolic Ah receptor in livers of mice nonresponsive to induction of cytochrome P1–450 by 3-methylcholanthrene. *Mol Pharmacol* 35(6):823–830
- Pendse SN, Maertens A, Rosenberg M et al (2016) Information-dependent enrichment analysis reveals time-dependent transcriptional regulation of the estrogen pathway of toxicity. *Arch Toxicol*. <https://doi.org/10.1007/s00204-016-1824-6>
- Perkins EJ, Ashauer R, Burgoon L et al (2019) Building and applying quantitative adverse outcome pathway models for chemical hazard and risk assessment. *Environ Toxicol Chem*. <https://doi.org/10.1002/etc.4505>
- Peters JM, Narotsky MG, Elizondo G, Fernandez-Salguero PM, Gonzalez FJ, Abbott BD (1999) Amelioration of TCDD-induced teratogenesis in aryl hydrocarbon receptor (AhR)-null mice. *Toxicol Sci* 47(1):86–92. <https://doi.org/10.1093/toxsci/47.1.86>
- Pohjanvirta R, Tuomisto J (1994) Short-term toxicity of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin in laboratory animals: effects, mechanisms, and animal models. *Pharmacol Rev* 46(4):483–549
- Poland A, Glover E, Kende AS (1976) Stereospecific, high affinity binding of 2,3,7,8 tetrachlorodibenzo *p* dioxin by hepatic cytosol Evidence that the binding species is receptor for induction of aryl hydrocarbon hydroxylase. *J Biol Chem* 251(16):4936–4946
- Poland A, Knutson JC (1982) 2,3,7,8-tetrachlorodibenzo-*p*-dioxin and related halogenated aromatic hydrocarbons: examination of the mechanism of toxicity. *Annu Rev Pharmacol Toxicol* 22:517–554
- Quandt K, Frech K, Karas H, Wingender E, Werner T (1995) MatInd and matInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res* 23(23):4878–4884
- R Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Shannon P, Markiel A, Ozier O et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–2504. <https://doi.org/10.1101/gr.1239303>
- Shen C, Huang Y, Liu Y et al (2011) A modulated empirical Bayes model for identifying topological and temporal estrogen receptor alpha regulatory networks in breast cancer. *BMC Syst Biol* 5(1):67. <https://doi.org/10.1186/1752-0509-5-67>
- Shen ES, Whitlock JP Jr (1992) Protein-DNA interactions at a dioxin-responsive enhancer: mutational analysis of the DNA-binding site for the liganded Ah receptor. *J Biol Chem* 267(10):6815–6819
- Simmons SO, Fan CY, Ramabhadran R (2009) Cellular stress response pathway system as a sentinel ensemble in toxicological screening. *Toxicol Sci* 111(2):202–225. <https://doi.org/10.1093/toxsci/kfp140>
- Sogawa K, Numayama-Tsuruta K, Takahashi T et al (2004) A novel induction mechanism of the rat CYP1A2 gene mediated by Ah receptor-Arnt heterodimer. *Biochem Biophys Res Commun* 318(3):746–755. <https://doi.org/10.1016/j.bbrc.2004.04.090>
- Sun YV, Boverhof DR, Burgoon LD, Fielden MR, Zacharewski TR (2004) Comparative analysis of dioxin response elements in human, mouse and rat genomic sequences. *Nucleic Acids Res* 32(15):4512–4523. <https://doi.org/10.1093/nar/gkh782>
- Supek F, Bošnjak M, Škunca N, Šmuc T (2011) Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0021800>
- Szepannek G (2013) orclus: ORCLUS subspace clustering. R package version 0.2-5
- Tyagi S, Gupta P, Saini AS, Kaushal C, Sharma S (2011) The peroxisome proliferator-activated receptor: a family of nuclear receptors role in various diseases. *J Adv Pharm Technol Res* 2(4):236–240. <https://doi.org/10.4103/2231-4040.90879>
- van der Meer DLM, Degenhardt T, Väisänen S et al (2010) Profiling of promoter occupancy by PPAR α in human hepatoma cells via ChIP-chip analysis. *Nucleic Acids Res* 38(9):2839–2850
- Walhout AJM (2006) Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. *Genome Res* 16(12):1445–1454. <https://doi.org/10.1101/gr.5321506>
- Wehrens R, Buydens LMC (2007) Self- and super-organizing maps in R: the kohonen package. *J Stat Softw* 21(5):1–19
- Whelan M, Andersen ME (2013) Toxicity pathways—from concepts to application in chemical safety assessment. Publications Office of the European Union, JRC, Luxembourg
- Whitelaw M, Pongratz I, Wilhelmsson A, Gustafsson JÅ, Poellinger L (1993) Ligand-dependent recruitment of the arnt coregulator determines DNA recognition by the dioxin receptor. *Mol Cell Biol* 13(4):2504–2514
- Wright EJ, Pereira De Castro K, Joshi AD, Elferink CJ (2017) Canonical and non-canonical aryl hydrocarbon receptor signaling pathways. *Curr Opin Toxicol* 2:87–92. <https://doi.org/10.1016/j.cotox.2017.01.001>
- Yu H, Luscombe NM, Qian J, Gerstein M (2003) Genomic analysis of gene expression relationships in transcriptional regulatory networks. *Trends Genet* 19(8):422–427. [https://doi.org/10.1016/s0168-9525\(03\)00175-6](https://doi.org/10.1016/s0168-9525(03)00175-6)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.