Mutation Research 746 (2012) 151–162

Contents lists available at SciVerse ScienceDirect

Mutation Research/Genetic Toxicology and Environmental Mutagenesis

journal homepage: www.elsevier.com/locate/gentox
Community address: www.elsevier.com/locate/mutres

A graphical systems model and tissue-specific functional gene sets to aid transcriptomic analysis of chemical impacts on the female teleost reproductive axis

Daniel L. Villeneuve[a,*], Natàlia Garcia-Reyero[b], Dalma Martinović-Weigelt[c], Zhenhong Li[d], Karen H. Watanabe[d], Edward F. Orlando[e], Carlie A. LaLone[a], Stephen W. Edwards[f], Lyle D. Burgoon[g], Nancy D. Denslow[h], Edward J. Perkins[i], Gerald T. Ankley[a]

[a] US EPA Mid-Continent Ecology Division, 6201 Congdon Blvd., Duluth, MN, 55804, USA
[b] Institute of Genomics, Biocomputing, and Biotechnology, Mississippi State University, Starkville, MS, 39762, USA
[c] University of St. Thomas, Department of Biology, 2115 Summit Ave., St. Paul, MN, 55105, USA
[d] Oregon Health & Science University, Division of Science and Engineering, West Campus, Beaverton, OR, 97006, USA
[e] University of Maryland, Department of Animal and Avian Sciences, College Park, MD, 20742, USA
[f] US EPA, NHEERL, Research Triangle Park, NC, 27711 USA
[g] US EPA, NCEA, Research Triangle Park, NC, 27711 USA
[h] University of Florida, Department of Physiological Sciences and Center for Environmental and Human Toxicology, Gainesville, FL, 32611, USA
[i] US Army Engineer Research and Development Center, 3909 Halls Ferry Rd., Vicksburg, MS, 39180, USA

ARTICLE INFO

Article history:
Received 14 December 2011
Accepted 15 December 2011
Available online 28 December 2011

Keywords:
Systems biology
Toxicogenomics
Gene set enrichment analysis
Fish
Endocrine

ABSTRACT

Oligonucleotide microarrays and other 'omics' approaches are powerful tools for unsupervised analysis of chemical impacts on biological systems. However, the lack of well annotated biological pathways for many aquatic organisms, including fish, and the limited power of microarray-based analyses to detect low level differential expression of individual genes can hinder the ability to infer and understand chemical effects based on transcriptomic data. Here we report on the supervised assembly of a series of tissue-specific functional gene sets intended to aid transcriptomic analysis of chemical impacts on the female teleost reproductive axis. Gene sets were defined based on an updated graphical systems model of the teleost brain–pituitary–gonadal-hepatic axis. Features depicted in the model were organized into gene sets and mapped to specific probes on three zebrafish (*Danio rerio*) and two fathead minnow (*Pimephales promelas*) microarray platforms. Coverage of target genes on the microarrays ranged from 48% for the fathead minnow arrays to 88% for the most current zebrafish platform. Additionally, extended fathead minnow gene sets, incorporating first degree neighbors identified from a Spearman correlation network derived from a large compendium of fathead minnow microarray data, were constructed. Overall, only 14% of the 78 genes queried were connected in the network. Among those, over half had less than five neighbors, while two genes, cyclin b1 and zona pellucida glycoprotein 3, had over 100 first degree neighbors, and were neighbors to one another. Gene set enrichment analyses were conducted using microarray data from a zebrafish hypoxia experiment and fathead minnow time-course experiments conducted with three different endocrine-active chemicals. Results of these analyses demonstrate the utility of the approach for supporting biological inference from ecotoxicogenomic data and comparisons across multiple toxicogenomic experiments. The graphical model, gene mapping, and gene sets described are now available to the scientific community as tools to support ecotoxicogenomic research.

Published by Elsevier B.V.

Abbreviations: EAC, endocrine-active chemical; GSEA, gene set enrichment analysis; GnIH, gonadotropin inhibitory hormone; SBML, systems biology markup language; MIRIAM, minimum information required in the annotation of models; ZFIN, zebrafish information network; LH, luteinizing hormone; FSH, follicle-stimulating hormone; GABA, gamma aminobutyric acid; GnRH, gonadotropin releasing hormone; GEO, gene expression omnibus; KEGG, Kyoto encyclopedia of genes and genomes; GO, gene ontology; HPG, hypothalamic-pituitary-gonadal; DAVID, database for annotation, visualization and integrated discovery.

* Corresponding author. Tel.: +1 218 529 5217; fax: +1 218 529 5003.
E-mail addresses: Villeneuve.dan@epa.gov (D.L. Villeneuve), Natalia@icnanotox.org (N. Garcia-Reyero), dalma@stthomas.edu (D. Martinović-Weigelt), zli@thehamner.org (Z. Li), watanabe@ebs.ogi.edu (K.H. Watanabe), eorlando@umd.edu (E.F. Orlando), lalone.carlie@epa.gov (C.A. LaLone), Edwards.stephen@epa.gov (S.W. Edwards), burgoon.lyle@epa.gov (L.D. Burgoon), ndenslow@ufl.edu (N.D. Denslow), Edward.J.Perkins@usace.army.mil (E.J. Perkins), ankley.gerald@epa.gov (G.T. Ankley).

1383-5718/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.mrgentox.2011.12.016

1. Introduction

Oligonucleotide microarrays and other 'omics' approaches are potentially powerful tools for unsupervised, discovery-driven, analysis of chemical impacts on biological systems. However, inferring and understanding chemical effects from transcriptomic, proteomic, and metabolomic data remains challenging. Due to the relatively small sample sizes typically used in (eco)toxicogenomic experiments and technical limitations of the microarray platform itself, statistical power to detect low level differential expression of individual genes is often poor. Additionally, well annotated biological pathways are lacking for many species (e.g., fish, frogs, invertebrates). Canonical pathways for a variety of functions critical to specific groups of organisms (e.g., vitellogenesis in oviparous animals, osmoregulation in aquatic organisms) or relevant to specific toxicological outcomes (e.g., impaired reproduction or development) are absent or poorly defined. Consequently, although they generate large amounts of data, many (eco)toxicogenomic experiments provide frustratingly little information and/-or biological/toxicological insight.

The present work was intended to enhance the utility of transcriptomic, and other 'omics' data for mechanistic ecotoxicology research with small fish. We had previously developed a graphical systems model of the teleost brain–pituitary–gonadal-hepatic axis to facilitate a systematic program of hypothesis-driven ecotoxicogenomics research concerning effects of endocrine-active chemicals (EACs) on fish reproduction [1,2]. The model was qualitatively and conceptually useful for hypothesis formulation and data interpretation. However, attempts to employ the model quantitatively in support of microarray data analyses were less successful. A compiled gene list based on the model was developed [3], but it contained features associated with critical reproductive functions in multiple different tissues while microarray analyses were conducted in a tissue-specific manner. Furthermore, the compiled gene list included a significant number of general cellular signal transduction-related features that are relatively uninformative. This is due to both the number of different pathways and functions they are involved with (i.e., lack of specificity), and the fact that, in many cases, they are functionally regulated primarily at the post-translational level as opposed to the transcriptional level. Therefore, we felt that quantitative analyses anchored to a priori knowledge represented in the graphical systems model could be improved by developing tissue-specific functional gene sets based on the graphical model.

To achieve that objective, we first updated our previous graphical model. Updates to the model were intended to add features and functional pathways related to either new information published since development of the original model or recent experimental results. For example, in recent years there has been increasing recognition of potential roles of kisspeptins and putative gonadotropin inhibitory hormones (GnIHs) in the neuroendocrine regulation of fish reproduction [4,5]. Thus, related features were incorporated into the revised model. Similarly, features related to cholesterol biosynthesis, prostaglandin signaling, and matrix metalloproteinases (relative to ovulation) were added based on recent microarray studies in which expression of related genes was modulated by EACs or other stressors [6–9]. Once the model was updated, features were mapped to corresponding probes on a number of commercial and custom microarray platforms for zebrafish (*Danio rerio*) and fathead minnow (*Pimephales promelas*). Using the model as a guide, features were organized into tissue-specific lists and then further subdivided based on functional categories. It is recognized that expression of many of these features is not truly tissue-specific. However, functional understanding of the role the feature plays is often restricted to specific tissues, as is reflected in the organization of the graphical model. Therefore lists were organized functionally and features were included in the gene lists for more than one tissue where the functional role in each tissue was understood. The gene lists developed were then used to conduct gene set enrichment analyses (GSEA). Gene set enrichment analysis is a computational method used to evaluate whether a set of genes, defined a *priori*, shows statistically significant concordant relationships between two biological states (e.g., phenotypes) [http://www.broadinstitute.org/gsea/index.jsp]. However, unlike other approaches that consider only those genes that exceed an arbitrary *p*-value cutoff, GSEA considers all the genes in the expression data set. As a result, it can facilitate detection of small changes in many genes within a set, or a large change in just a few genes [10,11]. To further broaden the scope of our analyses, we also extended the graphical model-based gene sets for ovary tissue by including nearest neighbors identified in a Spearman correlation-based transcriptional network. Finally, application of these tools was demonstrated using microarray data from several previously published studies.

2. Materials and methods

2.1. Graphical model updates

The graphical model of the teleost brain–pituitary–gonadal-hepatic axis, initially developed using CellDesigner 3.1 [1], was converted to CellDesigner 4.1 [www.celldesigner.org] (CellDesigner file provided as Supplementary Data). This new version of the software reads and writes Systems Biology Markup Language (SBML) level 2, version 4 format [http://sbml.org], supports MIRIAM (Minimum Information Required in the Annotation of Models) annotation [12] and conforms with Systems Biology Graphical Notation, process diagram level notation [13,14]. Where feasible, proteins and genes were named, or abbreviated, in a manner consistent with the nomenclature used by the Zebrafish Information Network (ZFIN; http://zfin.org/) [15]. Four distinct tissue compartments (i.e., brain, pituitary, generalized ovary, and liver), plus a blood compartment were included in the revised model (Fig. S1). Within the liver compartment, the primary additions were inclusion of the de novo cholesterol biosynthesis pathway and growth hormone-regulated expression of insulin-dependent growth factors (Fig. S2). As in the previous version, the generalized ovary compartment includes three important cell types, i.e., theca cells, granulosa cells, and oocytes. However, the revised model incorporates much greater detail concerning oocyte maturation and ovulation, associated prostaglandin signaling, and cholesterol homeostasis (Fig. S3). In the updated model, the pituitary compartment was collapsed from discrete luteininzing hormone (LH) and follicle-stimulating hormone (FSH) gonadotroph sub-compartments to a single generalized gonadotroph compartment in order to eliminate redundant depiction of signaling pathways involved in the regulation of LH and FSH β subunit expression (Fig. S4). Additionally, somatotroph and a generic, "other pituitary cells", sub-compartments were added to capture interactions between growth hormone and reproductive signaling, as well as other putative regulators. Finally, the brain compartment was reorganized into five neuronal cell-type sub-compartments which included kisspeptin and GnIH neurons, in addition to the gamma aminobutyric acid (GABAergic), dopaminergic, and generic gonadotropin releasing hormone (GnRH) neurons included in the original model (Fig. S5). Melatonin synthesis and signaling from the pineal gland was also added due to its putative role in GnIH regulation [5].

As with the previous version, the revised graphical model is chimeric. That is, where possible, it depicts relationships explicitly described for fish, or more precisely female fish. However, some aspects were inferred from mammalian and avian literature. Even in the case of fish-specific aspects, the model represents a synthesis of data from multiple species employing varying reproductive strategies and examined at different life stages. Consequently, the graphical model and the relationships it depicts should be regarded as a generalized, hypothesis-based, synthesis of diverse information from the published literature. Many of the relationships depicted need to be verified against species-specific literature and/or targeted experiments. Additionally, although certain aspects of the model may be common between males and females, the current model was oriented toward functional interpretation of data from females. While gene sets derived from the model can be applied to male data, functional interpretation of their significance may be quite different than for females. References related to the update of the graphical model are provided as Supplementary materials, Appendix A.

2.2. Gene lists

Tissue-specific gene lists were developed based on the graphical model. Because most of the graphical model nomenclature had been standardized to ZFIN, the ZFIN genes/markers/clones search tool (http://zfin.org/cgi-bin/webdriver?MIval=aa-newmrkrselect.apg) was used to match zebrafish annotation information to features in the model. Mapping of model features to specific targets in ZFIN was

conducted October–November, 2010. Specifically, zebrafish RNA accession numbers (generally Refseq), official gene symbols, gene names, and, when necessary, other synonyms were identified. These were used as query terms and searched against a database (Microsoft Office Access 2007) of microarray probe annotations. Initial searches were performed against the annotation file for Agilent's zebrafish microarray (V3, design ID 026437, December 2009; Agilent, Palo Alto, CA, USA). Additional searches were performed against two earlier versions of Agilent's zebrafish microarray (V2, design ID 109161, January 2008; V1, design ID 015064, 2006). As part of a previous comparative toxicogenomics effort, probes for homologous targets on two separate fathead minnow microarrays, a 4×44 K fathead minnow microarray developed by EcoArray (Gainesville, FL, USA; www.ecoarray.com; GEO Accession GPL7282) and a custom 8×15 K fathead minnow microarray (GEO Accession GPL9248) and Agilent's zebrafish microarray (V1, design ID 015064) were cross-referenced using a sequence-based reciprocal best hit approach [16,17]. Therefore, some fathead minnow microarray probes corresponding to graphical model features were identified using Agilent zebrafish V1 probe ids as a query against the fathead minnow-zebrafish cross-reference database. Additionally, because the fathead minnow microarray annotations rely heavily on the well-curated database of whole genome annotations for zebrafish, it was possible to match some additional fathead minnow microarray probes to graphical model features by using the zebrafish annotations (gene symbols, gene names, Refseq accessions) as query terms.

Certain features depicted in the model were purposely excluded from the gene lists. These were generally features associated with signal transduction processes known to be involved in a diversity of cellular signaling pathways (e.g., protein kinase A, stimulatory G proteins, adenylate cyclase). One rationale for this is that, functionally, many of these targets are thought to be primarily regulated at the post-translational level (e.g., via changes in phosphorylation state). Furthermore, based on previous experience with a compiled brain–pituitary–gonadal-hepatic axis gene list, differential expression of such features contributed little to our understanding of chemical impacts. Including them in the gene lists could actually obscure the significance of more informative targets. While this determination was somewhat arbitrary, we felt it would improve the utility of the lists.

There are some additional limitations and associated uncertainties that should also be considered when applying the lists. Due to genome duplication, teleost fish have two paralogous copies of many genes [18,19]. However, with the exception of a few well-studied teleost genes (e.g., *cyp19a1a*, *cyp19a1b*) neither the graphical model nor much of the literature it was derived from clearly differentiates among paralogous forms of the same gene, even though their regulation and/or functions may have diverged over evolutionary history. Similarly, for a number of generic features depicted in the model, there are multiple isoforms. For example, ovulation involves proteolysis of the extracellular matrix by matrix metalloproteinases [20]. Given the important role they play, matrix metalloproteinases are represented generically in the graphical model. However, there are over 25 distinct matrix metalloproteinase/metallopeptidase genes recognized in ZFIN. There is currently no information available in the literature regarding which of these are specifically involved in teleost ovulation. In the absence of information indicating which specific isoforms or paralogs were involved in the processes depicted in the model, we chose to include all isoforms and paralogs for which matching microarray probes could be identified in the current gene lists. We recognize that inclusion of multiple isoforms/paralogs could potentially bias GSEA; however the same could be said for exclusion, so we chose to err on the side of inclusiveness rather than exclusion. There were also a few of the features in the chimeric graphical model, that were included based on inference from mammalian literature, for which no appropriate zebrafish synonyms could be identified due to inconsistencies and uncertainties in the gene/protein nomenclature, or possibly lack of evolutionary conservation (e.g., constitutive androstane receptor, estrogen sulfotransferase). However, in the vast majority of cases putative orthologs could be identified and mapped to one or more zebrafish or fathead minnow microarray probes.

2.3. Extended gene lists

As part of a broad program of ecotoxicogenomic research related to effects of EACs on the brain–pituitary–gonadal-hepatic axis in fish [2] we have accumulated a large compendium of microarray data (i.e., >1200 microarrays from over 20 independent experiments using the 8×15 K custom fathead minnow microarray; GEO Accession GPL9248) for fathead minnow ovary tissue that can be used for correlation- and mutual information-based transcriptional network inference [21]. For the present work, we used a correlation-based network to identify genes whose expression was closely correlated with that of one or more features depicted in the graphical model, over a range of different biological perturbations. Microarray data from seven different time-course experiments in which fathead minnows were exposed to an EAC for eight days, followed by eight days of depuration, and samples collected on days 1, 2, 4, and 8 of both the exposure and depuration period (Table S1; following the same general design as [22–24]) were Fastlo [25] normalized within experiment. Spearman correlation was performed across all arrays ($n = 900$) and experiments ($n = 7$). An adjacency matrix of all the features on the fathead minnow microarray (GPL9248) was constructed, such that probe pairs for which the absolute value of the Spearman correlation coefficient (ρ) was greater than 0.80 were assigned a value of 1; while all others were assigned a value of 0. The adjacency matrix was read into igraph (an R library; http://igraph.sourceforge.net) generating a transcriptional network where an edge was created between any two probes (genes or nodes in the network) for which the absolute value of their Spearman correlation coefficient was greater than 0.80. First degree neighbors to the probes included in the graphical model-based ovary gene lists were then identified, where first degree neighbors are defined as nodes in the network that are directly connected to one another by a single edge (Fig. S6). A set of so-called "extended gene lists" were developed by taking the gene lists identified above (see Section 2.2) and adding the first degree neighbors to any of the genes in the original list. Redundancies were eliminated such that probes identified as first degree neighbors to more than one gene in the original list, or genes that were identified as first degree neighbors to one another, were represented only once.

Functional annotation clustering was performed on a subset of the genes identified through the network analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID [26]; http://david.abcc.ncifcrf.gov/home.jsp). Probe ids for the custom fathead minnow microarray have been annotated with the best hit zebrafish protein Refseq accession number (see GEO Accession GPL9248). Refseq accession numbers associated with the probes of interest were loaded directly into DAVID [26] as a gene list. Functional annotation clustering was then performed using default settings in DAVID with *Danio rerio* (zebrafish) as the background population for the analysis.

2.4. Gene set enrichment analyses

Gene set enrichment analyses were carried out using GSEA v2.0 (www.broad.mit.edu/gsea). Analyses were conducted using microarray data from previously published studies. The first of these was a hypoxia experiment in which zebrafish were exposed to three different concentrations of dissolved oxygen (1.0, 3.0, or 7.0 mg O_2/L) for either four or 14 d and transcriptional changes in the gonad of males and females were examined using a 21,000 gene microarray ([8], GEO Accession GSE10951). These were two-color microarray data in which all experimental samples were labeled with Cy5 and co-hybridized with a common reference pool of RNA labeled with Cy3. Microarray data from this study were reported previously [8], but no GSEA were conducted. To facilitate GSEA, the fold-change (experimental/reference) microarray data were median normalized, per gene, across all 60 samples (arrays) from the experiment using GeneSpring GX (Agilent). The resulting data matrix (.gct file), a .cls file that defined the experimental treatments, and a .gmt file defining 14 gene sets considered in the analysis were imported into the GSEA v2.0 software for GSEA.

Microarray data from experiments with three EACs, fadrozole (an aromatase inhibitor), prochloraz (an inhibitor of aromatase and other steroidogenesis enzymes), and 17β-trenbolone (a potent synthetic androgen), were also considered. Each experiment used a consistent experimental design in which fathead minnows were exposed to one of two concentrations of test chemical or control water for 1, 2, 4, or 8 days and then sampled or exposed continuously for 8 days followed by depuration in control water for an additional 1, 2, 4, or 8 days [22–24]. Microarray analyses were conducted on RNA extracted from ovary tissue using a custom 8×15 K fathead minnow microarray (GEO Accession GPL9248). Microarray data from these studies were deposited in the publically accessible CEBS (Chemical Effects in Biological Systems) database, managed by the United States National Institute of Environmental Health Sciences (NIEHS; http://www.niehs.nih.gov/research/resources/databases/cebs/index.cfm); accession numbers 010-00004-0002-000-7 (fadrozole), 010-00004-0003-000-8 (17β-trenbolone), 010-00004-0004-000-9 (prochloraz). Analyses of microarray data from these studies are on-going and only GSEA results for 23 gene sets derived as part of the present work are presented here. In all cases, the single color microarray data from each experiment was normalized, within experiment, using Fastlo [25] implemented in R (http://www.r-project.org). Following normalization, a series of data matrix (.gct file) and .cls files defining the phenotypes (experimental classes) were constructed using custom R code (developed by S. Edwards). These files, along with a .gmt file that included 23 gene sets were imported into the GSEA v2.0 software for GSEA.

All GSEA were conducted using 1000 permutations with the permutation type set to phenotype for the EAC studies (most sample sizes ≥ 7) and gene set permutation for the hypoxia experiment (generally $n = 5$ per phenotype). Binary combinations of phenotypes were considered in each analysis, where the first phenotype was a time point-matched control in the case of the fathead minnow experiments or time point- and sex-matched control in the case of the zebrafish experiment, and the second phenotype was a treatment defined by concentration × duration (× sex; zebrafish experiment). Enrichment in either phenotype one relative to phenotype two (positive enrichment score), or vice versa (negative enrichment score), was considered statistically significant at a nominal (uncorrected for false discovery rate) $p < 0.05$. Gene set size limits were set to a maximum of 500 features and a minimum of nine features, resulting in exclusion of two of the 23 gene sets included in the .gmt file used for the fathead minnow experiments. The total number of microarray features considered was 21,555 for the zebrafish experiment and 12,448 for the fathead minnow experiments, and probes were not collapsed to gene symbols due to incomplete annotation information.

3. Results

3.1. Liver gene list(s)

Sixty-eight genes were identified as being relevant to the liver compartment of the graphical systems model (Table 1). Those genes were organized into three functional categories (Table S2): steroid metabolism (27 genes), vitellogenesis (22 genes), and cholesterol biosynthesis (19 genes). Coverage of those 68 genes on the five small fish microarrays evaluated in the present study ranged from approximately 47% for EcoArray's fathead minnow microarray (GPL 2782) to 85% for the most recent Agilent zebrafish microarray design (026437) (Table 1).

3.2. Brain and pituitary gene lists

Graphical model features related to the brain and pituitary compartments of the model were associated, respectively, with 30 and 48 target genes identified in ZFIN (Table 1). Overall coverage of those targets on the small fish microarrays ranged from 40% to 93% for brain, and 46% to 88% for pituitary. As in the case of liver, Agilent's most recent zebrafish microarray design (026437) provided the greatest amount of coverage as well as the greatest number of probes per target. The custom fathead minnow microarray (GPL9248) had the poorest coverage of brain and pituitary targets. Brain targets were subdivided into eight functional groups (Table S3): dopaminergic neurons (six genes), gaba neurons (three genes), glia (one gene), gnih neurons (one gene), gnrh neurons (three genes), hypothalamic signaling (five genes), kiss neurons (four genes), and melatonin-related (seven genes). Pituitary targets were organized into four functional groups (Table S4): growth hormone-related (10 genes), gonadotropin (gth) subunits (three genes), regulators of gth expression or release (31 genes), and secondary regulators of gth/growth hormone (four genes). However, due to the small size of most of the functional subsets defined for these tissues, GSEA in the present study included only the entire tissue-specific lists.

3.3. Ovary gene lists

The greatest number of gene targets relevant to the graphical systems model was identified for the ovary compartment. Probes corresponding to around 85% of the 143 relevant target genes could be found on Agilent's zebrafish microarray design 026437 (Table 1), with the older designs covering 75% (019161) and 56% (015054) of the targets, respectively. Coverage on the two fathead minnow microarrays ranged from 48% (Ecoarray; GPL7282) to 54% (GPL9248). The functional categories for the ovary gene targets (Table S5) included: steroid biosynthesis (19 genes), prostaglandin-related (12 genes), ovulation (31 genes, all matrix metalloproteinases/peptidases or progestin receptors), regulators of oocyte maturation (35 genes), oocyte growth-related (18 genes), and genes involved in cholesterol uptake (nine genes). Although liver is generally regarded as the primary site of de novo cholesterol synthesis, the 19 genes associated with cholesterol biosynthesis (Tables S2 and S5) were also included in the overall ovary gene list. This inclusion was based on previous observations of altered ovarian expression of a number of cholesterol synthesis-related genes following exposure to EACs (e.g., [9]).

3.4. Extended gene lists

Because we had a large compendium of fathead minnow microarray (GPL9248) data available for ovary tissue that could be used for transcriptional network analysis (Table S1), we were able to create extended functional gene lists for ovary by including first degree neighbors to the model-derived targets, as identified from a Spearman correlation network. Among the 78 unique fathead minnow microarray (GPL9248) probes associated with the overall graphical model-derived gene list for ovary, 11 genes were identified as having at least one first degree neighbor in the Spearman correlation network (Table 2). Nearly half of those had just one or two first degree neighbors. However, two genes, cyclin b1 (*ccnb1*) and zona pellucida glycoprotein 3 (*zp3*) had greater than 100 first degree neighbors. Those two genes were first degree neighbors to one another, and shared 87 first degree neighbors in common in addition to one another (Fig. S7). None of the other genes were neighbors to one another, although apolipoprotein eb (*apoeb*) and activin A receptor type I like (*acvrl1*) shared one common first degree neighbor, a probe annotated as transgelin 2 (UF_Ppr_AF_108353; Fig. S7). Excluding duplicates, 151 first degree neighbors to the genes in the ovary list were identified.

Functional annotation clustering was performed on the set of genes identified as first degree neighbors of either *ccnb1* or *zp3* using DAVID [26]. Based on a Benjamini–Hochberg corrected p-value < 0.05 as the threshold for statistical significance, three significant annotation clusters were identified (Table S6). The first was a cluster of 10 annotation terms associated with ribosomes, ribosomal proteins, and translation. The second was a cluster of three terms related to zona pellucida, while the third group related to rRNA binding.

3.5. Gene set enrichment analyses

Gene set enrichment analysis was conducted on a set of zebrafish microarray data from a previously published hypoxia

Table 1
Overview of success rates mapping graphical model features to specific probes on five different small fish microarray platforms (i.e., Agilent zebrafish microarray designs 026437, 019161, and 015064; EcoArray's 4 × 22 K fathead minnow microarray, GEO Accession GPL7282; and a custom 8 × 15 K fathead minnow microarray designed by Nancy Denslow, University of Florida, GEO Accession GPL9248).

Tissue	Relevant[a] genes	Total number of probes[b]/target coverage[c]				
		zf 026437	zf 019161	zf 015064	fhm (GPL7282)	fhm (GPL9248)
Liver	68	167/58	94/51	57/39	41/32	38/36
Brain	30	82/28	45/26	28/20	18/13	19/12
Pituitary	48	76/42	60/36	34/27	25/23	26/22
Ovary	143	246/121	182/107	131/92	77/69	78/77
% coverage[d]		88 ± 4	78 ± 6	61 ± 5	48 ± 4	48 ± 6
Ave probes per target[e]		2.4 ± 0.6	1.7 ± 0.1	1.4 ± 0.1	1.2 ± 0.2	1.2 ± 0.3

[a] Relevant genes refers to model features for which a corresponding gene target could be identified using ZFIN (http://zfin.org/).
[b] The total number of microarray probes that matched to the relevant target genes based on available annotation information (i.e., genes may be represented by multiple probes on each microarray).
[c] Target coverage refers to the number of unique relevant target genes for which at least one corresponding microarray probe could be identified.
[d] Mean (±SD) percentage of relevant targets that could be mapped to at least one corresponding microarray probe across the four tissue-specific lists examined.
[e] Average (±SD) probe redundancy per target.

Table 2
Results of Spearman correlation network-based neighborhood analysis based on an initial list of 78 fathead minnow microarray (GEO Accession GPL9248) probes corresponding to features depicted in the ovary compartment of the graphical systems model (Fig. S1). Only genes with at least one first degree neighbor (see Fig. S6) in the network are listed.

Gene name	Gene symbol	Functional group	1st degree neighbors
Cyclin b1	ccnb1	Oocyte maturation	106
Zona pellucida glycoprotein 3	zp3	Oocyte growth	101
Activin A receptor type I like	acvrl1	Oocyte maturation	11
Lanosterol synthase	lss	Cholesterol biosynthesis	7
Apolipoprotein eb	apoeb	Cholesterol uptake	6
Inhibin, beta Aa	inhbaa	Oocyte maturation	3
bmp and activin membrane bound inhibitor homolog a	bambia	Oocyte maturation	2
Very low density lipoprotein receptor	vldlr	Oocyte growth	2
Matrix metalloproteinase 2	mmp2	Ovulation	1
Insulin-dependent Growth factor binding protein 5	igfbp5b	Oocyte maturation	1
Hydroxysteroid (17-beta) dehydrogenase 1	hsd17b1	Steroid biosynthesis	1

experiment [8]. These data provided an opportunity to compare novel GSEA results from the present study with those of previous analyses which were focused only on statistically differentially expressed genes [8]. Of the 14 graphical-model-based gene sets considered in the analysis, 12 were detected as either statistically over-expressed or under-expressed relative to the control (7.0 mg O_2/L) in one or more of the hypoxia treatment conditions (Table 3). Functional sets of prostaglandin-related and steroid metabolism-related genes were not significantly impacted in any of the hypoxia treatments. The specific gene sets affected by hypoxia treatment varied with exposure duration (4 d versus 14 d), severity of the hypoxia (1.0 versus 3.0 mg O_2/L), and sex. However, notably, the set of genes related to cholesterol uptake in the gonad was significantly under-expressed, relative to time-matched controls, in females under all four hypoxic conditions (duration × dissolved oxygen concentration) tested (Table 3). In females, the majority of significantly impacted gene sets were under-expressed relative to controls, with exceptions being over-expression of cholesterol synthesis-related genes in the ovary of females held at 3 mg O_2/L for 14 d, and genes related to oocyte growth (dominated by various isoforms of zona pellucida glycoproteins; Table S5). In males (testes), the gene set most consistently impacted by hypoxia was the set of 26 vitellogenesis-related genes identified for liver (Table 3). That gene set includes androgen receptor and estrogen receptor α, but is dominated by vitellogenin and zona pellucida glycoproteins (Table S2). It was also notable that that for both the 4 d hypoxia treatments in males both gene sets with a strong composition of genes coding for zona pellucida glycoproteins (i.e., "vitellogenesis_liver" and "oocyte_growth_gonad" gene sets) were identified as significantly over-expressed compared to control males (held at 7 mg O_2/L). The set of 12 genes related to cholesterol synthesis were significantly under-expressed in males exposed to hypoxia treatment for 14 d, both at 3 and 7 mg O_2/L. This contrasted with over-expression of cholesterol synthesis-related genes in the ovary

Table 3
Overview of gene set enrichment analysis (GSEA) results for microarray data from an experiment in which zebrafish were exposed to three different concentrations of dissolved oxygen (7.0, 3.0, 1.0 mg/L; abbreviated do7, do3, do1, respectively) for either 4 or 14 d [8]. Gonad was the target tissue analyzed using Agilent's zebrafish microarray (design 015064). Green shading indicates significant enrichment (nominal $p < 0.05$) in the control group compared to the treated, red shading indicates significant enrichment in the treated group relative to control, lack of shading indicates no significant enrichment. $n = 5$ unless otherwise noted (see also Table S8).

Gene Set ID	all_gonad	all_liver	all_pituitary	oocyte_maturation_gonad	all_brain	vitellogenesis_liver	steroid_biosynthesis_gonad	ooctye_growth_gonad	cholesterol_uptake_gonad	steroid_metabolism_liver	prostaglandin_related_gonad	ovulation_gonad	cholesterol_synthesis	all_hpgh[a]
Number of Probes	131	56	34	28	28	26	24	20	20	18	14	13	12	212
Treatment Class														
F, 14d, do1														
F, 14d, do3														
F, 4d, do1 [b]														
F, 4d, do3 [c]														
M, 14d, do1														
M, 14d, do3														
M, 4d, do1														
M, 4d, do3 [c]														

[a] The all_hpgh gene set includes all redundant probes from the all_gonad, all_liver, all_brain, and all_pituitary gene sets.
[b] $n = 6$.
[c] $n = 4$.

of females exposed to moderate hypoxia (3 mg O_2/L) for 14 d. The "all_hpgh" gene set that included all non-redundant probes from the tissue-specific gene lists was only detected as statistically altered in females held at 1 mg O_2/L for 14 d. Nonetheless, collectively, the results provide strong evidence suggesting that expression of genes coding for proteins known to play roles in the regulation of fish reproduction were impacted by hypoxia treatment.

Gene set enrichment analysis was also conducted on microarray data from three different fathead minnow experiments [22–24]. Our purpose here was to illustrate and explore the utility of GSEA for comparing across microarray experiments. The analysis considered 23 gene sets (although two were ultimately excluded due to insufficient size) defined based on the graphical systems model and/or first degree neighbors as identified from a Spearman correlation network constructed from data from these studies [22–24], as well as four others not considered here (Table S1). Impacts of three EACs including the aromatase inhibitor fadrozole, the steroidogenesis inhibitor prochloraz, and the synthetic androgen 17β trenbolone, over an eight day exposure period, followed by up to eight days of depuration in control water was examined. As in the zebrafish hypoxia experiment, all three chemical stressors significantly impacted the gene sets in a manner that varied as a function of stressor severity (concentration) and duration of exposure and/or depuration (Tables 4–6). Among the three chemicals, fadrozole impacted the greatest diversity of gene sets, with only two, one composed primarily of genes coding for zona pellucida glycoproteins (oocyte_growth) and the other composed of the gene coding for activin A receptor, type I-like (*acvrl1*) with its eleven first degree neighbors (Fig. S7), being unaffected under all treatment conditions considered (Table 4). In the case of prochloraz, five gene sets were unaffected (Table 5). These included the set of genes associated with the liver compartment of the graphical systems model (Table S2), the cholesterol synthesis-related gene set, the prostaglandin-related gene set and the overall "all_hpgh" gene set. The set of nine genes associated with cholesterol uptake was not affected, but an extended list that included the nine cholesterol uptake-related genes along with six first degree neighbors to *apoeb* (see Fig. S7) was significantly impacted, albeit only in fish that had been exposed to 30 μg prochloraz/L for eight days and

Table 4
Overview of gene set enrichment analysis (GSEA) results for microarray data from a 16 day time course experiment in which fathead minnows were exposed to 30 or 3.0 μg fadrozole (FAD)/L [22]. Ovary was the target tissue analyzed using a custom 8 × 15 K fathead minnow microarray (GEO Accession GPL9248). Fish were exposed for 1, 2, 4, or 8 days then sampled (d1, d2, d4, d8), or exposed continuously for 8 days then to control water only for 1, 2, 4, or 8 additional days (d9, d10, d12, d16). Gene set ids with black background indicate extended gene sets that include first degree neighbors from a Spearman correlation network. Green shading indicates significant enrichment (nominal $p < 0.05$) in the control group compared to the treated, red shading indicates significant enrichment in the treated group relative to control, lack of shading indicates no significant enrichment. $n = 8$ unless otherwise noted (see also Table S9).

Gene Set ID	EX-all_gonad	EX- oocyte_maturation	ccnb1, zp3, and neighbors[a]	EX-oocyte_growth	ccnb1 and neighbors	all_gonad	all_gonad_sans_chol_synth	all_liver	all_pituitary	oocyte_maturation	all_brain	liver_steroid_metabolism	EX-cholesterol_uptake	cholesterol_synthesis	acvrl1 and neighbors	EX-steroid biosynthesis	steroid_biosynthesis	oocyte_growth	prostaglandin-related	cholesterol_uptake	liver_vitellogenesis[b]	matrix_metalloproteinase[b]	all_hpgh[c]
Number of Probes	227	143	118	113	106	78	67	38	26	21	19	17	15	13	12	12	11	10	10	9	8	6	138
Treatment Class																							
FAD, 30, d1[d]																							
FAD, 30, d2																							
FAD, 30, d4[d]																							
FAD, 30, d8[d]																							
FAD, 30, d9																							
FAD, 30, d10[d]																							
FAD, 30, d12																							
FAD, 30, d16[d]																							
FAD, 3, d1																							
FAD, 3, d2																							
FAD, 3, d4																							
FAD, 3, d8																							
FAD, 3, d9																							
FAD, 3, d10																							
FAD, 3, d12[d]																							
FAD, 3, d16																							

[a] Non-redundant neighbors only; each probe represented only once.
[b] These gene sets, shaded gray, were excluded from the analysis because they did not meet the minimum size threshold (≥9 probes).
[c] The all_hpgh gene set includes all redundant probes from the all_gonad, all_liver, all_brain, and all_pituitary gene sets.
[d] $n = 7$.

Table 5
Overview of gene set enrichment analysis (GSEA) results for microarray data from a 16 day time course experiment in which fathead minnows were exposed to 300 or 30 μg prochloraz (PRO)/L [23]. Ovary was the target tissue analyzed using a custom 8 × 15 K fathead minnow microarray (GEO Accession GPL9248). Fish were exposed for 1, 2, 4, or 8 days then sampled (d1, d2, d4, d8), or exposed continuously for 8 days then to control water only for 1, 2, 4, or 8 additional days (d9, d10, d12, d16). Gene set ids with black background indicate extended gene sets that include first degree neighbors from a Spearman correlation network. Green shading indicates significant enrichment (nominal $p < 0.05$) in the control group compared to the treated, red shading indicates significant enrichment in the treated group relative to control, lack of shading indicates no significant enrichment. $n = 7$ unless otherwise noted (see also Table S10).

Gene Set ID	EX-all_gonad	EX- oocyte_maturation	ccnb1, zp3, and neighbors[a]	EX-oocyte_growth	ccnb1 and neighbors	all_gonad	all_gonad_sans_chol_synth	all_liver	all_pituitary	oocyte_maturation	all_brain	liver_steroid_metabolism	EX-cholesterol_uptake	cholesterol_synthesis	acvrl1 and neighbors	EX-steroid biosynthesis	steroid_biosynthesis	oocyte_growth	prostaglandin-related	cholesterol_uptake	liver_vitellogenesis[b]	matrix_metalloproteinase[b]	all_hpgh[c]
Number of Probes	227	143	118	113	106	78	67	38	26	21	19	17	15	13	12	12	11	10	10	9	8	6	138
Treatment Class																							
PRO, 300, d1 [d]									■	■					■								
PRO, 300, d2																							
PRO, 300, d4						■	■									■	■						
PRO, 300, d8 [d]											■					■	■						
PRO, 300, d9																							
PRO, 300, d10																							
PRO, 300, d12			■	■	■																		
PRO, 300, d16 [d]																		■					
PRO, 30, d1 [d]											■				■								
PRO, 30, d2																							
PRO, 30, d4	■	■	■	■	■													■					
PRO, 30, d8	■	■	■	■	■							■											
PRO, 30, d9																							
PRO, 30, d10																							
PRO, 30, d12	■	■	■	■	■																		
PRO, 30, d16													■		■								

[a] Non-redundant neighbors only; each probe represented only once.
[b] These gene sets, shaded gray, were excluded from the analysis because they did not meet the minimum size threshold (≥9 probes).
[c] The all_hpgh gene set includes all redundant probes from the all_gonad, all_liver, all_brain, and all_pituitary gene sets.
[d] $n = 6$.

then depurated for an additional eight days (Table 5). Compared to fadrozole and prochloraz, results of the GSEA for the 17β trenbolone were notable for the lack of significant effect on the gene sets involved in steroid biosynthesis or steroid metabolism (Table 6). Additionally, in the case of 17β trenbolone, the gene sets associated with all the gonad features in the graphical systems model (i.e., all_gonad; all_gonad_sans_chol_synth) as well as all brain features in the model (all_brain) were not significantly affected, even though functional subsets of the "all_gonad" list were significantly impacted.

While each chemical produced a unique response profile, there were some similarities. For example, the five gene sets that contained the large number of first degree neighbors to *ccnb1*, *zp3*, or both, were all significantly impacted under at least one experimental condition from each experiment. Based on the DAVID analysis of the neighbors to *ccnb1* and *zp3*, the results suggest that significant differences in the overall expression of ribosomal proteins and/or overall translational activity occurred in all three studies. Other than the sets sharing the *ccnb1* and *zp3* neighbor composition, the gene sets affected under at least one condition in all three experiments were the all_pituitary (Table S4), oocyte_maturation (Table S5), and the set of cholesterol uptake-related genes, including *apoeb*'s six first degree neighbors (EX-cholesterol_uptake). The latter of these three gene sets was impacted on d16 in all experiments, although only at the lower concentration tested in the case of the prochloraz and 17β-trenbolone experiments. This was the only time point-consistent positive effect (i.e., significant over- or under-expression of the gene set) observed across the three experiments. However, in terms of negatives, it was notable that in all three experiments, there was one time point during the depuration phase of the experiment where no significant over- or under-expression of any of the 23 (21 meeting size threshold) gene sets was detected. In general, this occurred on day 10, the second day of depuration following 8 d of continuous exposure (Tables 4–6). The only exception was for fish exposed to 3.0 μg fadrozole/L, in which case the no-effect time point in the depuration phase was on day 9 (24 h post-exposure) rather than day 10 (Table 4). It was also notable that in the prochloraz and 17β-trenbolone experiments, no significant effects were observed on day 2 (i.e., after 48 h of continuous exposure).

Table 6
Overview of gene set enrichment analysis (GSEA) results for microarray data from a 16 day time course experiment in which fathead minnows were exposed to 500 or 30 ng 17β-trenbolone (TRB)/L [24]. Ovary was the target tissue analyzed using a custom 8 × 15 K fathead minnow microarray (GEO Accession GPL9248). Fish were exposed for 1, 2, 4, or 8 days then sampled (d1, d2, d4, d8), or exposed continuously for 8 days then to control water only for 1, 2, 4, or 8 additional days (d9, d10, d12, d16). Gene set ids with black background indicate extended gene sets that include first degree neighbors from a Spearman correlation network. Green shading indicates significant enrichment (nominal $p < 0.05$) in the control group compared to the treated, red shading indicates significant enrichment in the treated group relative to control, lack of shading indicates no significant enrichment (see also Table S11).

Gene Set ID	EX-all_gonad	EX- oocyte_maturation	ccnb1, zp3, and neighbors[a]	EX-oocyte_growth	ccnb1 and neighbors	all_gonad	all_gonad_sans_chol_synth	all_liver	all_pituitary	oocyte_maturation	all_brain	liver_steroid_metabolism	EX-cholesterol_uptake	cholesterol_synthesis	acvrl1 and neighbors	EX-steroid biosynthesis	steroid_biosynthesis	oocyte_growth	prostaglandin-related	cholesterol_uptake	liver_vitellogenesis[b]	matrix_metalloproteinase[b]	all_hpgh[c]
Number of Probes	227	143	118	113	106	78	67	38	26	21	19	17	15	13	12	12	11	10	10	9	8	6	138
Treatment Class																							
TRB, 500, d1 [d]																							
TRB, 500, d2 [e]																							
TRB, 500, d4 [f]																							
TRB, 500, d8 [e]																							
TRB, 500, d9 [f]																							
TRB, 500, d10 [d]																							
TRB, 500, d12 [d]																							
TRB, 500, d16 [g]																							
TRB, 30, d1 [e]																							
TRB, 30, d2 [e]																							
TRB, 30, d4 [e]																							
TRB, 30, d8 [d]																							
TRB, 30, d9 [d]																							
TRB, 30, d10 [g]																							
TRB, 30, d12 [h]																							
TRB, 30, d16 [i]																							

[a] Non-redundant neighbors only; each probe represented only once.
[b] These gene sets, shaded gray, were excluded from the analysis because they did not meet the minimum size threshold (≥9 probes).
[c] The all_hpgh gene set includes all redundant probes from the all_gonad, all_liver, all_brain, and all_pituitary gene sets.
[d] $n = 7$; [e] $n = 8$; [f] $n = 6$; [g] $n = 4$; [h] $n = 5$.
[i] $n = 2$ and it was necessary to use gene set permutation instead of phenotype permutation.

4. Discussion

4.1. Graphical model and gene lists

Providing biological context in which to interpret the results of ecotoxicogenomics experiments remains challenging. Annotations and computational pathway analysis tools for fish, wildlife, invertebrates, etc., are substantially less developed than those for well characterized human health research models (e.g., rats, mice). There are examples where useful biological inference has been made based on detailed examination of individual transcript changes (i.e., differentially expressed genes). However, such analyses can be difficult to conduct in an efficient and unbiased manner that facilitates both objective interpretation of results and direct comparisons between multiple experiments. Ontology-based tools are widely employed, but often do not provide the level of specificity needed to gain much biological insight. Therefore, many ecotoxicogenomic experiments fall in the category of being data rich, but information poor.

The present study was aimed at improving our ability to extract information from toxicogenomic experiments related to the effects of chemical stressors on fish reproduction. In many respects, the pathway- and annotation-based data analysis tools commonly employed in toxicogenomic analyses are fundamentally based on synthesis and codification of accumulated knowledge (i.e., in the published literature) regarding the functions or roles of genes, proteins, metabolites, etc. in a biological system. The revised graphical model presented here provides that synthesis for a biological system not currently represented by available inventories of canonical pathways (e.g., the Kyoto Encyclopedia of Genes and Genomes; KEGG; http://www.genome.jp/kegg/), notably the intact brain–pituitary–gonadal-hepatic axis. For those unfamiliar with this system, the graphical model provides a useful biological context from which to draw connections and understand relationships between different features examined in an experiment. For those familiar with the system and the associated literature, it is a way to organize disparate sources of information into a generalized functional map that can serve as a basis for hypothesis testing as

well as comparing and contrasting among species, among chemical impacts, etc. Therefore, the graphical model in and of itself, can help derive information from toxicogenomic data.

Mapping the features depicted in the model to specific microarray probes allows for the quantitative application of this assembled biological context. At present, the primary limitation on this implementation is the quality and completeness of the annotation information for the microarray platform in question. However, comparison of the mapping efficiency/target coverage for three commercial zebrafish microarray platforms was illustrative. There was a marked increase in coverage of the features depicted in the graphical model with each successive update of the zebrafish microarray platform (Table 1). Coverage on the fathead minnow microarrays still lagged behind the commercial zebrafish platforms considered. Nonetheless, at the time the original graphical model was constructed, a 2000 gene microarray was the most comprehensive available for the species and only 30 relevant model features could be mapped to corresponding probes [1]. At present, over five times as many probes, accounting for roughly 50% genes/proteins relevant to the model could be mapped. This reflects a rapid and continuous improvement in the tools and annotations being applied in ecotoxicogenomics research.

4.2. Extended gene sets

Coupling the derivation of model-based gene sets with clusters of co-expressed genes identified from transcriptional network inference provides an effective means to integrate both knowledge-/hypothesis-directed and discovery-based approaches to microarray data analysis. In the present study, we used transcriptional networks in an effort to broaden the scope of our gene sets. On a practical level, this would have the benefit of increasing the number of genes represented in each set, particularly for functional categories with few members or low diversity of membership (e.g., dominated by multiple isoforms of a particular type of protein). The approach was partially successful in that regard. Some of the functional gene sets (e.g., those related to oocyte maturation, oocyte growth, cholesterol uptake) were substantially expanded by incorporating first degree neighbors from a Spearman correlation network. However, consistent with transcriptional network theory [27], the distribution of neighbors was very heterogeneous, with just a few of the model features having many first degree neighbors, while most had few or no first degree neighbors in the correlation network (Table 2). Consequently, the extent to which any particular functional subset was expanded was highly variable and at the level of the whole tissue list, addition of the neighbors weighted the list heavily toward certain functional subsets and not others. Ultimately, rather than replace the original gene sets with the extended sets that included first degree neighbors, we felt it was useful to include both the original and extended gene sets in the analyses. Useful inferences could be drawn by comparing and contrasting results for the original and expanded lists.

On a theoretical level, it is supposed that genes regulated in a coordinated or correlated fashion over a wide range of experimental conditions may be subject to similar regulation and/or play complementary functional roles [27]. Bearing in mind that the Spearman correlation network was agnostic as to whether the correlations between transcripts were positive or negative, results of the present study would seem to support that assumption. For example, correlations in the expression of *ccnb1* and *zp3* and with that of genes coding for ribosomal proteins and other zona pellucida glycoprotein isoforms seems biologically plausible. Based on the graphical model and prevailing understanding of oogenesis in fish and other vertebrates, we would hypothesize that the preovulatory period of oocyte growth would be characterized by elevated expression of zona pellucida glycoproteins [28,29]. High expression of transcripts coding for ribosomal proteins would support an increased number of ribosomes to support the production of abundant zona pellucida glycoproteins as well as translation of maternal RNAs whose expression is regulated post-transcriptionally during subsequent oocyte maturation and early embryonic development [30–32]. Cyclin b1 (*ccnb1*) is one of two critical components of the heterodimeric protein complex that makes up a maturation promoting factor (see Fig. S3) which triggers transition from the oocyte growth stage of oogenesis (characterized by a period of intensive RNA synthesis [30]) to final maturation (characterized by predominantly post-transcriptional regulation [31,32]). Therefore, it seems plausible that expression of *ccnb1* would be correlated (positively or negatively) with the expression of zona pellucida and ribosomal proteins. Examination of the annotations associated with the individual neighbors to *ccnb1* and *zp3* (Table S7) also reveals additional genes that align well with the functions of oocyte growth and maturation. For example, cathepsin B is involved in proteolysis of yolk components like vitellogenin during oocyte maturation [33]. Aurora kinase is associated with meiotic maturation and inhibitors of its activity have been shown to block meiotic maturation in mouse oocytes [34]. Similarly, genes annotated as oogenesis-related protein, similar to ovulatory protein 2, zygote arrest 1, also suggest a close relationship with oocyte growth and/or maturation. It is also notable that transgelin 2, identified as a first degree neighbor shared by *acvrl1* and *apoeb* was a gene identified in a separate microarray study (not included in the data compendium used to generate the current Spearman correlation network) to be a candidate marker of significant shifts in gross ovarian stage in fathead minnows [17]. We emphasize that the relationships represented in our Spearman correlation network and neighborhood analyses are simply correlations that exceed an arbitrary threshold. They are not necessarily functional relationships and attempts to reconstruct functional networks have met with rather limited success [35]. Nonetheless, these results provide encouragement that transcriptional network inference may be effectively employed to identify additional relevant molecular features to incorporate into the graphical model and associated gene sets. Additionally, the results illustrate how features depicted in the current graphical model can provide contextual anchors that can aid network-based analyses of data collected in toxicogenomic experiments.

4.3. GSEA

Microarray results from our hypoxia experiment with zebrafish [8] were previously examined in detail. However, those analyses relied primarily on evaluation of gene ontology (GO) categories identified as being statistically enriched among lists of differentially expressed genes derived for the various treatment conditions. Additionally, changes in the expression of individual genes among a list of 83 with known reproductive roles in the teleost brain–pituitary–gonadal-hepatic axis, based on our earlier version of the graphical model [1], were considered [8]. However, no GSEA was previously conducted on these data. Thus, the data set provided an opportunity to assess whether a GSEA analysis using the gene lists assembled in the present study could support useful inference regarding effects of hypoxia on the reproductive axis in exposed zebrafish. Martinovic et al. [8] noted that the majority of brain–pituitary–gonadal-hepatic axis genes differentially expressed between hypoxic and normoxic conditions were down-regulated. Consistent with that, GSEA generally showed enrichment in the control group (7.0 mg O_2/L) compared with the hypoxia treatments. One notable exception to this was the expression of vitellogenesis-related genes in the testis of males exposed to hypoxia. Martinovic et al. noted this as significant up-regulation of testicular vtg1 expression after 4 d

exposure to hypoxia [8]. Analyzing the entire set of vitellogenesis-related genes with GSEA, a significant effect was detected on day 4 as well as for the severe hypoxia treatment (1.0 mg O_2/L) on day 14 (Table 3). Based on GO analysis and evaluation of individual differentially expressed genes, Martinovic et al. also concluded that reorganization of lipid transport was involved in the response to hypoxia [8]. The present GSEA results support a similar conclusion, as gene sets associated with both cholesterol uptake and cholesterol synthesis were significantly impacted (Table S3). These results suggest that the GSEA-based analysis utilizing tissue-specific gene lists and their functionally organized sublists can, at least in some cases, be as effective as the more detailed examination of individual differentially expressed genes in supporting inference from ecotoxicogenomic data.

However, we expected that there would be additional advantages to a GSEA-based approach when comparing among experiments. Rather than the analyses depending on lists of differentially expressed genes, the composition of which can vary widely from one experiment to the next, GSEA considers the expression data for all probes on the array. If the same gene sets are used for GSEA, the domain of genes for comparison is largely equivalent, as reflected in the common structure of the GSEA results matrix for the three EAC experiments examined (Tables 4–6). By considering multiple individual genes as sets, there is a data reduction that facilitates comparison. Significance of the various sets can then be used to focus on those particular subsets of genes that could be considered in greater detail. This can improve both the efficiency of comparison across experiments, and stream-line the analysis of data within an experiment.

The intent of the present work was not to provide a comprehensive analysis of the microarray data from the EAC experiments, but rather to illustrate the application of the graphical model-based gene lists and their utility for GSEA-based comparisons among multiple experiments. Using this approach, we were able to rapidly and effectively compare the profiles of brain–pituitary–gonadal-hepatic axis-related transcriptional changes among three, relatively complicated toxicogenomic experiments. Similarities in the time-dependent effects (or lack thereof) could be discerned. For example, the general lack of significant effects on brain–pituitary–gonadal-hepatic axis-related gene sets compared to control was notable on both day 10 and day 2 of the experiments. One could hypothesize that this may reflect an increase in variability among individual fish as they initially responded to a change in their exposure condition (either initiation of chemical exposure [day 0] or cessation of chemical exposure [day 8.5]) and moved toward a different physiological state. However, based on available apical, real-time PCR, and/or metabolomics results, these time points did not appear consistently more variable than the others [22–24]. Nonetheless, the GSEA-based approach provides an efficient and comparable method to evaluate this observation across at least four additional EAC studies (see Table S1) to clarify whether this is a consistent trend, or simply a coincidence.

Comparing across chemicals, there was an interesting dichotomy between the two chemicals that directly inhibit one or more steroidogenic enzymes (i.e., fadrozole, prochloraz) and the synthetic androgen (17β-trenbolone). Specifically, significant over-expression of the gene set associated with steroid biosynthesis compared to controls was observed for both fadrozole and prochloraz (Tables 4 and 5). This was broadly consistent with the postulated compensatory increase in transcription of key steroidogenesis-related genes based on the results of selected real-time quantitative PCR analyses [22,23]. Steroid biosynthesis-related gene sets were not significantly impacted in the 17β-trenbolone experiment. Some alterations in expression of genes coding for steroidogenic enzymes were detected for 17β-trenbolone, but the responses were much more variable over time and included both significant up- and down-regulation over the time-course of the experiment [24]. This again provides focus for additional comparative investigation.

The GSEA analyses conducted as part of the present work were all based on microarray data for gonad tissue. Some significant effects were detected when the graphical model-related gene lists were combined into a single list that included targets relevant to all four tissues considered (i.e., all_hpgh). However, a greater number of significant impacts could be discerned when the overall list was broken into tissue-specific lists. Obviously, this also provides a finer level of biological resolution to support inference. We also recognize, however, that gene sets developed for other tissues (e.g., liver, brain, pituitary) were also impacted in gonad tissue following exposure to various EACs or hypoxia. Similarly, gene sets associated with female reproductive functions like oocyte growth and ovulation, were significantly impacted in males (testis tissue). Genes were assigned to tissue/sex-specific gene sets based on their well characterized/understood functions in those milieus. Some proteins like activin subunits, follistatin, insulin-dependent growth factors, etc. are known to be expressed and function in a variety of tissues and were thus associated with more than one tissue-specific list. However, transcriptomic analyses have also routinely uncovered differential expression of genes in tissues that would be considered atypical relative to the conventional understanding of their function. Increased expression of vitellogenin isoforms in the testis tissue of zebrafish exposed to hypoxia is one example (Table 3; [8]). Unexpected regulation of various genes and proteins in unconventional tissues can lead to important new avenues of research and an expanded understanding of biological function. For example, observations of expression of gonadotropins and GnRH isoforms and GnRH receptors in ovary tissue have led to a growing body of research on the intraovarian expression and function of those proteins (e.g., [36,37]) that is reshaping our understanding of GnRH and gonadotropin signaling. Our gene sets were designed to be tissue/sex-specific in order to leverage what we currently understand about their function to aid interpretation of toxicogenomic data and provide improved sensitivity and resolution. Nonetheless, including all gene sets for each tissue (or for a different sex) in the analysis is computationally efficient and may lead to discovery of important novel functions and/or responses for even seemingly well-characterized targets.

Overall, GSEA anchored to a graphical systems model should complement other types of microarray data analyses and provide another useful source of biological insight from which to draw inferences regarding effects of chemical stressors. As with transcriptomic data in general, the dynamic and variable nature of the transcriptomic responses still make these data rather difficult to interpret from graphical model-based GSEA results alone, without supporting phenotypic anchors. However, it is hoped that by more tightly coupling gene sets examined with the types of endpoints being measured in the experiment (in our case, indicators of reproductive endocrine function/status like plasma steroid and vitellogenin concentrations), we can enhance our interpretation of these results and more effectively compare among experiments. Due to the significant investment of time and expertise required to construct the graphical model and associated platform-specific gene sets, we feel that open sharing of such tools would benefit the scientific community. The updated graphical model, gene mapping, and associated gene lists (as .gmt files) are provided as supplementary files. We fully encourage other investigators to utilize these resources and/or adapt them for other purposes and it is expected that these tools will continue to evolve with the science.

Conflict of interest

None.

Acknowledgements

This manuscript has been subjected to review by the US Environmental Protection Agency (US EPA), National Health and Environmental Effects Research Laboratory and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does the mention of trade names or commercial products constitute endorsement or recommendation for use. The authors thank Jon Haselman and Rong-lin Wang for helpful comments on the manuscript. Primary funding for this work was provided by the US EPA and the US Army Environmental Quality Research Program (including BAA 11-4838). Additional contributions to this work were supported by grants from the Medical Research Foundation of Oregon [Grant #0634 to K.H.W.], US EPA Science to Achieve Results (STAR) program [Cooperative Agreement #RD-831848-01-0 to N.D.D., K.H.W., E.F.O.], the Division of Environmental and Biomolecular Systems at Oregon Health and Science University [Z.L.], and University of Maryland [E.F.O.].

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.mrgentox.2011.12.016.

References

[1] D.L. Villeneuve, P. Larkin, I. Knoebl, A.L. Miracle, M.D. Kahl, K.M. Jensen, E.A. Makynen, E.J. Durhan, B.J. Carter, N.D. Denslow, G.T. Ankley, A graphical systems model to facilitate hypothesis-driven ecotoxicogenomics research on the teleost brain–pituitary–gonadal axis, Environ. Sci. Technol. 41 (2007) 321–330.

[2] G.T. Ankley, D. Bencic, M. Breen, T.W. Collette, R. Connolly, N.D. Denslow, S. Edwards, D.R. Ekman, K.M. Jensen, J. Lazorchak, D. Martinovic, D.H. Miller, E.J. Perkins, E.F. Orlando, N. Garcia-Reyero, D. Villeneuve, R.L. Wang, K. Watanabe, Endocrine disrupting chemicals in fish: developing exposure indicators and predictive models of effects based on mechanisms of action, Aquat. Toxicol. 92 (2009) 168–178.

[3] R.L. Wang, D. Bencic, D.L. Villeneuve, G.T. Ankley, J. Lazorchak, S. Edwards, A transcriptomics-based biological framework for studying mechanisms of endocrine disruption in small fish species, Aquat. Toxicol. 98 (2010) 230–244.

[4] Y. Zohar, J.A. Munoz-Cueto, A. Elizur, O. Kah, Neuroendocrinology of reproduction in teleost fish, Gen. Comp. Endocrinol. 165 (2010) 438–455.

[5] K. Tsutsui, G.E. Bentley, G. Bedecarrats, T. Osugi, T. Ubuka, L.J. Kriegsfeld, Gonadotropin-inhibitory hormone (GnIH) and its control of central and peripheral reproductive function, Front. Neuroendocrinol. 31 (2010) 284–295.

[6] D.L. Villeneuve, I. Knoebl, P. Larkin, A.L. Miracle, B.J. Carter, N.D. Denslow, G.T. Ankley, Altered gene expression in the brain and liver of female fathead minnows exposed to fadrozole, J. Fish. Biol. 72 (2008) 1–60.

[7] L. Villeneuve, R.L. Wang, D.C. Bencic, A.D. Biales, D. Martinovic, J.M. Lazorchak, G. Toth, G.T. Ankley, Altered gene expression in the brain and ovaries of zebrafish (*Danio rerio*) exposed to the aromatase inhibitor fadrozole: microarray analysis and hypothesis generation, Environ. Toxicol. Chem. 28 (2009) 1767–1782.

[8] D. Martinovic, D.L. Villeneuve, M.D. Kahl, L.S. Blake, J.D. Brodin, G.T. Ankley, Hypoxia alters gene expression in the gonads of zebrafish (*Danio rerio*), Aquat. Toxicol. 95 (2009) 258–272.

[9] S.Y. Skolness, E.J. Durhan, N. Garcia-Reyero, K.M. Jensen, M.D. Kahl, E.A. Makynen, D. Martinovic-Weigelt, E. Perkins, D.L. Villeneuve, G.T. Ankley, Effects of a short-term exposure to the fungicide prochloraz on endocrine function and gene expression in female fathead minnows (*Pimephales promelas*), Aquat. Toxicol. 103 (2011) 170–178.

[10] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 15545–15550.

[11] W. Zhang, Z. Zeng, Y. Zhou, W. Xiong, S. Fan, L. Xiao, D. Huang, Z. Li, D. Li, M. Wu, X. Li, S. Shen, R. Wang, L. Cao, K. Tang, G. Li, Identification of aberrant cell cycle regulation in Epstein-Barr virus-associated nasopharyngeal carcinoma by cDNA microarray and gene set enrichment analysis, Acta Biochim. Biophys. Sin. (Shanghai) 41 (2009) 414–428.

[12] N. Le Novere, A. Finney, M. Hucka, U.S. Bhalla, F. Campagne, J. Collado-Vides, E.J. Crampin, M. Halstead, E. Klipp, P. Mendes, P. Nielsen, H. Sauro, B. Shapiro, J.L. Snoep, H.D. Spence, B.L. Wanner, Minimum information requested in the annotation of biochemical models (MIRIAM), Nat. Biotechnol. 23 (2005) 1509–1515.

[13] H. Kitano, A. Funahashi, Y. Matsuoka, K. Oda, Using process diagrams for the graphical representation of biological networks, Nat. Biotechnol. 23 (2005) 961–966.

[14] N. Le Novere, M. Hucka, H. Mi, S. Moodie, F. Schreiber, A. Sorokin, E. Demir, K. Wegner, M.I. Aladjem, S.M. Wimalaratne, F.T. Bergman, R. Gauges, P. Ghazal, H. Kawaji, L. Li, Y. Matsuoka, A. Villeger, S.E. Boyd, L. Calzone, M. Courtot, U. Dogrusoz, T.C. Freeman, A. Funahashi, S. Ghosh, A. Jouraku, S. Kim, F. Kolpakov, A. Luna, S. Sahle, E. Schmidt, S. Watterson, G. Wu, I. Goryanin, D.B. Kell, C. Sander, H. Sauro, J.L. Snoep, K. Kohn, H. Kitano, The systems biology graphical notation, Nat. Biotechnol. 27 (2009) 735–741.

[15] J. Sprague, L. Bayraktaroglu, D. Clements, T. Conlin, D. Fashena, K. Frazer, M. Haendel, D.G. Howe, P. Mani, S. Ramachandran, K. Schaper, E. Segerdell, P. Song, B. Sprunger, S. Taylor, C.E. Van Slyke, M. Westerfield, The Zebrafish Information Network: the zebrafish model organism database, Nucleic Acids Res. 34 (2006) D581–D585.

[16] D.L. Villeneuve, N. Garcia-Reyero, D. Martinovic, N.D. Mueller, J.E. Cavallin, E.J. Durhan, E.A. Makynen, K.M. Jensen, M.D. Kahl, L.S. Blake, E.J. Perkins, G.T. Ankley II, Effects of a dopamine receptor antagonist on fathead minnow dominance behavior and ovarian gene expression in the fathead minnow and zebrafish, Ecotoxicol. Environ. Saf. 73 (2010) 478–485.

[17] D.L. Villeneuve, N. Garcia-Reyero, D. Martinovic, J.E. Cavallin, N.D. Mueller, L.C. Wehmas, M.D. Kahl, A.L. Linnum, E.J. Perkins, G.T. Ankley, Influence of ovarian stage on transcript profiles in fathead minnow (*Pimephales promelas*) ovary tissue, Aquat. Toxicol. 98 (2010) 354–366.

[18] J.S. Taylor, I. Braasch, T. Frickey, A. Meyer, Y. Van de Peer, Genome duplication, a trait shared by 22,000 species of ray-finned fish, Genome Res. 13 (2003) 382–390.

[19] O. Jaillon, J.M. Aury, F. Brunet, J.L. Petit, N. Stange-Thomann, E. Mauceli, L. Bouneau, C. Fischer, C. Ozouf-Costaz, A. Bernot, S. Nicaud, D. Jaffe, S. Fisher, G. Lutfalla, C. Dossat, B. Segurens, C. Dasilva, M. Salanoubat, M. Levy, N. Boudet, S. Castellano, V. Anthouard, C. Jubin, V. Castelli, M. Katinka, B. Vacherie, C. Biemont, Z. Skalli, L. Cattolico, J. Poulain, V. De Berardinis, C. Cruaud, S. Duprat, P. Brottier, J.P. Coutanceau, J. Gouzy, G. Parra, G. Lardier, C. Chapple, K.J. McKernan, P. McEwan, S. Bosak, M. Kellis, J.N. Volff, R. Guigo, M.C. Zody, J. Mesirov, K. Lindblad-Toh, B. Birren, C. Nusbaum, D. Kahn, M. Robinson-Rechavi, V. Laudet, V. Schachter, F. Quetier, W. Saurin, C. Scarpelli, P. Wincker, E.S. Lander, J. Weissenbach, H. Roest Crollius, Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype, Nature 431 (2004) 946–957.

[20] M.F. Smith, W.A. Ricke, L.J. Bakke, M.P. Dow, G.W. Smith, Ovarian tissue remodeling: role of matrix metalloproteinases and their inhibitors, Mol. Cell. Endocrinol. 191 (2002) 45–56.

[21] E.J. Perkins, K. Chipman, S. Edwards, T. Habib, F. Falciani, R. Taylor, G. Van Aggelen, C. Vulpe, P. Antczak, A. Loguinov, Reverse engineering adverse outcome pathways, Environ. Toxicol. Chem. 30 (2010) 22–38.

[22] D.L. Villeneuve, N.D. Mueller, D. Martinovic, E.A. Makynen, M.D. Kahl, K.M. Jensen, E.J. Durhan, J.E. Cavallin, D. Bencic, G.T. Ankley, Direct effects, compensation, and recovery in female fathead minnows exposed to a model aromatase inhibitor, Environ. Health Perspect. 117 (2009) 624–631.

[23] G.T. Ankley, D. Bencic, J.E. Cavallin, K.M. Jensen, M.D. Kahl, E.A. Makynen, D. Martinovic, N.D. Mueller, L.C. Wehmas, D.L. Villeneuve, Dynamic nature of alterations in the endocrine system of fathead minnows exposed to the fungicide prochloraz, Toxicol. Sci. 112 (2009) 344–353.

[24] D.R. Ekman, D.L. Villeneuve, Q. Teng, K.J. Ralston-Hooper, D. Martinovic-Weigelt, M.D. Kahl, K.M. Jensen, E.J. Durhan, E.A. Makynen, G.T. Ankley, T.W. Collette, Use of gene expression, biochemical and metabolite profiles to enhance exposure and effects assessment of the model androgen 17beta-trenbolone in fish, Environ. Toxicol. Chem. 30 (2011) 319–329.

[25] K.V. Ballman, D.E. Grill, A.L. Oberg, T.M. Therneau, Faster cyclic loess: normalizing RNA arrays via linear models, Bioinformatics 20 (2004) 2778–2786.

[26] D. Huang, B. Sherman, R. Lempicki, Systematic, integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2009) 44–57.

[27] U. Alon, An Introduction to Systems Biology: Design Principles of Biological Circuits, Chapman and Hall/CRC, New York, 2007.

[28] P.M. Wassarman, Zona pellucida glycoproteins, J. Biol. Chem. 283 (2008) 24285–24289.

[29] E. Lubzens, G. Young, J. Bobe, J. Cerda, Oogenesis in teleosts: how fish eggs are formed, Gen. Comp. Endocrinol. 165 (2009) 367–389.

[30] H. Picton, D. Briggs, R. Gosden, The molecular basis of oocyte growth and development, Mol. Cell. Endocrinol. 145 (1998) 27–37.

[31] Y. Zhang, M.D. Sheets, Analyses of zebrafish and Xenopus oocyte maturation reveal conserved and diverged features of translational regulation of maternal cyclin B1 mRNA, BMC Dev. Biol. 9 (2009) 7.

[32] M.K. Kang, S.J. Han, Post-transcriptional and post-translational regulation during mouse oocyte maturation, BMB Rep. 44 (2011) 147–157.

[33] O. Carnevali, C. Cionna, L. Tosti, E. Lubzens, F. Maradonna, Role of cathepsins in ovarian follicle growth and maturation, Gen. Comp. Endocrinol. 146 (2006) 195–203.

[34] S.I. Lane, H.Y. Chang, P.C. Jennings, K.T. Jones, The Aurora kinase inhibitor ZM447439 accelerates first meiosis in mouse oocytes by overriding the spindle assembly checkpoint, Reproduction 140 (2010) 521–530.
[35] R.J. Prill, D. Marbach, J. Saez-Rodriguez, P.K. Sorger, L.G. Alexopoulos, X. Xue, N.D. Clarke, G. Altan-Bonnet, G. Stolovitzky, Towards a rigorous assessment of systems biology models: the DREAM3 challenges, PLoS One 5 (2010) e9202.
[36] T.T. Wong, Y. Zohar, Novel expression of gonadotropin subunit genes in oocytes of the gilthead seabream (*Sparus aurata*), Endocrinology 145 (2004) 5210–5220.
[37] C. Metallinou, B. Asimakopoulos, A. Schroer, N. Nikolettos, Gonadotropin-releasing hormone in the ovary, Reprod. Sci. 14 (2007) 737–749.