

Network-based Analysis of Time Series RNA-Seq Gene Expression Data by Integrating the Interactome and Gene Ontology Information

Yuji Zhang¹

¹Department of Epidemiology and Public Health

University of Maryland School of Medicine

Baltimore, Maryland, USA

yuzhang@som.umaryland.edu

Abstract— Monitoring the changes in gene expression patterns over time provides the distinct possibility of unraveling the mechanistic drivers characterizing cellular responses. Such time series gene expression data allow us to broadly “watch” the dynamics of the system. However, one challenge in the analysis of time series data is to establish and characterize the interplay between genes that are activated, deactivated or sustained in the context of a biological process or functional category. To address such challenges, novel algorithms are required to improve the interpretation of these data by integrating multi-source prior functional evidence. In this paper, we introduced a novel network-based approach to extract functional knowledge from time-dependent biological processes at a system level using time series mRNA deep sequencing data. First, a list of differentially expressed genes (DEGs) at each time point was identified. Second, GO terms that are enriched in each DEG list were identified. Third, the significance of interactions between DEGs in these GO terms at consecutive time points was measured. Finally, the significant interactions between DEGs in different GO terms were used to construct the interaction networks among GO terms between two consecutive time points, called GO networks. The proposed method was applied to investigate $1\alpha, 25(\text{OH})_2\text{D}_3$ -altered mechanisms in zebrafish embryo development. GO networks were constructed over 4 consecutive time points. Results suggest that biological processes such as cartilage development and one-carbon compound metabolic process are temporally regulated by $1\alpha, 25(\text{OH})_2\text{D}_3$. Such discoveries could not have been identified with canonical gene set enrichment analyses. These results demonstrate that the proposed approach can provide insight on the molecular mechanisms taking place in vertebrate embryo development upon treatment with $1\alpha, 25(\text{OH})_2\text{D}_3$. Our approach enables the monitoring of biological processes that can serve as a basis for generating new testable hypotheses. Such network-based integration approach can be easily extended to any temporal- or condition-dependent genomic data analyses.

Keywords — Network Analysis; Time Series; Functional Network; Gene Ontology; RNA Sequencing

I. INTRODUCTION (HEADING 1)

Biological systems are highly dynamic and responsive to the external environment. The gene expression in these systems is a temporal process. Different genes are required to play

different functional roles under different conditions. This is highly regulated by a complex regulatory system of diverse molecular interactions, such as protein-protein interactions (PPIs), protein-DNA interactions (PDIs), and metabolic signaling pathways [1]. Taking a snapshot of the gene expression profile in a biological system (e.g., cell cycle system and development) under a certain condition can reveal some of the genes that are specially expressed under this condition. However, to investigate how all the genes are regulated in the context of a biological system, and to determine the interaction relationships between these genes, it is necessary to measure the gene expression profile in a time series manner [2]. This can also provide the distinct possibility of unraveling the mechanistic drivers characterizing cellular responses [3]. Time series gene expression data have been widely applied to study a wide range of biological systems, including cell cycle [4], genetic interaction and knockouts [5, 6], and development [7]. Despite their unique features, many computational challenges still remain in analysing such gene expression profiles. For instance, it is difficult to study the relationships among differentially expressed genes (DEGs) at each time point in a case-control time series experiment, due to large number of DEGs and limited time points available. To address such challenges, algorithms are required that are specifically designed to improve the interpretability of these data by integrating multi-source prior biological evidence.

Molecular interactions such as PPIs and PDIs are essential for a wide range of cellular processes and form a network of astonishing complexity. Until recently, our knowledge of such complex networks was rather limited. The emergence of high-throughput technologies has given us possibilities to systematically survey and study the underlying biological system. The molecular interaction maps have been built in model organisms (e.g., *S.cerevisiae* [8], *D.melanogaster* [9] and *C.elegans* [10]), as well as in higher vertebrate organisms (e.g., zebrafish [11], mouse [12] and human [13]). Evidently, the generated interaction maps offer us a rich resource for systematic studies of molecular networks and complement other types of biological data. However, current interaction databases include a large amount of false positive and false negative interactions due to the unreliability of interaction mapping technologies available. In addition, these molecular

interactions are static. There is little direct information available on the temporal dynamics of these molecular interactions. To understand time-dependent biological processes at network level, molecular networks need to be considered as temporal and spatial rather than static information flow between molecules [14]. Recently, attempts have been made in integrating different types of biological data with molecular network interactions to reveal the dynamics of molecular networks [15]. However, only a few studies have investigated the dynamics of the molecular network interactions in time course gene expression data with limited success. For instance, Tang et al. [16] proposed to reconstruct time course protein interaction networks (TC-PINs) by incorporating time series gene expression into PPI networks. The functional modules from TC-PINs were enriched in related gene ontology (GO) biological processes than those from static PPI networks. However, the causal relationship between TC-PINs across time points could not be inferred. Such causal relationships are crucial to understand the underlying regulated biological processes in a time-dependent and context-specific manner. A propagation of such interactions from gene level to biological process/pathway level (e.g., gene ontology information) will help us identify the altered biological processes during the time in which these gene expressions are examined.

The gene ontology (GO) Consortium [17] has developed three separate ontologies—molecular function (MF), biological process (BP) and cellular component (CC) - to describe the attributes of gene products. Several studies have demonstrated that the molecular interactions and GO provide substantially congruent yet subtle different view of biological systems [18]. The hypothesis is that the interaction between any two proteins/genes indicates a general likelihood that these two proteins are functionally coupled or involved in the same biological process. Identifying enriched interactions between any two GO terms based on molecular interactions between genes assigned to these two GO terms are more statistically reliable: interactions reflect statistically enriched temporal connections between multiple genes of one GO term and multiple genes of another. However, this could not tell the temporal directionality in these connections. By incorporating time series gene expression data, the causal relations can be inferred in this GO network by highlighting information flow between GO biological processes enriched in DEGs at consecutive time points.

In this paper, we developed a novel network-based computational approach to study causal relationships between DEGs at consecutive time points in a case-control time series experiment. To overcome the limitation that the intervals of time series experiments usually would not fit the time scale of functional communications between most genes and the statistical power from only several time points would be too low for robust analysis, we constructed networks of GO biological process terms connected by significant interactions between DEGs on sequential time points. This enables us to understand the biological processes at GO scale, in which relations between nodes (representing GO terms) are more statistically stable. This is more statistically significant and biologically meaningful compared to single co-expressed links.

The detail of the proposed approach is presented in Figure 1. The proposed method was applied to time series mRNA-Seq data set to determine the influence of $1\alpha,25(\text{OH})_2\text{D}_3$ treatment on gene expression patterns in zebrafish embryo development and the causal relationship between DEGs at consecutive time points. The resulting networks suggest that well-studied as well as novel molecular mechanisms are regulated by $1\alpha,25(\text{OH})_2\text{D}_3$ treatment.

II. MATERIAL AND METHODS

A. mRNA-Seq gene expression data

The mRNA-Seq profiling in four biological replicate samples of $1\alpha,25(\text{OH})_2\text{D}_3$ - or ethanol-treated zebrafish, 2, 4, 6 and 7 days post-fertilization (hpf) was obtained by the Illumina HiSeq 2000 platform. The generated 50-bp FASTQ sequence reads were aligned to both the latest Zebrafish genome assembly (zv9) and our in-house exon junction database using BWA [19]. The aligned sequence tags were counted for each annotated genes/exons using custom scripts based on the UCSC genome binning approach [20]. A total of 14267 genes were annotated using RefSeq database and the raw read counts for genes were generated for further downstream analyses.

B. Zebrafish molecular interaction network

The zebrafish molecular interaction network was downloaded from FunCoup database (<http://FunCoup.sbc.su.se/>). In total, there are 1,999,529 interactions between 13033 proteins in the zebrafish interactome downloaded on January 3rd, 2012.

C. Gene ontology annotation in zebrafish

The gene ontology annotation was downloaded from the original website (<http://www.geneontology.org/>) on January 20th, 2012. In this paper, we used the biological process terms only since our goal is to identify the $1\alpha,25(\text{OH})_2\text{D}_3$ -altered mechanisms.

D. Differential gene expression analysis

For differential gene expression analysis between conditions, we eliminated genes without any reads across all samples. We used DESeq package in R to test for differential expression for all the remaining genes [21]. We conservatively accounted for multiple testing, employing a Bonferroni correction, yielding an adjusted p-value for differential expression for each gene. A strict adjusted P value cut-off of 0.01 was used to select significant DEGs

E. Construction of time-dependent GO-GO networks

A network of GO terms was generalized from the network of DEGs at different developmental stages in zebrafish embryos. At GO scale, relations between nodes (representing GO terms) are more statistically stable. Links reflect statistically enriched temporal connections between multiple genes in one specific GO term and multiple genes in another one. Thus, this GO-GO network highlights information flow between GO biological processes affected by $1\alpha,25(\text{OH})_2\text{D}_3$ at different developmental stages. If there were a significant

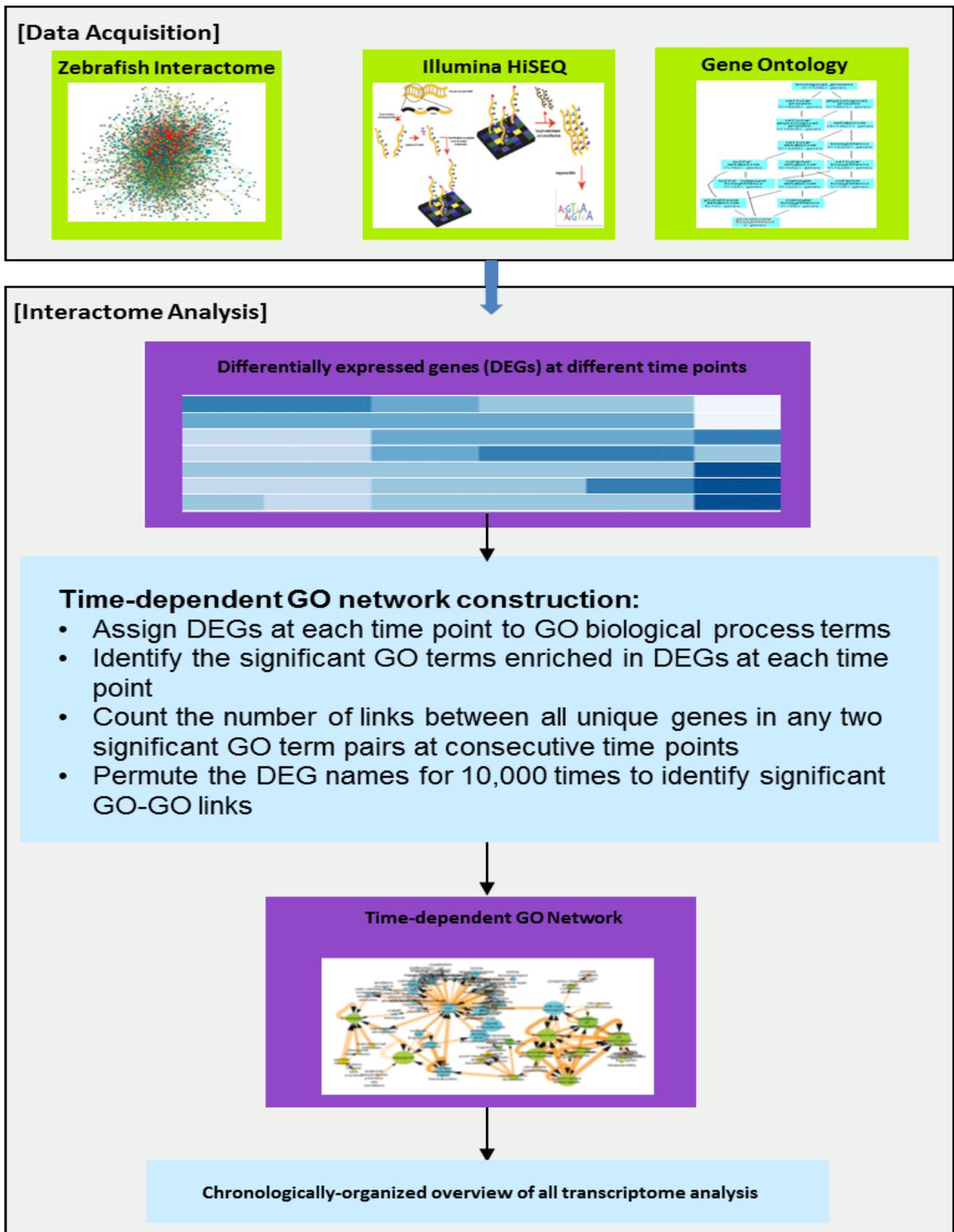


Figure 1. Overview of the proposed approach.

number of genes in GO term X first altered at one time point interacting with genes in GO term Y first altered on the next time point, we hypothesize that a causative relation exists $X \rightarrow Y$. We limited the network to only enriched GO-GO connections, i.e. one with significant more gene-gene interactions (given both genes were $1\alpha,25(\text{OH})2\text{D}_3$ -altered) than expected by chance. This allows us to focus on the major tendencies of propagation of $1\alpha,25(\text{OH})2\text{D}_3$ treatment and organismal response to it. Compared to the individual category enrichment, this approach yielded a much richer analysis for interpretation. The detailed reconstruction step is as follows:

- 1) For any two GO terms, a link was counted if any two DEGs in these two GO terms were connected in the original FunCoup network;
- 2) The GO-GO links were classified into time-dependent patterns according to the days when the gene were differentially expressed for the first time:
 - a. Day 2 \rightarrow Day 4: one gene was differentially expressed on Day 2, while the other on Day 4;
 - b. Day 4 \rightarrow Day 6: similar definition as in (a);
 - c. Day 6 \rightarrow Day 7: similar definition as in (a).
- 3) The GO For each candidate GO-GO network link, its statistical significance was evaluated by the permutation test, i.e. gene names were randomized in the FunCoup network for 10,000 times. The links between GO terms with P value less than 0.01 were considered statistically significant.

Enriched GO-GO links were kept in the GO-GO network, i.e. ones with P value less than 0.01. The network was visualized in the Cytoscape tool [22].

F. GoMiner analysis

The gene level Gene ontology enrichment analysis was performed using GoMiner [23] on the DEGs that were identified at each time point.

III. RESULTS

In this section, we present:

- 1) A description of generation and initial characterization of the mRNA-seq dataset obtained from zebrafish embryos altered by $1\alpha,25(\text{OH})2\text{D}_3$ treatment;
- 2) An overview the interactome-based analysis that we proposed;
- 3) Chronological-organized analysis of the transcriptome changes and interactome dynamics altered by $1\alpha,25(\text{OH})2\text{D}_3$ treatment during early zebrafish development.

Figure 1 illustrates the overview of the proposed analysis workflow.

A. Characterization of mRNA-seq dataset during zebrafish embryo development

Genome-wide transcriptional profiling were performed using Illumina HiSeq sequencing technique for four replicate cDNA libraries of $1\alpha,25(\text{OH})2\text{D}_3$ - or vehicle-treated zebrafish, 48, 96, 144, and 168 hours post fertilization (hpf) as described in our previous publication [24]. Overall, the RNA-seq data obtained from 32 independent zebrafish RNA libraries had comparable number of total reads [24]. These reads were mapped to the latest zebrafish genome assembly (zv9) from the UCSC website (<http://genome.ucsc.edu/>). The refFlat annotation file from the University of California Santa Clara (UCSC) Table Browser was used to generate raw reads mapped to each annotated gene in the annotation file. The genes altered by $1\alpha,25(\text{OH})2\text{D}_3$ treatment at each time point were identified using the negative binomial model as describe in [21]. A list of altered genes identified along with the days on which they were differentially expressed is presented in Supplemental Table 1. We also carried out the gene ontology (GO) enrichment analysis using the GOMiner tool [23]. However, due to the limited number of DEGs identified at each time point and the limitation associated with the Fisher's Exact Test, the results of these analyses could not provide much indication of the biological processes being moduled in response to $1\alpha, 25(\text{OH})2\text{D}_3$ treatment. To more efficiently derive biological insights from the genome-wide transcriptomic response to the treatment, we proposed a network-based analysis in the following sections.

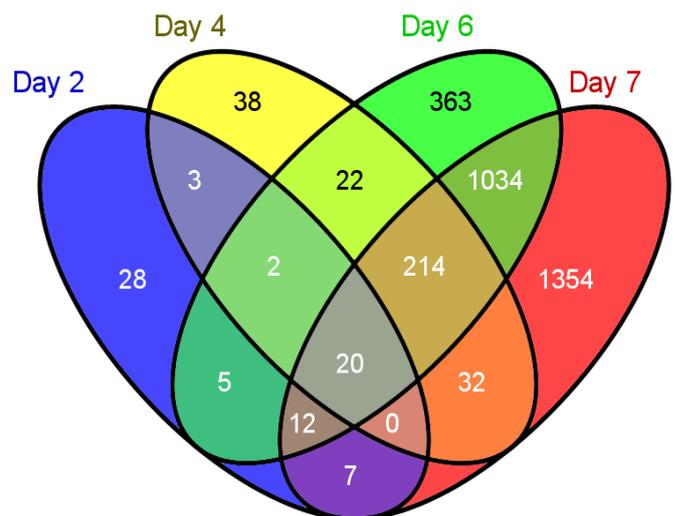
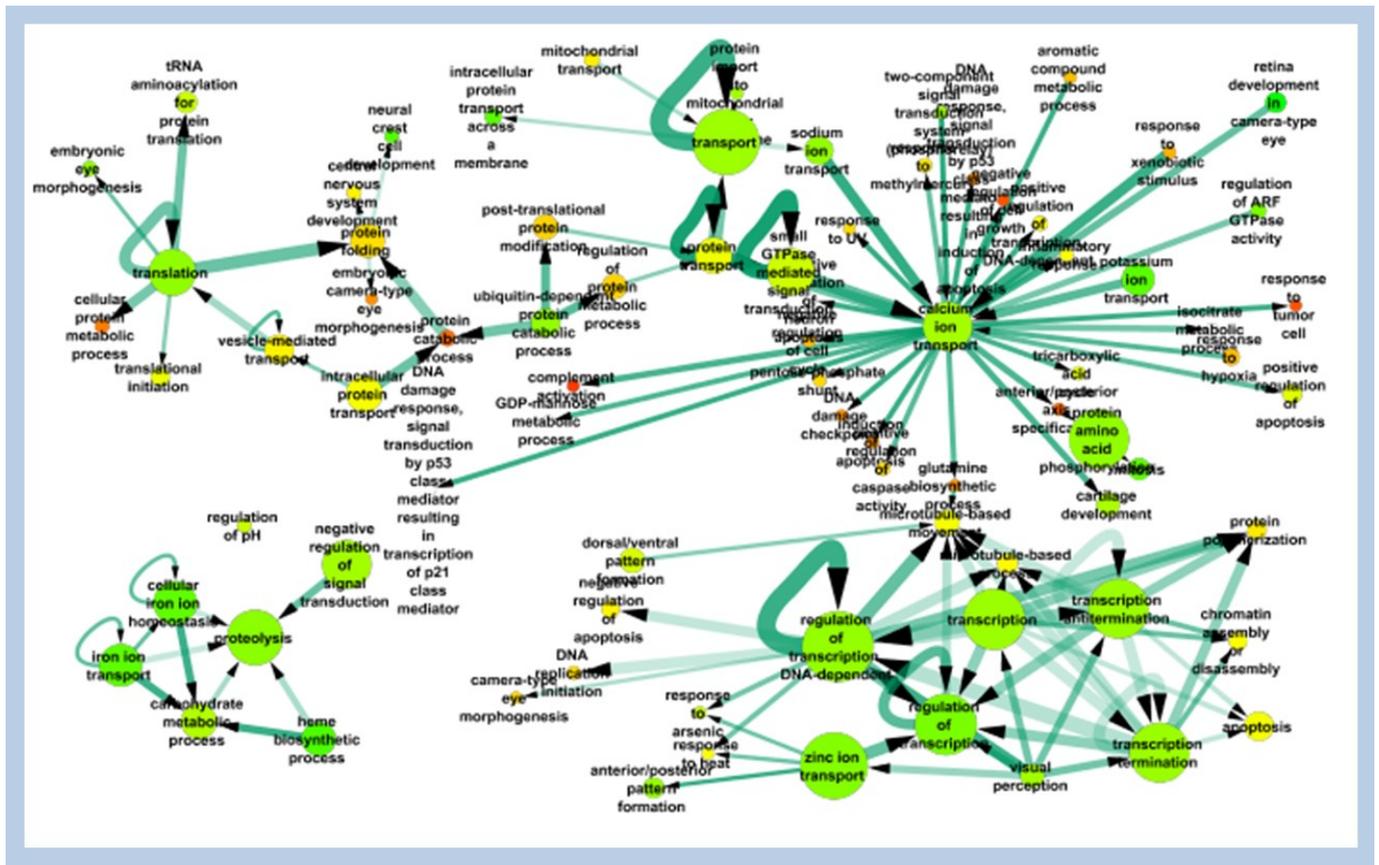


Figure 2. Venn diagram showing the overlap of DEGs at different developmental stages. Genes were grouped based on the day(s) they were differentially expressed. In four studied developmental stages 3134 genes were defined as differentially expressed during at least one stage.

B. Interactome-based analysis of differentially expressed genes during zebrafish development

We overlaid the DEGs onto the zebrafish functional interactome from the FunCoup database [11]. The DEGs were overlaid on their corresponding nodes in the interactome, and related functional interactions between genes were extracted



(C)

Figure 3. Network of GO terms enriched in $1\alpha,25(\text{OH})_2\text{D}_3$ -altered genes between consecutive days. (A) GO network of day 2 - > 4; (B) GO network of day 4 - > 6; (C) GO network of day 6 - > 7. Color represent the fraction of the gene in that node that were regulated by $1\alpha,25(\text{OH})_2\text{D}_3$ on any day (green is low and red is high). Edge thickness and opacity represent the number of gene-gene links between two GO terms and significance score ($-\log_{10}(\text{P value})$), respectively.

and reconstructed the $1\alpha,25(\text{OH})_2\text{D}_3$ specific interactome. Many network interactions connect the few genes altered on day 2 and many altered on later days. We found that there was a statistically significant enrichment in links between genes that were $1\alpha,25(\text{OH})_2\text{D}_3$ -altered earlier and genes regulated later in the course of experiment. This suggested that treatment affected signals were propagated along network routes from the initially affected genes (on day 2) towards network regions that were perturbed later.

Specifically, 3134 genes were up- or down-regulated by $1\alpha,25(\text{OH})_2\text{D}_3$ on at least one of the four days in the experiment (adjusted P value less than 0.01). On day 2, only 77 genes were changed. 331 genes on day 4, 1672 genes on day 6, and 2673 genes on day 7 differentially expressed in response to $1\alpha,25(\text{OH})_2\text{D}_3$ treatment (Figure 2). The property of these DEGs was investigated in the context of FunCoup network. The average degree of DEGs is significantly higher than non-DEGs (14.9 versus 5.8, the P value of one-way ANOVA less than 10⁻⁶). This indicated that DEGs were more enriched in hub genes (genes with higher node degree). This can partially explain the initially altered genes on days can

pass the changes to more interacted genes on later days through the network links/interactions.

To gain a better perspective on what this temporal pattern in enriched connections between $1\alpha,25(\text{OH})_2\text{D}_3$ -altered genes might mean, we analyzed the GO categories associated with the connected nodes in the context of interactome.

C. Network propagation analysis of differentially expressed genes during zebrafish development

The FunCoup network links among these genes can indicate a general likelihood how they are functionally related, but don't highlight the temporal directionality in these connections. Causal relations can be suggested by examining temporal changes, i.e., if information associated with gene A at time point t helps to predict the state of gene B at time point (t+1), then a causal relation A->B might be inferred [25, 26]. However, traditional network inference approaches could not identify such temporal regulatory relationship due to limited time points available. The statistical power from only four time points would be too low for robust analysis. To gain a better

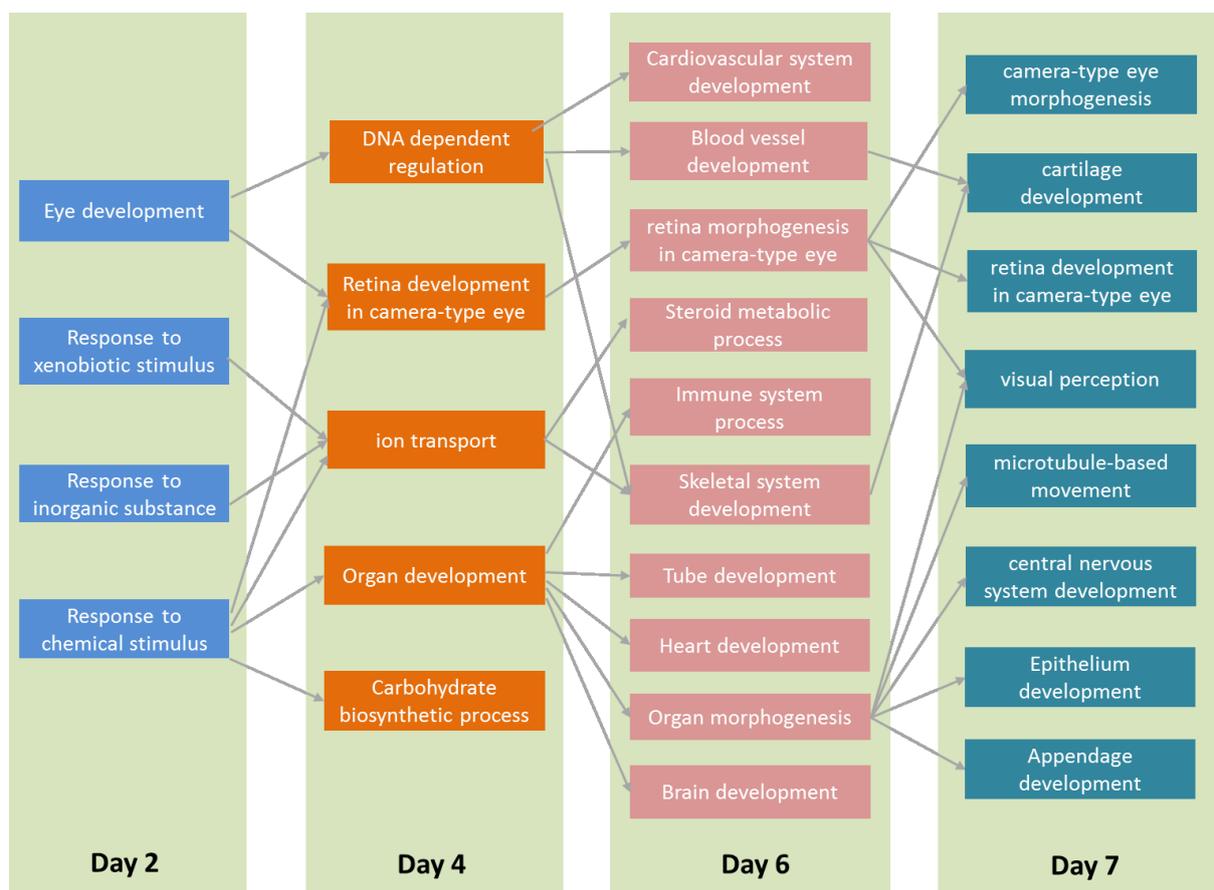


Figure 4. Network of GO terms enriched in $1\alpha,25(\text{OH})_2\text{D}_3$ -altered genes between consecutive days. (A) GO network of day 2 - > 4; (B) GO network of day 4 - > 6; (C) GO network of day 6 - > 7.

perspective on the temporal pattern among $1\alpha, 25(\text{OH})_2\text{D}_3$ -altered genes, we generalized a network of GO terms connected by the links between these DEGs on consecutive time points. At this broader scale, relations between nodes (GO biological processes) are statistically reliable: links reflect statistically enriched temporal connections between multiple genes of one node with multiple genes of another. Thus, this GO-GO network highlights flow between GO biological processes altered by $1\alpha, 25(\text{OH})_2\text{D}_3$ on different days.

$1\alpha, 25(\text{OH})_2\text{D}_3$ -altered genes in individual gene-gene interactions in FunCoup interactome were labeled with days when these genes were detected as differentially expressed. We were particularly interested in identifying the links in which one gene was altered earlier than the other. Thus, if there were a significant number of genes in GO category X altered on day d interacting with gene in GO category Y altered on day (d+1), we hypothesize that a causative relation X -> Y. Limiting the output to only enriched GO-GO connections allowed us to focus on the major changes of propagation of $1\alpha, 25(\text{OH})_2\text{D}_3$ and organismal response to it. Compared to the individual category enrichment approach such as GOMiner, our approach yielded a much richer analysis for interpretation of time series changes unique to time series gene expression data. The Figures 2-4 presented day-to-day enriched interactions at GO biological process

level. We provided a chronological interpretation on these findings below.

D. Chronological analysis of the interaction network altered by $1\alpha, 25(\text{OH})_2\text{D}_3$ at gene ontology level

The approach described above enabled flexible and deep monitoring of $1\alpha, 25(\text{OH})_2\text{D}_3$ altered changes in the transcriptome at GO level in the context of functional interactome. To show time-dependent information flow in embryonic development altered by $1\alpha,25(\text{OH})_2\text{D}_3$ treatment, GO networks of enriched GO-GO interactions were reconstructed.

Day 2 to Day 4 transition: the network of GO terms between DEGs on day 2 and 4 suggested a cascade initiated by changes in xenobiotic metabolism genes and leading to genes involved in ion transport and transcription regulation (Figure 3(A)). The “eye development” category is enriched on as early as day 2, indicating that eye development was changed by $1\alpha, 25(\text{OH})_2\text{D}_3$ treatment. The eye development of zebrafish starts as early as 28 hpf [27]. The vitamin D receptor has been shown to express in various tissues and organs including retina. This confirms the finding using our proposed approach.

Day 4 to Day 6 transition: day 4 was marked by the most significant increase of linkage from transcription

factors altered on day 4 towards others altered later. The most central node on day 4 is organ development, which became connected to multiple biological processes, such as cardiovascular system development, blood vessel development, immune system process, heart development, brain development, tube development, and others. This observation suggests that vitamin D treatment can alter biological processes involved in the development of many organs. One network of GO terms between day 4 and 6 is presented in Figure 3(B).

Day 6 to Day 7 transition: the organ morphogenesis was identified as a central node in the GO network of day 6 \rightarrow 7, connecting to multiple biological processes, such as nervous system development, circulatory system development, vasculature development, epithelium development, retina development in camera type eye, and many embryonic development terms including cartilage development and neuron generation. One network of GO terms between day 6 and 7 is presented in Figure 3(C).

To better interpretate the causal relationships between enriched GO categories on consecutive days, we presented a GO level information flow by combining the GO-GO networks across all four days (Figure 4). The interactome was altered in the regions scattered in the interactome to many biological processes that are clustered together in the interactome. This suggests that the effect of $1\alpha, 25(\text{OH})_2\text{D}_3$ treatment can be as early as 48 hpf in early zebrafish development.

IV. DISCUSSIONS

In this work, we have developed a network-based computational approach that analyzes time series mRNA-seq gene expression profiles in the context of molecular interactome and GO information to reveal temporal transcriptional changes altered by $1\alpha, 25(\text{OH})_2\text{D}_3$ in zebrafish embryo development. This enabled us to review the progression of $1\alpha, 25(\text{OH})_2\text{D}_3$ -induced changes in gene expression and the network structure itself in early zebrafish development. The efficiency of our analysis of $1\alpha, 25(\text{OH})_2\text{D}_3$ -altered global gene expression was enhanced by the interactome approach, as the network-based analysis approach were superior to their single-gene approach in terms of both statistical power and biological interpretability. A variety of interesting biological hypotheses were derived from our analysis. The significant biological processes include iron metabolism, neuronal and retinal development, and many organ development related pathways. Our approach is useful for discovering candidate biological processes that can serve as a basis for generating new testable hypotheses. Such network-based integration approach can be extended to any temporal- or condition-dependent genomic data analyses. Other types of interaction or ontology data can also be incorporated into this approach.

ACKNOWLEDGMENT

This work was supported by the Startup grant from the University of Maryland Greenbaum Cancer Center.

REFERENCES

- [1] J. Chen, B. J. Aronow, and A. G. Jegga, "Disease candidate gene identification and prioritization using protein interaction networks," *BMC Bioinformatics*, vol. 10, pp. 73, 2009.
- [2] Z. Bar-Joseph, "Analyzing time series gene expression data," *Bioinformatics*, vol. 20, no. 16, pp. 2493-503, Nov 1, 2004.
- [3] I. P. Androulakis, E. Yang, and R. R. Almon, "Analysis of time-series gene expression data: methods, challenges, and opportunities," *Annual review of biomedical engineering*, vol. 9, pp. 205-28, 2007.
- [4] I. Simon, J. Barnett, N. Hannett, C. T. Harbison, N. J. Rinaldi, T. L. Volkert, J. J. Wyrick, J. Zeitlinger, D. K. Gifford, T. S. Jaakkola, and R. A. Young, "Serial regulation of transcriptional regulators in the yeast cell cycle," *Cell*, vol. 106, no. 6, pp. 697-708, Sep 21, 2001.
- [5] T. R. Hughes, M. J. Marton, A. R. Jones, C. J. Roberts, R. Stoughton, C. D. Armour, H. A. Bennett, E. Coffey, H. Dai, Y. D. He, M. J. Kidd, A. M. King, M. R. Meyer, D. Slade, P. Y. Lum, S. B. Stepanians, D. D. Shoemaker, D. Gachotte, K. Chakraburty, J. Simon, M. Bard, and S. H. Friend, "Functional discovery via a compendium of expression profiles," *Cell*, vol. 102, no. 1, pp. 109-26, Jul 7, 2000.
- [6] A. P. Gasch, P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein, and P. O. Brown, "Genomic expression programs in the response of yeast cells to environmental changes," *Molecular biology of the cell*, vol. 11, no. 12, pp. 4241-57, Dec, 2000.
- [7] N. B. Ivanova, J. T. Dimos, C. Schaniel, J. A. Hackney, K. A. Moore, and I. R. Lemischka, "A stem cell molecular signature," *Science*, vol. 298, no. 5593, pp. 601-4, Oct 18, 2002.
- [8] J. L. Koh, H. Ding, M. Costanzo, A. Baryshnikova, K. Toufighi, G. D. Bader, C. L. Myers, B. J. Andrews, and C. Boone, "DRYGIN: a database of quantitative genetic interaction networks in yeast," *Nucleic acids research*, vol. 38, no. Database issue, pp. D502-7, Jan, 2010.
- [9] H. Nickla, "Interaction between pteridine synthesis and riboflavin accumulation in *Drosophila melanogaster*," *Canadian journal of genetics and cytology. Journal canadien de genetique et de cytologie*, vol. 14, no. 1, pp. 105-11, Mar, 1972.
- [10] M. I. Barrasa, P. Vaglio, F. Cavasino, L. Jacotot, and A. J. Walhout, "EDGEdb: a transcription factor-DNA interaction database for the analysis of *C. elegans* differential gene expression," *BMC Genomics*, vol. 8, pp. 21, 2007.
- [11] A. Alexeyenko, T. Schmitt, A. Tjarnberg, D. Guala, O. Frings, and E. L. Sonnhammer, "Comparative interactomics with Funcoup 2.0," *Nucleic acids research*, vol. 40, no. Database issue, pp. D821-8, Jan, 2012.
- [12] M. Shionyu, A. Yamaguchi, K. Shinoda, K. Takahashi, and M. Go, "AS-ALPS: a database for analyzing the effects of alternative splicing on protein structure, interaction and network in human and mouse," *Nucleic acids research*, vol. 37, no. Database issue, pp. D305-9, Jan, 2009.
- [13] D. G. Jamieson, M. Gerner, F. Sarafraz, G. Nenadic, and D. L. Robertson, "Towards semi-automated curation: using text mining to recreate the HIV-1, human protein interaction database," *Database : the journal of biological databases and curation*, vol. 2012, pp. bas023, 2012.
- [14] K. W. Kohn, "Molecular interaction map of the mammalian cell cycle control and DNA repair systems," *Molecular biology of the cell*, vol. 10, no. 8, pp. 2703-34, Aug, 1999.
- [15] F. Anthony Romero, N. B. Hastings, R. Moningga, Z. Guo, M. Wang, J. Di Salvo, Y. Lei, D. Trusca, Q. Deng, V. Tong, J. L. Terebetski, R. G. Ball, and F. Ujjainwalla, "The discovery of potent antagonists of NPBWR1 (GPR7)," *Bioorganic & medicinal chemistry letters*, vol. 22, no. 2, pp. 1014-8, Jan 15, 2012.
- [16] X. Tang, J. Wang, B. Liu, M. Li, G. Chen, and Y. Pan, "A comparison of the functional modules identified from time course and static PPI network data," *BMC Bioinformatics*, vol. 12, pp. 339, 2011.
- [17] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G.

Sherlock, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium," *Nature genetics*, vol. 25, no. 1, pp. 25-9, May, 2000.

- [18] A. Marco, and I. Marin, "Interactome and Gene Ontology provide congruent yet subtly different views of a eukaryotic cell," *BMC systems biology*, vol. 3, pp. 69, 2009.
- [19] H. Li, and R. Durbin, "Fast and accurate short read alignment with Burrows-Wheeler transform," *Bioinformatics*, vol. 25, no. 14, pp. 1754-60, Jul 15, 2009.
- [20] W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, and D. Haussler, "The human genome browser at UCSC," *Genome research*, vol. 12, no. 6, pp. 996-1006, Jun, 2002.
- [21] S. Anders, and W. Huber, "Differential expression analysis for sequence count data," *Genome Biol*, vol. 11, no. 10, pp. R106, 2010.
- [22] M. E. Smoot, K. Ono, J. Ruscheinski, P. L. Wang, and T. Ideker, "Cytoscape 2.8: new features for data integration and network visualization," *Bioinformatics*, vol. 27, no. 3, pp. 431-2, Feb 1, 2011.
- [23] B. R. Zeeberg, H. Qin, S. Narasimhan, M. Sunshine, H. Cao, D. W. Kane, M. Reimers, R. M. Stephens, D. Bryant, S. K. Burt, E. Elnekave, D. M. Hari, T. A. Wynn, C. Cunningham-Rundles, D. M. Stewart, D. Nelson, and J. N. Weinstein, "High-Throughput GoMiner, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of Common Variable Immune Deficiency (CVID)," *BMC Bioinformatics*, vol. 6, pp. 168, 2005.
- [24] T. A. Craig, Y. Zhang, M. S. McNulty, S. Middha, H. Ketha, R. J. Singh, A. T. Magis, C. Funk, N. D. Price, S. C. Ekker, and R. Kumar, "Research resource: whole transcriptome RNA sequencing detects multiple 1alpha,25-dihydroxyvitamin D(3)-sensitive metabolic pathways in developing zebrafish," *Molecular endocrinology*, vol. 26, no. 9, pp. 1630-42, Sep, 2012.
- [25] Y. Zhang, J. Xuan, B. G. de los Reyes, R. Clarke, and H. W. Resson, "Network motif-based identification of transcription factor-target gene relationships by integrating multi-source biological data," *BMC Bioinformatics*, vol. 9, pp. 203, 2008.
- [26] Y. Zhang, J. Xuan, B. G. de los Reyes, R. Clarke, and H. W. Resson, "Reconstruction of gene regulatory modules in cancer cell cycle by multi-source data integration," *PLoS One*, vol. 5, no. 4, pp. e10268, 2010.
- [27] S. S. Easter, Jr., and G. N. Nicola, "The development of vision in the zebrafish (*Danio rerio*)," *Developmental biology*, vol. 180, no. 2, pp. 646-63, Dec 15, 1996.