



Investigating the Combinatory Effects of Biological Networks on Gene Co-expression

Cheng Zhang¹, Sunjae Lee², Adil Mardinoglu^{2,3*} and Qiang Hua^{1,4*}

¹ State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai, China,

² Science for Life Laboratory, KTH-Royal Institute of Technology, Stockholm, Sweden, ³ Department of Biology and Biological Engineering, Chalmers University of Technology, Göteborg, Sweden, ⁴ Shanghai Collaborative Innovation Center for Biomanufacturing Technology, Shanghai, China

OPEN ACCESS

Edited by:

Xiaogang Wu,
Institute for Systems Biology, USA

Reviewed by:

Tunahan Cakir,
Gebze Technical University, Turkey
Byung-Kwan Cho,
Korea Advanced Institute of Science
and Technology, South Korea

*Correspondence:

Adil Mardinoglu
adilm@scilifelab.se;
Qiang Hua
qhua@ecust.edu.cn

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Physiology

Received: 11 February 2016

Accepted: 15 April 2016

Published: 02 May 2016

Citation:

Zhang C, Lee S, Mardinoglu A and
Hua Q (2016) Investigating the
Combinatory Effects of Biological
Networks on Gene Co-expression.
Front. Physiol. 7:160.
doi: 10.3389/fphys.2016.00160

Co-expressed genes often share similar functions, and gene co-expression networks have been widely used in studying the functionality of gene modules. Previous analysis indicated that genes are more likely to be co-expressed if they are either regulated by the same transcription factors, forming protein complexes or sharing similar topological properties in protein-protein interaction networks. Here, we reconstructed transcriptional regulatory and protein-protein networks for *Saccharomyces cerevisiae* using well-established databases, and we evaluated their co-expression activities using publically available gene expression data. Based on our network-dependent analysis, we found that genes that were co-regulated in the transcription regulatory networks and shared similar neighbors in the protein-protein networks were more likely to be co-expressed. Moreover, their biological functions were closely related.

Keywords: *Saccharomyces cerevisiae*, co-expression, co-regulation, transcriptional regulatory network, protein-protein interaction network

INTRODUCTION

Recent advances involving high-throughput measurements have facilitated the accumulation of extensive gene expression profile sets, and the gene expression patterns evident across multiple conditions have allowed a systems-level understanding of biology. Specifically, gene co-expression has been proposed as a useful methodology for uncovering gene functions (Stuart et al., 2003), and this approach prevails in many areas of systems biology (Carter et al., 2004; Aoki et al., 2007; Miller et al., 2008, 2010; Presson et al., 2008; Xulvi-Brunet and Li, 2010; Liao et al., 2011; MacDonald et al., 2015; Peake et al., 2015).

If the mRNA expression levels of two genes follow a similar pattern through multiple gene expression measurements, they are treated as co-expressed genes. Gene co-expression indicates that the transcription rates of these genes are modulated by similar molecules, and this underlying modularity has allowed us to uncover the functions of co-expressed genes. Transcription factors (TFs) bind to the promoters of target genes and this kind of binding modulates their expression at the mRNA level. Under different physiological conditions, two or more TFs can bind to the same target gene (Balaji et al., 2006), and this phenomenon is known as combinatorial regulation. Through combinatorial binding, a small set of TFs can dynamically modulate gene expression in response to diverse physiological conditions.

In the past decade, a large amount of high-throughput gene expression data has been made publically available, and this information has facilitated the development of global co-expression analyses. Consequently, various methods and algorithms have been developed for generating gene co-expression networks (Bar-Joseph et al., 2003; Tong et al., 2004; Zhang and Horvath, 2005; Luo et al., 2007; Langfelder and Horvath, 2008; Ruan et al., 2010; Savage et al., 2010; Roy et al., 2014).

Although co-expression analysis had been widely used, knowledge of the biological factors and topological characteristics that contribute to high gene co-expression remains limited. Currently, the knowledge of regulatory networks is increasing (Teixeira et al., 2006, 2014; Monteiro et al., 2008; Abdulrehman et al., 2011; Hughes and de Boer, 2013), but the combinatorial regulation by TFs makes it difficult to understand the contributing factors in co-expression. Through hundreds of expression measurements, a genome-scale study demonstrated that genes bound by similar TFs are highly co-expressed (Allocco et al., 2004). In addition, the expression profiles of genes coding for interacting protein pairs are more highly correlated than random pairs (Ge et al., 2001), but later genome-scale studies indicated that this correlation was significantly reduced and suggested that only those ones forming the same protein complex seem to be co-expressed (Bhardwaj and Lu, 2005; Xulvi-Brunet and Li, 2010). Furthermore, the topological properties of the protein coding genes in protein-protein interaction (PPI) networks may also contribute to the co-expression of genes (Han et al., 2004).

Although efforts have been made to explore the relationship between co-expression and biological networks, to the best of our knowledge, no study had exclusively analyzed the joint effects of different biological networks [e.g., transcriptional regulatory (TR) and protein-protein interactions (PPI) networks] on gene co-expression. In this study, we reconstructed TR and PPI networks based on well-established databases for biological interactions in *Saccharomyces cerevisiae*. We investigated the relationships between gene co-expression in each network as well as in multiple biological networks. Moreover, we constructed co-regulated biological networks and comprehensively studied their effects on gene co-expression.

MATERIALS AND METHODS

Microarray Data

Gene expression data for *S. cerevisiae* were retrieved from the Gene Expression Omnibus (GEO) database. All microarray data for *S. cerevisiae* using GPL2529 as platform and published before January 22th, 2014 in GEO database were selected. And to eliminate replicate size-based biases, only one replicate (replicate 1) were included in the analysis when multiple replicates are available. As a result, 1057 microarray datasets were selected. The microarray data were normalized with the Robust Multiarray Average (RMA; Bolstad et al., 2003; Irizarry et al., 2003a,b) and treated with the affy R package (Gautier et al., 2004). Open reading frame (ORF) ids were converted to gene ids. If a gene was mapped with more than one ORF, the mean value was used in

our analysis. Consequently, expression profiles for 5657 genes in 1057 different experimental conditions were obtained (Table S1). Finally, the pair-wise gene co-expression data were obtained for 5657 genes by calculating the Pearson correlation coefficients using MATLAB R2015a.

Reconstruction of Regulatory Networks

TR interactions for *S. cerevisiae* were retrieved from the YEASTRACT database (Teixeira et al., 2006, 2014; Monteiro et al., 2008; Abdulrehman et al., 2011). Two evidence types were presented in YEASTRACT: regulation with DNA binding evidence and regulation with expression evidence (i.e., expression evidence from TF knock-out or over-expression experiments). In this study, we treated these two regulation types independently. Consequently, we reconstructed two networks: a regulatory network with only the DNA binding evidence (regardless of the expression evidence), hereafter referred to as Bnet, and a regulatory network with only expression evidence (regardless of the binding evidence), hereafter referred to as Enet (Figure S1). Regulation data pertaining to genes that are not included in the microarray data were eliminated from the reconstructed networks in this study.

Reconstruction of a Protein-Protein Interaction Network

BioGRID is a database that includes comprehensive information about PPIs from diverse organisms (Stark et al., 2006; Chatr-aryamontri et al., 2013, 2015). All available PPI information for *S. cerevisiae* was retrieved from BioGRID, and a PPI network was reconstructed, hereafter referred to as Pnet. Similarly, interactions with genes that are not included in the microarray data were eliminated from the reconstructed Pnet.

Construction of “Co-Regulated” Networks

In this study, our use of the term “co-regulation” was based on the “regulator” similarities of gene pairs in biological networks. Since there’s no conventional definition for “regulators” in PPI network, we defined them as the first upstream neighbor proteins that were directly connected by PPIs as starting node of each interactions so they are comparable with regulators in TR networks. Here, “co-regulation” was quantified by calculating the similarity of the “regulators” in each network. We constructed a binary association matrix (i.e., adjacency matrix) for Bnet, Enet and Pnet, and quantified their gene-gene co-regulation similarities by calculating their Pearson correlation coefficients (r) [Spearman correlation was also tested and the results were almost identical (Figure S4)]. r values were calculated for each target gene pair based on their TF similarities in Bnet and Enet. In Pnet, the first upstream neighbors were used to examine similarities instead of TFs. The top 1 to 5% of co-regulated target-target interactions for each network were selected in Bnet, Enet and Pnet, and they were defined as co-regulation networks, namely BCRnet, ECRnet, and PCRnet, respectively. Self-interactions were excluded from the analysis during the co-regulated gene pair selection since their correlation should be one invariantly. The process of constructing co-regulation networks was exhibited for a toy network case in Figure 1.

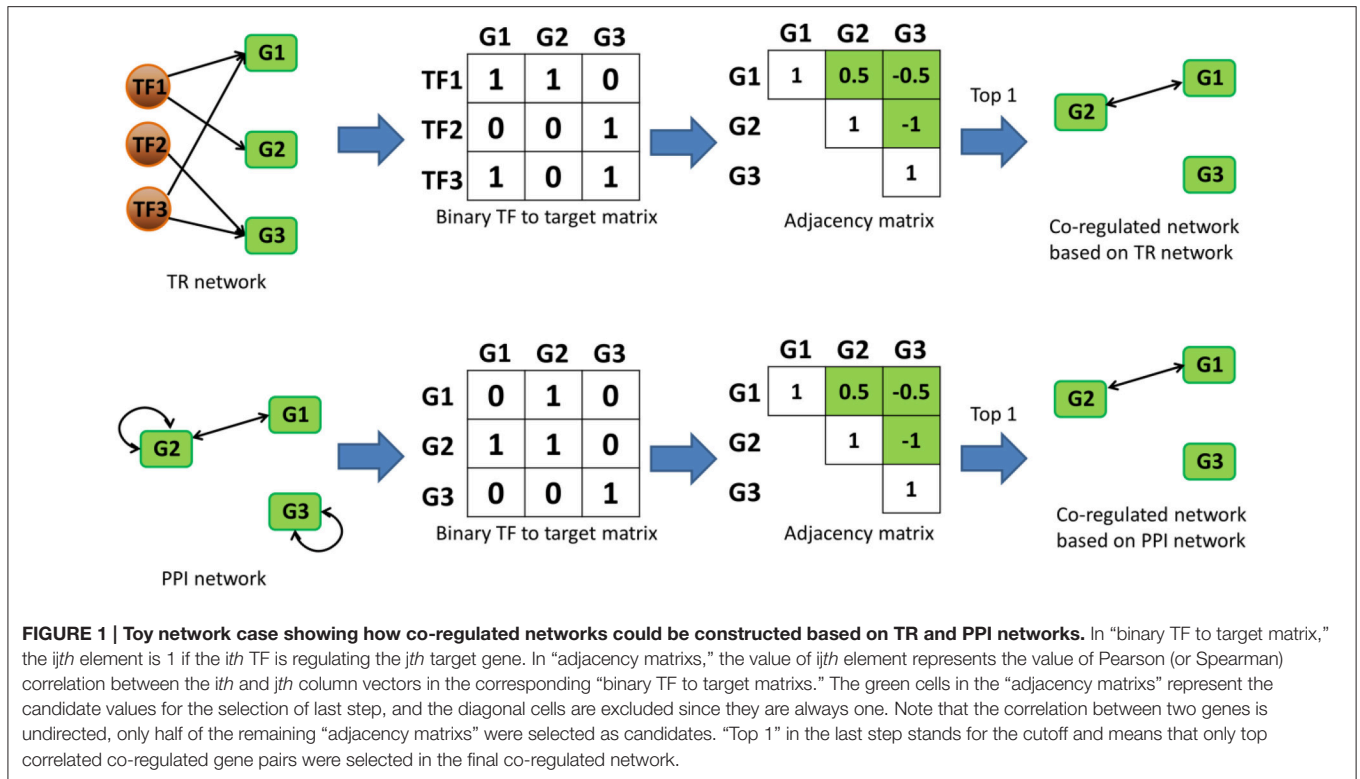


FIGURE 1 | Toy network case showing how co-regulated networks could be constructed based on TR and PPI networks. In “binary TF to target matrix,” the ij th element is 1 if the i th TF is regulating the j th target gene. In “adjacency matrices,” the value of ij th element represents the value of Pearson (or Spearman) correlation between the i th and j th column vectors in the corresponding “binary TF to target matrices.” The green cells in the “adjacency matrices” represent the candidate values for the selection of last step, and the diagonal cells are excluded since they are always one. Note that the correlation between two genes is undirected, only half of the remaining “adjacency matrices” were selected as candidates. “Top 1” in the last step stands for the cutoff and means that only top correlated co-regulated gene pairs were selected in the final co-regulated network.

Calculation of Average Co-Expression r and Sensitivity Analysis

Throughout this study, average co-expression r (ACER) was used to quantify the network co-expression. For the calculation of ACER, all self-interactions were excluded to eliminate bias because co-expression and co-regulation measurements from self-interactions were invariably scored as one (or 100%). All ACER values presented in this study were calculated based on the data from all 1057 microarrays. For sensitivity analysis, the ACER values were reevaluated by calculating the them for 100 randomly selected conditions from the 1057 microarray data 1000 times.

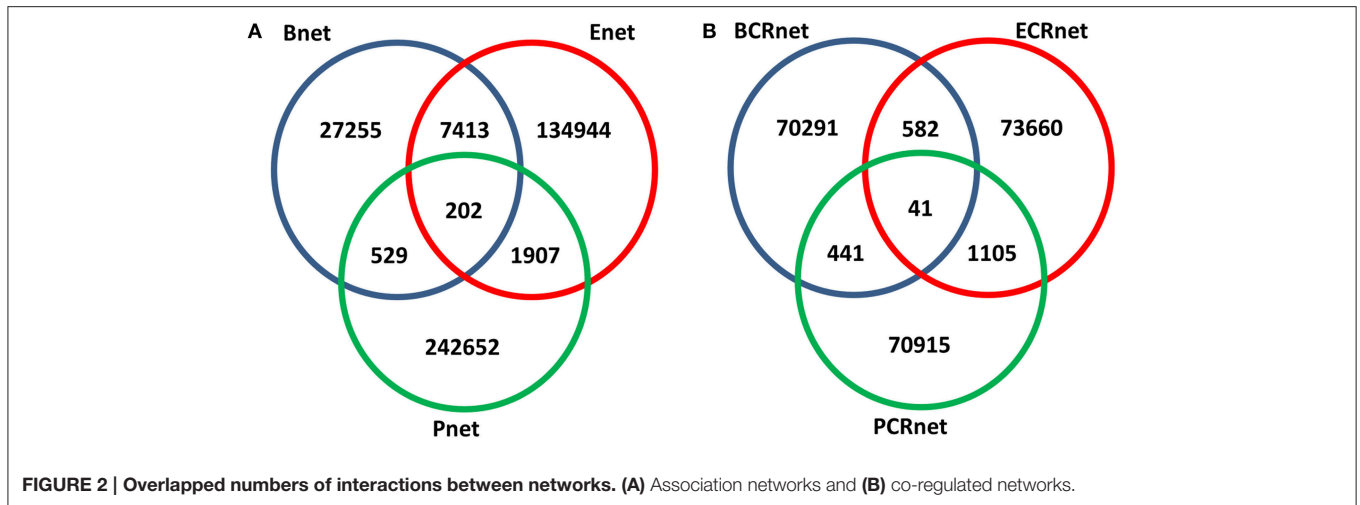
RESULTS

Reconstruction of Biological Networks and Their Co-Expression

Experiments examining the physical binding properties of TFs and those conducting TF-perturbed expression profiling have different characteristics, and combining those experiments together to infer regulatory interactions has been proposed (Blais and Dynlacht, 2005; Yang et al., 2010). Notably, “expression evidence” in the YEASTRACT database was collected from experiments where the TF was perturbed (e.g., knocked out) after which its targets were identified by scanning genes with significantly changed expression properties. Thus, we favored treating the regulatory interaction information from YEASTRACT separately.

We reconstructed three biological association networks with different biological backgrounds, namely Bnet, Enet, and Pnet, for *S. cerevisiae* based on the YEASTRACT and BioGRID databases (Table S2). Bnet was established from TF-binding experiments, such as chromatin immunoprecipitation (ChIP)-chip, whereas Enet was established using expression evidence. Pnet was formulated from PPIs. There were 5345, 5492, and 5392 genes in Bnet, Enet, and Pnet, respectively, and the number of interactions involved in Bnet, Enet, and Pnet were 35399, 144466, and 245078, respectively. The number of TFs in Bnet and Enet are 169 and 292, respectively. All three networks encompassed the majority of the genes (~95%) included in the microarray data, and their average connectivities were of the same magnitude (a ~7-fold difference between Bnet and Pnet), indicating that the network sizes were comparable to each other.

We evaluated the overlaps between the interactions within these three networks and found that the intersections between these three networks were relatively few (Figure 2A). This result suggested that these three networks with different biological backgrounds have their own characteristics. Interestingly, the overlap between Bnet and Enet was relatively small (21.5% for Bnet and 5.3% for Enet) despite the fact that they were both annotated as TR networks and had both originated from the same database. Even if the overlapped TFs and targets were considered, the overlap between them would remain small (Figure S2). This finding implied that the physical binding of TFs to target genes (Bnet) and the summarized effects of specific TF perturbation to expression changes of target genes (Enet) implicated different regulatory events as we proposed. Based on



the different biological backgrounds of Bnet/Enet and Pnet and their small overlaps (less than 2%), we assumed that the three networks were different from each other and embedded with different biological information throughout this study.

To investigate whether the gene pairs involved in the various biological interactions (regulatory interaction or PPI) were more likely to be co-expressed, we calculated the ACER values of all gene pairs for the three reconstructed networks separately and compared them to the reference ACER (Figure 3). The ACER values for Bnet, Enet, and Pnet were calculated as 0.0841, 0.0568, and 0.1392, respectively. Compared to the reference ACER, which was 0.0450, the ACER increases were statistically significant ($P < 0.001$), but their increases were small. The ACER was higher for Pnet compared with the other two networks, consistent with previous results demonstrating that genes encoding proteins with certain PPI types were more likely to be co-expressed (Jansen et al., 2002). Moreover, when we evaluated the overlapped associations, the ACER values were further increased compared with the associations from separated networks (except for the intersection between Enet and Pnet whose ACER values were slightly decreased compared to Pnet). However, these increased ACER values were also statistically significant ($P < 0.001$) but only slightly increased. These results suggested that although relevant, the gene involvement in both the TR and PPI networks did not necessarily contribute to high co-expression.

Construction of Co-Regulated Networks and Their Co-Expression

Because the involvement of various biological interactions did not contribute to high gene co-expression, we hypothesized that high co-expression may result from the topological similarities of genes in biological networks. We assumed that if two genes shared similar “regulators” (both in the TR and PPI networks), they tended to be highly co-expressed. To test our premise, we constructed three co-regulated networks (CRnets), including BCRnet, ECRnet, and PCRnet, by selecting the top 5% of co-regulated gene pairs from Bnet, Enet, and Pnet, respectively (Table S3). The resulting BCRnet, ECRnet, and

PCRnet contained 4520, 4673, and 4541 genes and included 71355, 75388, and 72502 gene-gene interactions, respectively. The overview of the intersections among BCRnet, ECRnet, and PCRnet is presented in Figure 2B. Notably, we generated three CRnets using identical cutoffs, and the resulting CRnets had similar network sizes and connectivities. Despite the similar sizes, the overlaps between the CRnets were small, indicating that the co-regulation networks from Bnet, Enet, and Pnet were different and independent from each other.

Next, we calculated the ACER for each CRnet and evaluated the ACER of the overlapped interactions among the three networks (Figure 4). Interestingly, we found that the CRnets were generally more likely to be co-expressed compared with their original networks (except for BCRnet, which was slightly less co-expressed). Additionally, significant ACER increases were noted when the overlapped gene pairs were examined. Intriguingly, when the intersection of all three CRnets was selected, the gene pair ACER increased to 0.7012, which strongly supported our hypothesis. Furthermore, by changing the cutoffs for CRnet generation, we found that as the cutoff became stricter, the ACER value increased accordingly. And when the strictest cutoff was selected, the r values of the only two gene pairs screened were 0.9012 and 0.8967 (Figure 5). To test the robustness of the positive association between these high correlations and topological similarities, we also performed a sensitivity analysis (see Methods section) for the highly co-expressed gene pairs observed at the intersection of all three CRnets. The results indicated that the calculated ACERs for the overlapped gene pairs of all three CRnets were highly conserved (Figures S3, S4). These results highlighted the consistency of the observed high correlation between co-expression and co-regulation patterns and demonstrated the collaborative outcomes among these biological networks in terms of similarity.

Gene Ontology (GO) Term Enrichment Analysis for the Co-Regulated Gene Pairs

To elucidate the underlying biological impact of these co-regulated gene pairs with high similarities, we checked the

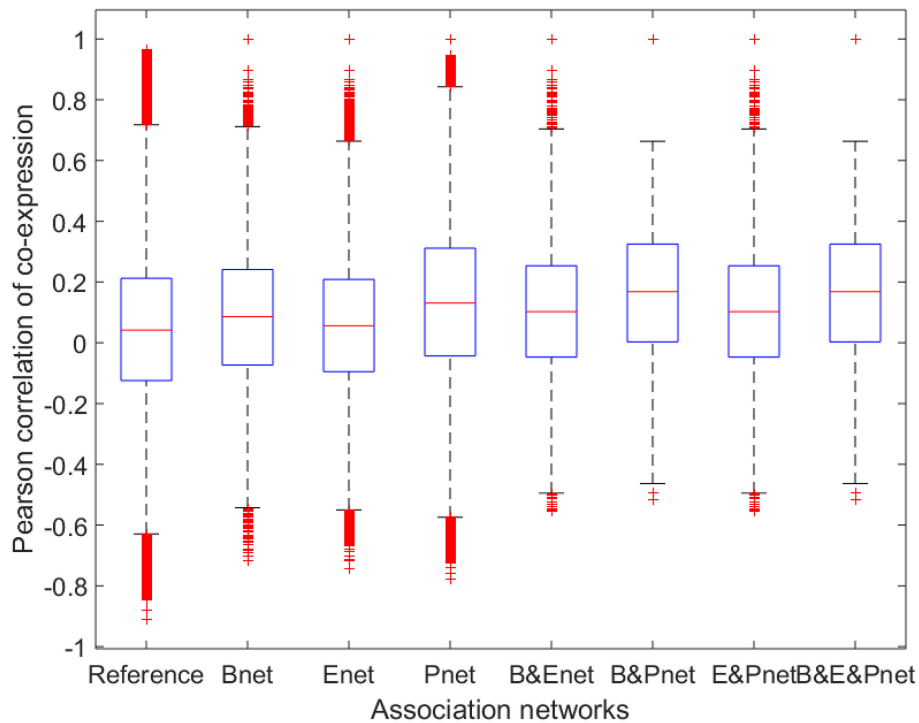


FIGURE 3 | Boxplot for Pearson correlations of co-expression for all gene pairs involved in reference and association networks. B&Enet, B&Pnet, and E&Pnet represent the overlaps between Bnet and Enet, Bnet and Pnet, and Enet and Pnet, respectively. B&E&Pnet represents the overlap among Bnet, Enet, and Pnet.

41 gene pairs shared in all three CRnets (top 5% **Table S4**). We first examined whether these 41 gene pairs were involved in known biological interactions from Bnet, Enet, and Pnet. Consequently, only 9 of the 41 gene pairs were presented in the interactions of Pnet and none were involved in biological interactions from Bnet or Enet. This finding indicated that we identified 32 genetic associations that were highly co-expressed without previous knowledge of their biological interactions.

Next, we performed a GO enrichment analysis ($P < 0.01$) using the *Saccharomyces* genome database (Cherry et al., 2012) to identify biological processes that were related to co-regulated gene pairs. We found that most gene pairs were either involved in the same process, sharing the same function, or localized to the same cellular component (**Table S5**). Additionally, these identified co-regulated gene pairs were highly enriched within the GO terms (**Figure 6**). Interestingly, many of the enriched terms, such as the structural constituents of ribosome-, methionine-adenosyltransferase- and nucleic acid-binding, were related to the regulation of transcription and translation. This finding indicated that the transcriptional and translational processes in yeast were strongly and comprehensively co-regulated. Recent studies have reported that the translational process is crucial to cell physiology, and the ribosome quantity is a key factor that affects proteome allocation (Basan et al., 2015; Hui et al., 2015). Thus, our result is consistent with these studies because it is axiomatic that vital biological

processes are highly co-regulated and co-expressed in the evolutionary context. In conclusion, we demonstrated that the highly co-regulated gene pairs identified here are consistently co-expressed, and they share important biologically significant functions.

DISCUSSION

In this study, we systematically evaluated the individual and combinatory effects of different biological networks on gene co-expression. A comprehensive and unbiased analysis was conducted based on the TR and PPI networks reconstructed here. We demonstrated that co-regulation in biological networks is more relevant to gene co-expression than biological associations. Additionally, we proposed a new method to evaluate the combinatory effects of different co-regulation networks on gene co-expression and found that two genes are highly co-expressed when they are co-regulated in both the TR and PPI networks. Moreover, these co-regulated genes were functionally relevant and involved in vital biological processes. Small-scale studies have demonstrated that co-regulation of gene pairs in the TR network is a key factor contributing to high gene co-expression (Allocco et al., 2004), but a large-scale investigation of co-regulation effects based on topologies provided by regulatory and PPI networks was still lacking. Therefore, our study is an important starting point to study the PPI effect on co-expression and

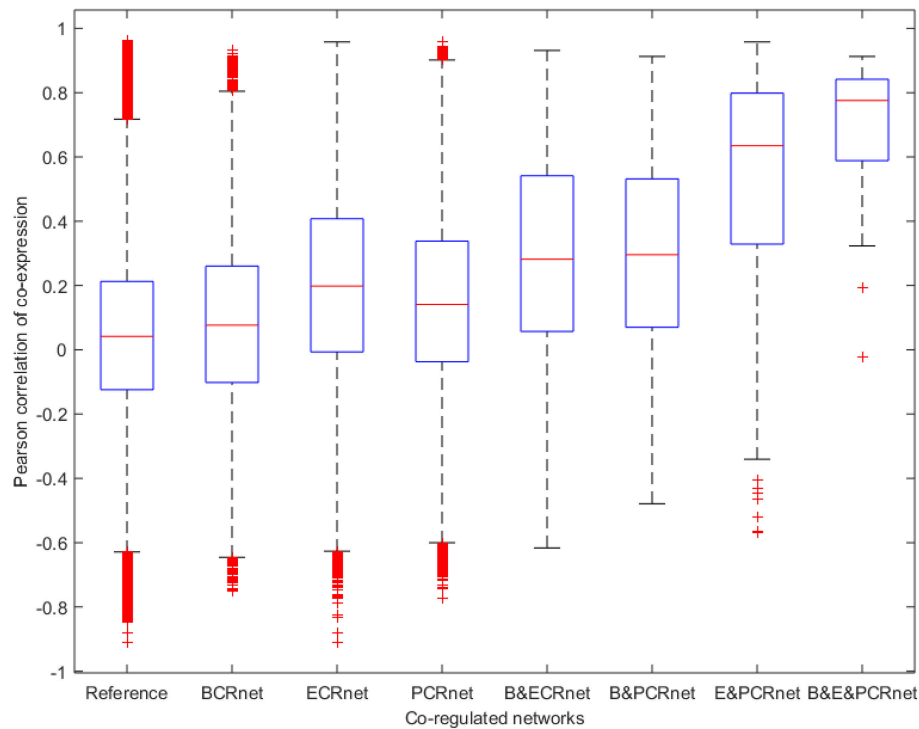


FIGURE 4 | Boxplot for Pearson correlations of co-expression for all gene pairs involved in reference and co-regulated networks. B&ECRnet, B&PCRnet, and E&PCRnet represent the overlaps between BCRnet and ECRnet, BCRnet and PCRnet, and ECRnet and PCRnet, respectively. B&E&PCRnet represents the overlap among BCRnet, ECRnet and PCRnet.

to explore collaborative events between the TR and PPI networks.

It's worth mentioning that the previous study reported a great correlation among genes co-regulated in TR networks (greater than 0.84 when they share more than 50% of their regulators) which is much higher than in our study. However, in our analysis, we also found the co-regulation in TR network is relevant to co-expression (significant but not dramatic), which is actually agreed with the main conclusion of the previous study. And we noted that the previous study only evaluated the co-regulation effect of 2284 genes which are around 40% of our gene set. In addition, they used an outdated version of the TR network, indicating smaller and biased datasets. These would together lead to a much higher correlation probably because of bias. Therefore, our study appeared to be more systematic and robust compared to the previous study.

Our analysis strongly suggests that combinatory effects do exist among biological networks with respect to co-expression, and the genes that are simultaneously co-regulated among the biological networks play important roles in regulating the mRNA expression levels of genes. Interestingly, we found that topological similarities in the PPI network, rather than the interactions themselves, played important roles in modulating gene expression levels. One plausible explanation may be that proteins interacting with the same proteins are functionally related and share similar signaling pathways. Previous genetic

epistatic studies have revealed that dysregulated expressions of interacting protein pairs of protein complexes and pathways showed detrimental effects to cell survivals and thus their coding genes might be more likely co-expressed for better genetic fitness (Kelley and Ideker, 2005; Collins et al., 2007). Another possible explanation is that similar PPI interactions of genes may result in similar folding process and/or other post-transcriptional modification of the protein coding genes. And these modification processes may modulate the level of signaling molecules that would normally influence gene regulation.

In our study, we distinguished the regulatory networks of different biological backgrounds separately because we assumed that regulatory interactions inferred from physical binding (i.e., regulations with binding evidence) and genetic perturbation experiments (i.e., regulations with expression evidence) revealed different regulatory events. This assumption is supported by a recent review which discussed that the transcription of genes depends on the combinatory effect of all its binding TFs, and many of them have only a minor effect. While expression evidence is more likely to be the major factor that regulating transcription of genes (Spivakov, 2014). Thus, separating these different networks would help us to make the embedded regulatory messages more apparent. Additionally, when we used merged regulatory networks and tested the shared gene pairs between the co-regulated network of the merged one and the PCRnet, we found that the shared gene pair correlations

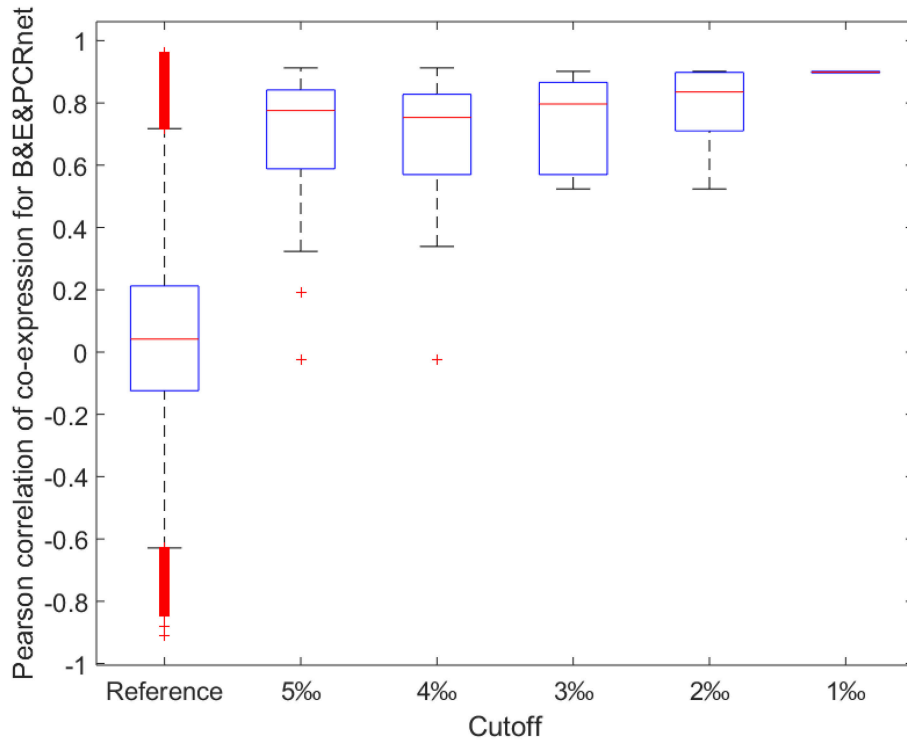


FIGURE 5 | Boxplot for Pearson correlations of co-expression for all gene pairs involved in reference and shared gene pairs among all three CRnets with different cutoffs. B&E&PCRnet represent the overlapped gene pairs among BCRnet, ECRnet, and PCRnet.

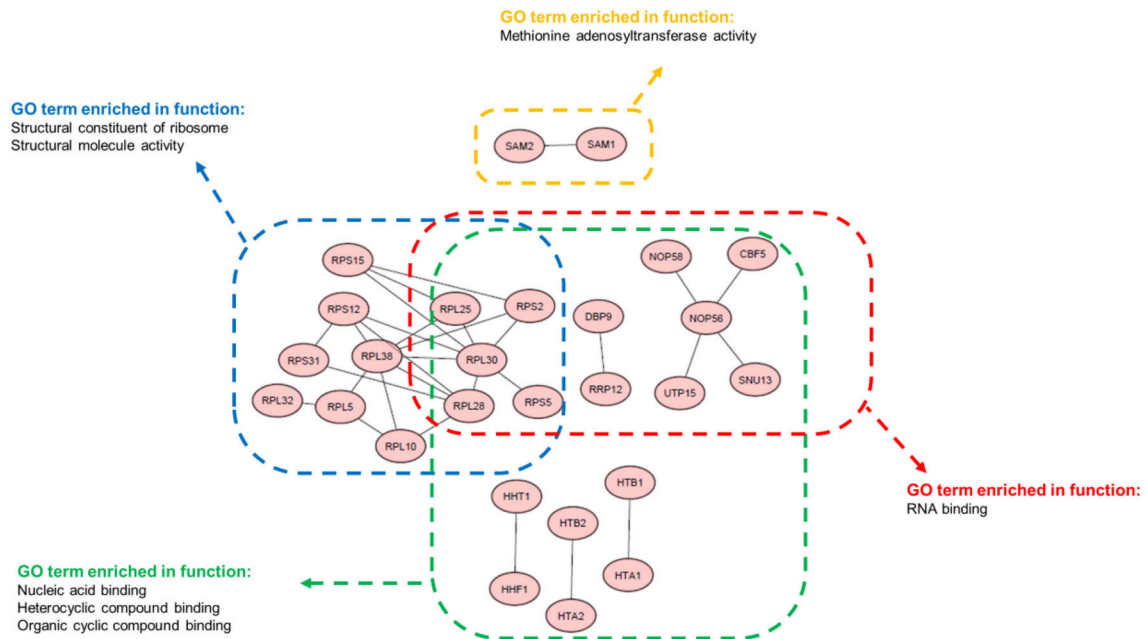


FIGURE 6 | Functional GO terms enriched among the top correlated gene pairs. 30 of 41 gene pairs that were enriched in the indicated functional GO terms are displayed. Genes are presented as nodes, and the co-regulated relationships are presented as edges.

were substantially decreased (~ 0.6 when the top 1% was selected).

Although these results are informative, several drawbacks to this study should be noted. First, binary networks were used, where the strength of the interactions, albeit important, was not considered. Additionally, there was no negative data for interactions (gene pairs with evidence of no interaction with each other) in publically available databases, impeding a more robust evaluation of co-regulation effects. Therefore, a systematic integration of this information in future studies would improve the correlation between co-expression and co-regulation and would facilitate the interpretation of mechanistic models of gene co-expression.

AUTHOR CONTRIBUTIONS

CZ participated in design and execution of the study, conducted data analysis and interpretation, and drafted the manuscript. SL participated in design of the study and data interpretation. AM and QH contributed to the design, execution and data interpretation. All authors participated in modifying and editing the manuscript, and the final manuscript had been read and approved by all authors.

FUNDING

This work was funded by the National Basic Research Program of China (973 Program) (2012CB721101), National Natural Science Foundation of China (21576089), and China Scholarship Council. This work was also supported by grant from the Knut and Alice Wallenberg Foundation.

REFERENCES

- Abdulrehman, D., Monteiro, P. T., Teixeira, M. C., Mira, N. P., Lourenço, A. B., dos Santos, S. C., et al. (2011). YEASTRACT: providing a programmatic access to curated transcriptional regulatory associations in *saccharomyces cerevisiae* through a web services interface. *Nucleic Acids Res.* 39(Suppl. 1), D136–D140. doi: 10.1093/nar/gkq964
- Allocco, D. J., Kohane, I. S., and Butte, A. J. (2004). Quantifying the relationship between co-expression, co-regulation and gene function. *BMC Bioinformatics* 5:18. doi: 10.1186/1471-2105-5-18
- Aoki, K., Ogata, Y., and Shibata, D. (2007). Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol.* 48, 381–390. doi: 10.1093/pcp/pcm013
- Balaji, S., Babu, M. M., Iyer, L. M., Luscombe, N. M., and Aravind, L. (2006). Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *J. Mol. Biol.* 360, 213–227. doi: 10.1016/j.jmb.2006.04.029
- Bar-Joseph, Z., Gerber, G. K., Lee, T. I., Rinaldi, N. J., Yoo, J. Y., Robert, F., et al. (2003). Computational discovery of gene modules and regulatory networks. *Nat. Biotechnol.* 21, 1337–1342. doi: 10.1038/nbt890
- Basan, M., Hui, S., Okano, H., Zhang, Z., Shen, Y., Williamson, J. R., et al. (2015). Overflow metabolism in *escherichia coli* results from efficient proteome allocation. *Nature* 528, 99–104. doi: 10.1038/nature15765
- Bhardwaj, N., and Lu, H. (2005). Correlation between gene expression profiles and protein–protein interactions within and across genomes. *Bioinformatics* 21, 2730–2738. doi: 10.1093/bioinformatics/bti398

ACKNOWLEDGMENTS

We thank Guodong Liu and Boyang Ji from Synthetic and Systems biology group, Chalmers University of Technology for fruitful discussions and constructive suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fphys.2016.00160>

Figure S1 | Sketch map of the retrieving of Bnet and Enet from YEASTRACT database.

Figure S2 | Venn diagram showing the number of shared/unique edges between Bnet and Enet within the same TFs and target genes.

Figure S3 | Boxplot showing the variation of ACERs for B&E&PCRnet through sensitivity analysis. Note here the y axis denotes the ACERs.

Figure S4 | Boxplot for Pearson correlations of co-expression for all gene pairs involved in reference and shared gene pairs among all three CRnets with different Spearman correlation cutoffs. Note that all the CRnets were generated by using Spearman instead of Pearson correlation as cutoff, and the result is identical with those using Pearson correlation (**Figure 5**).

Table S1 | Normalized microarray data used for the calculation of pairwise gene co-expressions.

Table S2 | Reconstructed Bnet, Enet and Pnet based on YEASERACT and BioGRID databases.

Table S3 | Co-regulated networks constructed based on the reconstructed Bnet, Enet and Pnet.

Table S4 | Detailed list of the identified top co-regulated 41 gene pairs.

Table S5 | Enriched GO terms of the selected 41 gene pairs in different categories.

- Blais, A., and Dynlacht, B. D. (2005). Constructing transcriptional regulatory networks. *Genes Dev.* 19, 1499–1511. doi: 10.1101/gad.1325605
- Bolstad, B. M., Irizarry, R. A., Åstrand, M., and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185–193. doi: 10.1093/bioinformatics/19.2.185
- Carter, S. L., Brechbühler, C. M., Griffin, M., and Bond, A. T. (2004). Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* 20, 2242–2250. doi: 10.1093/bioinformatics/bth234
- Chatr-aryamontri, A., Breitkreutz, B. J., Heinicke, S., Boucher, L., Winter, A., Stark, C., et al. (2013). The BioGRID interaction database: 2013 update. *Nucleic Acids Res.* 41, D816–D823. doi: 10.1093/nar/gks1158
- Chatr-aryamontri, A., Breitkreutz, B. J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., et al. (2015). The BioGRID interaction database: 2015 update. *Nucleic Acids Res.* 43, D470–D478. doi: 10.1093/nar/gku1204
- Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., et al. (2012). *Saccharomyces* genome database: the genomics resource of budding yeast. *Nucleic Acids Res.* 40, D700–D705. doi: 10.1093/nar/gkr1029
- Collins, S. R., Miller, K. M., Maas, N. L., Roguev, A., Fillingham, J., Chu, C. S., et al. (2007). Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* 446, 806–810. doi: 10.1038/nature05649
- Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). Affy—analysis of affymetrix genechip data at the probe

- level. *Bioinformatics* 20, 307–315. doi: 10.1093/bioinformatics/btg405
- Ge, H., Liu, Z., Church, G. M., and Vidal, M. (2001). Correlation between transcriptome and interactome mapping data from *Saccharomyces Cerevisiae*. *Nat. Genet.* 29, 482–486. doi: 10.1038/ng776
- Han, J. D., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., et al. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 430, 88–93. doi: 10.1038/nature02555
- Hughes, T. R., and de Boer, C. G. (2013). Mapping yeast transcriptional networks. *Genetics* 195, 9–36. doi: 10.1534/genetics.113.153262
- Hui, S., Silverman, J. M., Chen, S. S., Erickson, D. W., Basan, M., Wang, J., et al. (2015). Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol. Syst. Biol.* 11:784. doi: 10.15252/msb.20145697
- Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B., and Speed, T. P. (2003a). Summaries of affymetrix genechip probe level data. *Nucleic Acids Res.* 31:e15. doi: 10.1093/nar/gng015
- Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., et al. (2003b). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4, 249–264. doi: 10.1093/biostatistics/4.2.249
- Jansen, R., Greenbaum, D., and Gerstein, M. (2002). Relating whole-genome expression data with protein-protein interactions. *Genome Res.* 12, 37–46. doi: 10.1101/gr.205602
- Kelley, R., and Ideker, T. (2005). Systematic interpretation of genetic interactions using protein networks. *Nat. Biotechnol.* 23, 561–566. doi: 10.1038/nbt1096
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559
- Liao, Q. I., Liu, C., Yuan, X., Kang, S., Miao, R., Xiao, H., et al. (2011). Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. *Nucleic Acids Res.* 39, 3864–3878. doi: 10.1093/nar/gkq1348
- Luo, F., Yang, Y., Zhong, J., Gao, H., Khan, L., Thompson, D. K., et al. (2007). Constructing gene co-expression networks and predicting functions of unknown genes by random matrix theory. *BMC Bioinformatics* 8:299. doi: 10.1186/1471-2105-8-299
- MacDonald, M. L., Ding, Y., Newman, J., Hemby, S., Penzes, P., Lewis, D. A., et al. (2015). Altered glutamate protein co-expression network topology linked to spine loss in the auditory cortex of schizophrenia. *Biol. Psychiatry* 77, 959–968. doi: 10.1016/j.biopsych.2014.09.006
- Miller, J. A., Horvath, S., and Geschwind, D. H. (2010). Divergence of human and mouse brain transcriptome highlights alzheimer disease pathways. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12698–12703. doi: 10.1073/pnas.0914257107
- Miller, J. A., Oldham, M. C., and Geschwind, D. H. (2008). A systems level analysis of transcriptional changes in alzheimer's disease and normal aging. *J. Neurosci.* 28, 1410–1420. doi: 10.1523/jneurosci.4098-07.2008
- Monteiro, P. T., Mendes, N. D., Teixeira, M. C., D'Orey, S., Tenreiro, S., Mira, N. P., et al. (2008). YEASTRACT-DISCOVERER: new tools to improve the analysis of transcriptional regulatory associations in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 36(Suppl 1), D132–D136. doi: 10.1093/nar/gkm976
- Peake, J., Sampson, D., Broadbent, J., Bulmer, A., and Neubauer, O. (2015). Comparison of post-exercise muscle and neutrophil transcriptomes using weighted gene co-expression network analysis. *FASEB J.* 29(1_Supplement):675.10.
- Presson, A. P., Sobel, E. M., Papp, J. C., Suarez, C. J., Whistler, T., Rajeevan, M. S., et al. (2008). Integrated weighted gene co-expression network analysis with an application to chronic fatigue syndrome. *BMC Syst. Biol.* 2:95. doi: 10.1186/1752-0509-2-95
- Roy, S., Bhattacharyya, D. K., and Kalita, J. K. (2014). Reconstruction of gene co-expression network from microarray data using local expression patterns. *BMC Bioinformatics* 15(Suppl. 7):S10. doi: 10.1186/1471-2105-15-S7-S10
- Ruan, J., Dean, A. K., and Zhang, W. (2010). A general co-expression network-based approach to gene expression analysis: comparison and applications. *BMC Syst. Biol.* 4, 1–21. doi: 10.1186/1752-0509-4-8
- Savage, R. S., Ghahramani, Z., Griffin, J. E., de la Cruz, B. J., and Wild, D. L. (2010). Discovering transcriptional modules by bayesian data integration. *Bioinformatics* 26, 158–167. doi: 10.1093/bioinformatics/btq210
- Spivakov, M. (2014). Spurious transcription factor binding: non-functional or genetically redundant? Bioessays: news and reviews in molecular. *Cell Dev. Biol.* 36, 798–806. doi: 10.1002/bies.201400036
- Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34(Suppl. 1), D535–D539. doi: 10.1093/nar/gkj109
- Stuart, J. M., Segal, E., Koller, D., and Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302, 249–255. doi: 10.1126/science.1087447
- Teixeira, M. C., Monteiro, P., Jain, P., Tenreiro, S., Fernandes, A. R., Mira, N. P., et al. (2006). The YEASTRACT database: a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 34(Suppl. 1), D446–D451. doi: 10.1093/nar/gkj013
- Teixeira, M. C., Monteiro, P. T., Guerreiro, J. F., Gonçalves, J. P., Mira, N. P., dos Santos, S. C., et al. (2014). The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 42, D161–D166. doi: 10.1093/nar/gkt1015
- Tong, A. H., Lesage, G., Bader, G. D., Ding, H., Xu, H., Xin, X., et al. (2004). Global mapping of the yeast genetic interaction network. *Science* 303, 808–813. doi: 10.1126/science.1091317
- Xulvi-Brunet, R., and Li, H. (2010). Co-Expression networks: graph properties and topological comparisons. *Bioinformatics* 26, 205–14. doi: 10.1093/bioinformatics/btp632
- Yang, Y., Zhang, Z., Li, Y., Zhu, X.-G., and Liu, Q. (2010). Identifying cooperative transcription factors by combining chip-chip data and knockout data. *Cell Res.* 20, 1276–1278. doi: 10.1038/cr.2010.146
- Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4:17. doi: 10.2202/1544-6115.1128

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Zhang, Lee, Mardinoglu and Hua. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.