Essential Proteins and Functional Modules in the Host-Pathogen Interactions from Innate to Adaptive Immunity
C. albicans-zebrafish Infection Model

Chia-Chou Wu and Bor-Sen Chen
Control and Systems Biology Laboratory, Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan

Keywords: Computational Systems Biology, Network construction, Host-pathogen interaction, Protein-protein interaction network, Infection.

Abstract: The host and the pathogen are indispensable in the infectious diseases. Besides studying the host defensive and pathogen invasive mechanisms individually, the cross-species interactions, i.e., the host-pathogen interactions, become a novel and intense research subjects of the infectious diseases. In this study, two host-pathogen interaction networks are constructed for innate and adaptive immunity based on the time course microarray data of *C. albicans*-zebrafish infection model. The interaction variations in the host, pathogen, and host-pathogen regions are evaluated by comparing the two constructed networks. Those proteins of larger interaction variations stand for more pivotal roles in the transition from innate to adaptive immunity. Moreover, in the host-pathogen region, four significantly enriched functional modules are identified. Meanwhile, the interaction variations of these four functional groups imply the corresponding strategy shifts of the host and pathogen from innate to adaptive immunity. In view of these results, this study gives a systematic explanation about the transition from innate and adaptive immunity from functional modules perspective. Thus, this study provides potential targets for developing efficient therapies of the infectious diseases.

1 INTRODUCTION

The host and the pathogen are indispensable in the infectious diseases. In particular, the interplays between the host and pathogen shape the whole infection processes from the first pathogens exposure to the final outcomes of the infection (Tierney et al., 2012). After activating the first line of host defense mechanisms, the innate immunity recruits several types of cells (e.g., macrophages, dendritic cells, NK cells, etc.) to protect the host from the pathogen invasion and then strive to eliminate the threats from the pathogens. In turn, pathogens have evoked multiple strategies for surviving under the host immune-defense mechanisms. After several rounds of attacks and defenses between the host and pathogen, the host may eliminate pathogens and prepare for the next challenges or the pathogens may win the battle to cause chronic inflammation or death of the host. Despite the tremendous advances in the pathogenic mechanisms and the following triumph in the drug development (Arnold et al., 2012), the remaining issues (e.g., drug resistance) of infectious diseases become more troublesome. The dynamic and complex interactions between the host and pathogen may partially explain why those drugs are often not effective *in vivo* (Meijer and Spaink, 2011). Until a decade ago, the traditional viewpoint to treat the host and pathogen separately is shifted to a more holistic viewpoint on both players in the infection processes. This viewpoint transition results from (i) the realization of the indispensableness of the host-pathogen interactions (HPIs) in the infectious diseases and (ii) the advent of the OMICs biotechnology in measuring the genes, transcripts, and proteins at whole cell/organism levels (Schmidt and Volker, 2011). This permits a comprehensive interrogation of the pathogen at the whole-genome, transcriptome, and proteome levels as well as the host. From the molecular aspect, the infection processes can be viewed as the interference of pathogenic proteins with the hosts' interaction network (Arnold et al., 2012). Hence, to investigate infection processes from a systematic perspective, in this study we would construct dynamic host-pathogen protein-protein interaction (PPI) networks.

Regarding the success of the *C. albicans*-zebrafish infection model (Chao et al., 2010) as well as the

Wu C. and Chen B..
Essential Proteins and Functional Modules in the Host-Pathogen Interactions from Innate to Adaptive Immunity - C. albicans-zebrafish Infection Model.
DOI: 10.5220/0005201700170025
In *Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms* (BIOINFORMATICS-2015), pages 17-25
ISBN: 978-989-758-070-3

amenability to genetic manipulations (Gratacap and Wheeler, 2014), the zebrafish is a novel and potential model organism to study the immunity. Furthermore, the zebrafish and human immune systems are remarkably similar and more than 75% of human genes implicated in diseases have counterparts in zebrafish (Schier, 2013). This provides a strong connection between the zebrafish and human on the pathogenic mechanisms as well as immune responses, which are important for biomedical applications. The immune system of zebrafish as well as other vertebrates can further be divided into two subsystems, i.e., innate (unspecific) and adaptive (specific) immunity (Trede et al., 2004). Hence, the first dataset we used to construct the dynamic host-pathogen PPI network (HP-PPIN) measured the gene expression profiles during the first 18 hours after zebrafish is firstly exposed to a lethal dose of *C. albicans* (Chen et al., 2013). This dataset sampled the gene expression profiles at 9 time points (i.e., 0.5, 1, 2, 4, 8, 12, 16, 18-hour post-injection) with three replicates for *C. albicans* and zebrafish, respectively (Figure 1A). The outcomes of the interplay between the host and pathogen after their first contact are captured by time course microarray experiments. Immediately after the injection of *C. albicans* into zebrafish, the immune surveillance system of zebrafish senses the existence of the invaders. The recognition of the pathogen-associated molecular patterns (PAMPs) and/or damage-associated molecular patterns (DAMPs) by the pattern recognition receptors (PRRs) (e.g., toll-like receptors, C-type lectin receptors, etc.) (Trede et al., 2004; Romani, 2011) may be viewed as a starting point of a series of complex HPIs. Those PRRs would initiate downstream pathways that promote the activation of other parts of innate immune system and the clearance of pathogens (e.g., production and secretion of cytokines, chemokines, and chemotactic cues to recruit more immunocytes). Thus, the morphological transitions (yeast-to-hyphal form) (Kuo et al., 2013), required ions and small molecules transportation (Wang et al., 2014), and structures and components of molecules on the cell wall changes (Romani, 2011) are the strategies utilized by *C. albicans* to acquire nutrients and evade the clearance from the host innate defensive mechanisms. Those primary responses of the host and pathogen are recorded in the first dataset which is a result from a constant innate immune response and a delayed adaptive immune response (see Figure 1).

In addition to the activation of innate defensive mechanisms, the PRRs and antigen presenting cells would further activate the specific cells to clear the pathogens much more efficiently (Romani, 2011). To investigate the host-pathogen interactions in the specific defensive mechanisms (i.e., the adaptive immunity), we adopted the second dataset. The experimental design of the second dataset comprises two *C. albicans* injections into zebrafish. In the first injection, a nonlethal dose is used to cause the host primary responses and immunological memory in the host. And in 14 days after the first injection, a lethal dose is applied to zebrafish. Then, the gene expression profiles are recorded through microarray after the second injection and comprise 8 time points (2, 6, 12, 18, 24, 30, 36, 42-hour post-reinjection) with two replicates for *C. albicans* and zebrafish, respectively (Figure 1A). Due to the previous exposure to *C. albicans*, the immunological memory in zebrafish would be activated and force the activation of the adaptive immunity. Thus, the main portion of the information recorded in the second dataset reflects the outcomes of the HPIs in the adaptive immunity. In contrast to the innate immunity, the adaptive immunity of zebrafish and the responses of *C. albicans* to the adaptive immunity are less well-known. Here, in this study the combination of these two datasets provides an opportunity to investigate the HPIs and their roles in the innate and adaptive immunity.

The infection processes are often described as battles between the host and pathogen (Leroy and Raoult, 2010). The simultaneous considerations on the roles of the host and pathogen in the infection processes and the available genome-wide measurements make the possibility of the systematic viewpoints on the effects of the HPIs in the innate and adaptive immunity (Wang et al., 2013). In this study, the usage of the *C. albicans*-zebrafish infection model (Chao et al., 2010) shed light on the infectious diseases of the human host. As for the pathogen, the *C. albicans* is the most virulent member of the CUG clade of yeasts and a common cause of both superficial and invasive infections (Lohberger et al., 2014) which may cause life-threatening infections in immune-comprised host (Odds, 1979), such as HIV positive patients. Investigating the infection processes of *C. albicans* in detail can improve the knowledge of the pathogenic mechanisms and promote the control of infectious diseases. Hence, the two datasets (GSE32119 and GSE51603) are used for further analyses on the HPIs. To extract the interaction information from the time course microarray data, two dynamic HP-PPINs are built up for innate and adaptive immunity (Wang et al., 2013). The HP-PPINs consist of the PPIs between zebrafish and *C. albicans*, zebrafish and zebrafish, and *C. albicans* and *C. albicans*. Through examining the interaction variation between the innate and adaptive PPINs, the interactions of the largest difference in the net-

work indicates the occurrences of the most dramatic change. By evaluating the average interaction variation per edge, the critical proteins of the high interaction variations in the interface of the host and pathogen can be identified. Moreover, taking the advantage of advances in the ontology analysis, the significantly enriched functional modules in the interface of the host and pathogen can also be identified. Those functional modules may imply the strategies taken by the host and pathogen in the battles, the infection processes. Thus, those function modules and the proteins are potential drug targets of infectious diseases (Schmidt and Volker, 2011).

2 MATERIALS AND METHODS

2.1 Overview of Microarray Data

In this study, we adapted two microarray datasets: one is the temporal gene expression profiles of the host (zebrafish) and pathogen (*C. albicans*) in the period that they are firstly exposed to each other; the other is the temporal expression profiles of the host and pathogen in the period that they are secondly exposed to each other. In the first set of microarray data, the microarray experiments were performed to simultaneously profile genome-wide gene expressions in both *C. albicans* and zebrafish during the infection processes. Adult AB strain zebrafish were intraperitoneally injected with 1×10^8*C. albicans* (SC5314 strain) cells (a lethal dosage). Whereas the second microarray data measured the genome-wide gene expression level of the host and pathogen since their second contact (with 1×10^7 *C. albicans* injection, also a lethal dosage), that is, fourteen days after their first contact (with 1×10^5 *C. albicans*, a non-lethal dosage). Then, a two-step homogenization/mRNA extraction procedure was performed using the whole zebrafish infected with *C. albicans*. This approach could provide separate pools of gene transcripts from both the host and the pathogen, enabling individual estimation of specific gene expression profiles in either the host or the pathogen using sequence-targeted probes derived from the individual genome. Agilent in situ oligonucleotide microarrays, which cover 6,202 and 26,206 genes for *C. albicans* and zebrafish respectively, were used to profile time-course gene expression at 9 time-points (0.5, 1, 2, 4, 6, 8, 12, 16, 18 hours post-infection) with three replicates for both organisms in the first microarray dataset (Chen et al., 2013) and 8 time points (2, 6, 12, 18, 24, 30, 36, and 42 hours post-second infection) with two replicates for both organisms in the second microarray dataset. The first set of microarray was downloaded from the GEO database (GSE32119) and the second set of microarray (GSE51603) was prepared under the similar condition as the first set. Manipulation of the animal model was approved by the Institutional Animal Care and Use Committee of National Tsing Hua University (IRB Approval No. 09808).

2.2 Protein Pool Selection and Database Integration

There are two things to be completed before constructing the dynamic protein-protein interaction (PPI) network. The first is to have a protein pool from which the nodes in the resultant networks are chosen. And the second step is to have all possible PPIs among the proteins in the protein pool through integrating the interaction information from databases. Here, our protein pool is consisted of the union of the differentially expressed genes in the first and second set of microarray data and the differentially expressed genes between the first and second microarray datasets. The criterion to select the differentially expressed genes in the first and second microarray datasets is to compute the *p-value* of ANOVA test whether the average expression levels are different along the time (i.e., for the first dataset, the null hypothesis is $\mu_1 = \cdots = \mu_9$ and for the second dataset, the null hypothesis is $\mu_1 = \cdots = \mu_8$) and then to select those proteins with the corrected *p-value*<0.05 into the protein pool. Also the genes in the top 5% of the expression difference between the first and second datasets were chosen into the protein pool. Next, for the all possible interactions among the proteins in the protein pool, the interaction information of zebrafish-zebrafish, *C. albicans*-*C. albicans*, and zebrafish-*C. albicans* are needed. However, the lack of the information about these three kinds of interaction information makes it difficult to collect all possible interactions. Also it is impossible to consider full interactions among the proteins in the protein pool. To overcome the issue, the interaction information from the human and yeast are used due to their similarity to our studying subjects (zebrafish and *C. albicans*) and data availability. To infer the possible interactions of the studying subjects (zebrafish and *C. albicans*), the orthologs information in the Inparanoid is used to convert the interactions of human and yeast into the interactions of zebrafish and *C. albicans*. It should be noticed that the interactions inferred from the ortholog-based method were derived under many different experimental conditions, which cannot accurately reflect the actual condition of host-pathogen interactions during *C. albicans* infection processes; that

is, there exist false positives interactions in the all inferred possible interactions of zebrafish and *C. albicans*.

2.3 Host-Pathogen Protein-Protein Interaction Network (HP-PPIN) Construction

To construct the interspecies network from the protein pool and inferred interactions, the dynamic model of the protein-protein interaction is used to determine the realistic interaction network with one protein by one protein fashion. For a target protein i of the host, the dynamic interaction model is as follows (Wang et al., 2013):

$$\begin{aligned} p_i^{(h)}[k+1] =& \sigma_i^{(h)} p_i^{(h)}[k] + \sum_{n=1}^{N} \alpha_{in}^{(h)} p_n^{(h)}[k] \\ &+ \sum_{m=1}^{M} \gamma_{im} p_m^{(p)}[k] + \beta_i + \varepsilon_i[k+1] \end{aligned} \tag{1}$$

and of the pathogen

$$\begin{aligned} p_i^{(p)}[k+1] =& \sigma_i^{(p)} p_i^{(p)}[k] + \sum_{m=1}^{M} \alpha_{im}^{(p)} p_m^{(p)}[k] \\ &+ \sum_{n=1}^{N} \gamma_{in} p_n^{(h)}[k] + \beta_i + \varepsilon_i[k+1] \end{aligned} \tag{2}$$

where $p_i[k]$ denotes the protein activity level at time k, the superscript of p_i indicates the species of proteins (h: host, zebrafish; p: pathogen, *C.albicans*), $\varepsilon_i[k]$ denotes the environment noise at time k, σ_i denotes the self-regulation ability, α_{in} denotes the regulation ability from the regulator n of the same species as the target protein to the target protein i, γ_{im} denotes the regulation ability from the regulator m of the other species to the target protein i, and β_i denotes the basal activity level of target protein i. The biological meaning of this formulation is the protein activity level of target protein i in the future (at time $k+1$) is determined by the current protein activity level (at time k) of itself (σ_i), the regulations from other proteins of the same species (α_{in}) and the other species (γ_{im}), the basal activity, and the environmental noise. Due to the unavailability of the proteomic data, the expression levels measured by the microarray experiments are used to stands for the activity levels in the formulation. The dynamic model for the host target protein i can be further rewritten into a concise form as follows:

$$\mathbf{p}_i^{(h)} = \Phi_i \theta_i + \varepsilon \tag{3}$$

where

$$\mathbf{p}_i^{(h)} = \begin{bmatrix} p_i^{(h)}[1] & \cdots & p_i^{(h)}[K+1] \end{bmatrix}^T,$$

$$\theta_i = \begin{bmatrix} \alpha_{i1} & \cdots & \alpha_{iN} & \gamma_{i1} & \cdots & \gamma_{iM} & \sigma_i & \beta_i \end{bmatrix}^T,$$

$$\varepsilon_i = \begin{bmatrix} \varepsilon_i[1] & \cdots & \varepsilon_i[K+1] \end{bmatrix}^T,$$

and

$$\Phi_i = \begin{bmatrix} p_{i1}^{(h)}[0] & \cdots & p_{iN}^{(h)}[0] & p_{i1}^{(p)}[0] & \cdots & p_{iM}^{(p)}[0] & p_i[0] & 1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ p_{i1}^{(h)}[K] & \cdots & p_{iN}^{(h)}[K] & p_{i1}^{(p)}[K] & \cdots & p_{iM}^{(p)}[K] & p_i[K] & 1 \end{bmatrix}$$

Similarly, the dynamic model for the pathogen can also be rewritten into a similar form. The only unknown parameter θ_i can then be estimated by parameter estimation methods, such as least square estimation. However, due to the lack of large scale measurement of host and pathogen proteins, we alternatively used gene expression profiles as a substitute of protein activities to identify the parameters in the model. Furthermore, to make sure the model is unnecessarily complex, the Akaike information criterion (AIC) is introduced for model selection to balance the competing objectives of conformity to the data and parsimony, i.e., a trade-off between the model error and model complexity. Hence, the final network encompass the dynamic models of each protein with the minimum AIC values.

2.4 Relevance Score Calculation

To target the essential proteins in the host-pathogen protein-protein interaction network, the relevance scores are calculated for proteins or functional modules to correlate proteins with the evolution of the host-pathogen interactions from innate to adaptive immunity. The relevance score is basically a measurement of the variation of the regulation activity under a condition transition. According to the dynamic models, the constructed PPI network under a specific microarray experiment condition can be written as follows:

$$\begin{bmatrix} p_1^{(h)}[k+1] \\ \vdots \\ p_N^{(h)}[k+1] \\ p_1^{(p)}[k+1] \\ \vdots \\ p_M^{(p)}[k+1] \end{bmatrix} = \begin{bmatrix} \sigma_1^{(h)} & \cdots & \alpha_{1N}^{(h)} & \gamma_{11} & \cdots & \gamma_{1M} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{N1}^{(h)} & \cdots & \sigma_N^{(h)} & \gamma_{N1} & \cdots & \gamma_{NM} \\ \gamma_{11} & \cdots & \gamma_{1N} & \sigma_1^{(p)} & \cdots & \alpha_{1M}^{(p)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{M1} & \cdots & \gamma_{MN} & \alpha_{M1}^{(p)} & \cdots & \sigma_M^{(p)} \end{bmatrix} \times$$

$$\begin{bmatrix} p_1^{(h)}[k] \\ \vdots \\ p_N^{(h)}[k] \\ p_1^{(p)}[k] \\ \vdots \\ p_M^{(p)}[k] \end{bmatrix} + \begin{bmatrix} \beta_1^{(h)} \\ \vdots \\ \beta_1^{(p)} \\ \vdots \\ \beta_M^{(p)} \end{bmatrix} + \begin{bmatrix} \varepsilon_1^{(h)}[k+1] \\ \vdots \\ \varepsilon_N^{(h)}[k+1] \\ \varepsilon_1^{(p)}[k+1] \\ \vdots \\ \varepsilon_M^{(p)}[k+1] \end{bmatrix} \tag{4}$$

or in a more concise form:

$$\mathbf{p}[k+1] = A\mathbf{p}[k] + \beta + \varepsilon[k+1] \tag{5}$$

where A is a matrix representation of the network constructed under a specific microarray experiment condition. The regulation ability difference of two PPI networks between innate and adaptive immunity can be expressed as the following interaction difference matrix form (Wang et al., 2014):

$$D_{cond2-cond1} = A_{cond2} - A_{cond1} \tag{6}$$

In the condition transition, if the variation of the regulation abilities of a protein is larger, it may implies the protein plays a more important role in the condition transitions. So the relevance score of a protein can be defined as follows:

$$RS_p = \frac{\sum_{q=1}^{Q} \|d_{pq}\|}{\text{Degree of protein } p} \tag{7}$$

where d_{pq} is the pq-entry of $D_{cond2-cond1}$, that is, the average regulation ability variation of the protein p. The degree of protein p is the number of non-zero element in pth row of the difference matrix $D_{cond2-cond1}$. The relevance score is proposed to evaluate the interaction variations of proteins in the network.

3 RESULTS

3.1 Overview of the Host-Pathogen Protein-Protein Interaction Networks (HP-PPINs) for Innate and Adaptive Immunity

In this study, we aimed to understand the roles of host-pathogen interactions in the innate and adaptive immunity and the transition from innate to adaptive immunity with a systems biology approach. The outcomes of host-pathogen interactions are represented in the simultaneous measurements of the temporal gene expression profiles during the periods when innate and adaptive immunity are activated (Figure 1A). The HPIs in these two periods are then extracted by

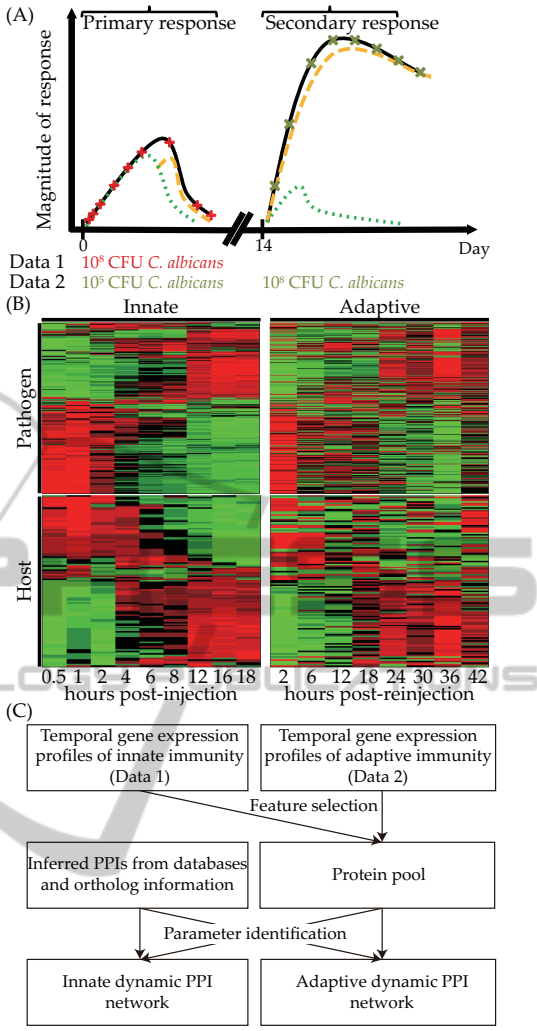

Figure 1: The overview of microarray experiments. (A) The schematic of two microarray datasets. In the first dataset, gene expression profiles of the *C. albicans* and zebrafish are recorded at 9 time points (+). It mainly represent the outcomes of HPIs in the innate immunity (green dot line). In the second dataset, gene expression profiles of the *C. albicans* and zebrafish are recorded at 8 time points (×). It mainly represents the outcomes of HPIs in the adaptive immunity (yellow dash line). (B) The overview of microarray data. (C) Flowchart for network construction. Features selected from the two microarray datasets consist the protein pool. The parameters in the dynamic models are identified based on the expression profiles and the interaction information of the selected features. In the end, two dynamic HP-PPINs are constructed for innate and adaptive immunity.

the dynamic models of the microarray data and further visualized as two HP-PPINs (Figure 1B). The protein pool encompassed 1620 proteins of interest including differentially expressed features and the top 5% of the expression level difference between innate and adaptive immunity (see Figure 1C) and there are

26060 PPI candidates. Then using dynamic network construction (see Figure 1C and methods for details), the resultant networks are consisted of 1512 proteins (1431 for the *C. albicans*; 81 for zebrafish) and 5722 PPIs (5510 for the intracellular region of *C. albicans*; 145 for the interspecies interaction; 66 for intracellular region of zebrafish) for innate immunity and 1578 proteins (1480 for the *C. albicans*; 98 for zebrafish) and 3755 PPIs (3577 for the intracellular region of *C. albicans*; 96 for the interspecies interaction; 82 for intracellular region of zebrafish) for adaptive immunity. The details of the amount of nodes and edges are summarized in the Figure 2B. In the amount variation of the nodes and edges of the pathogen, although there exists plenty of nodes shared by both innate and adaptive immunity in the host-pathogen and pathogen-pathogen regions (Figure 2B), the number of edges has changed from 5511 to 3577, that is, only 1203 edges are shared (Figure 2B). This implicates that the pathogen may use the almost the same set of protein (∼85%) but the different links to interact with the host and regulate functions within the pathogen itself under different challenges at innate and adaptive immunity. In contrast, the host might use a slightly different strategy for self-regulation. To efficiently identify and evaluate the importance of proteins in the innate and adaptive immunity, we aggregated the two networks (the innate and the adaptive immunity network) into an interaction difference network (IDN), i.e., the matrix D in the method (Figure 2A).

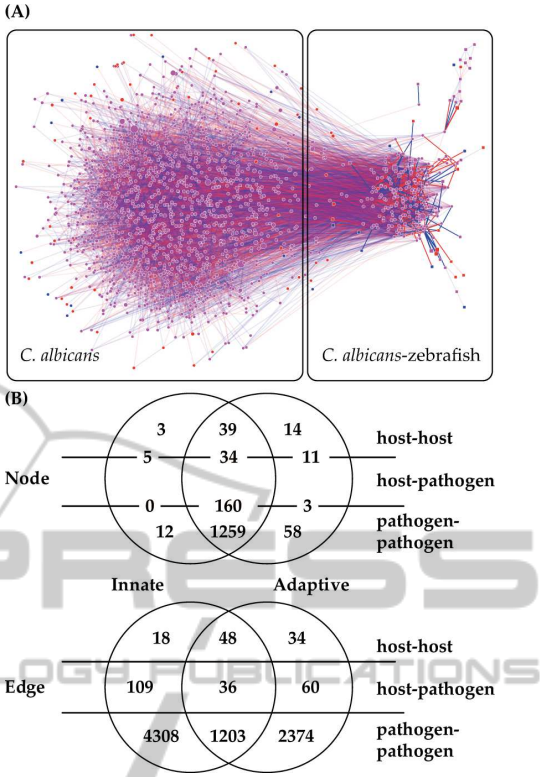

Figure 2: Result summary of dynamic HP-PPI network construction. (A) The difference of the constructed innate and adaptive networks (Node color: blue and red stand for exclusive existence in the innate and adaptive immune responses respectively and purple stands for coexistence in both immune responses. Edge color: blue and red stands for the attenuated and enhanced interaction, respectively.). (B) The number of the nodes and edges in the two dynamic HP-PPINs.

3.2 The Essential Proteins and Functional Modules based on the Relevance Scores

The relevance score stated in the previous section is a quantity to represent the average interaction variations per links of a protein, that is, the ratio of the total interaction variation of a protein to the number of links possessed by the protein. Hence, the relevance score is considered to evaluate the extent of the interaction variations which may be critical in this study for the observation of the amount of the nodes and edges in the IDN (Figure 2A), the difference between innate HP-PPIN and adaptive HP-PPIN. Comparing to the other similar calculation (Wang et al., 2014), our proposed relevance score is more proper to evaluate the interaction variation of proteins in the network since the consideration of degree of protein excludes the proteins have many links of little variations. In the following we would focus on the proteins of the top ten in relevance scores at three regions, that is, the host-host, host-pathogen, and pathogen-pathogen region, respectively.

3.2.1 The Host-host Region

In the region of host-host interaction (Figure 3), the top ten proteins in the relevance scores showed their close relationships with innate and adaptive immune responses. Extracting the ten proteins and their first neighbors from the IDN, there are five components in the host-host region. The biggest one is consisted of *f2*, LOC798231, LOC793315, *ace2*, *gnai1*, and their first neighbors. Starting from *gnat2*, a host G-protein, also one end of HPIs has connections with chemokines-related proteins (*cck-c5a* and si:dkey-269d20.3) and chemotaxis-related proteins (ENSDARP00000105159 and ENSDARP00000111107). Then angiogenesis- and coagulation-related proteins (*agt*, *ace2*, *f2*, and ENSDARP00000098661) are connected to the chemokines-related proteins. In the final part of the components, there are three more proteins, serine proteinase inhibitor (*serpinc1*), proki-

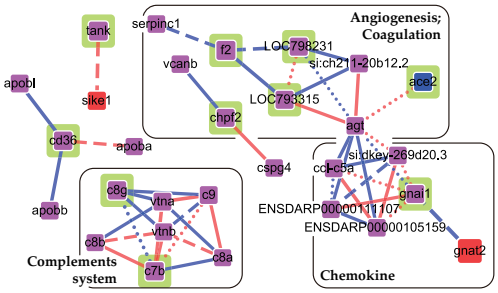

Figure 3: The ten proteins highest in relevance scores in the host-host region and their first neighbors. The nodes with shadow are the proteins of higher relevance scores (The meaning of node color, edge color, and line style are the same as Figure 2.).

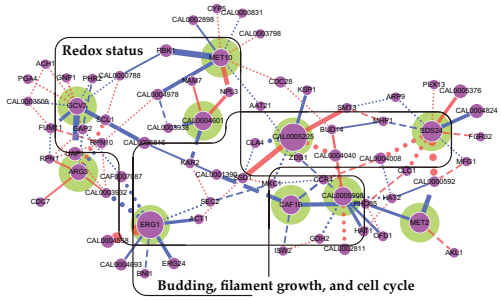

Figure 4: The ten proteins highest in relevance scores in the pathogen-pathogen region and their first neighbors (The meaning of node color, edge color, and line style are the same as Figure 2.).

neticin (ENSDARP00000109666), and suppressor of IKBKE 1 (*sike1*). In this component, the roles of angiogenesis and coagulation are manifest in the innate and adaptive immunity. The second component mainly consists of complements (*c7b*, *c8g*, *c8a*, *c8b*, *c9*) and vitronectins (*vtna* and *vtnb*). Given the well-known roles of complement system in immunity, the vitronectins got researchers' attention in the field of immunity recently (Gerold et al., 2008). The *cd36* and apolipoproteins form the third component. CD36 plays a pivotal role in macrophage foam-cell formation and atherogenesis, which is reduced by apolipoproteins. Although the last two components are less documented, the versican (*vcanb*) and tank are reported their roles on the inflammation (Wight et al., 2014).

3.2.2 The Pathogen-Pathogen Region

In the pathogen-pathogen region (Figure 4), the ten proteins with the highest relevance scores and their first neighbors form a single component. In this component, the importance of redox status in the innate and adaptive immune responses is emphasized again (Wang et al., 2014). ERG1, CAL0005908, MET10, and GCV3 are all related to the redox status of *C. albicans*. Also CAL0005225, ERG1, and SDS24 are responsible to the expansion of *C. albicans* due to their functions on the budding, filament growth, and cell cycle, respectively. Especially, MET10 is also responsible to the responses to the stress from the host and environment. Another major function in this component is the transferase activity. MET2 is a homoserine acetyltransferase which can transform homoserine, a toxin for *C. albicans*, to another compound. ARG3 would facilitate the production of citrulline, which can induce the pseudohyphal morphogenesis. The morphology transformation of *C. albicans* has been proven to be important in the pathogenesis of *C. albicans*. In the end, the hydrolase, CAF16, exerts its influence on the RNA polymerase II although the specific affected genes are still unknown.

3.2.3 The Host-Pathogen Region

In the host-pathogen region, we also selected 10 proteins from the host and the pathogen, respectively. Those interspecies proteins form more complicated interaction networks in the host-pathogen region (Figure 5). Meanwhile, the functional structure is slightly different comparing to the host-host and pathogen-pathogen regions. A possible mechanism how redox status in the host and pathogen correlates is shown in the extracted IDN, i.e., the interaction between thioredoxin (*txn*) and RiboNucleotide Reductase 1 (RNR1). In addition to the redox role of RNR1, RNR1 also has impact on the iron utility, filament growth, and cell cycle. This implicates the effect of redox status on the pathogen is multifaceted. One of the interactions of RNR1, the link between RNR1 and CAL0003932, is attenuated in the adaptive immune response. This implies the ability of adaptive immunity to attenuate the deubiquitination and degradation of proteasome in the pathogen. However, CAL0003932 has not been well-characterized. A group of chemokine-related host proteins constitutes another similar function in the host-pathogen region. Comparing to the chemokine-related function in the host-host region, its roles in the HPIs are more interesting. CAG1, the entry how chemokine-related functions affect the pathogen, is related to the hyphal growth, mating, and biofilm formation of pathogen, which are all important in the pathogenesis. Besides the same functions (redox status and chemokines) in the pathogen-pathogen and host-host regions, there are several functions which can only be seen in the host-pathogen region, that is drug responses, fatty acid, glycine metabolism, and circadian. The interactions among TAF60, *gtf2a2*, *polr2e* emerge in the adaptive immunity. TAF60, a transcription factor, is

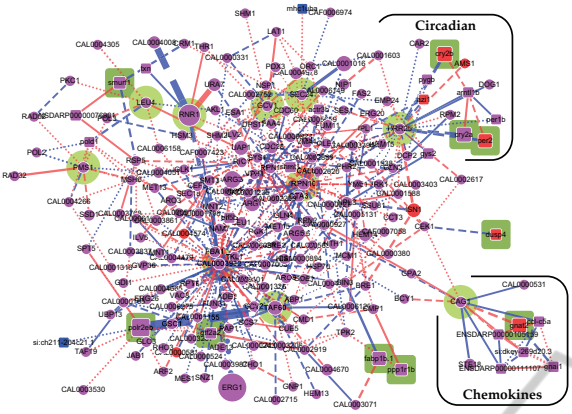

Figure 5: The ten host and pathogen proteins highest in relevance scores in the pathogen-pathogen region and their first neighbors (The meaning of node color, edge color, and line style are the same as Figure 2.).

responsible for the drug responses in the pathogen and *gtf2a2* and *polr2e* are related to the gene transcription in the host. Their interactions stand for a possible mechanism how the effects of HPIs are exerted into the gene level. Other proteins for drug response are SEC24 and LEU4, which are not well-characterized. The fatty acid binding protein (*fabp1b.1*) of the host shows the involvement of fatty acid in the innate and adaptive immune responses. This is another proof for the hypothesis of a local and systemic crosstalk between adipocytes and monocytes mediated by fatty acids (Kopp et al., 2009). This fatty acid binding protein links to the protein of the pathogen, SIN3, related to the filament growth. The interaction between fatty acid binding protein and SIN3 becomes more negative in the adaptive immunity. The third function, glycine metabolism, has been implicated the contribution to the infectious capacity of the pathogen (Flynn et al., 2010). The final and interesting function is the circadian rhythm in the host and pathogen. The circadian rhythm-related proteins of the host (*cry2a*, *cry2b*, and *per2*) and the pathogen (HRR25) form a sub-network in the host-pathogen region. The circadian rhythms in the host and pathogen are correlated and there are plenty of the pathogens functions (yeast-hyphal switch, gene transcription, pathogenesis, etc.) are affected through HRR25.

4 CONCLUSION

In this study, the dynamic network modeling is used to identify the complex and dynamic HPIs during two different types of immune system. Based on the high-throughput expression level measurements, two HP-PPINs are constructed and then compared with each other. We found that the pathogen may change the interactions between proteins rather than recruit a whole new set of proteins to react with the host. Hence, we proposed the relevance score to quantify the difference of the regulation of a protein between the innate and adaptive HP-PPIN. The relationships between the HPIs and several proteins of higher relevance scores are verified by literatures. Also, several not well-known proteins but with higher relevance scores are suggested their roles in the HPIs. Moreover, the circadian-, redox status-, angiogenesis-, and coagulation-related functions are correlated in the HPIs. The interplays among the four functional modules cause the changes in the mechanisms of the pathogenesis, responses to stresses, and cell cycle in the *C. albicans*. In the end, these proteins of higher relevance scores and the four functional modules are regarded as the essential components in the HPIs as well as the potential targets for future drug developments.

ACKNOWLEDGEMENTS

Ministry of Science and Technology, R.O.C. (Taiwan) [MOST 103-2745-E-007-001-ASP].

REFERENCES

Arnold, R., Boonen, K., Sun, M. G. F., and Kim, P. M. (2012). Computational analysis of interactomes: Current and future perspectives for bioinformatics approaches to model the host-pathogen interaction space. *Methods*, 57(4):508–518.

Chao, C. C., Hsu, P. C., Jen, C. F., Chen, I. H., Wang, C. H., Chan, H. C., Tsai, P. W., Tung, K. C., Wang, C. H., Lan, C. Y., and Chuang, Y. J. (2010). Zebrafish as a model host for *Candida albicans* infection. *Infection and Immunity*, 78(6):2512–2521.

Chen, Y. Y., Chao, C. C., Liu, F. C., Hsu, P. C., Chen, H. F., Peng, S. C., Chuang, Y. J., Lan, C. Y., Hsieh, W. P., and Wong, D. S. H. (2013). Dynamic transcript profiling of *Candida albicans* infection in zebrafish: A pathogen-host interaction study. *PloS One*, 8(9).

Flynn, N. E., Patyrak, M. E., Seely, J. B., and Wu, G. Y. (2010). Glycine oxidation and conversion into amino acids in *Saccharomyces cerevisiae* and *Candida albicans*. *Amino Acids*, 39(2):605–608.

Gerold, G., Abu Ajaj, K., Bienert, M., Laws, H. J., Zychlinsky, A., and de Diego, J. L. (2008). A Toll-like receptor 2-integrin β_3 complex senses bacterial lipopeptides via vitronectin. *Nature Immunology*, 9(7):761–768.

Gratacap, R. L. and Wheeler, R. T. (2014). Utilization of zebrafish for intravital study of eukaryotic pathogen-

host interactions. *Developmental and Comparative Immunology*, 46(1):108–115.

Kopp, A., Gross, P., Falk, W., Bala, M., Weigert, J., Buechler, C., Neumeier, M., Scholmerich, J., and Schaffler, A. (2009). Fatty acids as metabolic mediators in innate immunity. *European Journal of Clinical Investigation*, 39(10):924–933.

Kuo, Z. Y., Chuang, Y. J., Chao, C. C., Liu, F. C., Lan, C. Y., and Chen, B. S. (2013). Identification of infection- and defense-related genes via a dynamic host-pathogen interaction network using a *Candida albicans*-zebrafish infection model. *Journal of Innate Immunity*, 5(2):137–152.

Leroy, Q. and Raoult, D. (2010). Review of microarray studies for host-intracellular pathogen interactions. *Journal of Microbiological Methods*, 81(2):81–95.

Lohberger, A., Coste, A. T., and Sanglard, D. (2014). Distinct roles of *Candida albicans* drug resistance transcription factors TAC1, MRR1, and UPC2 in virulence. *Eukaryotic Cell*, 13(1):127–142.

Meijer, A. H. and Spaink, H. P. (2011). Host-pathogen interactions made transparent with the zebrafish model. *Current Drug Targets*, 12(7):1000–1017.

Odds, F. C. (1979). *Candida and candidosis*. University Park Press, Baltimore.

Romani, L. (2011). Immunity to fungal infections. *Nature Reviews Immunology*, 11(4):275–288.

Schier, A. F. (2013). Genomics zebrafish earns its stripes. *Nature*, 496(7446):443–444.

Schmidt, F. and Volker, U. (2011). Proteome analysis of host-pathogen interactions: Investigation of pathogen responses to the host cell environment. *Proteomics*, 11(15):3203–3211.

Tierney, L., Kuchler, K., Rizzetto, L., and Cavalieri, D. (2012). Systems biology of host-fungus interactions: turning complexity into simplicity. *Current Opinion in Microbiology*, 15(4):440–446.

Trede, N. S., Langenau, D. M., Traver, D., Look, A. T., and Zon, L. I. (2004). The use of zebrafish to understand immunity. *Immunity*, 20(4):367–379.

Wang, Y. C., Lin, C., Chuang, M. T., Hsieh, W. P., Lan, C. Y., Chuang, Y. J., and Chen, B. S. (2013). Interspecies protein-protein interaction network construction for characterization of host-pathogen interactions: a *Candida albicans*-zebrafish interaction study. *BMC Systems Biology*, 7.

Wang, Y. C., Tsai, I. C., Lin, C., Hsieh, W. P., Lan, C. Y., Chuang, Y. J., and Chen, B. S. (2014). Essential functional modules for pathogenic and defensive mechanisms in *Candida albicans* infections. *Biomed Research International*.

Wight, T. N., Kang, I., and Merrilees, M. J. (2014). Versican and the control of inflammation. *Matrix Biology*, 35:152–161.