

Crosstalk Network Biomarkers of a Pathogen-Host Interaction Difference Network from Innate to Adaptive Immunity

Chia-Chou Wu and Bor-Sen Chen^(✉)

Control and Systems Biology Laboratory, Department of Electrical Engineering,
National Tsing Hua University, Hsinchu, Taiwan
d9761820@oz.nthu.edu.tw, bschen@ee.nthu.edu.tw
<http://www.ee.nthu.edu.tw/bschen>

Abstract. Crosstalks between host and pathogen are crucial in the infection process. To obtain insight into the defense mechanisms of the host and the pathogenic mechanisms of the pathogen, pathogen-host interactions in the infection process have become a novel and promising research subject in the field of infectious disease. In this study, two pathogen-host dynamic crosstalk networks were constructed to investigate the transition of pathogenic and defensive mechanisms from the innate to adaptive immune system in the entire infection process based on two-sided time course microarray data of *C. albicans*-zebrafish infection model and database mining. Potential crosstalk network biomarkers for the transition from innate to adaptive immunity were identified based on proteins with larger interaction variations inside the host and pathogen, and at the interface between the host and pathogen. The crosstalk network biomarkers consist of proteins with larger interaction variation scores in the pathogen-host interaction difference network. From the crosstalk network biomarkers, the molecular mechanisms of innate and adaptive immunity were successfully investigated from a systems biology perspective. In view of these results, the proposed crosstalk network biomarkers may serve as potential therapeutic targets of infectious diseases.

Keywords: Pathogen-host interaction · Dynamic crosstalk network · Crosstalk network biomarkers · Interaction difference network · Interaction variation score

1 Introduction

Our immune system protects us from deadly threats from pathogens. To fulfil the requirements, the immune system has to detect the invasion of exogenous pathogens, watch for the pathogenic conversion of endogenous microbes, communicate the threats to the other systems in our bodies, e.g., the nervous [1–4] and digestive system [5–7], and then coordinate the systems to evade the threats. Obviously, the immune system cannot function alone. In the past, the studies

regarding the immune system [8–10] focused on the molecular functions and cellular constitution of the immune system itself, and on the physiological effects of immune-related molecules and cells. However, the immune system is one part of a biological organism. Hence, from a systematic perspective, we should consider all systems as a whole, and not view the immune system in isolation.

Immune-related molecules (e.g., chemokines, cytokines, interferons, etc.) and cell types (e.g., lymphocytes, monocytes, mast cells, etc.) are commonly studied with respect to the molecular functions and cellular constitution of the immune system. After activating the first line of the host defense mechanisms (i.e., innate immunity), several cell types (e.g., macrophages, dendritic cells, natural killer cells, etc.) are recruited to protect the host from pathogen invasion and eliminate the threats from pathogens. The recognition of pathogen-associated molecular patterns (PAMPs) and/or damage-associated molecular patterns (DAMPs) by pattern recognition receptors (PRRs) (e.g., toll-like receptors, C-type lectin receptors, etc.) [10,11] can be viewed as a starting point in a series of the following complex mechanisms. The PRRs initiate downstream pathways that promote the activation of other parts of the innate immune system and the clearance of pathogens (e.g., production and secretion of cytokines, chemokines, and chemotactic cues to recruit more leukocytes). Meanwhile, the macrophages and dendritic cells are responsible for presenting antigens to induce the synthesis of the antibodies specific to the presented antigens if it is the first exposure of the host to the pathogen (Fig. 1A). If it was not the first exposure of the host to the pathogen, existing immunological memory cells proliferate and induce the synthesis of antibodies (Fig. 1A). In short, the interplays between T cells, B cells, macrophages, dendritic cells, etc. have been elaborated in detail at the physiological level. For the treatment of infectious diseases, current drug targets focus on some key molecules rather than the cellular level. Therefore, investigation of the systematic offensive and defensive mechanisms at a molecular level is the most important topic from a drug discovery and design perspective.

Compared with the host immune system, pathogenic mechanisms, not to mention interspecies protein-protein interactions (PPIs) between the host and pathogen, have attracted less attention. The battle is a two-sided affair, that is, the interplays between the host and pathogen shape the whole infection process, from the first exposure to the pathogen to the final outcomes of the infection [12]. Therefore, about a decade ago, the traditional viewpoint to treat the host and pathogen separately shifted to a more holistic viewpoint that included both players in the infection process. This viewpoint transition resulted from (i) the realization of the indispensability of pathogen-host interactions (PHIs) in infectious diseases and (ii) the advent of omics biotechnology to quantify genes, transcripts, and proteins at whole cell/organism levels [13]. This permitted a comprehensive interrogation of both the pathogen and host at the whole-genome, transcriptome, and proteome levels. Despite tremendous advances in understanding pathogenic mechanisms and the subsequent triumphs in drug development [14], the remaining issues (e.g., drug resistance) of infectious diseases have become more troublesome. The dynamic and complex interactions between

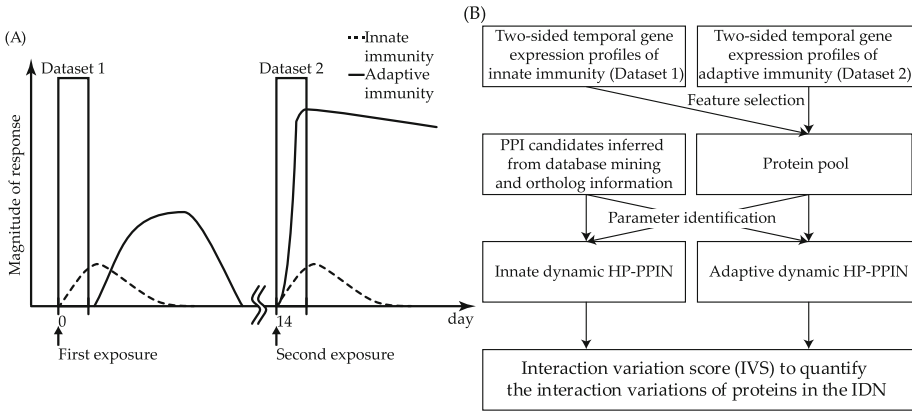


Fig. 1. Study design and the flowchart of PH-PPIN construction. (A) The first and second exposure induced the innate and adaptive immune responses, respectively, and the two-sided temporal gene expression profiles were recorded by microarray experiments (the rectangles are the observation windows of microarray experiments). (B) The flowchart delineates the procedures used in this study. Selected proteins of interest based on the microarray data formed a protein pool. The PPI candidates collected from the database mining and ortholog information were further pruned into the innate and adaptive dynamic PH-PPINs by the dynamic interaction model, system order detection method, and microarray data. Finally, the interaction variation scores were used to evaluate the significance of proteins in the interaction difference network, which was derived from the two constructed PH-PPINs.

the host and pathogen may partially explain why certain drugs are often not effective *in vivo* [15]. Hence, to investigate infection processes from a systematic perspective, in this study we constructed dynamic pathogen-host PPI networks (PH-PPINs) of innate and adaptive immunity.

To obtain systematic molecular interaction networks for targeted therapy, we utilized the *C. albicans*-zebrafish infection model [16]. We measured the temporal gene expression profiles of *C. albicans* and zebrafish during the infection process, constructed the interspecies PPIs using a dynamic interaction model, and identified the crosstalk network biomarkers with proposed interaction variation scores (Fig. 1B). Given the success of the *C. albicans*-zebrafish infection model [16] as well as its amenability to genetic manipulation [17], the zebrafish is a novel and potential model organism to study 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 [18]. This provides a strong connection between the zebrafish and humans with respect to pathogenic mechanisms as well as immune responses, which are important for biomedical applications. The immune system of zebrafish as well as other vertebrates can be divided into two subsystems: i.e., innate and adaptive immunity [11]. The first dataset (GSE32119, [19]) we used to construct a dynamic PH-PPIN measured the two-sided gene expression profiles during the first 18 hours after zebrafish was first exposed to *C. albicans* to induce primary responses.

The second dataset (GSE51603, [20]) measured the two-sided gene expression profiles during the first 42 hours after zebrafish was secondarily exposed to *C. albicans* to induce secondary responses. To extract the interaction information from the time course microarray data, two dynamic PH-PPINs were constructed for innate and adaptive immunity in the infection process. By evaluating interaction variations based on the corresponding interaction variation scores, critical proteins and crosstalk network biomarkers of larger interaction variations in the infection process were identified. These crosstalk network biomarkers suggest the strategies taken by the host and pathogen during the infection process. Thus, these crosstalk network biomarkers could be potential drug targets when battling infectious diseases [13].

2 Material and Methods

2.1 Overview of Microarray Data

In this study, we used two microarray datasets: one was the two-sided temporal gene expression profiles of the host (zebrafish) and pathogen (*C. albicans*) in the period after first exposure (GSE32119, [19]), which were used to record the pathogen-host interaction information of innate immunity; the other was the two-sided temporal expression profiles of the host and pathogen in the period after secondary exposure (GSE51603, [20]), which were used to record the pathogen-host interaction information of adaptive immunity. For the first dataset, an experiment was performed to simultaneously profile the genome-wide gene expressions of innate immunity in both *C. albicans* and zebrafish during the infection process. *C. albicans* (SC5314 strain) was intraperitoneally injected into Adult AB strain zebrafish. The second dataset measured the genome-wide gene expressions of adaptive immunity in both *C. albicans* and zebrafish after the second exposure to *C. albicans*, fourteen days after the first exposure. Then, a two-step homogenization/mRNA extraction procedure was performed using the whole zebrafish infected with *C. albicans*. This approach can provide separate pools of gene transcripts from both the host and the pathogen, which provides individual estimates of specific gene expression profiles in either the host or pathogen using sequence-targeted probes derived from the individual genomes. Agilent *in situ* oligonucleotide microarrays, which cover 6,202 and 26,206 genes for *C. albicans* and zebrafish respectively, were used to profile temporal gene expressions; the first dataset consisted of three replicates of each organism measured at 9 time points (0.5, 1, 2, 4, 6, 8, 12, 16, and 18 h post-injection), and the second dataset consisted of two replicates of each organism measured at 8 time points (2, 6, 12, 18, 24, 30, 36, and 42 h post-re-injection). Both datasets were prepared under similar experimental conditions.

2.2 Protein Pool Selection and Database Integration

There are two steps that need to be completed before a dynamic protein-protein interaction (PPI) network with a dynamic interaction model can be constructed.

The first step is to have a protein pool from which the nodes in the resultant networks will be selected, and the second step is to obtain all possible PPIs among the proteins in the protein pool by integrating the interaction information from database mining. Here, our protein pool consisted of the union of the differentially expressed genes in the first and second datasets and the differentially expressed genes between the first and second microarray datasets. The criterion used to select the differentially expressed genes in the first and second microarray datasets was to compute the p -value of the ANOVA test to determine whether the average expression levels differed over time (i.e., for the first dataset, the null hypothesis was $\mu_1 = \dots = \mu_9$, and the average expression levels were the same for all 9 time points; for the second dataset, the null hypothesis was $\mu_1 = \dots = \mu_8$, and the average expression levels were the same for all 8 time points), and then to select those proteins with a Bonferroni corrected p -value < 0.05 for inclusion in the protein pool. In addition, the genes in the top 5% of the expression difference between the first and second datasets were selected for the protein pool. Next, to know all possible interactions between the proteins in the protein pool, interaction information for the zebrafish-zebrafish, *C. albicans*-*C. albicans*, and zebrafish-*C. albicans* pairs are needed. However, the lack of information about these three kinds of interactions makes it difficult to collect all possible interactions. In addition, it is impossible to consider all interactions between the proteins in the protein pool. To overcome this issue, interaction information from human and yeast was used because of their similarity to our study subjects (zebrafish and *C. albicans*) and data availability. To infer the possible interactions of the study subjects (zebrafish and *C. albicans*), the ortholog information in the Inparanoid database [21] was used to convert the interactions of human and yeast [19,22] into the interactions of zebrafish and *C. albicans*. It should be noted that the interactions inferred from the ortholog-based method were derived under different experimental conditions. Consequently, the data do not accurately reflect the actual biological condition of the pathogen-host interactions during the *C. albicans* infection process; that is, false positive interactions exist in the complete set of inferred possible interactions of zebrafish and *C. albicans*, and these false positive interactions need to be validated and removed using real microarray data. Therefore, the false positive interactions were deleted from the candidate PPIs and realistic pathogen-host crosstalk PPI networks in innate and adaptive immunity were constructed using the two-sided microarray data and the dynamic model of PPI interaction in the following section.

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

To construct the interspecies PPI network from the protein pool and candidate PPIs, the dynamic protein-protein interaction model was used to determine the realistic PH-PPIN using individual proteins in succession. Given that the total numbers of the host and pathogen proteins are N and M , respectively, then for a host target protein i in the PH-PPIN, the dynamic interaction model is as follows [26]:

$$p_i^{(h)} [k + 1] = \sigma_i^{(h)} p_i^{(h)} [k] + \sum_{n=1}^N \alpha_{in}^{(h)} p_n^{(h)} [k] + \sum_{m=1}^M \gamma_{im} p_m^{(p)} [k] + \beta_i^{(h)} + \epsilon_i^{(h)} [k + 1] \quad (1)$$

where $p_i^{(h)} [k]$ denotes the protein level of the host target protein i at time k , $\epsilon_i^{(h)} [k]$ denotes the environmental noise at time k , $\sigma_i^{(h)}$ denotes the self-regulation ability of the host target protein i , $\alpha_{in}^{(h)}$ denotes the interaction strength between the host protein n and the host target protein i , γ_{im} denotes the interaction strength between the pathogen protein m and the host target protein i , and $\beta_i^{(h)}$ denotes the basal level of the host target protein i . Similarly, the dynamic interaction model of a pathogen target protein j can be written as follows:

$$p_j^{(p)} [k + 1] = \sigma_j^{(p)} p_j^{(p)} [k] + \sum_{m=1}^M \alpha_{jm}^{(p)} p_m^{(p)} [k] + \sum_{n=1}^N \gamma_{jn} p_n^{(h)} [k] + \beta_j^{(p)} + \epsilon_j^{(p)} [k + 1] \quad (2)$$

The biological significance of this formulation is that the protein level of the host (pathogen) target protein i (j) in the future (at time $k + 1$) is determined by its current protein level (at time k) with self-regulation ability $\sigma_i^{(h)}$ ($\sigma_j^{(p)}$), the interaction strength between the host (pathogen) target protein i (j) and the proteins of the same species $\alpha_{in}^{(h)}$ ($\alpha_{jm}^{(p)}$) and the other species γ_{im} (γ_{jn}), the basal level $\beta_i^{(h)}$ ($\beta_j^{(p)}$), and the environmental noise $\epsilon_i^{(h)}$ ($\epsilon_j^{(p)}$) in the future. Due to the unavailability of proteomic data, the expression levels measured by the two-sided microarray experiments were used to represent the protein levels in the dynamic interaction model. The dynamic interaction model for the host target protein i can be further rewritten into a concise form as follows:

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

where $\mathbf{p}_i^{(h)} = [p_i^{(h)} [1] \cdots p_i^{(h)} [K]]^T$, $\theta_i = [\alpha_{i1}^{(h)} \cdots \alpha_{iN}^{(h)} \gamma_{i1} \cdots \gamma_{iM} \sigma_i^{(h)} \beta_i^{(h)}]^T$, $\epsilon_i^{(h)} = [\epsilon_i^{(h)} [1] \cdots \epsilon_i^{(h)} [K]]^T$, and

$$\Phi_i = \begin{bmatrix} p_1^{(h)} [0] & \cdots & p_N^{(h)} [0] & p_1^{(p)} [0] & \cdots & p_M^{(p)} [0] & p_i^{(h)} [0] & 1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ p_1^{(h)} [K - 1] & \cdots & p_N^{(h)} [K - 1] & p_1^{(p)} [K - 1] & \cdots & p_M^{(p)} [K - 1] & p_i^{(h)} [K - 1] & 1 \end{bmatrix}.$$

The dynamic model for the pathogen can also be rewritten into a similar form. The only unknown parameter $\theta_i^{(h)}$ can then be estimated using parameter estimation methods, such as the least-squares estimation. However, due to the lack of large-scale measurements of host and pathogen protein levels, we used the temporal gene expression profiles as a substitute of protein activities to identify the parameter $\theta_i^{(h)}$ in the model. Furthermore, to ensure the model was not unnecessarily complex, the Akaike information criterion (AIC) was introduced to detect the true model order (the number of interactions). The true model

order with minimum AIC was considered as the criterion to delete false positive interactions in the candidate PH-PPINs. Hence, the final dynamic PH-PPINs encompass the dynamic interaction model of each protein with the minimum AIC value to remove the false positive PPIs. Finally, after identifying the parameters for each protein in the protein pools, the identified interactions parameters $(\alpha_{in}^{(h)}, \alpha_{jm}^{(p)}, \gamma_{im}, \text{ and } \gamma_{jn})$ formed the final dynamic PH-PPIN.

2.4 Interaction Variation Score (IVS) Calculation

To target the network biomarkers in the PH-PPINs, the IVSs were calculated for proteins to correlate proteins with the transition of the pathogen-host interactions from innate to adaptive immunity. The proteins in the PH-PPINs with the largest PPI variations from innate to adaptive immunity can be considered as crosstalk network biomarkers in the entire infection process and are considered as significant drug targets. Therefore, we investigated these crosstalk network biomarkers as follows. The IVS is a measurement of the variation of the interaction strength under a biological condition transition. According to the dynamic interaction models, the constructed PH-PPIN under a specific condition (innate or adaptive) can be written as follows:

$$\begin{bmatrix} p_1^{(h)} [k + 1] \\ \vdots \\ p_N^{(h)} [k + 1] \\ p_1^{(p)} [k + 1] \\ \vdots \\ p_M^{(p)} [k + 1] \end{bmatrix} = \begin{bmatrix} \sigma_1^{(h)} \cdots \alpha_{1N}^{(h)} & \gamma_{11} & \cdots & \gamma_{1M} \\ \vdots & \ddots & \vdots & \vdots \\ \alpha_{N1}^{(h)} \cdots \sigma_N^{(h)} & \gamma_{N1} & \cdots & \gamma_{NM} \\ \gamma_{11} & \cdots & \gamma_{1N} & \sigma_1^{(p)} \cdots \alpha_{1M}^{(p)} \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{M1} \cdots \gamma_{MN} & \alpha_{M1}^{(p)} \cdots \sigma_M^{(p)} \end{bmatrix} \begin{bmatrix} p_1^{(h)} [k] \\ \vdots \\ p_N^{(h)} [k] \\ p_1^{(p)} [k] \\ \vdots \\ p_M^{(p)} [k] \end{bmatrix} + \begin{bmatrix} \beta_1^{(h)} \\ \vdots \\ \beta_N^{(h)} \\ \beta_1^{(p)} \\ \vdots \\ \beta_M^{(p)} \end{bmatrix} + \begin{bmatrix} \epsilon_1^{(h)} [k + 1] \\ \vdots \\ \epsilon_N^{(h)} [k + 1] \\ \epsilon_1^{(p)} [k + 1] \\ \vdots \\ \epsilon_M^{(p)} [k + 1] \end{bmatrix} \tag{4}$$

where the notations are the same as in the dynamic interaction models. The above equation can be written in a more concise form:

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

where A is a systematic interaction matrix of the PH-PPIN constructed under a specific condition. The interaction difference of two PH-PPINs between innate and adaptive immunity can be expressed in the following interaction difference matrix form:

$$D_{\text{adaptive-innate}} = A_{\text{adaptive}} - A_{\text{innate}} . \tag{6}$$

If the variation of the interaction strength of a protein is larger during a biological condition transition (innate→adaptive immunity in this study), this may imply the protein has a more important role in the transition from innate to adaptive immunity. Therefore, the IVS used to evaluate the interaction variability of a protein in the transition from innate to adaptive immunity can be defined as follows:

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

where d_{pq} is the pq -entry of $D_{\text{adaptive-innate}}$, that is, the average interaction variation of the protein p in the transition from innate to adaptive immunity. The degree of protein p is the number of non-zero elements in the p th row of the interaction difference matrix $D_{\text{adaptive-innate}}$. Those proteins with larger IVSs are considered as significant proteins that play an important role in the transition from innate to adaptive immunity in the infection process.

3 Results

3.1 The Pathogen-Host Protein-Protein Interaction Networks (PH-PPINs) of Innate and Adaptive Immunity

In this study, we aimed to investigate the systematic offensive and defensive mechanisms of pathogen and host at the molecular level. In particular, we aimed to understand the roles of pathogen-host interactions (PHIs) in innate and adaptive immunity from a systems biology perspective. The outcomes of interactions between the host and pathogen were recorded based on the two-sided temporal gene expression profiles of *C. albicans* and zebrafish that were simultaneously measured during the primary and secondary response periods in the infection process. During the two periods (the rectangles in Fig. 1A), the observed variations in the gene expression levels were mainly due to innate and adaptive immunity, respectively. We further selected 1620 proteins of interest for the protein pool, including those with differentially expressed features and the top 5% of the expression level difference between the two datasets. The comparison of their temporal profiles (Fig. 2A) implied that their expression patterns changed: the activation of a group of pathogens genes was delayed and the repression of a group of hosts' genes was advanced. The changes in the gene expression patterns implied the PHIs in these two periods should have corresponding variations. To determine the variations of the underlying PHIs, 26060 PPI candidates inferred from the database mining and ortholog information were further pruned using the dynamic interaction models, model order detection method, and two-sided microarray data (Fig. 1B) and then the innate and adaptive dynamic PH-PPINs were formed (Fig. 2B). In particular, the two constructed PH-PPINs were the underlying mechanisms used to explain the observed changes in the gene expression patterns in the infection process.

The resultant PH-PPINs consisted of 1512 proteins (1431 *C. albicans* proteins; 81 zebrafish proteins) and 5721 PPIs (5510 intracellular interactions inside *C. albicans*; 145 interspecies interactions; 66 intracellular interactions inside zebrafish) for innate immunity, and 1578 proteins (1480 *C. albicans* proteins; 98 zebrafish proteins) and 3755 PPIs (3577 intracellular interactions inside *C. albicans*; 96 interspecies interactions; 82 intracellular interactions inside zebrafish) for adaptive immunity. Looking at the amount of variation in the nodes and edges of the pathogen, although most of the pathogenic nodes are shared between innate and adaptive immunity, the number of edges changed from 5511 to 3577: that is, only 1203 edges are shared. This implies that the pathogen may use almost the same set of proteins (85%) but with different links to interact with

the host and to regulate functions within the pathogen itself in response to various challenges from innate and adaptive immunity. In contrast, the host may use a different strategy since a different distribution of node and edge numbers was found compared with the pathogen. In the zebrafish, there are three more

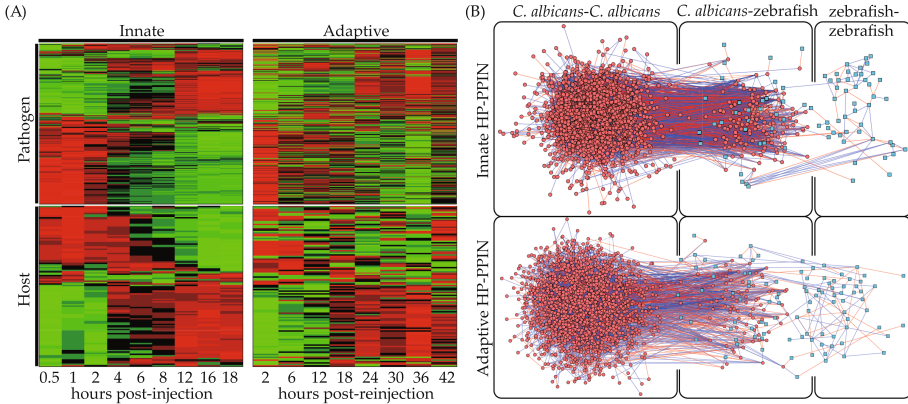


Fig. 2. Temporal gene expression profiles of the proteins of interest and the constructed innate and adaptive dynamic PH-PPINs. (A) The horizontal axis indicates the sampling time points in the microarray experiments. The vertical axis shows the genes clustered according to their expression patterns in innate immunity. (B) The innate and adaptive PH-PPINs consist of PPIs in three domains: pathogen-pathogen, pathogen-host, and host-host.

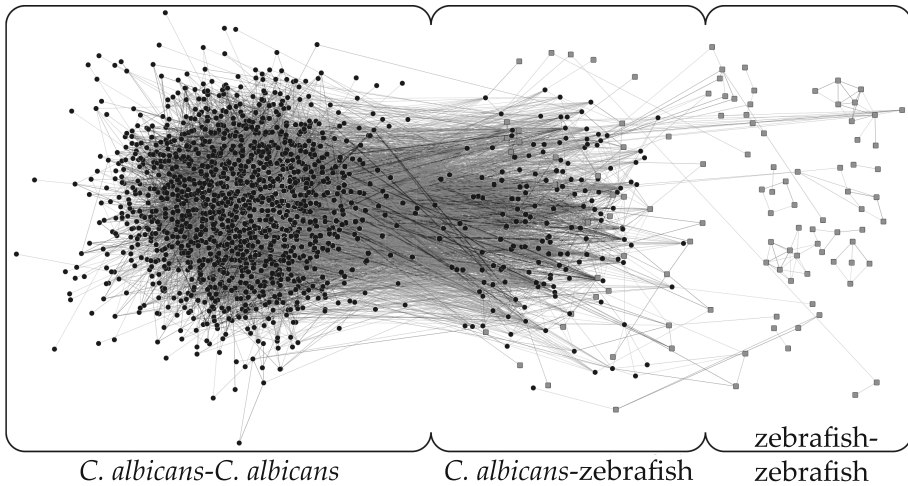


Fig. 3. The interaction difference network between innate and adaptive immunity. The IDN between innate and adaptive immunity consists of interactions in the three domains: pathogen-pathogen, pathogen-host, and host-host domains. The round and square nodes indicate the pathogen and host proteins, respectively.

significantly enriched functions (angiogenesis, coagulation, and circadian clock) in the adaptive PH-PPIN compared with the innate PH-PPINs (metabolic processes, immune responses, and apoptosis). In addition, in *C. albicans*, there are two more significantly enriched functions (circadian clock and filament growth) compared with the innate PH-PPINs (response to stimulus, redox status, and budding). The new functions in the adaptive PH-PPIN compared with the innate PH-PPIN indicated changes in the response strategies of the host and pathogen. To efficiently identify and evaluate the significance of proteins in the innate and adaptive dynamic PH-PPINs, we differentiated the two PH-PPINs into an interaction difference network (IDN) (Fig. 3), i.e., the matrix D in Eq. (6), and then used interaction variation scores (IVSs) to evaluate the interaction variations of proteins in the IDN.

3.2 Identifying Crosstalk Network Biomarkers in the IDN Using IVS

Cell signaling depends on dynamic PPINs [23]. Hence, the interaction variation in the PPINs indicates the change in cell signaling and the corresponding consequences in the cellular functions. To illustrate the variation of PPINs, we adopted the notations of node color, edge color, and edge line style as shown in Fig. 3 to illustrate the existence of proteins and their interactions, and the variation of interactions from innate to adaptive dynamic PH-PPINs. Further, to focus on the proteins with significant variations, the IVS stated in Eq. (7) was used to evaluate the average interaction variation 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 IVS can quantify the extent of the interaction variations, which may signify the importance of the proteins in the transition from innate to adaptive immunity, i.e., the IDN between innate and adaptive PH-PPINs (Fig. 3). In the following, we focused on the proteins with the ten highest IVSs in the three domains, that is, the host-host (zebrafish-zebrafish), pathogen-pathogen (*C. albicans*-*C. albicans*), and pathogen-host (*C. albicans*-zebrafish) domain, and determined the crosstalk network biomarkers in these domains.

The Crosstalk Network Biomarkers in the Host-Host Domain. In the host-host domain of the IDN, the ten proteins with the highest IVSs show 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 domain (Fig. 4A). The largest component consists of *f2*, LOC798231, LOC793315, *ace2*, *gnai1*, and their first neighbors (the left part of Fig. 4A). *gnat2*, a host G-protein that formed one end of the inter-species interaction, has connections with chemokine-related proteins (*ccl-c5a* and si:key-269d20.3) and chemotaxis-related proteins (ENSDARP00000105159 and ENSDARP0000111107). The angiogenesis- and coagulation-related proteins (*agt*, *ace2*, *f2*, and ENSDARP00000098661) are connected to these chemokine-related proteins. This component also consists of two other proteins: i.e., serine

proteinase inhibitor (*serpincl1*) and prokineticin (ENSDARP00000109666). The roles of angiogenesis, coagulation, and chemokines are manifested in innate and adaptive immunity in this component. The second component mainly consists of complements (*c7b*, *c8g*, *c8a*, *c8b*, and *c9*) and vitronectins (*vtna* and *vtnb*). Given the well-known roles of the complements system in immunity, vitronectins have recently attracted much attention in the field of immunity [24]. The *cd36* and apolipoproteins (*apob1*, *apoba*, and *apobb*) form the third component (the lower right part of Fig. 4A). CD36 plays a pivotal role in macrophage foam-cell formation and atherogenesis, which is reduced by apolipoproteins. Although the last two components are much less documented, the roles of versican (*vcamb*) and tank in inflammation have been reported [25].

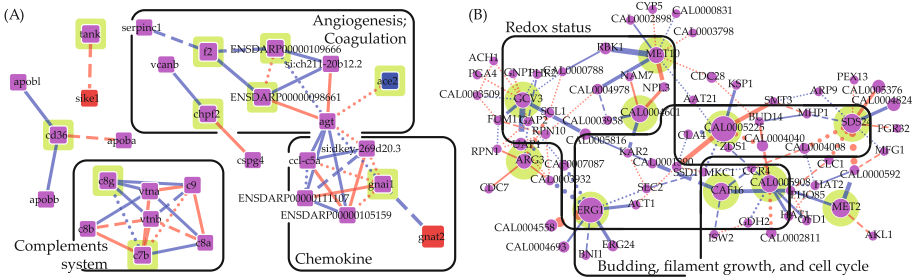


Fig. 4. The crosstalk network biomarkers in the host-host and pathogen-pathogen domains. (A) Chemokines, the complements system, and angiogenesis and coagulation are the three major crosstalk network biomarkers in the host-host domain of the IDN owing to the higher IVSs of their members. (B) Redox status and pathogen expansion are the two major crosstalk network biomarkers in the pathogen-pathogen domain of the IDN owing to the higher IVSs of their members. The shadowed nodes represent the proteins with the ten highest IVSs in their domains.

The Crosstalk Network Biomarkers in the Pathogen-Pathogen Domain.

In the pathogen-pathogen domain, the ten proteins with the highest IVSs and their first neighbors form a single component (see Fig. 4B). In this component, the importance of redox status in the innate and adaptive immune responses is re-emphasized [26]. ERG1, CAL0005908, MET10, and GCV3 are all related to the redox status of *C. albicans*. In addition, CAL0005225, ERG1, and SDS24 are related to the expansion of *C. albicans* due to their functions in budding, filament growth, and the cell cycle, respectively. In particular, MET10 is related to the response to stress from the host and environment. Another major function in this component is transferase activity. MET2 is a homoserine acetyltransferase that can transform homoserine, a toxin for *C. albicans*, and is important for *C. albicans* survival. ARG3 facilitates the production of citrulline, which can induce pseudohyphal morphogenesis. The morphological transformation of *C. albicans* has been proven to be important in its pathogenesis. The hydrolase CAF16 exerts its influence on RNA polymerase II although the specific targeted genes are still unknown.

The Crosstalk Network Biomarkers in the Pathogen-Host Domain. In the pathogen-host domain, we also selected ten proteins from both the host and pathogen. These interspecies proteins form crosstalks that are more complicated than those in the pathogen-pathogen domain (Fig. 5). A possible mechanism for the correlation between redox status in the host and pathogen is shown in the pathogen-host domain, i.e., the interaction between thioredoxin (*txn*) and ribonucleotide reductase 1 (RNR1). In addition to its role in redox status, RNR1 also influences the iron utility, filament growth, and cell cycle of *C. albicans*. This implies that the effect of redox status on the pathogen is multifaceted. Compared with chemokine-related functions in the host-host domain, the role of chemokine-related functions in the PHIs of the pathogen-host domain are more interesting. CAG1, one protein involved in how chemokine-related functions affect the pathogen, is related to the hyphal growth, mating, and biofilm formation of the pathogen, which are all important in pathogenesis. In contrast to the appearance of redox status and chemokines in the pathogen-pathogen and host-host domains, respectively, gene transcription and the circadian clock can only be seen in the pathogen-host domain. Interactions between TAF60,

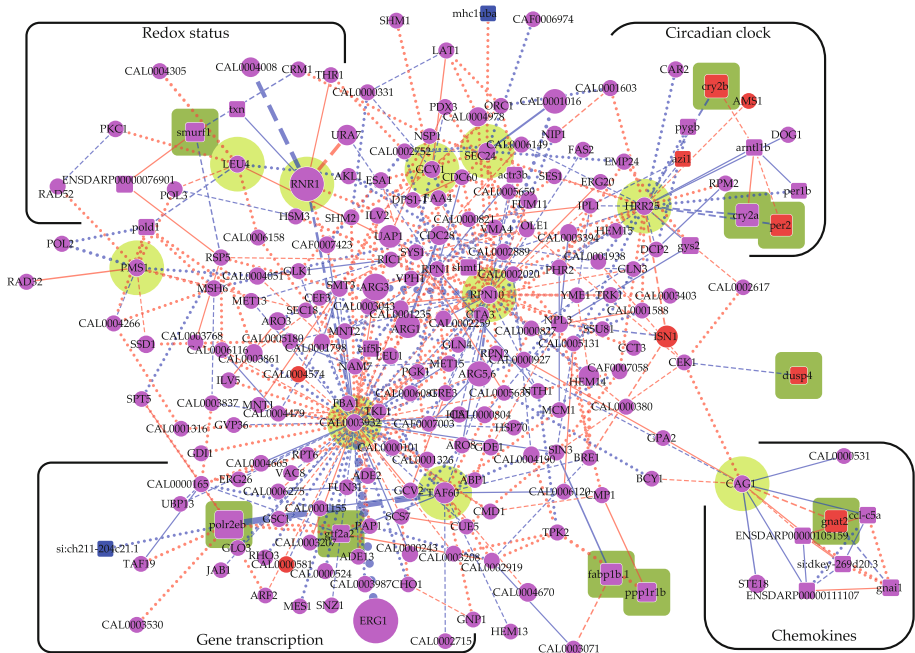


Fig. 5. The crosstalk network biomarkers in the pathogen-host domain. Redox status, circadian rhythm, gene transcriptions, and chemokines are the four major crosstalk network biomarkers in the pathogen-host domain. The shadowed nodes represent the proteins with the ten highest IVSs in the pathogen-host domain. The round and square nodes indicate the pathogen and host proteins, respectively.

gtf2a2, and *polr2e* emerged in adaptive immunity. TAF60, a transcription factor, is responsible for the drug responses in the pathogen, and *gtf2a2* and *polr2e* are related to gene transcription in the host. Their interactions indicate a possible mechanism as to how the PHIs affect the gene expression level. In addition to gene transcription, the circadian clock has an interesting function in the host and pathogen. The circadian clock-related proteins of the host (*cry2a*, *cry2b*, and *per2*) and pathogen (HRR25) form a sub-network in the host-pathogen domain. The circadian rhythms in the host and pathogen are correlated and numerous functions of the pathogen (yeast-hyphal switch, gene transcription, pathogenesis, etc.) are affected through HRR25.

In summary, we found that the proteins of the most variable interactions in the IDN are the elements related to chemokines, angiogenesis, coagulation, redox status, pathogen expansion, gene transcription, and circadian clock functions: i.e., the so-called crosstalk network biomarkers. Thus, these crosstalk network biomarkers change considerably in the transition from innate to adaptive immunity in the infection process and are potential targets for treatment and vaccination. To further evaluate the plausibility of the crosstalk network biomarkers, we selected angiogenesis, coagulation, redox status, and the circadian clock due to their systemic influence and investigated the interplay between these biomarkers based on the IDN.

4 Discussion

We presented pathogen-host protein-protein interaction networks (PH-PPINs), which were generated by dynamic interaction models and two-sided microarray data during innate and adaptive response periods. The dramatic changes in the number of PHIs from innate to adaptive immunity (145 interactions in innate immunity, 96 interactions in adaptive immunity, and 36 shared interactions) and almost the same nodes appearing in both innate and adaptive immunity suggest that the strategy used by the pathogen are characterized by the use of almost the same subset of proteins to respond to the two different defense mechanisms of the host (i.e., innate and adaptive immunity) but with different interactions. On the other hand, although the strategies used by the host were quite subtle, we can tell that the expression patterns of the host genes change from innate to adaptive immunity based on the temporal gene expression profiles (Fig. 2A). Once the lack of PPI information for zebrafish is remedied, the resultant PH-PPIN can provide further insight into the responding strategies of the host and pathogen. For now, we were able to quantitatively investigate the interaction variations from innate to adaptive immunity in the infection process by following the clues regarding the variations in the number of interaction and the interaction strengths identified in the dynamic interaction models.

To focus our investigation over a smaller and more meaningful range, we utilized IVS to evaluate the average interaction variation of a protein in the transition from innate to adaptive immunity in the infection process. The IVS rules out the possibility of a large IVS being caused by many small interaction variations, a weakness of the carcinogenesis relevance value (CRV) of [28].

Hence, the IVS could better focus on proteins with large interaction variations. Further, we visualized the interaction difference matrix from innate to adaptive immunity as an IDN (Fig. 3B), which can be divided into three domains according to the types of interactions involved. For the three domains, we focused on the proteins with the ten highest IVSs to investigate their interaction variations and determined the crosstalk network biomarkers. Not surprisingly, several immune-related and pathogenic crosstalk network biomarkers emerged: chemokines, cytokines, the complement system, pathogen expansion, and redox status. Nevertheless, three additional crosstalk network biomarkers—circadian clock, angiogenesis, and coagulation—were found for the larger interaction variations of their components. Although these functions are not totally new in immunity research, crosstalk among these crosstalk network biomarkers is a novel contribution of this study. In particular, the influences of circadian clock, redox status, angiogenesis, and coagulation are systemic. The samples and sampling time points of the microarray data provided us an opportunity to gain insight into the mechanisms of how these systemic crosstalk network biomarkers interact. The whole fish body samples provided a holistic view of the systemic variations of the transcriptomes from innate to adaptive immune response. The observation windows of the microarray experiments (Fig. 1A) revealed the involvement of the circadian clock in innate and adaptive immunity, which may be concealed if there are not enough sampling points over several days. Thus, we identified several significant proteins and crosstalk network biomarkers in the three domains based on their larger interaction variations from innate to adaptive immunity and then explored them by taking a closer look at the IDN.

In summary, our findings underpin the criticality of the circadian clock crosstalk network biomarker in terms of the type of immune response generated by an organism [27] and further show how the circadian clock, redox status, angiogenesis, and coagulation crosstalk network biomarkers are tightly coupled with pathogenesis and the host immune systems. This provides an opportunity to design new and efficient therapeutic guidelines for drug targets and the time window for treatments.

References

1. Ren, K., Dubner, R.: Interactions between the Immune and nervous systems in pain. *Nat. Med.* **16**, 1267–1276 (2010)
2. Capuron, L., Miller, A.H.: Immune system to brain signaling: Neuropsychopharmacological implications. *Pharmacol. Ther.* **130**, 226–238 (2011)
3. Ransohoff, R.M., Brown, M.A.: Innate immunity in the central nervous system. *J. Clin. Invest.* **122**, 1164–1171 (2012)
4. Nataf, S.: The sensory immune system: a neural twist to the antigenic discontinuity theory. *Nat. Rev. Immunol.* **4**, 280 (2014)
5. Sekirov, I., Russell, S.L., Antunes, L.C.M., Finlay, B.B.: Gut Microbiota in health and disease. *Physiol. Rev.* **90**, 859–904 (2010)

6. Esplugues, E., Huber, S., Gagliani, N., Hauser, A.E., Town, T., Wan, Y.S.Y., O'Connor, W., Rongvaux, A., Van Rooijen, N., Haberman, A.M., Iwakura, Y., Kuchroo, V.K., Kolls, J.K., Bluestone, J.A., Herold, K.C., Flavell, R.A.: Control of T_H17 cells occurs in the small intestine. *Nature* **475**, 514–518 (2011)
7. Lee, W.J., Hase, K.: Gut Microbiota-generated metabolites in animal health and disease. *Nat. Chem. Biol.* **10**, 416–424 (2014)
8. Sorokin, L.: The impact of the extracellular matrix on inflammation. *Nat. Rev. Immunol.* **10**, 712–723 (2010)
9. Pulendran, B., Ahmed, R.: Immunological mechanisms of vaccination. *Nat. Immunol.* **12**, 509–517 (2011)
10. Romani, L.: Immunity to fungal infections. *Nat. Rev. Immunol.* **11**, 275–288 (2011)
11. Trede, N.S., Langenau, D.M., Traver, D., Look, A.T., Zon, L.I.: The use of zebrafish to understand immunity. *Immunity* **20**, 367–379 (2004)
12. Tierney, L., Kuchler, K., Rizzetto, L., Cavalieri, D.: Systems biology of host-fungus interactions: turning complexity into simplicity. *Curr. Opin. Microbiol.* **15**, 440–446 (2012)
13. Schmidt, F., Volker, U.: Proteome analysis of host-pathogen interactions: investigation of pathogen responses to the host cell environment. *Proteomics* **11**, 3203–3211 (2011)
14. Arnold, R., Boonen, K., Sun, M.G.F., Kim, P.M.: Computational analysis of interactomes: current and future perspectives for bioinformatics approaches to model the host-pathogen interaction space. *Methods* **57**, 508–518 (2012)
15. Meijer, A.H., Spaink, H.P.: Host-pathogen interactions made transparent with the zebrafish model. *Curr. Drug Targets* **12**, 1000–1017 (2011)
16. 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., Chuang, Y.J.: Zebrafish as a model host for *Candida albicans* infection. *Infect. Immun.* **78**, 2512–2521 (2010)
17. Gratacap, R.L., Wheeler, R.T.: Utilization of zebrafish for intravital study of eukaryotic pathogen-host interactions. *Dev. Comp. Immunol.* **46**, 108–115 (2014)
18. Schier, A.F.: Genomics: zebrafish earns its stripes. *Nature* **496**, 443–444 (2013)
19. 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., Wong, D.S.H.: Dynamic transcript profiling of *Candida albicans* infection in zebrafish: a pathogen-host interaction study. *PLoS One* **8**, e72483 (2013)
20. Lin, C., Lin, C.N., Wang, Y.C., Liu, F.Y., Chuang, Y.J., Lan, C.Y., Hsieh, W.P., Chen, B.S.: The role of TGF- β signaling and apoptosis in innate and adaptive immunity in zebrafish: a systems biology approach. *BMC Syst. Biol.* **8**, 116 (2014)
21. Ostlund, G., Schmitt, T., Forslund, K., Kostler, T., Messina, D.N., Roopra, S., Frings, O., Sonnhammer, E.L.L.: InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* **38**, D196–203 (2010)
22. Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G.M., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M.R., Jassal, B., Jupe, S., Matthews, L., May, B., Palatnik, S., Rothfels, K., Shamovsky, V., Song, H., Williams, M., Birney, E., Hermjakob, H., Stein, L., D'Eustachio, P.: The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–477 (2014)
23. Yasui, N., Findlay, G.M., Gish, G.D., Hsiung, M.S., Huang, J., Tucholska, M., Taylor, L., Smith, L., Boldridge, W.C., Koide, A., Pawson, T., Koide, S.: Directed network wiring identifies a key protein interaction in embryonic stem cell differentiation. *Mol. Cell.* **54**, 1034–1041 (2014)

24. Gerold, G., Ajaj, K.A., Bienert, M., Laws, H.J., Zychlinsky, A., de Diego, J.L.: A toll-like receptor 2-integrin β_3 complex senses bacterial lipopeptides via vitronectin. *Nat. Immunol.* **9**, 761–768 (2008)
25. Wight, T.N., Kang, I., Merrilees, M.J.: Versican and the control of inflammation. *Matrix Biol.* **35**, 152–161 (2014)
26. Wang, Y.C., Lin, C., Chuang, M.T., Hsieh, W.P., Lan, C.Y., Chuang, Y.J., Chen, B.S.: Interspecies protein-protein interaction network construction for characterization of host-pathogen interactions: a *Candida albicans*-zebrafish interaction study. *BMC Syst. Biol.* **7**, 79 (2013)
27. Curtis, A.M., Bellet, M.M., Sassone-Corsi, P., O’Neill, L.A.J.: Circadian clock proteins and immunity. *Immunity* **40**, 178–186 (2014)
28. Wang, Y.C., Chen, B.S.: A network-based biomarker approach for molecular investigation and diagnosis of lung cancer. *BMC Med. Genomics* **4**, 2 (2011)