OPEN ACCESS Freely available online

PLOS ONE

Dynamic Transcript Profiling of *Candida albicans* Infection in Zebrafish: A Pathogen-Host Interaction Study

Yan Yu Chen[1☯], Chun-Cheih Chao[2☯], Fu-Chen Liu[2], Po-Chen Hsu[3], Hsueh-Fen Chen[3], Shih-Chi Peng[4], Yung-Jen Chuang[2], Chung-Yu Lan[3], Wen-Ping Hsieh[5], David Shan Hill Wong[1]*

1 Department of Chemical Engineering, National Tsing Hua University, Hsinchu, Taiwan, R.O.C., **2** Department of Medical Science & Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, Taiwan, R.O.C., **3** Institute of Molecular and Cellular Biology, National Tsing Hua University, Hsinchu, Taiwan, R.O.C., **4** Nuclear Medicine and Molecular Imaging Center, Chang Gung Memorial Hospital, Taoyuan, Taiwan, R.O.C., **5** Institute of Statistics, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.

Abstract

Candida albicans is responsible for a number of life-threatening infections and causes considerable morbidity and mortality in immunocompromised patients. Previous studies of *C. albicans* pathogenesis have suggested several steps must occur before virulent infection, including early adhesion, invasion, and late tissue damage. However, the mechanism that triggers *C. albicans* transformation from yeast to hyphae form during infection has yet to be fully elucidated. This study used a systems biology approach to investigate *C. albicans* infection in zebrafish. The surviving fish were sampled at different post-infection time points to obtain time-lapsed, genome-wide transcriptomic data from both organisms, which were accompanied with in sync histological analyses. Principal component analysis (PCA) was used to analyze the dynamic gene expression profiles of significant variations in both *C. albicans* and zebrafish. The results categorized *C. albicans* infection into three progressing phases: adhesion, invasion, and damage. Such findings were highly supported by the corresponding histological analysis. Furthermore, the dynamic interspecies transcript profiling revealed that *C. albicans* activated its filamentous formation during invasion and the iron scavenging functions during the damage phases, whereas zebrafish ceased its iron homeostasis function following massive hemorrhage during the later stages of infection. Most of the immune related genes were expressed as the infection progressed from invasion to the damage phase. Such global, inter-species evidence of virulence-immune and iron competition dynamics during *C. albicans* infection could be crucial in understanding control fungal pathogenesis.

Citation: Chen YY, Chao C-C, Liu F-C, Hsu P-C, Chen H-F, et al. (2013) Dynamic Transcript Profiling of *Candida albicans* Infection in Zebrafish: A Pathogen-Host Interaction Study. PLoS ONE 8(9): e72483. doi:10.1371/journal.pone.0072483

Editor: Joy Sturtevant, Louisiana State University, United States of America

Received March 19, 2013; **Accepted** July 10, 2013; **Published** September 3, 2013

Funding: This study was supported by grants NSC 100-2627-B-007-001, NSC100-2627-B-007-002 and NSC100-2627-B-007-003 from the National Science Council (Taiwan, Republic of China). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: dshwong@che.nthu.edu.tw

☯ These authors contributed equally to this work

Introduction

Candida albicans is a major fungal human pathogen [1,2]. Although *C. albicans* exists as a commensal organism on the cutaneous and mucosal surfaces of healthy individuals, it is also an opportunistic pathogen that can cause systemic and chronic infections. In immunocompromised patients, *C. albicans* is responsible for a number of life-threatening infections, causing considerable morbidity and mortality. Prior research has identified several *C. albicans* virulence factors, including yeast-to-hyphae morphogenesis, the expression of cell surface adhesins, and extracellular hydrolytic activities [3–5]. However, the mechanism that triggers *C. albicans* to transform from yeast to hyphae has yet to be fully elucidated [6].

Previous studies of *C. albicans* pathogenesis have suggested several steps that may lead to virulent infection, including early adhesion, invasion, and late tissue damage [7–9]. *C. albicans* may first colonize the host's epithelial surface and persist on expanding, leading to invasion and tissue damage. In the invasion stage, morphogenesis plays a significant role that involves cell transit from ovoid yeast (blastospore) to a filamentous (pseudohyphae and hyphae) form. Finally, the fungus penetrates deep into tissues and organs, causing severe damage and possible host fatality. The underlying molecular events for such a progression are expected to be regulated by the pathogen-host interaction.

The interaction between *C. albicans* and the host during infection is a complex and dynamic process. *C. albicans* utilizes different strategies to cope with the host's environmental cues for morphological change, proliferation, and escape from the host's immune defenses. The environmental conditions within the host include nutrient and chemical factors, such as the availability of metal ions and amino acids, reactive oxygen and nitrogen species, host-secreted antimicrobial peptides, antifungal drugs used in therapy, or changes in temperature and pH [10]. Several genome-wide studies have investigated these complex pathogen-host interactions. They have typically utilized a cell line-based system

to represent the host environment within an organ or tissue. These approaches include *C. albicans* infections in oral epithelial cells [8], macrophages [11], neutrophils [12], dendritic cells [13], THP-1 human mononuclear cells [14], vascular endothelial cells [15], and human blood [16]. Moreover, there are some studies on *C. albicans* infections in living hosts, e.g. mice [17,18] and rabbits [19]. However, none of these prior studies provided the whole genome response of a living host during *C. albicans* infection.

Recently, researchers have demonstrated that zebrafish (*Danio rerio*) can be a useful model for invasive fungal research. Chao et al. 2010 [20] performed two-step homogenization/mRNA extraction of a whole zebrafish infected with *C. albicans* could offer a pool of gene transcripts from both the host and the pathogen, enabling the individual estimation of specific gene expression profiles using sequence-targeted probes derived from individual genomes. Brothers et al. 2011 [21] also used zebrafish to study candidemia and that revealed how the lethal disease takes effect in zebrafish and found NADPH oxidase is required for regulation *C. albicans* filamentous formation. However, those aforementioned studies did not provide genome-wide transcript profiling data to uncover the comprehensive dynamic interaction in either the host or the pathogen. Microarray dataset of fungal infection surviving zebrafish at different post-infection time points was provided by our research group (NCBI Gene Expression Omnibus Database (GSE32119)). Kuo et al. 2013 [22] used this microarray data set to construct infection-related protein-protein interaction (PPI) network and intracellular PPI networks of the two species, the division of stages was based on histological analysis. In this present work, the same dataset was analyzed by a different method. This principal component analysis-based approach allows us to identify significant expression profiles, and classify infection stages based on transcript profiling of both zebrafish host and fungal pathogen that are consistent with the histological observations. The results also indicated that progression of infection and emergence of virulence is highly correlated with iron competition between the host and the fungal pathogen.

Results and Discussion

Dynamic gene expression analysis of *C. albicans* infection in zebrafish

Time-course gene expression analyses in both *C. albicans* and zebrafish elucidated the molecular basis of *C. albicans* infection within a living host. A total of 6,202 genes were found to exist in the expression profile of *C. albicans*. One-way ANOVA detected significant variations across nine time points. Figure S1A illustrates the numbers of significant genes identified at different *p*-values. Using a *p*-value of ≤ 0.01 as the cut-off threshold, the data filter procedures selected a total of 4,827 significant genes for *C. albicans*. As mentioned previously, hyphal morphogenesis is an important virulence factor in *C. albicans*. As a validation to the proposed approach, the present study's findings identified all key genes related to hyphal formation [10], except *NRG1*. Applying a more stringent criterion of a *p*-value less than $0.01/6202 = 1.61 \times 10^{-6}$ (with Bonferroni correction), the target gene pool was further reduced to total of 1,307 significant genes (Table S1). Gene ontology (GO) analysis of these 1,307 significant genes, based on the *Candida* Genome Database (http://www.candidagenome.org) [23], revealed that the most significant gene functions in terms of their biological process were related to transport, and the most active cellular component is the cell membrane(Table S2). This indicated that membrane transport could represent the key activated function in *C. albicans* as it grows and transforms within zebrafish during infection.

Principal component analysis (PCA) was applied to analyze the dynamic variations in these significant genes using different time points as variables. The scores of the first, second, and third PCs accounted for 79% of the observed variations in the data. The second PC accounted for an additional 12% of the variations, the third PC accounted for 5%, the fourth PC accounted for 2%, and the others accounted for less than 1%. Hence, the first two PCs could account for more than 91% of the variations in the expression dataset.

Figure 1A displays the relative contribution of different time points to the first two PCs. The first PC featured the differential expression between time points 0.5 and 1, and 12, 16 and 18 hours post infection (hpi). The second PC featured the differences between time points 2, 4, 6 and 8 h, and 0.5, 1, 12, 16 and 18 hpi. Histological analysis of the host zebrafish was performed at the exact time points corresponding with the array study. Using the host's liver as a focus, it was found that the pathogen's morphological features correlated well with the three different stages of fungal infection. In Figure 1C, within 1 hpi, *C. albicans* remained predominantly in yeast form and attached to the surface of the zebrafish liver. No tissue invasion was observed in the samples collected at 0.5 and 1 hpi. This period was thus defined as the adhesion phase. *C. albicans* began to form hyphae around 2 hpi, which marked the starting point for the morphological transition *C. albicans*. The fungal invasion into the liver tissue became evident from 4 to 8 hpi; therefore, the time points from 2 to 8 hpi were defined as the invasion phase. Tissue damage and extensive penetration became evident beyond 12 hpi, and the host survival rate decreased rapidly from this time point onward. This result was consistent with the observations reported in the group's previous study [20], and thus, the time points from 12, 16, to 18 hpi were defined as the damage phase. Here PCA analysis and the histological findings were able to classified, in a phenomenological manner, experimental time points into the adhesion, invasion, and damage phases. Nevertheless, the molecular events that occurred in and between stages may be more complex and remained to be investigated.

The scores of the first two PCs accounted for more than 90% of the variations, and Figure S2A displays the scatter plot of the two scores for these significant genes. These genes were then classified into two major groups (Groups 1 and 2) by clustering principal component scores using the K-mean method and validity index. Figure 2A shows their expression profiles. There were three subgroups (Subgroups I, II, and III) with different expression patterns in Group 1. In Subgroup I, the genes were highly expressed in the adhesion phase, but their expression levels decreased when *C. albicans* entered the invasion and damage phases. The genes in Subgroup II had high positive second PC scores, and those in Subgroup III had high negative second PC scores. The expression of the genes in Subgroup II was lowest in the invasion phase, while the expression of the genes in Subgroup III was highest in the invasion phase.

There were also three subgroups (Subgroups IV, V, and VI) in Group 2. In comparison with the aforementioned data, *C. albicans* did not significantly express most of the identified genes in Subgroup IV in the adhesion phase; instead, their expression increased in the invasion and damage phases. The genes in Subgroup V demonstrated the lowest expression levels in the invasion phase, while the genes in Subgroup VI demonstrated the highest expression levels in the invasion phase.

As described previously, *C. albicans* expressed distinct groups of genes during its interactions with zebrafish. Table 1 lists the key biological processes that demonstrated an association bias toward either Group 1 or Group 2 (A complete list shown in Table S3)

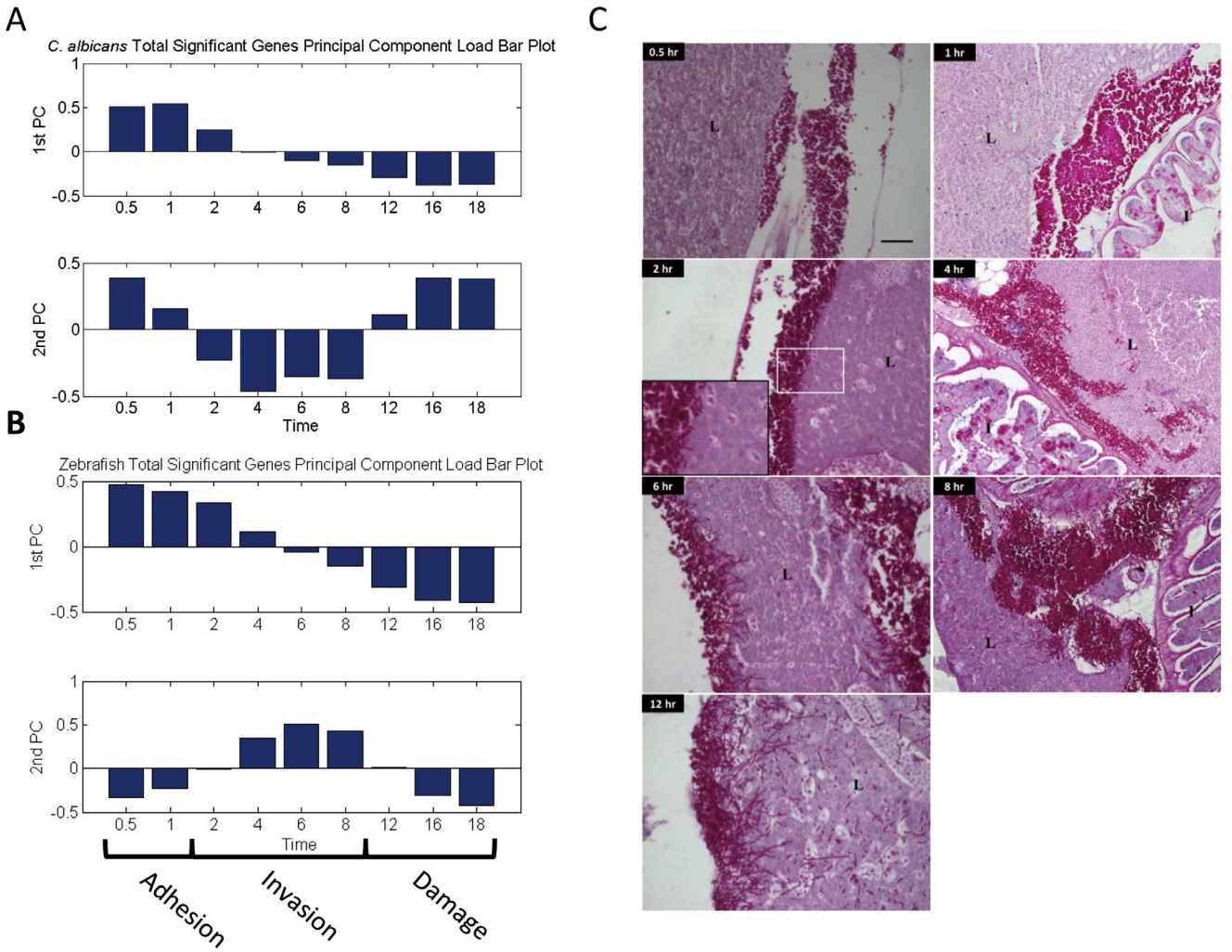

Figure 1. PC load bar plot and histology of zebrafish infected by *C. albicans*. (A) The pathogen's load contributions to the first principal component indicated that the differences in expression between 0.5 and 1 and 12, 16 and 18 hpi represented the axes of largest variation. The load contributions to the second principal components featured the gene expression differences between time points 2, 4, 6 and 8 hpi and 0.5, 1, 12, 16 and 18 hpi. Two principal components sufficiently explained more than 93% of the variations of the zebrafish expression profiles. (B) The host's load contributions to the first principal component indicated that the differences in expression between time points 0.5, 1, 2 and 4 hpi and 12, 16 and 18 hpi represented the axes of largest variation. The load contributions to the second principal components featured the gene expression differences between time points 6, 8 and 12 hpi and 0.5, 1, 2, 4, 16 and 18 hpi. (C) The histology of zebrafish liver indicated that *C. albicans* mainly grew in yeast form at 1 hpi and began to attach to the zebrafish liver. 0.5 and 1 hpi were thus defined as the adhesion phase. *C. albicans* began to transit to the hyphal form at 2 hpi. Invasion to the liver was evident at 4 to 8 hpi. Hence, time points 4 to 8 hpi were defined as the invasion phase, with 2 hpi defined as a transition point for morphogenesis. Beyond 12 hpi, tissue damage, more extensive penetration, and fish death occurred. Time points 12, 16 and 18 hpi were defined as the damage phase.
doi:10.1371/journal.pone.0072483.g001

with a *p*-value of less than 0.1. Biofilm formation, which is important in fungal pathogenesis, showed no specific bias towards group 1 or group 2 was also included as a reference. In particular, the genes related to hyphal morphogenesis, response to chemical stimuli, response to stress, organelle organization, and cell wall organization demonstrated a significant bias toward Group 1. However, the genes with a significant association bias toward Group 2 tended to function in cellular homeostasis and amino acid metabolic processes. Moreover, the response to chemical stimuli, the response to stress, organelle organization and cell walls are very common GO functions highlighted in *C. albicans*.

Analysis of *C. albicans* genes on cell surface. The list of significant *Candida* genes contained 135 cell surface-related genes, which can be divided into two groups in the PC subspace. Figure 3A showed the expression profiles of these 135 genes throughout the time course of the infection. A compiled list of the genes is shown in Table S1.

There were 17 genes (subgroups II, III, V, VI in Figure 3A) that had expression profiles particularly featured in the invasion phase. Among them, previous studies have demonstrated that *ALS3* and *INT1* encode cell surface proteins related to *C. albicans* adhesion to the host cells [24,25]. *PRA1* encodes a cell surface antigen, and disruption of the *PRA1* gene protects *C. albicans* against leukocyte killing, thus increasing *C. albicans* virulence and organ invasion *in vivo* [26]. *NGT1* encodes an N-acetylglucosamine transporter induced during hyphal growth [27], and *PGA32* encodes a putative GPI-anchored adhesion-like protein that iron transcriptionally regulates [28]. Taken together, these findings indicated the importance of iron uptake and utilization in *C. albicans* virulence.

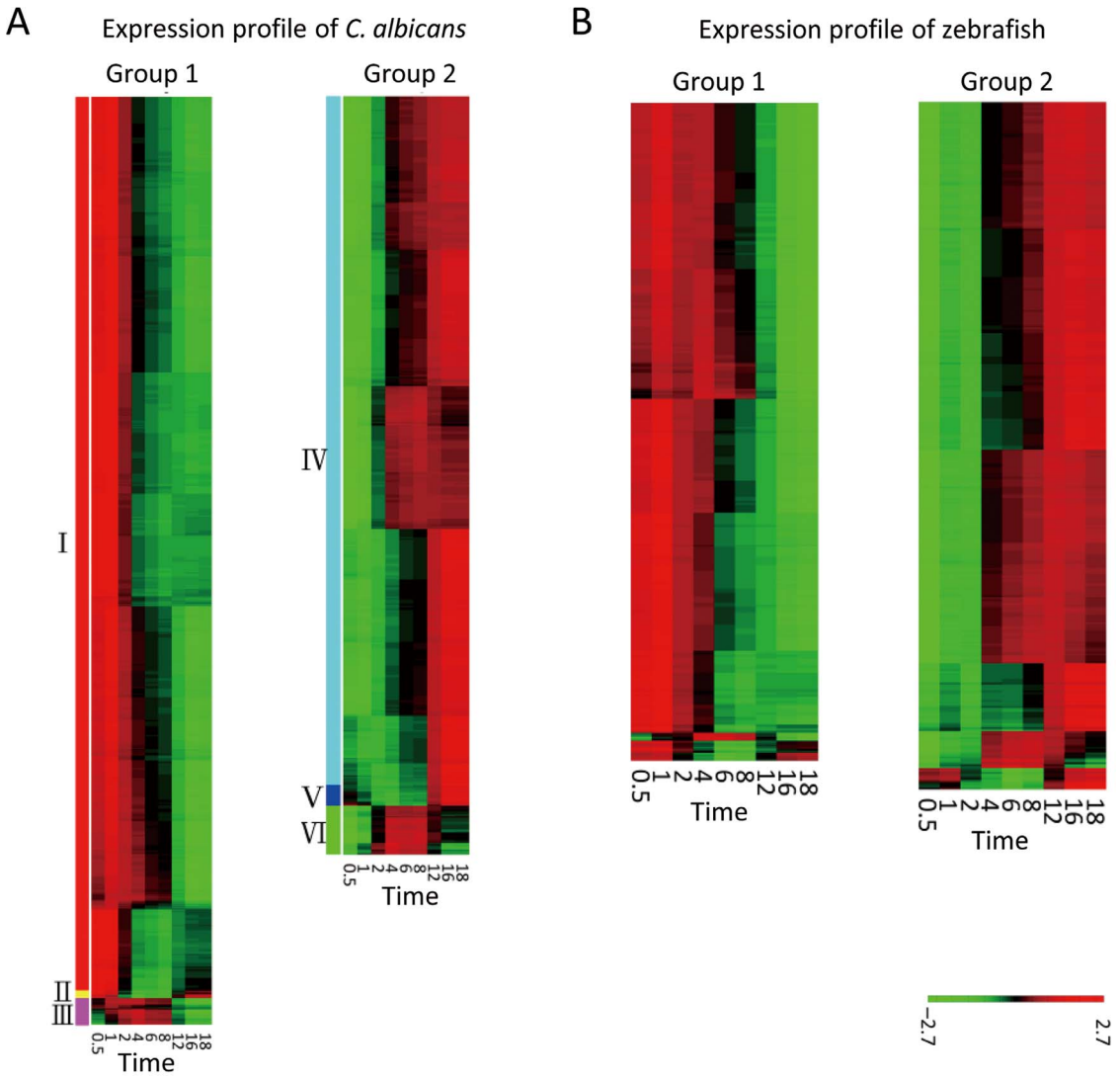

Figure 2. Significant genes profile of *C. albicans* and zebrafish. The PCA result of *C. albicans* could be classified into two main groups. The heat maps (A) revealed that *C. albicans* progressively suppressed most genes in Group 1 and progressively expressed most genes in Group 2. The genes near the top and bottom of the circle were either over-expressed or suppressed in the invasion phase. The PCA result of the zebrafish could be classified into two groups. The heat maps (B) of the two groups revealed that, in Group 1, the zebrafish progressively suppressed most genes and, in Group 2, progressively expressed most genes.
doi:10.1371/journal.pone.0072483.g002

As shown in Table S4, GO analysis of the biological processes linked to the cell surface-related genes revealed a number of biological processes that were related to hyphal formation and cell wall reorganization, which demonstrated a significant association bias toward Group 1. However, cellular homeostasis was found to demonstrate a significant bias toward Group 2. Seven out of the nine identified genes involved in this process were *RBT5*, *PGA10*, *CSA1*, *SMF12*, *HMX1* and *FTR1*, all of which are known to be involved in iron homeostasis, and *CRP1*, which has a known function in copper homeostasis. These results suggested that iron transport and scavenging are important activities of *C. albicans* during the infection process.

The above analysis was summarized in Figure 4, which includes the method used, key findings and the related tables and figures. First, significant genes were first singled out by ANOVA and Bonferroni correction. GO analysis of these genes revealed that the key function is related to membrane transport. On the other hand, PCA showed that the infection can be classified into three stages according to gene expression profiles. K-mean clustering suggested that the genes profile can be cluster into two main groups: progressively expressed and progressively suppressed. Binomial tests were used to identify important GO-terms in the two main groups. For both total significant genes and cell surface-related genes, hyphal formation and cell wall reorganization demonstrated a significant association bias toward Group 1. Cellular homeostasis was found to demonstrate a significant bias toward Group 2, and most cell surface-related genes in cellular homeostasis are iron related. Therefore, we next move to identify the profiles of hyphal-related genes and iron-related genes during infection.

Hyphal-related genes of *C. albicans*. The analysis identified hyphal growth as the fungal pathogen's primary process in the invasion and damage phases. There were a total of 122 hyphal-related genes. In agreement with prior research findings [6,8,10]

Table 1. *C. albicans* genes cluster by PCA subspace.

GO term	Group 1		Group 2	
	No. of Genes	p-value	No. of Genes	p-value
DNA metabolic process	53	0.000	2	1.000
vesicle-mediated transport	82	0.000	16	1.000
response to stress	82	0.000	33	0.999
filamentous growth	86	0.000	36	0.999
transport	152	0.000	76	0.999
hyphal growth	44	0.001	14	0.999
response to drug	55	0.001	21	0.997
response to chemical stimulus	86	0.003	41	0.995
translation	15	0.003	3	0.985
ribosome biogenesis	11	0.006	2	0.968
protein catabolic process	37	0.015	16	0.970
cellular amino acid metabolic process	19	0.943	25	0.030
cellular homeostasis	8	0.987	17	0.004

This table lists the key biological processes that demonstrated an association bias toward either Group 1 or Group 2 with a *p*-value of less than 0.05 (A complete list was shown in Table S3).

doi:10.1371/journal.pone.0072483.t001

and CGD annotation, most of these genes have been reported to contribute to *C. albicans* virulence in animal models other than zebrafish, and have thus been identified as essential virulent genes for *C. albicans*. These genes could also be divided into two groups in the PC subspace (Table S1). Figure 3B shows their expression profiles throughout the time course of the infection.

Group 1 contained 86 genes, and Group 2 had 36 genes. There were six genes that had profiles featured particularly in the invasion phase. Among them, it was found that *C. albicans* transiently suppresses *SHA3*, *FCR1* and *DFG16* in the invasion phase. In terms of gene function, *SHA3* has a known involvement in glucose transport [29], while *FCR1* encodes a putative zinc cluster transcription factor [29]. *DFG16* is required for host tissue invasion [30]. Conversely, *C. albicans* transiently expresses *CDC14*, *CLB2*, and *BUB2* in the invasion phase. In terms of gene function, *CDC14* is related to cell cycle function, *CLB2* is required for mitotic exit [31], and *BUB2* participates in the mitotic cell cycle spindle orientation checkpoint [31].

Multiple signaling and regulatory pathways control *C. albicans* yeast-to-hyphal transition. This study identified 122 significant filamentous related genes. Previous studies have mapped many of these genes onto pathways known to control morphological transition, and these genes expressed significantly during infection are arranged in Table 2. *CST20* and *CEK1* encode components of the mitogen-activated protein kinase (MAPK) signaling pathway [10]. Moreover, *CDC42*, *CDC24*, and *BEM3* gene products function upstream of the MAPK pathway [32] and are expressed during the adhesion phase. *HOG1*, a MAP kinase involved in osmotic, heavy metal, and core stress responses [33,34], is also activated during the adhesion phase. *SSK1* and *CHK1*, which are response regulators of the two-component system, are activated during the damage phase [35,36].

C. albicans expressed the downstream target of the protein kinase C (PKC) pathway, *MKC1*, which is involved in contact-induced invasive filamentation, during the adhesion phase. Of the remaining essential biofilm formation genes, *C. albicans* up-regulated *RBT5* and transiently expressed *ALS3* during the invasion phase. Of the genes involved in the pH response signaling pathway, *C. albicans* was found to express *RIM21* during the damage phase and *PHR1* during the adhesion phase. Rim21 may function as a sensor of alkaline pH [37], and Phr1 is a putative cell surface glycosidase with involvement in cross-linking of the beta-1,3- and beta-1,6-glucans [38].

The results of this study further revealed the modulation of several components in the *C. albicans* TOR pathway. This was significant, as the target of the rapamycin (TOR) signaling pathway regulates a wide variety of cellular processes in response to nutrients. Of these, *C. albicans* expressed *GLN3* during the damage phase, and expressed *RHB1*, *GAP1*, *MEP2* and *SIT4* during the adhesion phase.

C. albicans activated other hypha-related genes, such as *SIN3* and *SET3*, during the adhesion phase. *SIN3* is known to encode a key component of a specific histone deacetylase complex. It can down-regulate true hyphae formation via binding to the *EFG1* promoter [39]. Efg1 is a transcriptional regulator and activates filamentous growth in *C. albicans*. On the other hand, the gene product of *SET3* forms a histone deacetylase complex that is involved in regulating *C. albicans* morphogenesis and virulence [40].

The present study's analyses of *C. albicans* also identified genes related to calcium. It was found that *C. albicans* would express calmodulin (*CMD1*) and the catalytic subunit of calcineurin (*CMP1*). *C. albicans* also up-regulated the expression of *MID1*, which encodes a putative calcium channel of the high affinity calcium uptake system, during the damage phase. *MID1* plays a role in thigmotropism (contact sensing) and the directional hyphae tip response of *C. albicans* [41].

Taken together, the results indicated that *C. albicans* exhibits dynamic gene regulatory controls in many well-known signaling pathways to coordinate various functions during the adhesion phase, but that it will then progressively suppress them. This result was consistent with the finding of yeast-to-hyphal transition occurring in the early stage of infection (Figure 1C).

Iron-related genes of *C. albicans*. The analysis of cell surface-related genes identified iron transport and scavenging as the primary processes. The expression of these iron-related genes mainly occurred during the invasion and damage phases. Using the

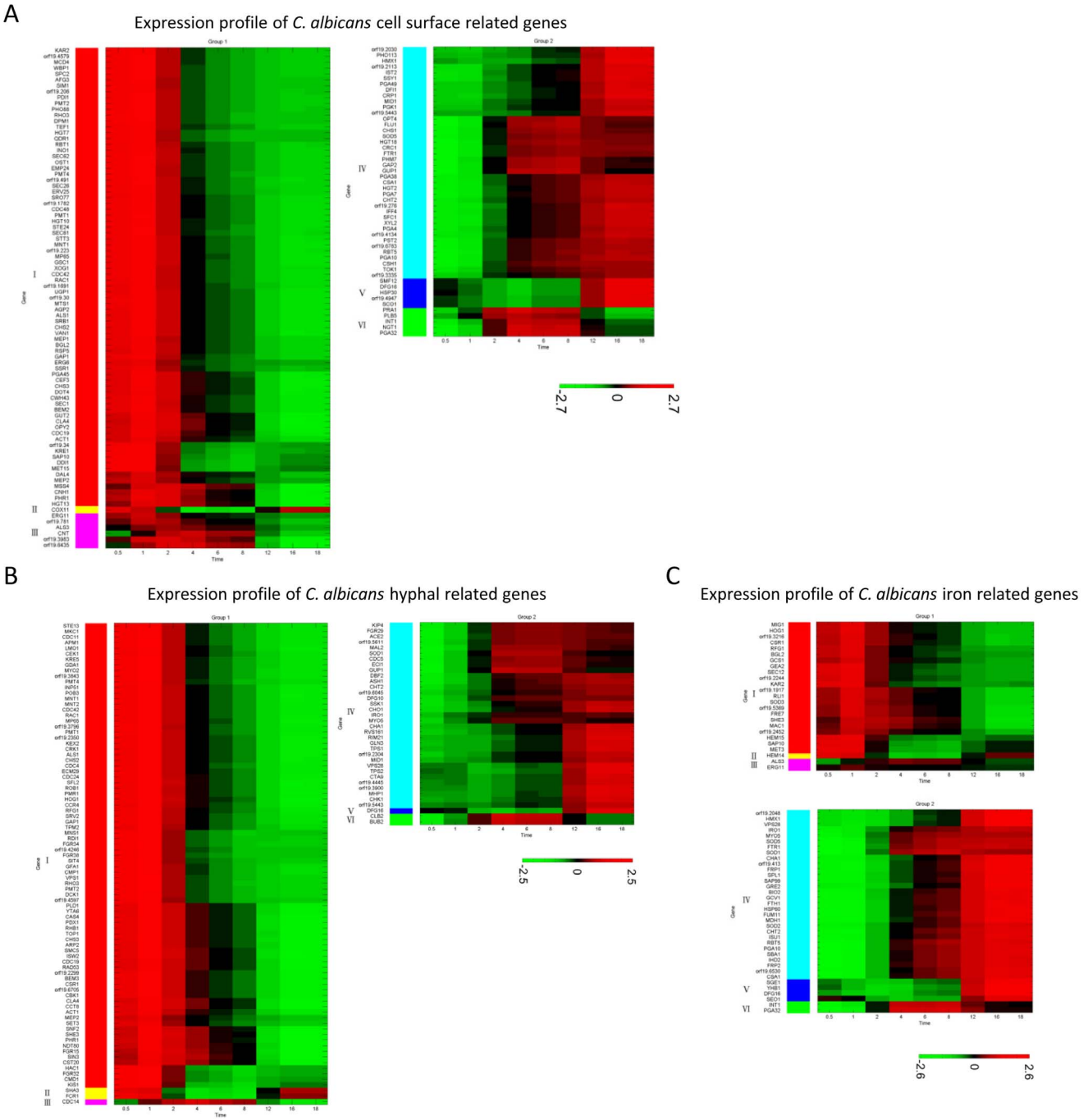

Figure 3. Expression profiles of the cell surface-related, hyphal-related, and iron-related genes in *C. albicans*. The list of significant genes contained 135 cell surface-related genes. PCA divided these into two main groups. (A) shows the cell surface-related gene expression profile. 16 genes, *COX11*, *ERG11*, *orf19.781*, *ALS3*, *CNT*, *orf19.3983*, *orf19.6435*, *SMF12*, *DFG16*, *orf19.4947*, *SCO1*, *PRA1*, *PLB5*, *INT1*, *NGT1* and *PGA32* were particularly featured in the invasion phase. The list of significant genes also contained 122 hyphal-related genes. PCA divided these into two main groups. (B) shows the hyphal-related gene expression profile. Six genes were particularly featured in the invasion phase. Of these, *C. albicans* transiently suppressed *SHA3*, *FCR1* and *DFG16* during the invasion phase. Analysis of the cell surface-related genes identified cellular hemostasis as a primary process on the cell surface, and most of these genes were iron homeostasis-related. The list of significant genes contained 61 iron-related genes. PCA divided these into two main groups. (C) shows the iron-related gene expression profile. Nine genes were particularly featured in the invasion phase. *C. albicans* transiently suppressed *HEM14*, *SGE1*, *YHB1*, *DFG16* and *SEO1*, and transiently expressed *ALS3*, *ERG11*, *INT1* and *PGA32* in the invasion phase.
doi:10.1371/journal.pone.0072483.g003

literature search approach described in Material and Methods under the specific functional gene list section, the iron-related genes list was found to contain 238 genes, 61 of which had significant expressions. Of these, 25 were in the Group 1 cluster, while 36 were in the Group 2 cluster within the PC subspace (Table S1). Figure 3C provides a list of the genes in these two groups.

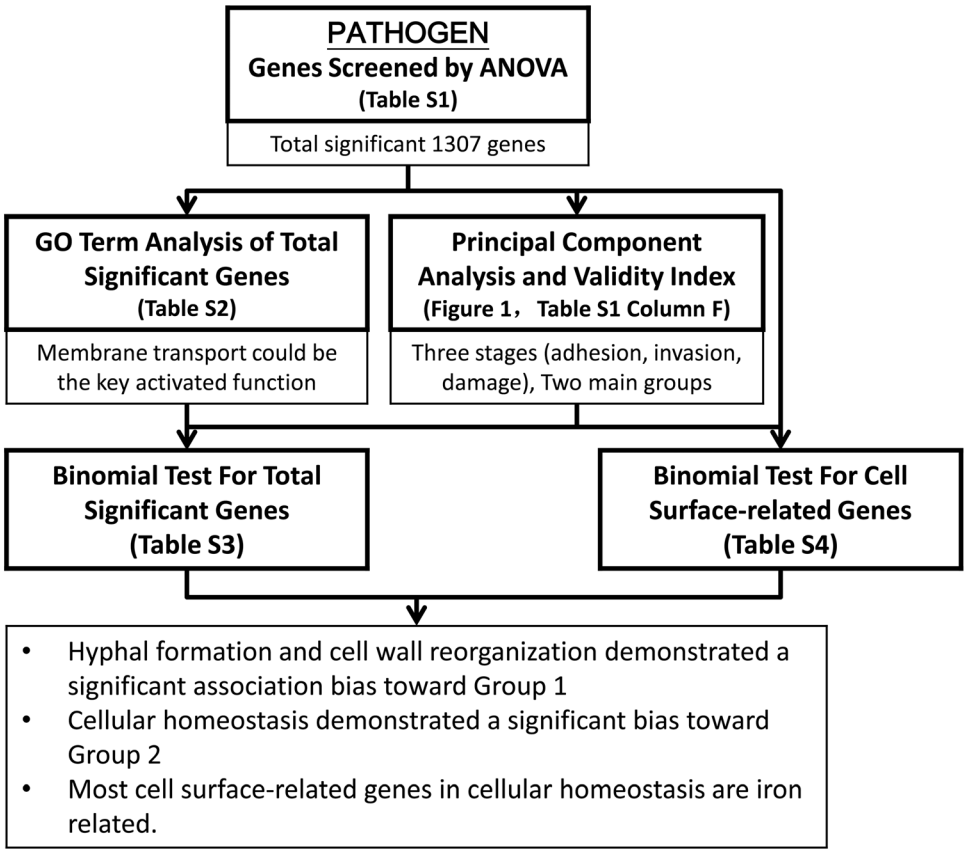

Figure 4. *C. albicans* genes analysis procedure diagram. The analysis of *C. albicans* was summarized in this figure, which includes the method used, key findings and the related Tables and Figures. First significant genes were first singled out by ANOVA and Bonferroni correction. GO analysis of these genes revealed that the key function is related to membrane transport. PCA showed that the infection can be classified into three stages according to gene expression profiles. K-mean clustering suggested that the genes profile can be cluster into two main groups: progressively expressed and progressively suppressed. Binomial tests were used to identify important GO-terms in the two main groups. For both total significant genes and cell surface-related genes, hyphal formation and cell wall reorganization demonstrated a significant association bias toward Group 1. Cellular homeostasis was found to demonstrate a significant bias toward Group 2, and most cell surface-realted genes in cellular homeostasis are iron related.
doi:10.1371/journal.pone.0072483.g004

C. albicans is known to regulate different genes to utilize and uptake iron by sensing the iron concentration in its environment. Accumulated evidence to link these genes' induction to the surrounding iron concentration has been described in previous studies [28,42–49]. Genes related to iron uptake and utilization that were expressed significantly during infection [18,28,42,45–58] are arranged in Table 3 according to whether they are expressed in high/low *ex vivo* experiments, and whether they belong to group 1 or 2 in an *in vivo* experiments. There is no correlation between two classifications. This shows the importance of understanding the *in vivo* response of iron-related genes during infection.

We also mapped present significant genes into iron-related pathway to identify the key factor during infection. There are at least three known strategies for *C. albicans* to acquire iron from its host: the siderophore uptake system, the reductive system, and the heme-iron uptake system. The primary iron source appears to be the iron-storage protein, ferritin [42]. Many of these genes have been mapped onto these pathways controlling morphological

Table 2. Significant *C. albicans* genes in signaling pathway of yeast-to-hyphae transition.

Path	Group 1		Group 2
	Progressively suppressed	**Transient**	**Progressively exppressed**
MAPK	*CST20, CEK1, CDC42, CDC24, BEM3, HOG1*		*SSK1, CHK1*
PKC	*MKC1*	*ALS3*	*RBT5*
pH	*PHR1*		*RIM21*
TOR	*RHB1, GAP1, MEP2, SIT4,*		*GLN3*

Map significant filamentous related genes onto well-known controlling morphological transition pathways.
doi:10.1371/journal.pone.0072483.t002

Table 3. Profile of iron-related genes during inection and reference.

Reference iron environment	Progressively suppressed	Transiently suppressed	Progressively expressed	Transiently expressed
High	*KAR2, BGL2, RFG1, SEC12, orf19.2452*	*HEM14, YHB1*	*MDH1, PGA10, CHT2, ISU1, orf19.2048, BIO2, orf19.413, FUM11, SBA1, SPL1, HSP60*	*ERG11, PGA32*
Low	*SAP10, GEA2, GCS1, MET3, orf19.1917, orf19.2244, orf19.3216*	*SGE1, SEO1*	*HMX1, RBT5, FTR1, SOD5, CSA1, FRP1, FTH1, CHA1, GRE2, GCV1, IHD2, SAP99, orf19.6530*	*INT1*
Unknown	*FRE7, CSR1, MIG1, MAC1, HOG1, SHE3, RLI1, SOD3, HEM15, orf19.5369*	*DFG16*	*VPS28, MYO5, FRP2, IRO1, SOD1, SOD2*	*ALS3*

Compare profile of iron-related genes with those previous researches, which link these genes' induction to the surrounding iron concentration.
doi:10.1371/journal.pone.0072483.t003

transition. Als3, which is known to play an important role in hyphae formation and invasion, is needed to acquire iron from the host's ferritin. Based on the dynamic gene expression data, it was found that *C. albicans* only transiently expressed *ALS3* in the early invasion phase, while *FTR1* and other ferric reductases (*FRP1*, *FRP2*, and *GRE2*) were progressively induced during the invasion phase. Their expressions were then sustained in the damage phase. These data indicated that the uptake of iron had already started in the early invasion phase. On the other hand, *C. albicans* expressed *HMX1* in the damage phase, and up-regulated *PGA10* and *RBT5* during the invasion and damage phases. This indicated the possible uptake of iron from the host's hemoglobin as the zebrafish suffered massive hemorrhaging during the damage phase. The microarray profiles of *ALS3*, *FTR1* (reductive system), *RBT5*, and *HMX1* (heme-iron uptake system) are shown in Figure 5. This summarized the finding that the reductive uptake was turned on during the adhesion and invasion stages, while the heme-iron uptake become important in the damage phase.

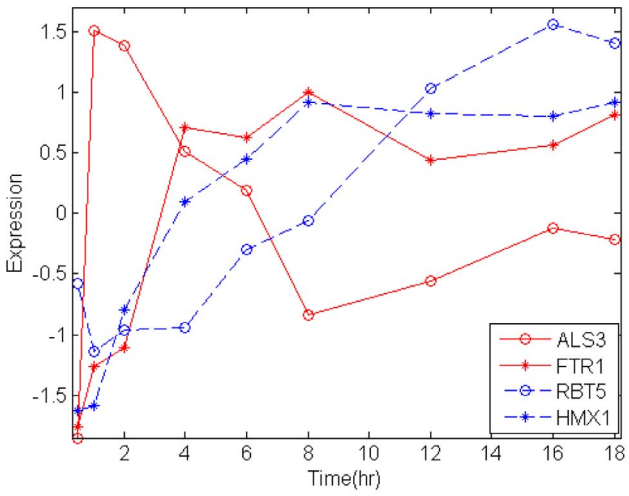

Figure 5. Expression profiles of significant *C. albicans* genes on well-known iron uptake pathway. In reductive system, *ALS3*, needed to acquire iron from the host's ferritin, transiently expressed in the early invasion phase. *FTR1*, a high-affinity iron permease, was progressively induced during the invasion phase, and expression was then sustained in the damage phase. In heme-iron uptake system, *C. albicans* up-regulated *RBT5* during the invasion and damage phases and expressed *HMX1* in the damage phase. These data indicated that the uptake of iron had already started in the early invasion phase and the possible uptake of iron from the host's hemoglobin as the zebrafish suffered massive hemorrhaging during the damage phase.
doi:10.1371/journal.pone.0072483.g005

Dynamic gene expression analysis of zebrafish host response during *C. albicans* infection

The present study applied the same statistical methods to analyze the zebrafish host's dynamic gene expression profile during *C. albicans* infection. The expression profile contained a total of 43,803 probes. Similarly, one-way ANOVA was applied to detect significant gene expression variations across the nine-point time course. Figure S1B illustrates the numbers of significant genes identified at different *p*-values. There were a total of 10,826 significant probes with a *p*-value less than 0.01. Gene annotation was carried out using the Database of Visualization and Integrated Discovery (DAVID) v6.7 [59] and ZFIN [60]. It shall be noted that, unlike the *Candida* pathogen database, there was less bioinformation available for zebrafish. The resulting list of significant genes contained 6,150 genes. Many known genes associated with host defense or immune response mechanisms were in this list. However, the following discussion focused on a list of 576 significant genes (Table S5) that fulfilled the stringent criterion of a *p*-value less than 0.01/43803 (Bonferroni correction).

PCA was applied to analyze the dynamic variations in these significant genes using different time points as variables. The scores of the first PC accounted for 84% of the observed variations in the data, while the second PC explained an additional 9% of the variations. The third PC explained 2%, and the others explained less than 1%. Therefore, the first two PCs could account for more than 93% of the variations in the expression dataset.

Figure 1B displays the contribution plots of these two principal components. The first PC featured the differences in gene expression between time points 0.5, 1, 2 and 4 hpi, and 12, 16 and 18 hpi. The second PC featured the expression differences between time points 4, 6, 8 and 12 hpi, and 0.5, 1, 2, 16 and 18 hpi. This result indicated that the invasion phase in zebrafish was slightly delayed compared to that in *C. albicans*.

Figure S2B illustrates the loci of significant zebrafish genes in the PC subspace. These could be clustered into two main groups according to the validity index. Figure 2B shows the expression profiles of the two groups. The genes in Group 1 (334 probes, 267 genes) demonstrated high expression in the adhesion phase, but this expression started to decline upon entering the invasion and damage phases. The genes in Group 2 (349 probes, 309 genes) demonstrated low expression in the adhesion phase, but this expression increased in the invasion and damage phases.

Gene ontology analysis of these 576 significant zebrafish host genes was performed using DAVID. The analysis revealed that the cluster of genes with the highest enrichment score was related to iron binding (Table S6). The cluster with the second highest enrichment score was related to hemostasis. The high enrichment score relating to hemostasis was supported by observations of frequent internal bleeding during the experiments.

Iron-related genes of zebrafish. The previously mentioned analyses of the *C. albicans* gene expression profiles identified iron transport and scavenging as important processes during infection. Interestingly, GO analysis of the significant genes in the host zebrafish also indicated that iron homeostasis was important biological process too. Of the 576 significant zebrafish genes, 24 displayed GO associated with iron binding and iron transport.

Figure 6A shows the expression profiles of the iron-related genes throughout the infection time course. Group 1 contained 18 genes that the zebrafish progressively suppressed, and Group 2 contained six genes that were progressively expressed, including.

It was worth mentioning that the zebrafish progressively suppressed most of the iron-related genes during infection. It should also be noted that most of the iron-related genes were categorized based on informatic annotations instead of experimental data. Nonetheless, one previous wet bench study has identified *SLC25A37* to be related to iron transporter activity and hemeprotein synthesis [61].

Immune-related genes of zebrafish. The immune response is a key process for zebrafish to defend themselves against *C. albicans* infection. Of the 576 significant genes identified, 13 had GO related to immune response and immune system development.

Figure 6B shows the expression profiles of these genes throughout the time course of the infection. Group 1 contained three genes. Previous studies have demonstrated that *HBL4* and *VTNB* encode components of the extracellular matrix [62–64]. *HBL4* and *VTNB* are involved in damaged cell repair and are significantly expressed during the adhesion phase, but become progressively suppressed in the damage phases. Group 2 contained the other 10 genes. Among them, *IL1B, TNFA* and *TNFB* are well-known marker genes of the immune response. Further analysis also identified the expression of the growth factor gene *FGF21*, which implied that zebrafish may activate a healing mechanism to repair the damage caused by *C. albicans* infection. This was supported by the observed up-regulation of *RUNX1*, which may promote the production of blood cells following tissue penetration by *C. albicans*. It was noteworthy that the observed immune response occurred immediately after the initial adhesion phase of the infection.

In contrast to previous studies that used cell lines, tissues and organs to pass as a host system for microbial infection, this study utilized a whole organism with complete immune capability. The GO analysis revealed that only a small set of known immune-related genes were significantly expressed the zebrafish. This indicated that the immune response was only one of numerous simultaneous biological activities vigorously regulated in the host during infection. The complex interaction between the immune system and the other physiological processes in the host warrants further investigation in the future.

The above analysis was summarized in Figure 7, which includes the method used, key findings and the related Tables and Figures. First, the significant genes were singled out by ANOVA and Bonferroni correction. GO analysis of these genes revealed that the key function is related to iron-related function. PCA and K-mean clustering suggested that the genes profile can be cluster into two main groups: progressively expressed and progressively suppressed. Finally, binomial tests were used to identify the bias of specific functions. Iron-related genes demonstrated a significant

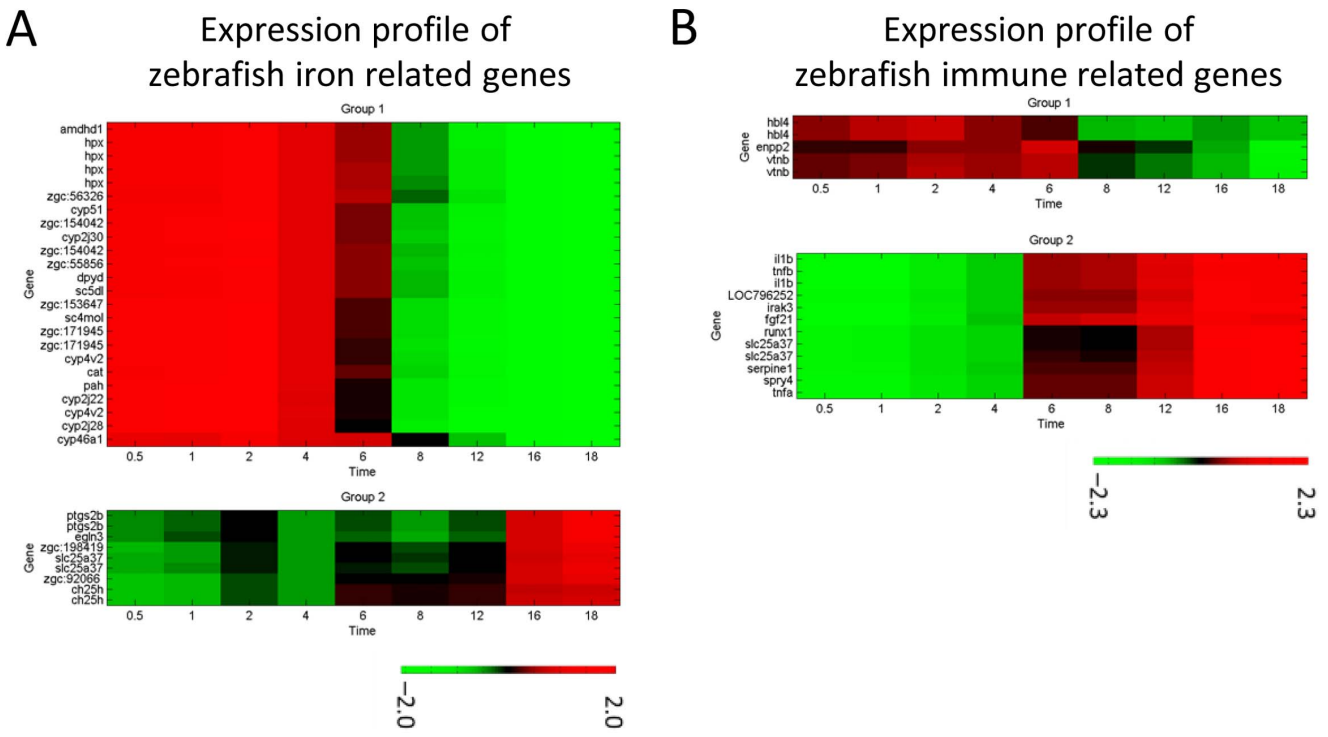

Figure 6. Expression profiles of the zebrafish immune-related and iron-related genes. The immune response is a key process for zebrafish to defend themselves against *C. albicans* Of the 576 significant genes identified, 13 genes had GO related to the immune response and immune system development. PCA divided these into two main groups. (A) shows the immune-related gene expression profile. Analysis of the *C. albicans* expression profiles identified iron transport and scavenging as important factors during infection. GO analysis also revealed that iron homeostasis was an important biological process among the significant genes. There were 24 genes in the 576 significant zebrafish genes that had GO associated with iron binding and iron transport. PCA divided these into two main groups. (B) shows the iron-related gene expression profile.
doi:10.1371/journal.pone.0072483.g006

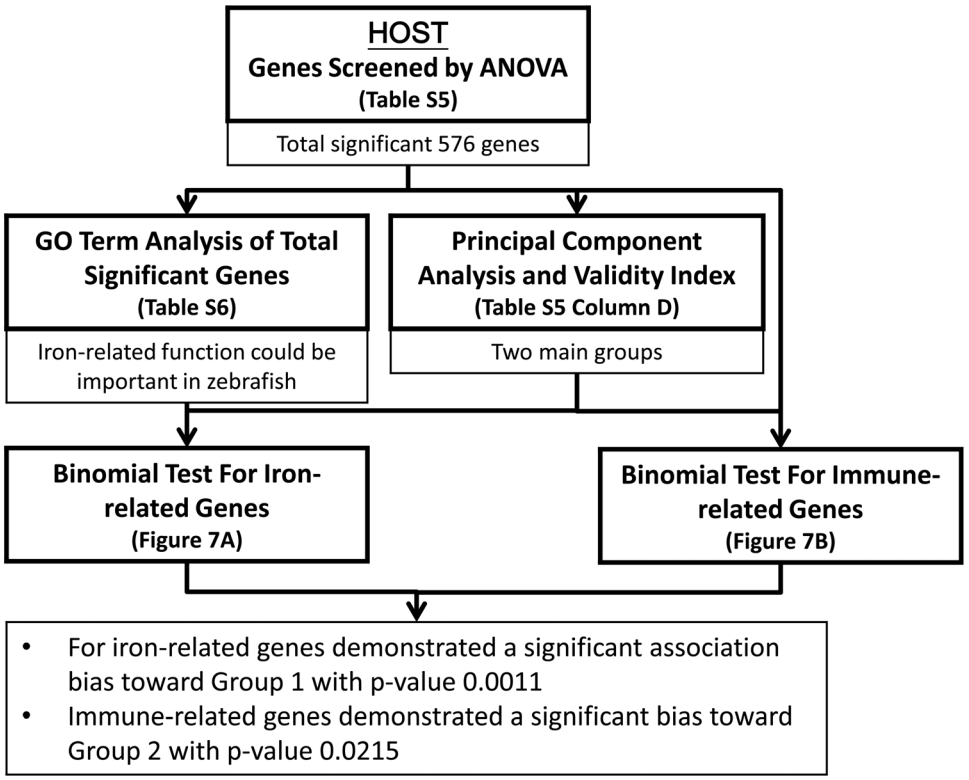

Figure 7. Zebrafish genes analysis procedure diagram. The analysis of zebrafish was summarized in this figure, which includes the method used, key findings and the related Tables and Figures. First significant genes were first singled out by ANOVA and Bonferroni correction. GO analysis of these genes revealed that the key function is related to iron-related function. PCA and K-mean clustering suggested that the genes profile can be cluster into two main groups: progressively expressed and progressively suppressed. Finally, binomial tests were used to identify the specific function, iron-related and immune-related. Iron-related genes demonstrated a significant bias toward Group 1 p-value 0.0011. Immune-related genes demonstrated a significant association bias toward Group 2 with p-value 0.0215.
doi:10.1371/journal.pone.0072483.g007

bias toward Group 1 p-value 0.0011. Immune-related genes demonstrated a significant association bias toward Group 2 with p-value 0.0215. These results indicated that the onset of immune in the host is highly correlated with failure of iron regulation.

Comparison between *Candida* gene expression profiles in different hosts

This present study also compared Candida genome expression profiles in other host models to investigate possible virulence genes. There are several published studies providing accessible transcription profiles of *C. albicans* during infection in different living hosts, such as mice and rabbits [17–19,65]. This study chose the *C. albicans* genome-wide data set from Thewes et al. (2007) [18], whose microarray dataset has a complete Candida gene list and a 4 time point (0, 0.5, 3, 5hpi) series profile generated from samples collected from the livers of mice infected with Candida. To ensure consistency, this study applied the proposed method to re-analyze Thewes's gene expression dataset and identify two groups of pathogen genes, including 288 up-regulated genes and 185 down-regulated genes. Figure 8A shows that the total intersection of the conserved genes from this work and Thewes et al. consisted of 130 genes. Figure 8B and Figure 8C show 39 up-regulated genes and 45 down-regulated genes had the same expression profile in both the zebrafish and the mouse models. The entire list is shown in Table S7.

Among the 84 genes with common dynamic expression profiles, the functions of the genes with the top 15 fold change were studied. In this list, *SAP5* is associated with the utilization of protein as nitrogen source and virulence [66]. *RHR2*, regulated by *HOG1* and *NRG1*, is a response to biofilm and hyphal [67]. *SNZ1*, *CHT2*, *ZRT1*, and *PCK1* are associated with hyphal growth [68–70]. *SIM1*, described as a macrophage-down-regulated gene, was repressed both in mice and fish. In terms of the iron-related genes, *SOD5*, *PGA7*, *FTR1*, and *RBT5* were identified. *FTR1* is a ferric reductase, *RBT5* is known as a hemoglobin receptor, and *PGA7* is described as part of a hemoglobin-receptor gene family [71]. *orf19.6084*, *orf19.449*, and *orf19.5848* are described as biofilm-induced genes [72]. There is no description for *orf19.215*, but it is an ortholog to *ScAVO2*, which is related to Tor2p kinase and other proteins.

Interestingly, there were 29 genes out of 1307 genes related to temperature in the fish model, and 14 genes out of 473 genes in mouse model. There were only five temperature-related genes found in both models (*MYO2*, *SRV2*, *PKC1*, *SMT3* and *MKC1*). A close examination of their expression profiles revealed that only *MYO2* was progressively suppressed in both models, and that *C. albicans* up-regulated the other four genes in mice and down-regulated them in fish. This may have been caused by the body temperature difference between mice (36.9 degree Celsius) and zebrafish (28 degree Celsius). In a previous study, *MYO2* was found to be not only related to temperature but also to cytoskeletal polarity, migration, and hyphal growth, therefore *MYO2* may not be directly affected by temperature [73,74].

To summarize, *C. albicans* will turned on a set of the specific virulence genes, which includes genes related to pathogenesis and iron homeostasis, in both fish and mouse models. These genes not

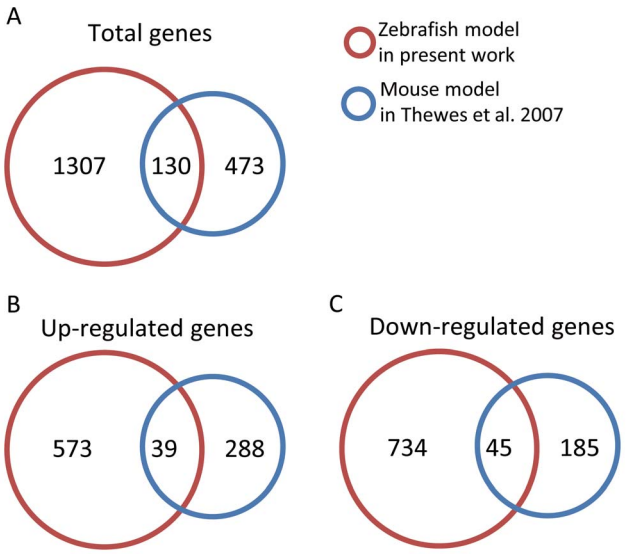

Figure 8. Comparison of gene expressions of mouse and zebrafish models during *C. albicans* infection. The present work was compared with the *C. albicans* genome wide data set in Thewes et al. (2007) [18]. (A) shows that the intersection of this work and that of Thewes et al. had 130 genes. The proposed method was used to re-analyze the gene expression profiles and cluster them into two groups that contained288 up-regulated genes and 185 down-regulated genes. (B) and (C) show that 39 up-regulated genes and 45 down-regulated genes had the same expression profile in the zebrafish model.
doi:10.1371/journal.pone.0072483.g008

only activated the hyphal growth, but also help the pathogen to uptake iron in the host's microenvironment. It should be pointed out that our dataset include the whole body host response to pathogen, but the dataset of Thewes et al. (2007) was an organ-specific response. Hence there must be caution against too much emphasis on the similarity between the datasets. However, these *Candida* genes with common expression profiles in different host species may indeed be truly important during *C. albicans* infection.

Conclusions

This study analyzed a set of time-lapse genome expression data between the *Candida* pathogen and zebrafish that captured snapshots of dynamic molecular events during infection. It was found that *C. albicans* turned on yeast-to-hyphae morphogenesis to invade the host, and that many cell surface-related and hyphal-related genes were activated during the early stages of infection.

As expected, the zebrafish progressively induced their immune systems to defend against the aggressor. One of the host's immune defenses was to restrict the iron available to the pathogens [75–77]. Many of the iron-related genes were suppressed in the late phase. A contrary example, *SLC25A37*, which is responsible for pumping iron out of the cell for hemooprotein synthesis [61], was consistent with the hypothesis that the zebrafish host was trying to sequester iron at the initial phase of infection, and that its late expression in the damage phase was caused by massive hemorrhaging. Also, the zebrafish expressed immune related genes much more significantly during the late stage of infection. In response to this defense mechanism, the reductive and heme-iron uptake were found to be important strategies for *C. albicans* to acquire iron from the host [42]. The genes responsible were progressively induced, especially after the collapse of the iron restricting mechanism in the fish host. The expressions of *FTR1*, a reductase in the reductive system, and *RBT5*, a hemoglobin receptor in the heme-iron system, were found to be similar in both the fish and mouse models. The microarray profiles were validated using real-time PCR. The expression patterns of the *RBT5*, *CHT2*, *MNT1* and *PMT1 C. albicans* genes (Figure 9A) and the *TNFA*, *SLC25A37*, *VTNB* and *CYP4V2* zebrafish genes (Figure 9B) were consistent with those found in the DNA microarray analyses.

Kuo et al. 2013 [22] study the host-pathogen interaction dynamics, based on available data-base information of PPI network. A time series modeling method and Adaike Information Criterion were used to determine the active inter-species and intra-species networks in adhesion stage and invasion stage, respectively. It shall be noted that the division of the two stages was phenomenologically based. The top ranked proteins found for *C. albicans* are mainly related to hyphal growth. The top ten proteins with active interactions in the host are mainly related to apoptosis and innate immunity. In contrast, the present study used the gene expression profiles to support statistically the phenomenological classification of the fungal infection processes into distinct phases (adhesion, invasion, and damage) throughout the pathogen's infection episode. Furthermore, we suggest iron competition is important for the emergence and progression of pathogen virulence. The pathogen invasion process is accompanied by the failing of iron homeostasis function in the host and the switching of iron uptake pathway of the pathogen. Although the interaction between the immune system and other physiological processes of zebrafish require further investigation in the future, this present work demonstrated the importance of correlation of intra-species regulatory activities from stage to stage. Such knowledge may facilitate the development of novel approaches to attenuate *C. albicans* virulence.

Materials and Methods

Candida albicans and zebrafish

Adult AB strain zebrafish were used for the experiments. Their maintenance and care after infection were performed according to the procedures described previously [20]. All of the zebrafish use and experimental protocols in this research were reviewed and approved by the Institutional Animal Care and Use Committee of National Tsing Hua University (IRB Approval NO. 09808). *C. albicans* SC5314 strain was used for injection and was prepared as described previously [20].

Infection and histological analysis

For the microarray analysis, each fish was intraperitoneally injected with 1×10^8 *C. albicans* cells suspended in 10 µl sterile 1X phosphate-buffered saline (PBS) after being anesthetized using 0.17 g/ml of tricaine (Sigma). At each time point (0.5, 1, 2, 4, 6, 8, 12, 16 and 18 hpi), three infected fish were collected, sacrificed by immersion in ice water, frozen in liquid nitrogen, and then maintained at –80°C. Although using whole infected fish to obtain transcriptional profiles could result in the loss of specific information, such as tissue and organ data, it allowed the investigation of the entire genome response during *C. albicans* infection. For the survival assay, the total injected yeast form cell number was 10^8 cells. For histological analysis, fish were collected at designated time points, and then fixed, sectioned, stained and observed according to the methods used previously [20].

C. albicans and zebrafish RNA purification

The zebrafish infected with *Candida albicans* were treated with Trizol® reagent (Invitrogen, USA), ground in liquid nitrogen using a small mortar and pestle, and then disrupted using a MagNA Lyser system (Roche) that used glass beads (cat. no. G8772-100G,

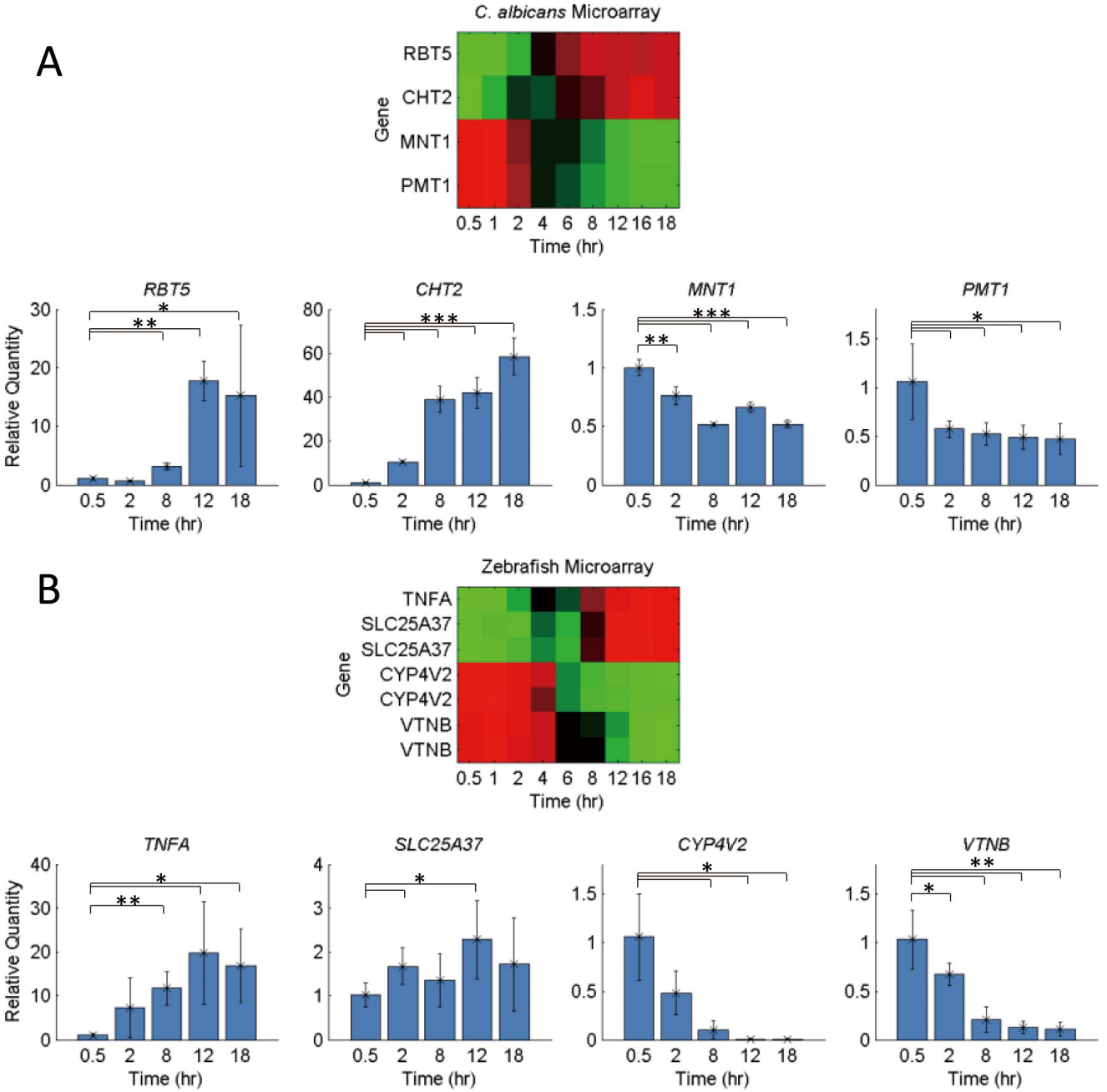

Figure 9. Validation of expression patterns by real-time polymerase chain reaction. (A) The *C. albicans* gene expression patterns (*RBT5*, *CHT2*, *MNT1*, and *PMT1*) revealed by the DNA microarray are shown on the top. Real-time PCR of *C. albicans* genes (RNA isolated from three infected zebrafish per group) are shown in the bottom. *C. albicans* expressed *RBT5* to use iron in the invasion and damage phase (after 6 hpi), expressed *CHT2* in the late stage (after 12 hpi), and activated *MNT1* and *PMT1* in the adhesion phase (0.5 and 2 hpi). (B) The zebrafish gene expression patterns (*TNFA*, *SLC25A37*, *VTNB*, and *CYP4V2*) revealed by the DNA microarray are shown on the top. Real-time PCR of the zebrafish genes (RNA isolated from three infected zebrafish per group) are shown on the bottom. The zebrafish progressively expressed *TNFA* and *SLC25A37* during the infection and turned on *VTNB* and *CYP4V2* in the adhesion phase (0.5, 1 hpi). The profiles of the real time PCR were identical and could validate the present genome microarray in both the host and the pathogen. The asterisks above the bar indicate that the level of target gene expression was significantly different from the levels of expression in the other time points: * = $P<0.2$; ** = $P<0.05$; *** = $P<0.01$.
doi:10.1371/journal.pone.0072483.g009

Sigma) shaken at 5,000 rpm for 15 sec. After phase separation was performed by adding chloroform, the total RNA was purified using an RNeasy Mini Kit (Qiagen, Germany). Purified RNA was quantified at OD260nm using an ND-1000 spectrophotometer (Nanodrop Technology, USA) and qualitated using a Bioanalyzer 2100 (Agilent Technology, USA) with an RNA 6000 nano labchip kit (Agilent Technologies, USA).

Preparation of *C. albicans* custom array and zebrafish array

For the *C. albicans* gene expression analysis, probes were designed using eArray of Agilent technologies. Along the design process, 6,205 *Candida albicans* transcripts (Assembly 21, from the *Candida* genome database,) were uploaded to eArray and designed using base composition methodology. Duplicated target sequences were removed and the resulting 6,202 probes were designed. The custom microarray was manufactured in an 8×15k format by in situ synthesis of the oligonucleotide probes (Agilent Technology). Each array consisted of 6,202 *Candida albicans* specific probes, which were printed in duplicate. The zebrafish gene expression analysis used Agilent zebrafish V2 oligo arrays (Agilent Technology), which were manufactured in a 4×44K format. Control experiments were performed to that ensure the *C. albicans* samples would not cross-hybridize to the zebrafish microarrays and that the zebrafish samples would not cross-hybridize to the *C. albicans* microarrays. In this experiment, the cross-hybridization between different samples and microarrays was not significant (data not shown).

Microarray experiments

A total of 1 μg of the total RNA was amplified using a QuickAmp labeling kit (Agilent Technologies, USA) and labeled with Cy3 (CyDye, PerkinElmer, USA) during the *in vitro* transcription process. A total of 0.625 μg of Cy3 cRNA for the *C. albicans* array and 1.65 μg of Cy3 cRNA for the zebrafish array was fragmented to an average size of approximately 50 to 100 nucleotides through incubation in a fragmentation buffer at 60°C for 30 minutes. The fragmented labeled cRNA was then hybridized to an oligo microarray at 60°C for 17 h. After washing and drying using a nitrogen gun, the microarrays were scanned using an Agilent microarray scanner (Agilent Technologies, USA) at 535 nm for Cy3. The scanned images were analyzed using Feature extraction 9.5.3 software (Agilent Technologies, USA), and image analysis and normalization software was employed to quantify the signal and background intensities for each feature. The raw data were uploaded onto the NCBI Gene Expression Omnibus Database (GSE32119).

Microarray data analysis

The set of zebrafish arrays were processed separately from the *C. albicans* arrays. The Loess normalization curve for the *C. albicans* data was fitted using every probe on the array, while the Loess curve for the zebrafish data was fitted using the 485 control probes. After the pre-processing step with Loess normalization, the mean intensity levels for each of the 6,202 pairs of duplicated probes on the *C. albicans* array were calculated as representatives. For each of the probes, one-way ANOVA was adopted to screen for significant time variations across the nine time points in either set of data. The significance level was adjusted using the Bonferroni correction, which is used for data sets with extremely large number of samples, to achieve a family-wise error rate of 0.01. The adjustment resulted in 1,307 significant probes (out of the total 6,202) for the *C. albicans* data (Table S1) and 683 probes (out of the total 43,803) for the zebrafish data (Table S5).

Principal component analysis

Principal component analysis (PCA) uses an orthogonal transformation to convert a set of sample observations of potentially correlated variables into a set of values of uncorrelated variables called principal components [78]. Let microarray as $[X]_{I\times J}$ be a set of $\mathcal{J}$-dimension multivariate data with sample points I: $[X]_{I\times J}=[t]_{I\times M}[p]_{M\times J}+[\varepsilon]_{I\times J}$.

It can be compressed into a set of M-dimension ($M<\mathcal{J}$) reduced-order data $[t]_{I\times M}$ known as scores by a transformation defined by the matrix $[p]_{M\times J}$, and ε is noise. There were nine PCs in p that could explain the microarray variations. In this study PCA was used to analyze the variations of significant genes by treating different time points as variables and different genes as samples. The relative contributions of the the m^{th} PC can be computed as:

$$relative\,\text{contribution}=\frac{tr\left(\mathbf{t}_m\mathbf{t}_m^T\right)}{tr\left(\mathbf{X}_m\mathbf{X}_m^T\right)}\times 100\%$$

The principle component analysis first identified that load vector with the largest contribution, and other load vectors with decreasing contributions in a sequential manner. Using this method, the key dynamic features of the data set can be captured.

K-means clustering and validity index

The distributions of genes in the M-dimension PC subspace can be clustered into different groups by standard K-mean clustering analysis of the score $[t]_{I\times M}$. In K-mean analysis the number of clusters must be specified a priori. The validity index [79] was used to determine the number of cluster required. Let $z_j,k=1...K$ be the centers of K clusters $\{C_1...C_K\}$ determined in the K-mean analysis. The validity ratio is defined as the ratio of average intra-group distance, i.e. the average distance between sample points and their cluster centers, and intergroup distance, i.e. the minimum distance betrween two cluster centers:

$$\Psi(K)\equiv\frac{\frac{1}{\mathcal{J}}\sum_{k=1}^{K}\sum_{\mathbf{t}_j\in C_k}\left\|\mathbf{t}_j-\mathbf{z}_k\right\|}{\min\limits_{\substack{k,k'\in\{1\cdots K\}\\ k<k'}}\left(\left\|\mathbf{z}_k-\mathbf{z}_{k'}\right\|\right)^2}$$

$$\text{inter}=\min_{k<k'}\left(\|z_k-z_{k'}\|\right)^2,k,k'=1,2,..,K$$

The optimum number of cluster K^* is determined by minimalizing the validity index $K^*=\arg\min_K\psi(K)$.

Binomial test for group gene ontology annotation

We have determined that the microarray data can categorized into two major groups in terms of dynamic response. The binomial tests were used to determine whether a process was substantially biased toward a certain group. Let q represent the probability of finding a gene in Group 1. For example, there were 1,307 significant genes in *C. albicans*, of which 734 were in Group 1 and 573 were in Group 2; hence, $q=734/1307$. Let n represent the total number of genes related to a specific biological process, e.g. $n=55$ for the DNA metabolic process as shown in Table 1. Let m represent the number of genes related to this process found in Group 1, e.g. $m=53$ for or the DNA metabolic process as shown in Table 1. The p-value of bias toward Group 1 was defined as the cumulative probability of m or fewer genes being distributed into Group 1, when n genes are randomly distributed into Group 1 and Group 2 according to the probabilities q and $1-q$:

$$p-value = 1-\sum_{i=0}^{m}\binom{n}{i}q^{i}(1-q)^{n-i}$$

For the example of DNA metabolic process, the p-value is 0.000. The smaller the p-value, the smaller is the probability of observing the given distribution by chance. Hence the bias of DNA metabolic process towards group 1 is definite, as well as the bias of cellular homeostasis towards group 2 (a p-value of 0.004 in Table 1). On the other hand, data in Table 1 showed that we cannot determine whether genes related to biofilm process is bias toward group 1 (p-value 0.430) or group 2 (p-value 0.383). For regular statistical testing with small number of samples, a p-value of 0.05 is used. This p-value was thus used in the binomial test for identification of bias distributions of a small set of genes among two different groups.

Specific functional gene list

To collect all possible receptors and sensors on the *C. albicans* membrane, text-mining in the GCD was performed, using the keywords "plasma membrane", "cell surface", and "cell wall", in order to research the GO terms "transport" and "membrane" from Table S1. The specific keywords identified a total of 533 genes, which were denoted as cell surface-related genes.

Hyphal-related biological processes were also researched by text-mining using the keywords "filamentous growth", "hyphal growth", and "pseudohyphal growth". This identified 431 genes, which were named filamentous related genes.

As shown in Table S4, the significant biological processes demonstrating the bias toward Group 2, progressive expression, was cellular homeostasis. Conducting searches for possible previously identified iron-induced genes, iron phenotypes genes, predicted iron-related genes [18,46,50–55], iron-responsive regulators [56–58], iron uptake genes [42], iron utilization genes [45,47–49] and iron-modulated genes [28] obtained 238 genes, which were named iron-related genes.

RNA isolation and quantitative real-time PCR

The control fish and the infected fish (infected with 10 µl 1×10^{8} CFU/ml *C. albicans* cells) were sacrificed by immersion in ice water at different time points and homogenized in liquid nitrogen. The total RNA was extracted using TRIzol reagent according to the manufacturer's instructions (Invitrogen), and the RNA quality was analyzed by gel electrophoresis. For cDNA synthesis, 6.25 µg of the total RNA was pretreated with DNase I (Invitrogen) and then reverse transcribed using the SuperScript III enzyme (Invitrogen) in a 50 µl reaction mixture. Quantitative real-time PCR was carried out using the 7500 real-time PCR system (Applied Biosystems). The primers used in this study are listed in Table S8. Briefly, each 15 µl reaction mixture contained 25 ng cDNA, 7.5 µl SYBR green PCR master mixture (Applied Biosystems), 0.2 µM forward primer, and 0.2 µM reverse primer. The reactions were performed using 1 cycle of 95°C for 10 min followed by 40 cycles of 95°C for 15 s and 60°C for 1 min.

Supporting Information

Figure S1 *C. albicans* and zebrafish genes at different p-values. (A) Venn diagram of the significant genes in *C. albicans*. There were a total of 4,827 significant genes with a *p*-value of less than 0.01, and there were 4,827 significant genes with a *p*-value of less than 0.1/6202 (Bonferroni correction). This study focused on 1,307 genes that fulfilled the stringent criterion of a *p*-value less than 0.01/6202. This list included several genes related to *C. albicans* hyphal formation, such as *MIG1*, *RFG1*, *ECE1*, and *CEK1*. (B) Venn diagram of the significant genes in zebrafish. There were a total of 36,292 significant genes with a *p*-value of less than 0.01, and there were 6,150 significant genes with a *p*-value of less than 0.1/43803. This study focused on 683 genes that fulfilled the same criterion used in the *C. albicans* analyses; namely, a *p*-value less than 0.01/43803 (Bonferroni correction). This gene list included key genes related to the immune response, such as *IL1B*, *TNFA*, and *TNFB*.
(TIF)

Figure S2 Clustering of significant genes in the PC subspace. (A) A scatter plot of the PC scores of the first two principal components demonstrating that the significant *C. albicans* genes could be classified into two main groups and six sub-groups. (B) A scatter plot of PC scores of the first two principal components demonstrating that the significant *C. albicans* genes could be classified into two main groups.
(TIF)

Table S1 *C. albicans* significant genes list. Applying a stringent criterion of a *p*-value less than $0.01/6202 = 1.61\times10^{-6}$ (with Bonferroni correction) further reduced the target gene pool and identified a total of 1,307 significant genes. The gene list also shows classifications according to groups in the PC subspace and cell-surface, iron and hyphal related functions.
(XLS)

Table S2 *C. albicans* gene ontology analysis. Gene ontology annotation of 1,307 genes performed using the Candida Genome Database and the GO Slim Mapper terms "biological process" and "cellular components". The most significant biology process GO set related to transport. The most significant cellular component GO set related to the cell membrane.
(XLS)

Table S3 GO comparison of total *C. albicans* significant genes group. The table shows the GO of identified biological processes and demonstrating a significant bias in Group 1 or Group 2. Binomial distribution was used to calculate the *p*-value, assuming that significant genes of a specific process were randomly distributed between two groups.
(XLS)

Table S4 GO comparison of *C. albicans* cell surface related genes group. The table shows the GO of identified biological processes of *C. albicans* cell surface related genes and demonstrating a significant bias in Group 1 or Group 2.
(XLS)

Table S5 Zebrafish significant genes list. A list of 576 significant genes that fulfilled the stringent criterion of a *p*-value less than 0.01/43803 (Bonferroni correction); the same criterion used in *C. albicans* data analyses. The compiled list also shows classifications according to groups in the PC subspace.
(XLS)

Table S6 Zebrafish gene ontology analysis. Gene ontology analysis result of 576 significant zebrafish host genes using DAVID.
(XLS)

Table S7 Identical expression profile of *C. albicans* genes in mice host. The proposed method was used to re-analyze the gene expression profiles in the *C. albicans* genome wide data set in Thewes et al. (2007) [18] and cluster them into two

groups of 288 up-regulated genes and 185 down-regulated genes. Among them, 39 up-regulated genes and 45 down-regulated genes had the same expression profile in the zebrafish model. The 84 genes are shown in this table.
(XLS)

Table S8 Sequences of primers used for real-time quantitative PCR. The primers used in this study are listed in this table.
(XLS)

References

1. Calderone RA (2002) Candida and candidiasis. Washington, D.C.: ASM Press. 451 p.
2. Odds FC (1979) Candida and candidosis. Baltimore: University Park Press. 382 p.
3. Gow NA, Brown AJ, Odds FC (2002) Fungal morphogenesis and host invasion. Curr Opin Microbiol 5: 366–371.
4. Navarro-Garcia F, Sanchez M, Nombela C, Pla J (2001) Virulence genes in the pathogenic yeast Candida albicans. FEMS Microbiol Rev 25: 245–268.
5. Sundstrom P (2002) Adhesion in Candida spp. Cell Microbiol 4: 461–469.
6. Hube B (2004) From commensal to pathogen: stage- and tissue-specific gene expression of Candida albicans. Curr Opin Microbiol 7: 336–341.
7. Naglik JR, Challacombe SJ, Hube B (2003) Candida albicans secreted aspartyl proteinases in virulence and pathogenesis. Microbiol Mol Biol Rev 67: 400–428, table of contents.
8. Wachtler B, Wilson D, Haedicke K, Dalle F, Hube B (2011) From attachment to damage: defined genes of Candida albicans mediate adhesion, invasion and damage during interaction with oral epithelial cells. PLoS One 6: e17046.
9. Wilson D, Thewes S, Zakikhany K, Fradin C, Albrecht A, et al. (2009) Identifying infection-associated genes of Candida albicans in the postgenomic era. FEMS Yeast Res 9: 688–700.
10. Biswas S, Van Dijck P, Datta A (2007) Environmental sensing and signal transduction pathways regulating morphopathogenic determinants of Candida albicans. Microbiol Mol Biol Rev 71: 348–376.
11. Lorenz MC, Bender JA, Fink GR (2004) Transcriptional response of Candida albicans upon internalization by macrophages. Eukaryotic Cell 3: 1076–1087.
12. Rubin-Bejerano I, Fraser I, Grisafi P, Fink GR (2003) Phagocytosis by neutrophils induces an amino acid deprivation response in Saccharomyces cerevisiae and Candida albicans. Proc Natl Acad Sci U S A 100: 11007–11012.
13. Huang Q, Liu DY, Majewski P, Schulte LC, Korn JM, et al. (2001) The plasticity of dendritic cell responses to pathogens and their components. Science 294: 870–875.
14. Barker KS, Liu T, Rogers PD (2005) Coculture of THP-1 human mononuclear cells with Candida albicans results in pronounced changes in host gene expression. J Infect Dis 192: 901–912.
15. Barker KS, Park H, Phan QT, Xu L, Homayouni R, et al. (2008) Transcriptome profile of the vascular endothelial cell response to Candida albicans. J Infect Dis 198: 193–202.
16. Fradin C, Kretschmar M, Nichterlein T, Gaillardin C, d'Enfert C, et al. (2003) Stage-specific gene expression of Candida albicans in human blood. Mol Microbiol 47: 1523–1543.
17. Andes D, Lepak A, Pitula A, Marchillo K, Clark J (2005) A simple approach for estimating gene expression in Candida albicans directly from a systemic infection site. J Infect Dis 192: 893–900.
18. Thewes S, Kretschmar M, Park H, Schaller M, Filler SG, et al. (2007) In vivo and ex vivo comparative transcriptional profiling of invasive and non-invasive Candida albicans isolates identifies genes associated with tissue invasion. Mol Microbiol 63: 1606–1628.
19. Walker LA, Maccallum DM, Bertram G, Gow NA, Odds FC, et al. (2009) Genome-wide analysis of Candida albicans gene expression patterns during infection of the mammalian kidney. Fungal Genet Biol 46: 210–219.
20. Chao CC, Hsu PC, Jen CF, Chen IH, Wang CH, et al. (2010) Zebrafish as a model host for Candida albicans infection. Infect Immun 78: 2512–2521.
21. Brothers KM, Newman ZR, Wheeler RT (2011) Live Imaging of Disseminated Candidiasis in Zebrafish Reveals Role of Phagocyte Oxidase in Limiting Filamentous Growth. Eukaryotic Cell 10: 932–944.
22. Kuo Z-Y, Chuang Y-J, Chao C-C, Liu F-C, Lan C-Y, et al. (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: 137–152.
23. Skrzypek MS, Arnaud MB, Costanzo MC, Inglis DO, Shah P, et al. (2010) New tools at the Candida Genome Database: biochemical pathways and full-text literature search. Nucleic Acids Res 38: D428–432.
24. Bendel CM, Kinneberg KM, Jechorek RP, Gale CA, Erlandsen SL, et al. (1999) Systemic infection following intravenous inoculation of mice with Candida albicans int1 mutant strains. Mol Genet Metab 67: 343–351.
25. Hoyer LL, Payne TL, Bell M, Myers AM, Scherer S (1998) Candida albicans ALS3 and insights into the nature of the ALS gene family. Curr Genet 33: 451–459.
26. Soloviev DA, Jawhara S, Fonzi WA (2011) Regulation of Innate Immune Response to Candida albicans Infections by alpha(M)beta(2)-Pra1p Interaction. Infection and Immunity 79: 1546–1558.
27. Alvarez FJ, Konopka JB (2007) Identification of an N-acetylglucosamine transporter that mediates hyphal induction in Candida albicans. Mol Biol Cell 18: 965–975.
28. Lan CY, Rodarte G, Murillo LA, Jones T, Davis RW, et al. (2004) Regulatory networks affected by iron availability in Candida albicans. Mol Microbiol 53: 1451–1469.
29. Uhl MA, Biery M, Craig N, Johnson AD (2003) Haploinsufficiency-based large-scale forward genetic analysis of filamentous growth in the diploid human fungal pathogen C.albicans. EMBO J 22: 2668–2678.
30. Barwell KJ, Boysen JH, Xu W, Mitchell AP (2005) Relationship of DFG16 to the Rim101p pH response pathway in Saccharomyces cerevisiae and Candida albicans. Eukaryot Cell 4: 890–899.
31. Bensen ES, Clemente-Blanco A, Finley KR, Correa-Bordes J, Berman J (2005) The mitotic cyclins Clb2p and Clb4p affect morphogenesis in Candida albicans. Mol Biol Cell 16: 3387–3400.
32. Yamaguchi Y, Ota K, Ito T (2007) A novel Cdc42-interacting domain of the yeast polarity establishment protein Bem1. Implications for modulation of mating pheromone signaling. J Biol Chem 282: 29–38.
33. Alonso-Monge R, Navarro-Garcia F, Molero G, Diez-Orejas R, Gustin M, et al. (1999) Role of the mitogen-activated protein kinase Hog1p in morphogenesis and virulence of Candida albicans. J Bacteriol 181: 3058–3068.
34. San Jose C, Monge RA, Perez-Diaz R, Pla J, Nombela C (1996) The mitogen-activated protein kinase homolog HOG1 gene controls glycerol accumulation in the pathogenic fungus Candida albicans. J Bacteriol 178: 5850–5852.
35. Calera JA, Zhao XJ, Calderone R (2000) Defective hyphal development and avirulence caused by a deletion of the SSK1 response regulator gene in Candida albicans. Infect Immun 68: 518–525.
36. Chauhan N, Inglis D, Roman E, Pla J, Li D, et al. (2003) Candida albicans response regulator gene SSK1 regulates a subset of genes whose functions are associated with cell wall biosynthesis and adaptation to oxidative stress. Eukaryot Cell 2: 1018–1024.
37. Davis D (2003) Adaptation to environmental pH in Candida albicans and its relation to pathogenesis. Curr Genet 44: 1–7.
38. Fonzi WA (1999) PHR1 and PHR2 of Candida albicans encode putative glycosidases required for proper cross-linking of beta-1,3- and beta-1,6-glucans. J Bacteriol 181: 7070–7079.
39. Tebarth B, Doedt T, Krishnamurthy S, Weide M, Monterola F, et al. (2003) Adaptation of the Efg1p morphogenetic pathway in Candida albicans by negative autoregulation and PKA-dependent repression of the EFG1 gene. J Mol Biol 329: 949–962.
40. Hnisz D, Majer O, Frohner IE, Komnenovic V, Kuchler K (2010) The Set3/Hos2 histone deacetylase complex attenuates cAMP/PKA signaling to regulate morphogenesis and virulence of Candida albicans. PLoS Pathog 6: e1000889.
41. Brand A, Shanks S, Duncan VM, Yang M, Mackenzie K, et al. (2007) Hyphal orientation of Candida albicans is regulated by a calcium-dependent mechanism. Curr Biol 17: 347–352.
42. Almeida RS, Wilson D, Hube B (2009) Candida albicans iron acquisition within the host. FEMS Yeast Res 9: 1000–1012.
43. Arnaud MB, Costanzo MC, Skrzypek MS, Binkley G, Lane C, et al. (2005) The Candida Genome Database (CGD), a community resource for Candida albicans gene and protein information. Nucleic Acids Res 33: D358–363.
44. Arnaud MB, Costanzo MC, Shah P, Skrzypek MS, Sherlock G (2009) Gene Ontology and the annotation of pathogen genomes: the case of Candida albicans. Trends Microbiol 17: 295–303.
45. Baek YU, Li M, Davis DA (2008) Candida albicans ferric reductases are differentially regulated in response to distinct forms of iron limitation by the Rim101 and CBF transcription factors. Eukaryot Cell 7: 1168–1179.
46. Homann OR, Dea J, Noble SM, Johnson AD (2009) A phenotypic profile of the Candida albicans regulatory network. PLoS Genet 5: e1000783.

Acknowledgments

The authors would also like to thank the NTHU-NHRI Zebrafish Core Facility for its support.

Author Contributions

Conceived and designed the experiments: CCC FCL PCH HFC YJC CYL. Performed the experiments: CCC FCL PCH HFC. Analyzed the data: YYC SCP DSHW WPH. Contributed reagents/materials/analysis tools: CCC FCL PCH HFC YYC. Wrote the paper: YYC YJC CYLDSHW.

47. Hsu PC, Yang CY, Lan CY (2011) Candida albicans Hap43 is a repressor induced under low-iron conditions and is essential for iron-responsive transcriptional regulation and virulence. Eukaryot Cell 10: 207–225.
48. Johnson DC, Cano KE, Kroger EC, McNabb DS (2005) Novel regulatory function for the CCAAT-binding factor in Candida albicans. Eukaryot Cell 4: 1662–1676.
49. Murad AM, d'Enfert C, Gaillardin C, Tournu H, Tekaia F, et al. (2001) Transcript profiling in Candida albicans reveals new cellular functions for the transcriptional repressors CaTup1, CaMig1 and CaNrg1. Mol Microbiol 42: 981–993.
50. Albrecht A, Felk A, Pichova I, Naglik JR, Schaller M, et al. (2006) Glycosylphosphatidylinositol-anchored proteases of Candida albicans target proteins necessary for both cellular processes and host-pathogen interactions. J Biol Chem 281: 688–694.
51. Almeida RS, Brunke S, Albrecht A, Thewes S, Laue M, et al. (2008) The Hyphal-Associated Adhesin and Invasin Als3 of Candida albicans Mediates Iron Acquisition from Host Ferritin. Plos Pathogens 4.
52. Ramanan N, Wang Y (2000) A high-affinity iron permease essential for Candida albicans virulence. Science 288: 1062–1064.
53. Sarthy AV, McGonigal T, Coen M, Frost DJ, Meulbroek JA, et al. (1997) Phenotype in Candida albicans of a disruption of the BGL2 gene encoding a 1,3-beta-glucosyltransferase. Microbiology 143 (Pt 2): 367–376.
54. Weissman Z, Kornitzer D (2004) A family of Candida cell surface haem-binding proteins involved in haemin and haemoglobin-iron utilization. Mol Microbiol 53: 1209–1220.
55. Weissman Z, Shemer R, Conibear E, Kornitzer D (2008) An endocytic mechanism for haemoglobin-iron acquisition in Candida albicans. Mol Microbiol 69: 201–217.
56. Enjalbert B, Smith DA, Cornell MJ, Alam I, Nicholls S, et al. (2006) Role of the Hog1 stress-activated protein kinase in the global transcriptional response to stress in the fungal pathogen Candida albicans. Mol Biol Cell 17: 1018–1032.
57. Kadosh D, Johnson AD (2001) Rfg1, a protein related to the Saccharomyces cerevisiae hypoxic regulator Rox1, controls filamentous growth and virulence in Candida albicans. Mol Cell Biol 21: 2496–2505.
58. Zaragoza O, Rodriguez C, Gancedo C (2000) Isolation of the MIG1 gene from Candida albicans and effects of its disruption on catabolite repression. J Bacteriol 182: 320–326.
59. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44–57.
60. Bradford Y, Conlin T, Dunn N, Fashena D, Frazer K, et al. (2011) ZFIN: enhancements and updates to the Zebrafish Model Organism Database. Nucleic Acids Res 39: D822–829.
61. Shaw GC, Cope JJ, Li L, Corson K, Hersey C, et al. (2006) Mitoferrin is essential for erythroid iron assimilation. Nature 440: 96–100.
62. Jima DD, Shah RN, Orcutt TM, Joshi D, Law JM, et al. (2009) Enhanced transcription of complement and coagulation genes in the absence of adaptive immunity. Molecular Immunology 46: 1505–1516.
63. Preissner KT, Jenne D (1991) Structure of vitronectin and its biological role in hemostasis. Thrombosis and Haemostasis 66: 123–132.
64. Jackson AN, McLure CA, Dawkins RL, Keating PJ (2007) Mannose binding lectin (MBL) copy number polymorphism in Zebrafish (D. rerio) and identification of haplotypes resistant to L. anguillarum. Immunogenetics 59: 861–872.
65. MacCallum DM (2009) Massive induction of innate immune response to Candida albicans in the kidney in a murine intravenous challenge model. FEMS Yeast Res 9: 1111–1122.
66. Dunkel N, Morschhauser J (2011) Loss of heterozygosity at an unlinked genomic locus is responsible for the phenotype of a Candida albicans sap4Delta sap5Delta sap6Delta mutant. Eukaryot Cell 10: 54–62.
67. Murad AM, Leng P, Straffon M, Wishart J, Macaskill S, et al. (2001) NRG1 represses yeast-hypha morphogenesis and hypha-specific gene expression in Candida albicans. EMBO J 20: 4742–4752.
68. Nantel A, Dignard D, Bachewich C, Harcus D, Marcil A, et al. (2002) Transcription profiling of Candida albicans cells undergoing the yeast-to-hyphal transition. Mol Biol Cell 13: 3452–3465.
69. Dunkler A, Walther A, Specht CA, Wendland J (2005) Candida albicans CHT3 encodes the functional homolog of the Cts1 chitinase of Saccharomyces cerevisiae. Fungal Genet Biol 42: 935–947.
70. Chen Y, Yang L, Feng C, Wen LP (2005) Nano neodymium oxide induces massive vacuolization and autophagic cell death in non-small cell lung cancer NCI-H460 cells. Biochem Biophys Res Commun 337: 52–60.
71. Almeida RS, Wilson D, Hube B (2009) Candida albicans iron acquisition within the host. Fems Yeast Research 9: 1000–1012.
72. Bonhomme J, Chauvel M, Goyard S, Roux P, Rossignol T, et al. (2011) Contribution of the glycolytic flux and hypoxia adaptation to efficient biofilm formation by Candida albicans. Mol Microbiol 80: 995–1013.
73. Woo M, Lee K, Song K (2003) MYO2 is not essential for viability, but is required for polarized growth and dimorphic switches in Candida albicans. Fems Microbiology Letters 218: 195–202.
74. Singh RP, Prasad HK, Sinha I, Agarwal N, Natarajan K (2011) Cap2-HAP Complex Is a Critical Transcriptional Regulator That Has Dual but Contrasting Roles in Regulation of Iron Homeostasis in Candida albicans. Journal of Biological Chemistry 286: 25154–25170.
75. Nairz M, Schroll A, Sonnweber T, Weiss G (2010) The struggle for iron – a metal at the host-pathogen interface. Cell Microbiol 12: 1691–1702.
76. Schaible UE, Kaufmann SH (2004) Iron and microbial infection. Nature reviews Microbiology 2: 946–953.
77. Wang L, Cherayil BJ (2009) Ironing out the wrinkles in host defense: interactions between iron homeostasis and innate immunity. Journal of innate immunity 1: 455–464.
78. Jolliffe IT (2002) Principal component analysis. New York: Springer. xxix, 487 p.
79. Ray S, Turi RH (1999) Determination of number of clusters in K-means clustering and application in colour image segmentation. Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques (ICAPRDT'99): 27–29.