



universität
wien

DISSERTATION

Titel der Dissertation

„A dual systems approach to decipher host-
pathogen interactions“

Verfasserin

Lanay Tierney, B.Sc.

angestrebter akademischer Grad

Doktorin der Naturwissenschaften (Dr.rer.nat.)

Wien, 2012

Studienkennzahl lt. Studienblatt:

A 091 490

Dissertationsgebiet lt. Studienblatt:

Molekulare Biologie

Betreuer:

Prof. Dr. Karl Kuchler

We have to continually be jumping off cliffs and developing our wings on the way down.

- Kurt Vonnegut

Table of contents

1. Introduction.....	5
1.1 <i>Candida</i> species as opportunistic fungal pathogens.....	5
1.1.1 Phylogeny and genome evolution of <i>Candida</i> species.....	5
1.1.2 Host recognition of <i>Candida</i> species.....	8
1.1.3 Immune evasion strategies of <i>Candida</i> species.....	11
1.2 Post-transcriptional regulation <i>Candida albicans</i> genes.....	12
1.2.1 Introduction to splicing.....	13
1.2.2 Intron evolution in <i>Candida</i> species.....	15
1.2.3 Splicing regulation in fungi.....	15
1.3 Origins of systems biology.....	16
1.3.1 Applying systems biology to translational research.....	17
1.3.2 Investigating fungal pathogenesis using systems biology.....	20
2. Aims of this thesis.....	23
3. Results.....	24
3.1 Paper 1: <i>Frontiers in Microbiology</i> . 2012; 3:85.....	24
3.2 Paper 2: <i>Current Opinion in Microbiology</i> . 2012; 06:22.....	39
3.3 Paper 3: (Tierney <i>et al.</i> , submitted).....	47
4. Unpublished results.....	87
4.1 Establishing simultaneous RNA-seq with fungi.....	87
4.2 Putative genes are among the highest enriched during infection.....	89
4.3 A putative <i>C. albicans</i> splicing component is regulated during infection.....	91
4.4 <i>C. pintolopesii</i> as a model to investigate <i>Candida</i> pathogenesis.....	93
5. Additional materials and methods.....	96
5.1 Fungal strains and growth conditions.....	96
5.2 <i>Candida pintolopesii</i> <i>de novo</i> sequencing.....	96
5.3 <i>Candida</i> hyphal induction.....	97
5.4 Strain construction.....	97
5.5 Mice.....	97
5.6 Neutrophil killing assay.....	97
5.7 Gastrointestinal tract colonization with <i>Candida</i> species.....	98
5.8 qPCR.....	99

6. General discussion.....	100
6.1 Major results presented in this thesis.....	100
6.1.1 Using simultaneous RNA-seq to investigate host-pathogen interactions...101	
6.1.2 Predicting host-pathogen interactions using network inference.....102	
6.1.3 Novel spliceosome proteins may account for infection specific splicing.....102	
6.1.4 Studying the commensal to pathogen transition using <i>C. pintolopesii</i>104	
7. References.....	106
8. Appendix.....	118
8.1 Additional publications.....	118
8.2 Summary.....	119
8.3 Zusammenfassung.....	120
8.4 Abbreviations.....	121
8.5 Curriculum Vitae.....	123
8.6 Acknowledgements.....	125

1. Introduction

1.1 *Candida* species as opportunistic fungal pathogens

Pathogenic fungi are a major cause of death among long-term hospitalized patients. The most common of these pathogenic fungi are *Candida spp.*, which naturally colonize the mucosal surfaces of the majority of the human population. *Candida* infections come in many forms, the most common of which include oral, gastrointestinal, and vaginal candidiasis. Women are especially prone to *Candida* infections, with over 75% of women experiencing at least one bout of vaginal candidiasis during their reproductive years (1). Although the most common *Candida spp.* affecting patients is by far *C. albicans*, at least 16 species are regularly collected from patients in the clinic (2), with *C. glabrata*, *C. tropicalis*, *C. parapsilosis*, and *C. krusei* species maintaining the highest incident rate behind *C. albicans*. Numerous risk factors have been linked with the development of Candidemia. Some of the most frequent causes include: diabetes, organ transplantation surgery, long-term use of broad-spectrum antibiotics, cancer therapy, and immunosuppression (3). In general, it is thought that imbalances in the immune system as a result of these risk factors, as well as others, can cause this commensal species to switch to the pathogenic state leading to dissemination of the organism into the blood stream and systemic infection (4). The severity of an infection is highly variable among patients and can be dependent on the infecting *Candida spp.*, the extent of any underlying health concern including a genetic predisposition, and the starting location of the infection on the body itself. With the increase in the number of people affected by these risk factors over the past decade, especially with immunocompromised patients, Candidemia has become a significant cost burden to the public health care system, where the average cost to treat a *Candida* infection in the ICU in North America is estimated at over 20,000\$ per patient, and leads to an increase the average length of a hospital stay by 400% (5). Therefore, *Candida* research is becoming an increasingly important field in infection biology.

1.1.1 Phylogeny and genome evolution of *Candida* species

The fungal kingdom is diverse group of eukaryotic species composed of over 1 million species, the vast majority of which are non-pathogenic (6). Those fungi that have been associated with infection in humans belong primarily to the *Candida* genera. *Candida spp.* are members of the fungal phylum *Ascomycota* and include both haploid and diploid fungal genomes (Figure 1). *Candida* are thought to have diverged from its yeast ancestor, *S. cerevisiae* approximately 200 million years ago (7). With the exception of *C. glabrata* and *C. krusei*, all *Candida spp.*, belong to the CUG clade. This clade is characterized by the translation of the codon CUG to serine instead of leucine (8), posing challenges for heterologous markers and functional gene studies between *Candida spp.* and their ancestral species.

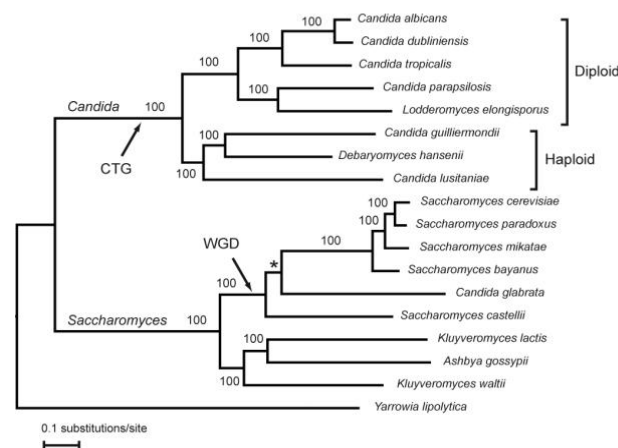


Figure 1. Phylogenetic relationship among sequenced *Candida* species. (Adapted from Butler *et al.*, 2009)

For the majority of human pathogenic fungi, the host is exposed to the spore of the fungus by inhaling them the natural environment. *Candida spp.* however are a rare exception to this in that they are found almost exclusively in association with the human host. *Candida spp.* are opportunistic pathogens of the human gut microbiota. The close association of the species with its human host is moreover thought to be a driving factor in its evolution. For this reason, a major focus of *Candida* research is to elucidate the mechanisms responsible for the commensal to pathogen transition (9). Comparative genome analysis has aided in the identification of traits important for the evolution of pathogenicity. A phylogenomic approach was employed across multiple *Candida* and *Saccharomyces* genomes in order to identify gene families, which are enriched in the pathogen genome compared to its non-pathogenic ancestor. Significant enrichment was found for over 20 gene families of not present in *S. cerevisiae* with

particular expansion of genes previously associated with drug resistance, biofilm functions, adhesions, and morphogenesis of the cell wall for filamentous growth (10). Adhesions are especially important for the association of fungal cells with host surfaces and the acquisition of iron required for infectious processes (11) and is absent from the *Saccharomyces* clade.

The morphological switching of *Candida* is a major virulence trait, and the diversity of forms it can acquire is important to *Candida* growth, survival, and dissemination with its host. *S. cerevisiae* grows exclusively by budding and is found either in the yeast or pseudohyphal form. Budding cells that separate freely from one another characterize yeast cells. Pseudohyphal cells are an ellipsoid shape where the mother and daughter cells remain connected at the point of septation. Hyphal cells lose this constriction and have long polarized filaments. *C. albicans*, is capable of growing in all three forms. Interestingly, *C. albicans* cells that have been modified to grow only in the yeast (12) or hyphal (13) forms only have a significant reduced virulence capacity, emphasizing the necessity of morphological switching for the overall virulence of the species. Hyphal and pseudohyphal formation is triggered in response to a variety of environmental conditions (14). Importantly, not all *Candida* spp. share this morphological diversity, including the second most common *Candida* spp., *C. glabrata*, which can only grow in the yeast form although it is nevertheless pathogenic. Therefore virulence and tissue invasion is not entirely dependent on filamentous growth (15).

Significant mechanical barriers of the host in the form of epithelial cells as the first line of defense exist against infectious agents and dissemination of fungal cells. The morphological diversity of *C. albicans* in conjunction with the evolution of the adhesion family aids in overcoming this barrier (16) (Figure 2). First, *C. albicans* cells adhere to the epithelial surface through interactions between the fungal cell wall and the epithelial cell membranes. For *C. albicans*, two of the most important proteins involved in adhesion are *ALS3* and *HWP1*. *ALS3* is a member of the cell wall adhesion family, which induces endocytosis by host cells by binding to host cadherins (17), has pleiotropic roles in iron acquisition (11) and is expressed in a hyphal specific manner (18). *HWP1* is also a hyphal specific protein whose N-terminal domain is important for linking proteins of the epithelial cells to the *Candida* cell surface (19). Invasion into host tissue takes place via two main mechanisms including induced endocytosis or active penetration. Induced endocytosis is characterized by interactions between hyphal

filaments and host cell membranes whereby hyphal form *Candida* triggers a reorganization of the host cytoskeleton and subsequent pathogen uptake (20). On the other hand, active penetration does not require interaction with the host cell machinery penetration of the hyphae into host cells results in cell and tissue damage (21). Regardless of the mechanism utilized, once fungal cells have breached the epithelial layer it is capable of disseminating via the blood stream to the major body organs and is indeed capable of proliferating in almost all host niches. This proliferation results in a variety of clinical manifestations, the most severe being acute renal failure and septic shock (22).

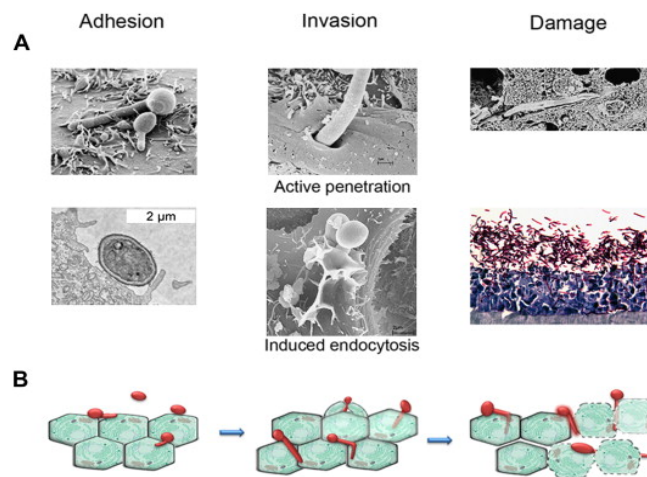


Figure 2. Stages of *C. albicans* tissue invasion. *C. albicans* Infection progresses from adhesion, invasion (via active penetration or induced endocytosis), and finally tissue damage. **(A)** Electron micrographs of *C. albicans* invading oral epithelial tissue. **(B)** Schematic drawing of the described infection stages. (Adapted from Naglik *et al.*, 2011)

1.1.2 Host recognition of *Candida* species

Fungi are recognized by germ line encoded receptors of the innate immune system referred to as pathogen recognition receptors (PRR) (23). These PRRs include Toll -like receptors (TLRs), C-type and S-type lectins and are expressed on innate immune cells (24). There are 13 TLRs identified in humans and mice, and each one is important for the activation of different arms of the immune response, primarily through the common adaptor molecule MyD88. The major ones associated with fungal infection include TLR4, TLR6, TLR2, and TLR9 (25). The lectin receptors are known to

recognize pathogens in a calcium dependent manner and the main ones associated with fungal infection Dectin-1, Dectin-2, and Mincle. PRRs as a whole are important in the activation of innate and adaptive immune responses of the host. The downstream signaling pathways for PRRs are partially redundant signaling cascades to activate a pathogen specific host response. PRRs recognize pathogen associated molecular patterns (PAMPs) present on the fungus. In many cases, fungal PAMPs are localized to the cell wall. The cross talk between fungal PAMPs and PRR determines the immune response and eventually host survival (Figure 3).

The fungal cell wall maintains cell wall integrity and is the first point of contact between the fungus and the host environment and the general structure of the cell is conserved among many fungi. The cell wall itself is primarily composed of carbohydrates, which form a protective meshwork surrounding the plasma membrane. The *Candida* cell wall can be split into two layers: the outer layer and the inner layer. The outer layer is made up of the mannoproteins and cell wall proteins. The cell wall proteins are anchored to the cell wall by glycosylphosphatidylinositol (GPI)-anchor-dependent cell wall proteins (GPI-CWPs) to the β , 1-6-glucans and β , 1-3-glucans of the inner layer. Mannoproteins are highly glycosylated with mannose-containing polysaccharides, and have a low permeability to the external environment. The inner layer is composed primarily of the glucans including β , 1-6-glucans and β , 1-3-glucans and chitin, which form a meshwork with one another by hydrogen bonding adjacent polysaccharide chains (26). Based on the morphological stage of the fungus, different components of the cell wall will be exposed to the immune cells therefore will influence recognition of the pathogen (27) .

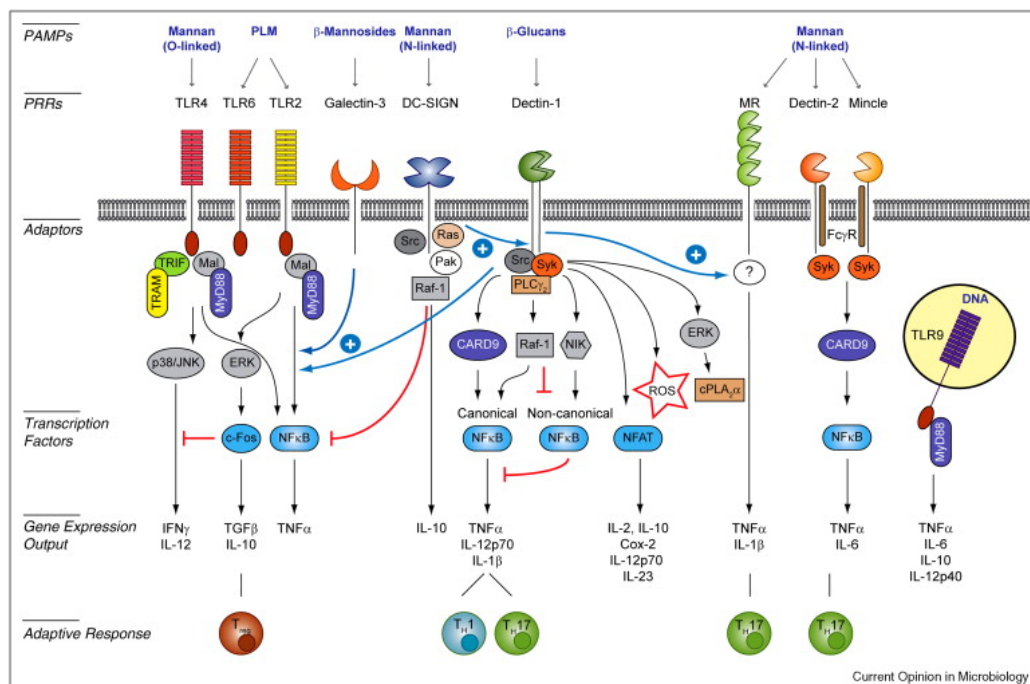


Figure 3. Fungal pathogens are recognized by PAMPs and trigger downstream signaling of the immune system. The recognition of fungal pathogens begins with surface receptors and endosomal TLRs which recognize components of the *C. albicans* cell wall, such as mannans and β -glucans. These receptors trigger a signaling cascade of transcription factors and signaling adaptors which shape the activation of the adaptive response and host survival. (Adapted from Bourgeois *et al.*, 2010)

The changing morphology of *C. albicans* during infection alters the exposure of cell wall components to immune cells and thereby influences both fungal recognition and the eventual activation immune response to the fungus (28). Therefore, the immune reactivity of specific components of the *C. albicans* cell wall is an active area of research in understand the commensal to pathogen transition (29). So far it has been shown that glucans are pro-inflammatory mediators and its exposure to immune cells is increased during the yeast to hyphal transition and treatment with echinocandins, increasing cytokine induction by immune cells (30). Chemical differences between yeast and hyphal mannans also seem to play a role in immune activation to the fungus by triggering distinct toll-like receptor signaling pathways (31). Chitin has been shown to activate both pro- and anti-inflammatory signaling and its role in recognition remains to be elucidated (32). Overall, understanding the composition of *C. albicans* cell wall at differing infection stages will be important for the elucidation of morphology dependent PAMPs.

1.1.3 Immune evasion strategies of *Candida* species

Fungal pathogens have developed an arsenal of strategies to avoid recognition and killing by host immune cells (33). The shielding of β -glucans, which are known to have a pro-inflammatory effect, from recognition by host Dectin-1 is one of the best-studied immune evasion strategies responses of *Candida spp.* (34). β -glucans are presented on the cell wall in a morphology dependent manner: they are highly present on yeast form cells but are mostly hidden in hyphal form cells, therefore the yeast to hyphal transition has been proposed to be mechanism of immune evasion (35). Additionally, the formation of hyphae has been shown to be an important process to impairing intercellular killing by macrophages (36). Many fungal species have also been shown to inhibit the formation of the phagolysosome including *Candida spp.* (37)(Figure 4). The development of biofilms also has been suggested to provide physical protection for proliferating layers of *C. albicans* cells from antifungal treatment and attack by host cells (38). *C. albicans* avoids host killing by counteracting reactive oxygen species (ROS) produced by immune cells. This has been shown for *C. albicans* degrading host-derived ROS using cell-surface superoxide dismutases (39) or detoxifying host-derived nitric oxide (40).

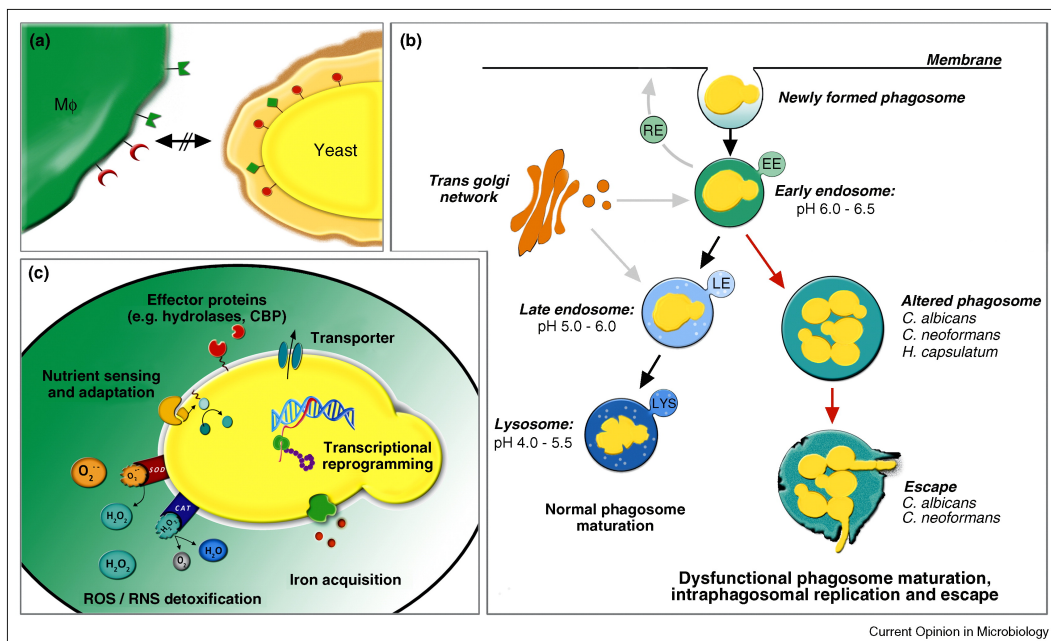


Figure 4. Immune modulation and evasion strategies of *Candida spp.* (A) The *C. albicans* cell wall hinders the identification of pathogen associated molecular patterns (PAMPs). (B) The mature phagolysosome digests invading pathogens. Some pathogens can inhibit the maturation process and

escape from the maturing phagosome. **(C)** Once phagocytosed, *C. albicans* attempts to manipulate the internal environment to proliferate. (Adapted from Seider *et al.*, 2010)

1.2 Post-transcriptional regulation of *C. albicans* genes

Post-transcriptional modification of RNA is an essential process for gene regulation in response to changing environmental conditions, the biogenesis of the ribosome, and during cellular growth. These modifications primarily include splicing, 3' polyadenylation, and 5' capping to generate mature mRNA transcripts. In mammals, these processes are well studied *in vitro* and *in vivo* (41) especially because of the almost ubiquitous prevalence of splicing and alternative splicing for mammalian genes (42). However, comparative genome analysis for 5 fungi including *S. cerevisiae*, *Schizosaccharomyces pombe*, *Aspergillus nidulans*, *Neurospora crassa*, and *Cryptococcus neoformans* revealed that fungi have comparatively shorter introns as well as having differences in the information content for the introns themselves than mammalian species (43). Especially in the case of *S. cerevisiae* and *C. albicans*, which have undergone significant intron loss over their evolutionary history (44), splicing and indeed alternative splicing in some fungi can be relatively rare events and for this reason are often poorly studied. Nevertheless, the identification of post-transcriptional mechanisms may reveal additional adaptation mechanisms important for fungal to the host environment and their overall virulence based on their ability to quickly transcriptionally respond to changing environmental conditions.

Several examples of post-transcriptional modifications for fungi have recently been described. Two major snoRNAs families for the filamentous fungus *Neurospora crassa* were identified which are highly conserved among species in the fungal kingdom. These include genes that are important for alternative splicing (45). SnoRNA are known to be important for the modification of rRNAs and snRNAs in eukaryotes. Recently, the *C. albicans* gene Dicer (CaDcr1) was proposed to be the only RNase III enzyme in the fungus and functionally plays a role in biogenesis of the ribosome and spliceosome (46). The loss of CaDcr1 results in a severe growth phenotype. Additionally, CaDcr1 was found to be required for cleavage of the 3' external transcribed spacer for pre-rRNA and for the U4 snRNA tail of the spliceosome. Anaerobic glycolysis maintains cellular homeostasis and in mammals, the transcriptional and post-transcriptional regulations of associated genes are regulated in

response to hypoxia (47). Similarly, in response to hypoxia, it was suggested that glycolysis in *A. fumigatus* (48) and the anaerobic *S. cerevisiae* (49) is post-transcriptionally regulated. Taken together, these data shown that post-transcriptional regulation takes place during diverse facets of fungal development.

1.2.1 Introduction to Splicing

The spliceosome is responsible for catalyzing intron removal from pre-mRNA (50). It is composed of five small nuclear ribonucleoprotein particles (snRNPs), which form the core of the spliceosome, in addition to hundreds of accessory proteins, many which are still functionally uncharacterized. The spliceosome recognizes conserved intron motifs, referred to as the 5' and 3' splice site, and the branch site. The assembly of the spliceosome is a highly regulated process for efficient intron removal (Figure 5). RNA sequence elements and protein regulators including numerous exonic and intronic splicing enhancers and silencers influence intron removal. Secondary structures of the pre-mRNA itself have also been shown to influence intron selection (51) by hindering the recognition of splice sites, or exposing sequences for protein binding (52, 53).

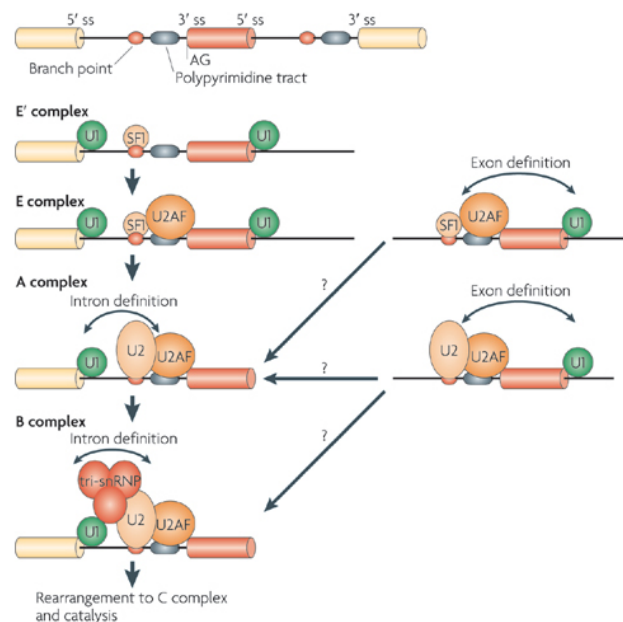


Figure 5. Pre-mRNA splicing by the spliceosome. Intron removal is catalyzed by the spliceosome consisting of two transesterification reactions involving two sequential nucleophilic attacks on the terminal phosphodiester bonds of the intron. The spliceosome begins this catalysis with the binding to the branch point of the U1 snRNP to the 5' splice site, and splicing factor 1 (SF1) in an ATP-independent manner. These components combined form the E prime complex. This complex is then converted to the E complex

via the recruitment of the U2 auxiliary factor (U2AF) to the terminal splice site. The A complex is then formed in an ATP-dependent manner by replacing SF1 by U2 snRNP at the branch point. The addition of the U4/U6–U5 tri-snRNP forms the B complex, which contains all of the necessary components to carry out splicing. Conformational changes of the B complex and the loss of U1 and U4 snRNPs results in the formation of the catalytically active spliceosome, known as the C complex. (Adapted from Chen *et al.*, 2009)

The selection of specific introns and exons for inclusion in mature mRNA transcript is known as alternative splicing (AS), generating multiple protein products from a single pre-mRNA transcript (54)(Figure 6). AS is responsible for the increasing protein diversity without requiring an increase genome size. It is not uncommon for a human gene to contain hundreds of possible isoforms (55). One of the most famous examples isoform diversity for a single pre-mRNA is for the *Down syndrome cell-adhesion molecule* Dscam gene that can theoretically generate 38,016 unique transcripts (56). The regulation of these specific forms must be tightly regulated, as mistakes in the generation of the correct AS isoform has been directly linked to the development of human diseases (57).

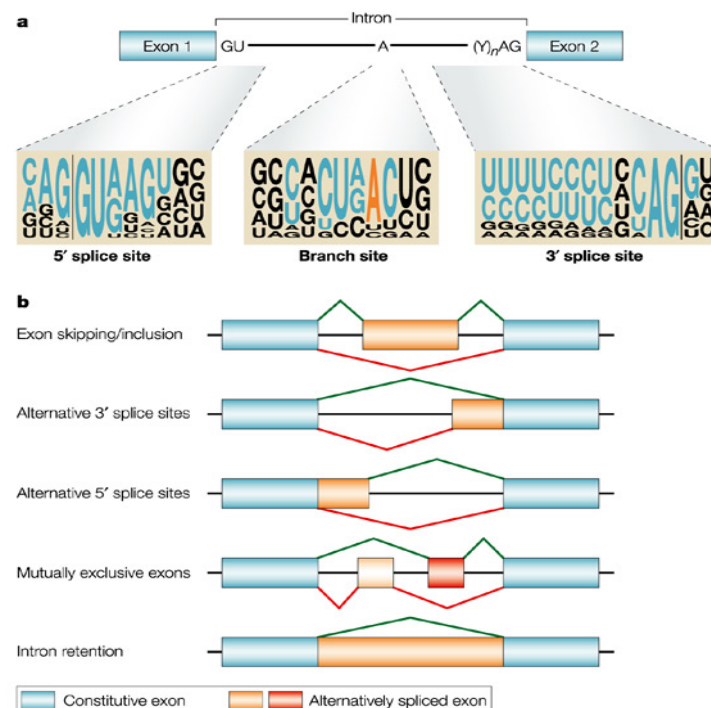


Figure 6. Forms and consensus sequences of alternative spliced genes. (A) Conserved intron motifs at the 5' and 3' splice site, and the branch point. **(B)** Common types of alternative splicing in mammals (alternative path is shown in green) which can function separately or in combination for genes with multiple introns. (Adapted from Cartegni *et al.*, 2002)

1.2.2 Intron evolution in *Candida* species

Historically introns were thought of as junk DNA sequences, but they have gained significant research attention based on discovery of the high prevalence of AS in humans and in gene and protein regulation implicated in human diseases. The high variation of introns between eukaryotic organisms has generated considerable debated among scientists for the evolutionary forces responsible for intron loss and gain. Two of the most popular theories that exist are the “intron early” and the “intron late” hypothesis. The intron early hypothesis states that the evolutionary ancestor of both eukaryotes and prokaryotes were intron rich, and introns were eventually lost among prokaryotes. However, the intron late hypothesis argues that the evolutionary ancestor of both eukaryotes and prokaryotes where intron less, and they later evolved into the genome of eukaryotes (58) . The average human gene contains around 7 introns per gene (58). Unlike humans, the hemiascomycetous yeasts, including *C. albicans* and *S. cerevisiae* have undergone significant amounts of intron lost over their evolutionary history, with less than 5% of the genome is now estimated as intron containing, and the majority of these genes contain only one intron per gene (59).

Selective environmental pressures are believed to be the driving force in intron evolution (59). However, the mechanism and conditions under which this intron loss itself has taken place is still not known. Several models exist to describe how intron loss could have take place. In the classic model of intron loss, a gene is converted by recombining with a reverse-transcribed copy of a spliced mRNA transcript, which results in the deletion of the intron (60). Introns can also be lost simply by genomic deletion (61). Most recently, an alternative mechanism of intron loss was proposed for the hemiascomycetous yeasts based on the prevalence of yeast noncoding RNAs (62). They predicted that almost half of snoRNAs are located within introns and that intron loss took place primarily though the degeneration of the splicing signals, so they are no longer capable of being recognized by the yeast spliceosome.

1.2.3 splicing regulation in fungi

Although the number of intron containing genes is indeed quite low among the hemiascomycetous yeasts, the regulation of their splicing and alternative splicing have demonstrable influences on the adaptation of the fungi to host environments. For *S.*

cerevisiae, splicing regulation has been shown to take place for numerous genes in response to environmental and developmental conditions including the meiosis gene *MER2* is regulated during the cell cycle (63), a post-transcriptional feedback loop was identified for *CRY2* (64) and the ribosomal gene *RPL30* (65) influencing ribosome assembly, and RNA export (66). Interestingly, there was a higher prevalence of genes with non-canonical splice sites for examples of regulated splicing compared to intron containing genes in general (64-66). In *C. albicans*, several cases of regulated AS have been identified in response to temperature stress and changing nutrient availability (44). We find that both splicing and alternative splicing in *C. albicans* is regulated under infection conditions (Tierney *et al.*, submitted), suggesting a greater role for splicing regulation for adaptation to the host environment.

1.3 Origins of systems biology

The roots of modern systems biology have been proposed to originate as early as the 1940s with Schrödinger's paradox: molecular biology describes a progressive ordering and decrease of chaos in biological systems, whereas paradoxically, many biochemical processes necessitate the production of chaos. Unlike reductionist strategies, this school of thought aimed to discover and describe general biological principles in a quantitative manner (67). Today, systems biology has evolved into two major schools of thought: functional genomics and mathematical modeling. In general, systems biology approaches follow iterative cycles of modeling and data generation, based on a given hypothesis. It is a technique to investigate the behavior of components in a biological system in response to perturbations; these data are then integrated from a variety of sources in order to mathematically model a response pattern (68). The term systems biology has only been in use for just over a decade. Since information flow within biological systems is hierarchical in nature, systems biology provides insight into biological questions using information at increasing levels of evidence. The central dogma of molecular biology is a good example of this structure, where the unidirectional flow of information in biological systems is from DNA, to RNA, to protein. These biological levels provide different information about an organisms response to a specific environment and techniques have arisen to investigate each step on a genome-wide level including: genome sequencing (DNA), microarray and RNA sequencing technology (RNA-seq) (RNA), proteomics (protein) in

addition to an ever increasing number of -omics screens (69). From an experimental point of view, there are fundamental differences between systems and classical biology techniques in regards to the scale of the information used (Figure 7); Nevertheless, both approaches converge on their use of the data to perform predictive, hypothesis-driven experiments regarding a specific biological question regardless of the starting scale of the experimental data (70).

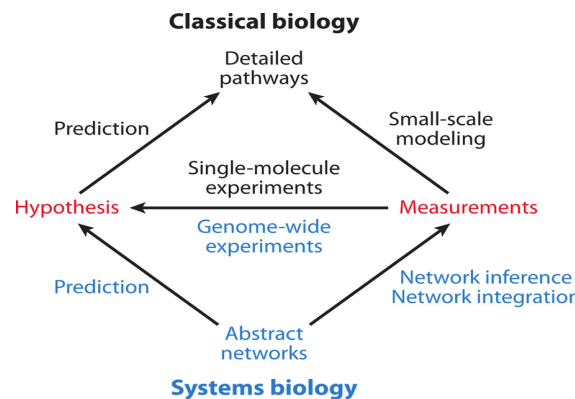


Figure 7. Experimental differences between systems biology and classical biology (Adapted from Chuang *et al.*, 2010)

1.3.1 Applying systems biology to translational research

One of the main goals of systems biology is to improve drug target discovery and thereby the drug development process as a whole. Therefore, a strong emphasis has been placed on the application of systems biology experiments to translational research. Translational research is the application of basic research to the treatment of human diseases. Despite the increased use of genome-wide approaches, the application of these data to increase the development of new drugs remains stagnant because of major hurdles in data analysis and the development of cost effective methods. Nevertheless, significant advancements have been made in “bench to bedside” approaches to infectious diseases at different levels, including genomic, proteomic, and metabolomic (71) and each level will be addressed specifically in turn.

It is generally accepted that the start of the genomics era was heralded in by the Human Genome Project in 1985 (72). Since then, genome and transcriptome sequencing has revolutionized the ability to investigate adaptation and evolution using genome-wide expression profiling and genomic sequencing for a steadily increasing

number of organisms using high density microarrays and RNA-seq. To distinguish closely related pathogenic species epidemiologists have started using RNA-seq technology to follow disease transmission and its evolution. For example, during a recent cholera outbreak in Haiti, using comparative sequence analysis, the Haitian strain was found to be more closely related to strains from Bangladesh and therefore was likely to have been introduced from human activity (73). Additionally, after the swine flu outbreak in 2009, the H1N1 flu strain was sequenced and compared to the seasonal human flu strains to track its virulence evolution at the genomic level (74). Sequencing technologies have also sparked the emergence of reverse vaccinology, a term that describes the use of computer-facilitated predictions of antigen prevalence and immunogenicity *in silico*, an approach that significantly increases the predictive power of vaccine targets (75). Additionally, reverse genetics methods had been applied for pathogenesis studies. This method uses the modification of the genomic sequence either through insertion, mutagenesis, or deletion of a sequence of interest, in a plasmid based system, which is then transfected in the virus of interest in order to investigate the role of the modification on viral infectivity (76). Comparative genomics projects for fungi have also taken advantage of the new sequencing technology and increased in their frequency over recent years, including the Fungal Genome Initiative from the Broad Institute (77), with over 50 sequenced fungal species. These projects focus on the identification and the prevalence of antifungal drug targets among related species in order to investigate the broader application of differing treatment strategies.

Proteomics uses mass spectrometry (MS) based techniques to profile large-scale protein turn over in the evaluation of cellular function for a specific organism. Although many proteomics techniques are now in use, one common approach is so called “shotgun proteomics” which combines MS and high performance liquid chromatography (HPLC) to profile thousands of individual proteins with limited prior knowledge of the component peptides. This technique delivers a high sensitivity of protein detection even for low concentration proteins (78). In the field of infectious diseases, shotgun proteomics was used to quantify protein expression for the malaria parasite, *Plasmodium falciparum* in a time dependent manner and its association with its host (79). To date, applications of proteomics include global protein profiling, the investigation of protein modifications, protein interactions and protein-gene interactions. These data increase drug target identification for functional proteins identified during infectious processes.

Metabolomics profiles small molecule metabolites including amino acids, nucleotides, lipids, and sugars, of an organism of interest. This is the most recent addition to the field of systems biology and is important because it profiles the level at which many pharmaceuticals are functionally active (80). Metabolomics principally uses nuclear magnetic resonance (NMR) spectroscopy or MS to identify metabolites. So far it has been used to profile mechanism of *C. albicans* increasing drug susceptibility to environmental stress (81). In cytomegalovirus, pathways upregulated by viral infection were identified using changes in metabolic flux in uninfected and infected human cells (82). Metabolomics provides a difference perspective on the investigation of disease pathogenesis.

Haplo-insufficiency screens and fitness profiling in fungi against different antifungal agents and other environmental stresses has been an effective approach to identify novel drug targets (83). The availability of large-scale heterozygous knockout collections for some of the major fungi including *S. cerevisiae* (84) and *C. albicans* (85), among others, has facilitated this screening process.

The success of the individual -omics approaches highlights the applicability of such technologies to pathogenesis research and holds the promise of advancing the field of personalized medicine (86). Modeling using computational approaches is an emerging method for the combination and simplification of multiple -omics data sets (87). These models can incorporate topological (protein-protein interactions), qualitative (regulatory networks), and quantitative (co-regulation) data (88). Computational and network models aim to predict biological behaviors for an organism. Major challenges still lay ahead in identifying ways to combine and integrate multiple -omics data sets in the identification of therapeutic targets, and theoretical analysis pipelines for such research are already being established (Figure 8).

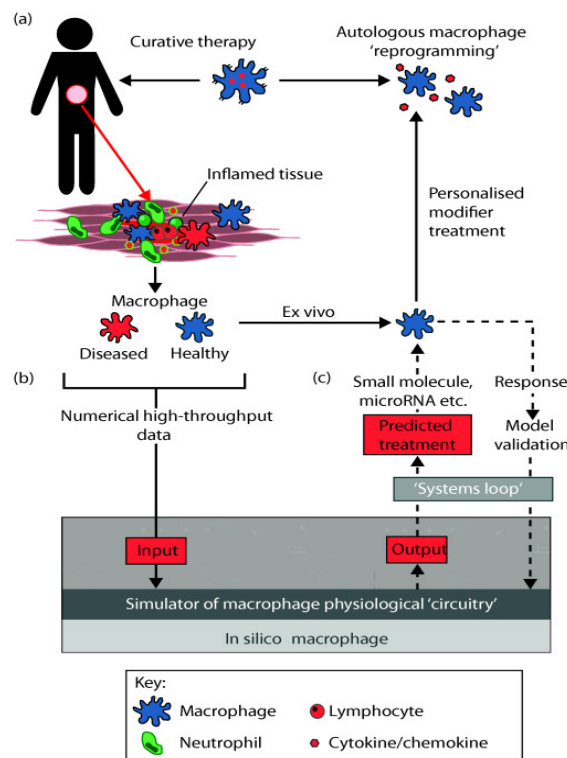


Figure 8. Integrating *in silico* modeling strategies with data generated from a wet lab environment for infection situations using macrophages as a model. (A) *Ex vivo* reprogram of patient macrophages in order to treat inflammatory disease. (B) Integrating high-throughput pathway-based analyses with *in silico* simulations. (C) Combining data gained from unhealthy patient macrophages in an iterative learning loop to validate *in silico* model and tailor treatment options on a personalized basis. (Adapted from Ghazal *et al.*, 2011)

1.3.2 Investigating fungal pathogenesis using systems biology

The application of systems biology to fungal infection poses additional challenges compared to other pathogens. Unlike diseases caused by bacteria or viruses, fungi are eukaryotic and the similarity of some of their membrane structures to host cells can present additional challenges for the immune system to distinguishing self from non-self compared to other pathogenic organisms. The diagnosis of fungal infections is hindered by the vast diversity of adaptive strategies of different organisms to their environment and to their host. Therefore, pathogenesis research benefit from genome-wide analysis techniques to encompass this complexity. Especially in the case of opportunistic fungal pathogens, such as *C. albicans*, which have co-evolved with their human host, complex strategies to evade the immune system have been

developed. Understanding the balance between the commensal and pathogenic state for fungi and the physical and genetic causes that tip the balance from one state to the other is quintessential for the further understanding of pathogenic processes (89). Modeling these different approaches can span orders of magnitude. A popular approach to simplify such data sets is to incorporate different biological levels in order to get a global view of adaptation to the host (Figure 9). These strategies generally can be viewed from two perspectives: a “top-down”, in which the information flow is from the general (i.e. genome or population-wide) to the specific (the regulation of individual gene or genes) or vice versa, which is referred to as a “bottom-up” approach (90).

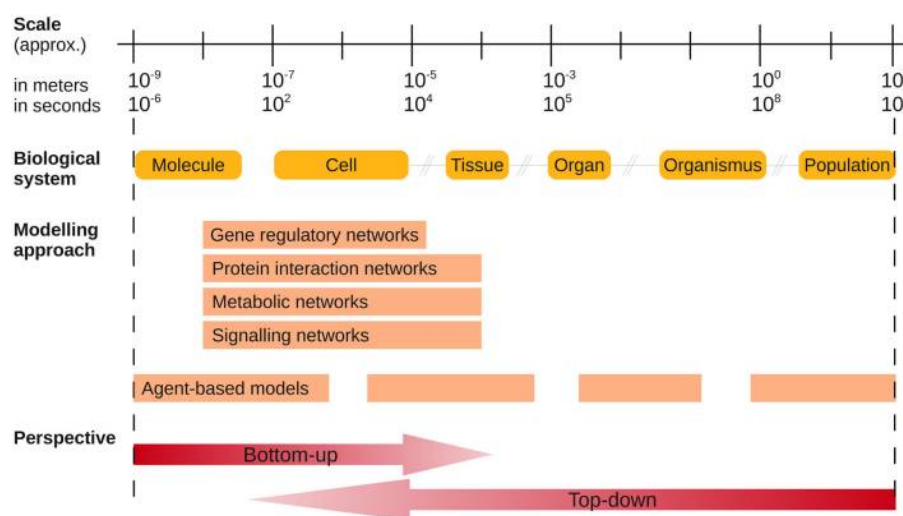


Figure 9. Modeling levels of infectious diseases with systems biology. Schematic diagram shows the orders of magnitude spanned by biological systems and the overlap in terms of scale for current mathematical modeling networks. (Adapted from Horn *et al.*, 2012)

These different global perspectives of gene regulation are being increasingly applied to fungal pathogenesis research. For example, top-down perspective methods such as reverse engineering of gene regulatory networks using genome-wide transcriptional data sets such as microarray and RNA-seq; can be used to predict adaptive responses to the host cells. Based on microarray data of *C. albicans* cells under iron starvation, a transcriptional network was reconstructed, suggesting a mechanism for iron regulation during the infection process with human oral epithelial cells (91). Most recently, a modeling approach termed network inference was used in the investigation of host-pathogen interactions using *C. albicans* as a model over a

time course of infection (92). Transcriptome analysis of the host response to fungal infection has also been used to investigate global response patterns of human monocytes to *Aspergillus fumigatus* (93). The human monocyte-derived dendritic cells immune responses to components of the fungal cell wall were profiled using *C. albicans* or the nonpathogenic *S. cerevisiae* (94). The fungal response of *A. fumigatus* to monocyte-derived immature DCs was further analyzed in a time resolved fashion, identifying genes involved in fermentation, drug transport, pathogenesis and response to oxidative stress (95).

Proteomics analysis for fungi is a rapidly expanding field. To date, it has focused primarily on understanding the contribution of the fungal cell wall, morphology, and the effects of antifungal drugs. The response of the *C. albicans* cell wall on the protein level to amino acid starvation has been recently investigated, suggesting a role for it as a morphogenetic signal in mucus membranes (96). Changes in the cell wall proteome during the yeast to hyphal transition of *C. albicans* and the integration of this data with prior knowledge of protein abundance, allowed the authors to hypothesize on metabolic reorganization as a result of these changes in morphology (96). Based on collected *in vitro* and *in vivo* data sets from fungal as well as bacterial and viral pathogens, a network of protein interactions from human proteins in response to these different pathogens was created, to investigate common pathway responses to multiple pathogens (97). These data combined provide insight into the functional response of the quick transcriptional changes to differing environmental conditions encountered within the host.

Emerging -omics such as metabolomics, and secretomics, are becoming available for many fungal species. The effect of the antifungal drug fluconazole on the *C. albicans* cell wall was recently profiled, highlighting its effects on cell wall integrity (98). Furthermore, modeling metabolic networks recently became available for *A. fumigatus* (99). The steadily increasing number of analysis levels to investigate fungal and host response will be useful to elucidate the mechanism of adaptation responsible for fungal virulence.

2. Aims of this Thesis

The success of pathogens is largely attributed to the speed and diversity of strategies they evoke to both evade and invade their host. The transcriptional flexibility of *Candida spp.* is quintessential for invasion into host tissues, however the connection between these transcriptional changes and proliferation within the host is poorly characterized under both *in vitro* and *in vivo* conditions.

To this end, we used an *in vitro* model of BMDCs with different *Candida spp.* to address the following aims:

- I. To investigate host-pathogen cross-talk real time using simultaneous RNA-seq of the both transcriptomes *in vitro*
- II. To identify and characterize adaptive mechanisms of *Candida spp.* to the host environment and vice versa which were not capable of being detected using previous technologies
- III. To develop a method to model host-pathogen cross-talk *in silico*
- IV. To adapt an *in vivo* colonization model in mice for simultaneous RNA-seq experiments with *Candida spp.*

3. Results

3.1 Paper 1: *Frontiers in Microbiology*. 2012; 3:85.

An Interspecies Regulatory Network Inferred from Simultaneous RNA-seq of *Candida albicans* Invading Innate Immune Cells

Lanay Tierney^{1#}, Jörg Linde^{2#}, Sebastian Müller², Sascha Brunke^{3,4}, Juan Camilo Molina³, Bernhard Hube^{3,5}, Ulrike Schöck⁶, Reinhard Guthke² & Karl Kuchler^{1*}

Recent advances in genome-wide integrative approaches have facilitated the use of inference of network models based on deep sequencing data. However, inferred networks maps had previous considered only a single species. This study presents the first adaptation of network inference to model host-pathogen interactions. Using a time course of infection with and BMDCs, we performed RNA-seq of both host-pathogen transcriptomes and identified novel host-pathogen interaction partners using *Netgenerator*. These putative interactions were experimentally verified, showing the applicability of network inference to the study of multiple species simultaneously. Biologically, the experimental verification of this subnetwork identified a novel role for the interaction between the iron regulatory transcription factor Hap3 in and Ptx3 and Mta2 in *M. musculus*, a soluble receptor and component of the nucleosome remodeling and histone deacetylase complex NuRD, respectively. Technically, this work is also the first implementation of simultaneous RNA-seq of a host-pathogen interaction, an approach that I developed during my time in the laboratory for *Candida spp. in vitro* with high applicability to other species and model systems.

For this study, I conceived and designed all wet lab experiments, performed, analyzed, and interpreted the data, and wrote the manuscript.



An interspecies regulatory network inferred from simultaneous RNA-seq of *Candida albicans* invading innate immune cells

Lanay Tierney^{1†}, Jörg Linde^{2†}, Sebastian Müller², Sascha Brunke^{3,4}, Juan Camilo Molina³, Bernhard Hube^{3,5}, Ulrike Schöck⁶, Reinhard Guthke² and Karl Kuchler^{1*}

¹ Christian Doppler Laboratory for Infection Biology, Max F. Perutz Laboratories, Medical University of Vienna, Vienna, Austria

² Research Group Systems Biology and Bioinformatics, Leibniz-Institute for Natural Product Research and Infection Biology – Hans-Knoell-Institute, Jena, Germany

³ Department Microbial Pathogenicity Mechanisms, Leibniz-Institute for Natural Product Research and Infection Biology – Hans-Knoell-Institute, Jena, Germany

⁴ Center for Sepsis Control and Care, Jena University Hospital, Jena, Germany

⁵ Friedrich Schiller University, Jena, Germany

⁶ GATC Biotech AG, Konstanz, Germany

Edited by:

Franziska Mech, Leibniz-Institute for Natural Product Research and Infection Biology – Hans-Knoell-Institute, Germany

Reviewed by:

Thomas Dandekar, University of Würzburg, Germany
Oliver Kurze, Friedrich Schiller University Jena, Germany

*Correspondence:

Karl Kuchler, Christian Doppler Laboratory for Infection Biology, Max F. Perutz Laboratories, Medical University of Vienna, Campus Vienna Biocenter, Dr. Bohr-Gasse 9/2, A-1030 Vienna, Austria.
e-mail: karl.kuchler@meduniwien.ac.at

[†]Lanay Tierney and Jörg Linde have contributed equally to this work.

The ability to adapt to diverse micro-environmental challenges encountered within a host is of pivotal importance to the opportunistic fungal pathogen *Candida albicans*. We have quantified *C. albicans* and *M. musculus* gene expression dynamics during phagocytosis by dendritic cells in a genome-wide, time-resolved analysis using simultaneous RNA-seq. A robust network inference map was generated from this dataset using NetGenerator, predicting novel interactions between the host and the pathogen. We experimentally verified predicted interdependent sub-networks comprising Hap3 in *C. albicans*, and Ptx3 and Mta2 in *M. musculus*. Remarkably, binding of recombinant Ptx3 to the *C. albicans* cell wall was found to regulate the expression of fungal Hap3 target genes as predicted by the network inference model. Pre-incubation of *C. albicans* with recombinant Ptx3 significantly altered the expression of Mta2 target cytokines such as IL-2 and IL-4 in a Hap3-dependent manner, further suggesting a role for Mta2 in host–pathogen interplay as predicted in the network inference model. We propose an integrated model for the functionality of these sub-networks during fungal invasion of immune cells, according to which binding of Ptx3 to the *C. albicans* cell wall induces remodeling via fungal Hap3 target genes, thereby altering the immune response to the pathogen. We show the applicability of network inference to predict interactions between host–pathogen pairs, demonstrating the usefulness of this systems biology approach to decipher mechanisms of microbial pathogenesis.

Keywords: host–pathogen, RNA-seq, network inference, modeling, reverse engineering, *Candida*, dendritic cells

INTRODUCTION

Both host and pathogenic species have evolved a plethora of strategies to rapidly adapt to the changing environmental dynamics within the infection milieu. However, the extent of this complexity has only recently been investigated through the use of system biology approaches (reviewed in Rizzetto and Cavalieri, 2011). On the molecular level, these adaptations are mediated by complex interaction networks, which sense these environmental changes and transmit the information throughout the cell, leading to a cascade of changes in gene and eventually protein expression. Understanding these underlying interaction networks is important to elucidate how organisms and defense mechanisms interact during microbial infection processes. Genome-wide integrative approaches for modeling have become increasingly popular (Rizzetto and Cavalieri, 2011) due to the availability of high-throughput sequencing technologies, including RNA sequencing (RNA-seq). These technologies now allow for the parallel sequencing of millions of nucleotide sequences simultaneously (Wang

et al., 2009; Zhang et al., 2011). One major advantage to using sequencing approach rather than microarrays is that it is a species-independent platform, allowing for an in-depth investigation of non-model organism species, as well as multiple organisms from a single experiment.

In many cases, the underlying interaction networks between the organisms of interest are unknown. Network inference uses reverse engineering techniques (Hecker et al., 2009b; Marbach et al., 2010) to predict unknown interaction networks based on high-throughput gene expression data. A number of approaches have been established to predict inference networks including Bayesian network modeling (Friedman et al., 2000), information theoretical approaches (Butte and Kohane, 2000; Faith et al., 2007), regression based models (D'Haeseleer et al., 1999; Hecker et al., 2009a), and differential equation models (Holter et al., 2001; Guthke et al., 2005, 2007). Biological networks are scale free networks composed of nodes and edges, where nodes represent the objects of interest and edges show the relations between those objects (Le Novère

et al., 2009). Biological interaction networks often use nodes to represent genes or proteins, and edges to show either a direct or indirect interaction, such as protein binding or transcriptional regulation (Barabasi and Oltvai, 2004). Network inference has been successfully applied to a variety of biological scenarios, including the modeling of immune diseases (Guthke et al., 2005; Hecker et al., 2009a), full-genomic models of *Escherichia coli* (Faith et al., 2007), and more recently, small scale networks describing fungal infections (Linde et al., 2010). So far these model have only focused on a single species and have not addressed host–pathogen interactions.

In the present work, we have generated the first interspecies computational model of molecular host–pathogen interactions. We used RNA-seq expression data from an infection time course of *Candida albicans* and bone marrow-derived dendritic cells (BMDs) from *M. musculus*. *C. albicans* is one of the most prevalent opportunistic human fungal pathogens. Although *C. albicans* normally colonizes the human host, a variety of factors, most notably immune suppression, can lead to dissemination of fungal cells throughout the body. This dissemination can lead to a wide range of diseases, from thrush to multi-organ failure (Gudlaugsson et al., 2003). We focused on dendritic cells as our model host based of their function as antigen-presenting cells, their specialization in pathogen recognition, and their greater role in activating and modulating adaptive immune responses (Netea et al., 2008; Bourgeois et al., 2010). We experimentally verified predicted sub-networks of the interspecies inferred regulatory network, which identifies a role of the transcription factor Hap3 in *C. albicans* during *in vitro* infection. We find that fungal Hap3 is regulated by murine Ptx3, a soluble pattern recognition receptor acting as an opsonin for pathogens (Diniz et al., 2004). We show that Ptx3 binding to *C. albicans* regulates fungal Hap3 target genes, altering the immune response in dendritic cells. Based on the regulation of downstream cytokines and the regulation of *MTA2* mRNA in a Hap3-dependent manner, we provide indirect evidence for a role for Mta2, a member of the nucleosome remodeling and histone deacetylase complex NuRD (Manavathi et al., 2007). We propose a mechanism whereby Ptx3 binding to *C. albicans* leads to cell wall remodeling via fungal Hap3 target genes, thereby changing the ability of the fungi to be recognized by immune cells. The experimental verification of the predicted interspecies interactions is proof-of-principle that network inference can be used to investigate microbial pathogenesis. We suggest that this could be a useful method to identify potential antifungal target genes.

MATERIALS AND METHODS

CANDIDA STRAINS AND GROWTH CONDITIONS

All strains were routinely grown on YPD plates (1% yeast extract, 2% peptone, 2% glucose, 2% agar) and in standard rich media YPD (1% yeast extract, 2% peptone, 2% glucose) for liquid culture at 30°C. Fungal cells were collected in the logarithmic growth phase by a brief centrifugation, washed in sterile PBS, and diluted for all interaction studies. The following strains were used in this study: *C. albicans* clinical isolate SC5314 (Gillum et al., 1984) and homozygous knock-out of Hap3 (*hap3Δ/hap3Δ*) and revertant strain (*hap3Δ/hap3Δ + Clp10 (HAP3, URA3)*, abbreviated in the text as *hap3Δ/hap3Δ + HAP3*), were generated from the

strain BWP17 (*ura3::imm434/ura3::imm434 iro1/iro1::imm434 his1::hisG/his1::hisG arg4::hisG/arg4::hisG*) by stepwise deletion of both alleles using PCR-amplified *HAP3::ARG4* and *HAP3::HIS1* cassettes (Gola et al., 2003) and a cIP10 plasmid containing *HAP3* and its promoter and terminator sequences integrated at the RP10 locus (Murad et al., 2000). The homozygous knock-out of *cda2* and revertant were kindly provided by Neil Gow (Aberdeen, UK).

CELL CULTURE OF PRIMARY IMMUNE CELLS FROM MOUSE BONE MARROW

Bone marrow was differentiated to either BMDs or bone marrow-derived macrophages (BMDMs) from the femurs of 7- to 9-week-old wild type C57BL/6 mice and assessed for homogeneity as previously described using a panel of marker antibodies (Bourgeois et al., 2009).

FUNGAL-MAMMALIAN CELL CO-CULTURE

Fungal-mammalian cell co-cultures were performed as previously described (Bourgeois et al., 2009). Briefly, immune cells were plated at a density of 1.0×10^5 cells/cm² in sterile cell culture dishes and incubated with fungal cells at a multiplicity of infection (MOI) of five fungal cells per immune cell. Samples were incubated at 37°C in 5% CO₂, 95% humidity for up to 24 h.

cDNA PREPARATION FOR RNA-seq

Total RNA was isolated from immune cells and *C. albicans* using the SV total RNA isolation system (Promega, Madison, MI, USA) following manufacturers instructions. To obtain RNA mixtures from both *C. albicans* and BMDs, cells were first scraped in the provided lysis buffer, followed by homogenization with 200 μl of 0.5 mm acid-washed glass beads (Sigma-Aldrich, St. Louis, MO, USA) in a Fast Prep-24 cooling block at 4°C (MP Bio-medicals Europe, Illkirch, France) for 45 s at 5 m/s. Ribosomal RNA was depleted from 10 μg of pooled total RNA samples using the RiboMinus eukaryote kit for RNA-seq (Invitrogen, Carlsbad, CA, USA) and concentrated using the corresponding RiboMinus Concentration Module (Invitrogen) following manufactures instructions for three independent biological repeats. For each sample, 1 μg of ribosomal-depleted RNA was converted into cDNA using the SMARTer PCR cDNA Synthesis kit and the Advantage 2 polymerase mix (Clontech, Mountain View, CA, USA). PCR amplifications were performed on 1/10 of the first strand synthesis reaction for 18 cycles of 90°C for 1 min, 95°C for 15 s, 65°C for 30 s, and 68°C for 6 min on a GeneAmp PCR system 9700 (Applied Biosystems, Carlsbad, CA, USA), and purified on ChromaSpin columns (Clontech, Mountain View, CA, USA). The resulting cDNAs were sequenced on the Genome Analyzer IIx at GATC (Konstanz, Germany) using 36 bp, single run, indexed read mode.

SEQUENCE READ MAPPING, PRE-PROCESSING, AND DATA NORMALIZATION

All sequencing reads were mapped using TopHat 1.2.0 (Trapnell et al., 2009) against the SC5314 *C. albicans* assembly 21 (Skrzypek et al., 2010) and the *M. musculus* UCSC version mm9 from the ENSEMBL database (Flicek et al., 2011). Mapping was carried out using the default settings in which only unique hits were kept for

further analysis. The gene expression and normalization analysis was performed as previously described (Mortazavi et al., 2008). Genes were tested for differential expression using the bioconductor package baySeq (Hardcastle and Kelly, 2010) relative to the 0-min infection time point. The analysis was carried out for *C. albicans* and *M. musculus* genes individually.

CLUSTERING AND OVER-REPRESENTED GENE ONTOLOGY TERMS

Fuzzy c-means clustering (Bezdek, 1992) was applied to the two expression matrices of differentially expressed genes from *C. albicans* and *M. musculus*. The optimal number of clusters was estimated as previously described (Guthke et al., 2005; Linde et al., 2010). Functional categorization and significantly over-represented categories were identified using the tool FungiFun (Priebe et al., 2011). All four hierarchical levels of Funcat (Ruepp et al., 2004) and Gene Ontology (Ashburner et al., 2000) categorization were used in this study.

NETWORK INFERENCE PREDICTION AND MEASURING INTERACTION ROBUSTNESS

Network inference was performed as previously described using the NetGenerator tool (Guthke et al., 2005; Linde et al., 2010). Briefly, NetGenerator is based on a set of linear differential equations and models the temporal change of the expression intensity $x_i(t)$ of gene i ($i = 1 \dots n$) at time t as the weighted sum of the expression intensities of all other genes and an external stimulus $u(t)$ at time t . The external stimulus $u(t)$ is modeled as a stepwise constant function representing the change from no host–pathogen interaction to the onset of the interaction. The tool aims to identify a network structure, which best fits to the measured RNA-seq data, while it minimizes the number of predicted interactions (Guthke et al., 2005). Thus, a sparse network is inferred.

NetGenerator offers the possibility to integrate prior knowledge (i.e., putative regulatory interactions based on additional data besides the initial time series expression data). Based on the confidence of the prior knowledge source, it is possible to score each proposed interaction. The confidence of the prior knowledge is based on the level of experimentation used to verify a specific interaction and the number of independent experiments showing the same interaction. Since different data sources might be contradictory, prior knowledge is softly integrated, i.e., if a proposed interaction contradicts the measured data too much it can be removed by NetGenerator. Furthermore, the tool may add new interactions not covered by the prior knowledge in order to fit to the measured data. In this study, prior knowledge from public databases was softly integrated (Guthke et al., 2005). Each proposed interaction was scored in an additive manner based on the confidence of the prior knowledge source as follows: direct evidence that a gene is involved in a host–pathogen interaction (confidence score = 0.5), co-expression of two genes (confidence score = 0.25), and the occurrence of the respective transcription factor binding motif in the upstream intergenic regions of genes (confidence score = 0.125). Prior knowledge was obtained from GeneMania¹, IntAct (Aranda et al., 2010), BioGrid (Stark et al., 2011), the *C. albicans* database (Skrzypek et al., 2010), the mouse genome database

(Blake et al., 2011), and a number of peer-reviewed publications (Lane et al., 2001; Doedt et al., 2004; Martchenko et al., 2004; Zhao et al., 2004; Fradin et al., 2005; Oberholzer et al., 2006; Wang et al., 2006; Spira et al., 2007; Thewes et al., 2007; Zakikhany et al., 2007; Almeida et al., 2008; Baek et al., 2008; Nobile et al., 2008; Frohner et al., 2009; Griffin et al., 2009; Raman et al., 2009; Sellam et al., 2009; Hinze et al., 2010; Hou et al., 2010; Smith et al., 2010; Wachtler et al., 2011) summarized in Figure 1C. Putative regulatory interactions were tested for robustness using two methods. First, Gaussian noise was introduced with a mean of 0 and SD 0.05 to the estimated mRNA concentrations for 1000 iterations. Secondly, predicted interactions were screened for robustness against changes in prior knowledge by iterating the modeling approach 1000 times while randomly skipping 10% of all interactions in the set of prior knowledge for each run. Only edges that were confirmed by more than 50% of the iterations were considered to be robust and used in the resultant model.

REAL-TIME qPCR ANALYSIS

RNA sample preparation, reverse transcription, and real-time PCR were performed previously described (Bourgeois et al., 2009) using the following primers: mouse β -Actin, forward 5'-GCGTGACATCAAAGAGAG-3' reverse 5'-AGGAGCCAGAGCAGTAATC-3' (RTPrimerDB)² mouse MTA2, forward 5'-CACTGCTATAGCCTCAGGCC-3', reverse 5'-GCTAGGAGCTGGAACCTCAC-3', mouse PTX3, forward 5'-CCTGCTTTGTGCTCTCTGGT-3', reverse 5'-TCTCCAGCATGATGAACAGC-3' (Diniz et al., 2004), *C. albicans* TUP1, forward 5'-GACTACGCCTCAAACGAAGC-3' reverse 5'-TGGTGCCACAATCTGTTGTT-3', *C. albicans* FRE6 forward 5'-CCGGTAAACATCCATTCCAC-3', reverse 5'-TTGATCCAAATGCCATT-CAA-3', *C. albicans* SEF1, forward 5'-GTGGAGGACTCGTTCATGGT-3', reverse 5'-TGAACCAGCAGATTCAGAG-3', *C. albicans* RPI1, forward 5'-TGCTGACAGAGTCAAGA-AACC-3' reverse 5'-GAACCAACCACCGAAATCAC-3' as determined using the sequence analysis software Vector NTI (Invitrogen, Carlsbad, CA, USA). Results were calculated using the $\Delta\Delta ct$ method and are expressed as the fold of the gene expression of interest versus the expression of a housekeeping gene in *M. musculus* (β -Actin) or *C. albicans* (RPI1) in treated versus untreated conditions.

CYTOKINE QUANTIFICATION FROM CO-CULTURE SUPERNATANTS

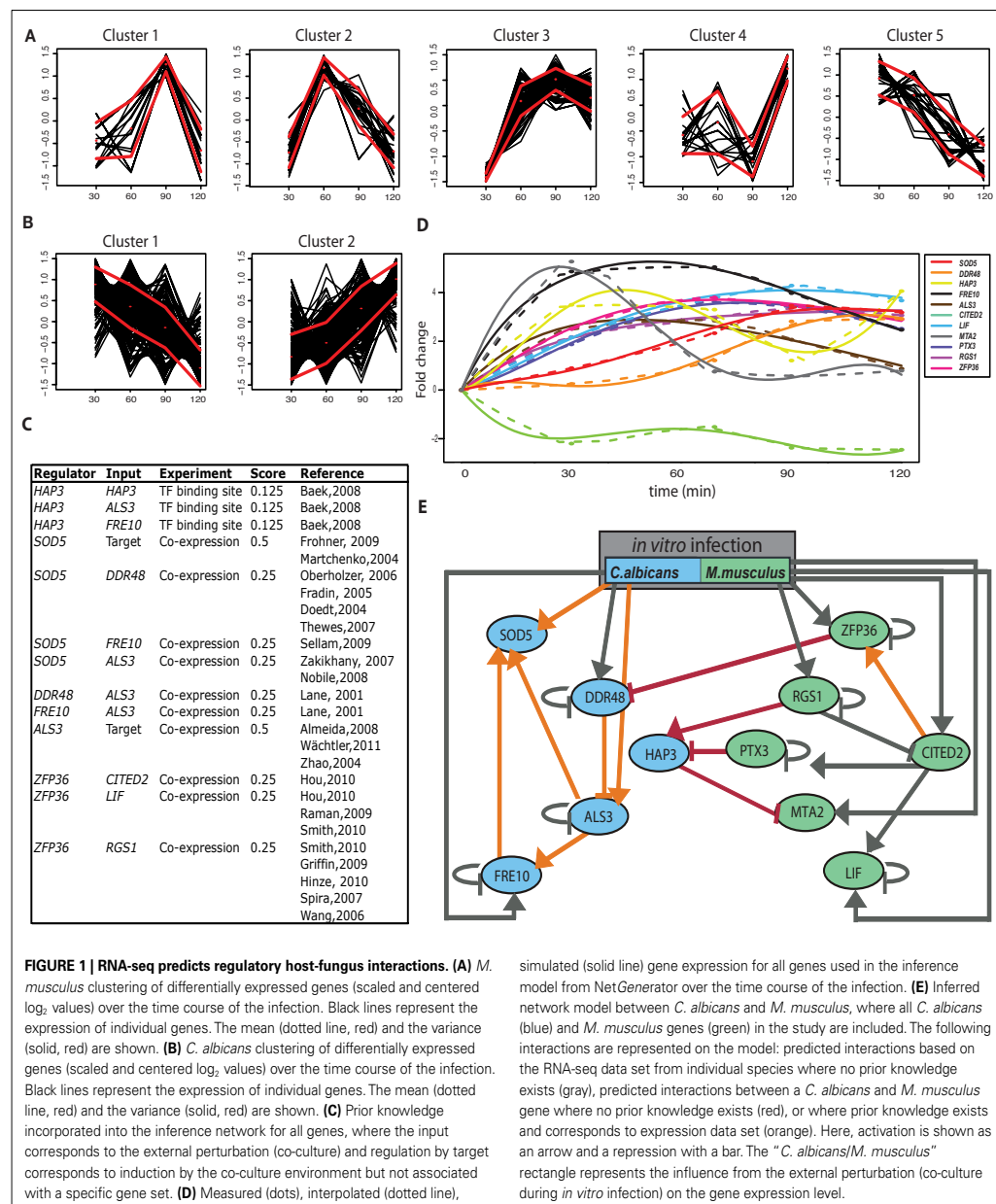
The amount of cytokines released into cell culture supernatants by immune cells during *in vitro* interaction studies with heat killed *C. albicans* were assayed after 24 h of co-culture using the mouse IL-2, IL-4, or TNF α Ready-set-go ELISA kit (R&D Systems, Minneapolis, MN, USA) or the Mouse Cytokine Array Panel A kit (R&D Systems) according to the manufacturers instructions.

BINDING AND LABELING Ptx3 IN VITRO

Recombinant mouse Ptx3 (rmPtx3) protein (R&D Systems) was reconstituted in sterile PBS and diluted for all experiments. Some 0.5×10^6 fungal cells were incubated for 1 h at 37°C with 5 μ g reconstituted rmPtx3. Ptx3 was labeled with the primary antibody against Ptx3 (Abnova, Taiwan) and secondarily

¹ <http://genemania.org/data/>

² <http://medgen.ugent.be/rtprimerdb/index.php>



labeled with goat-anti-rabbit 649 Dylight (Thermo Scientific, Rockford, Illinois). Fungal cell wall chitin was labeled using 10 μ M of Calcofluor White (Sigma-Aldrich). Intracellular labeling of Ptx3 was performed using the BD Cytofix/Cytoperm Fixation/Permeabilization kit with BD GolgiPlug protein transport inhibitor (BD Biosciences, Heidelberg) after 6 h of *C. albicans* infection following manufactures instructions. Preparations were

assessed by flow cytometry or visualized on an Olympus Cell-R live imaging unit (Olympus, Essex, UK) for all experiments.

STATISTICAL ANALYSIS FOR INFERENCE MODEL VERIFICATION

Statistical analysis of data was performed using the GraphPad Prism graphing and analysis software (GraphPad Software, San Diego, USA) for all *in vitro* experiments excluding the RNA-seq

analysis described above. Statistical significance was assessed using with the Student *t*-test and a *p*-value <0.05 was considered significant.

RESULTS

INFERRED REGULATORY NETWORK IDENTIFIES NOVEL INTERSPECIES HOST–PATHOGEN INTERACTIONS

We used massively parallel RNA sequencing of cDNA (RNA-seq) obtained from co-cultures of *C. albicans* and *M. musculus* BMDCs over 2 h to model an infection time course from fungal adhesion to early host cell lysis. In total, we obtained approximately 120 million reads, which were mapped to the *C. albicans* 21 assembly or *M. musculus* mm9 genome and analyzed each for differential gene expression relative to the pre-infection state. We identified 545 differentially expressed genes for *C. albicans* and 240 for *M. musculus* over the complete time course.

The small number of measured data points for each gene over the time course restricts the modeling approach to a limited number of genes. If there was no pre-selection of the genes, or a large number of genes were to be used, it would result in an over-fitting of the measured data that would not produce a robust inference model. For this reason, it is necessary to select a set of relevant genes to be represented by nodes in the network model. To identify candidate genes in *C. albicans* and *M. musculus*, all differentially expressed genes were first clustered (Bezdek, 1992) by their kinetics during the time course (Figures 1A,B). From each cluster, one or more representative genes were chosen for use within the model. Several considerations were taken into account for the selection of candidate genes. In *C. albicans*, we preferentially chose genes that have been either annotated as virulence genes (i.e., adhesion, hyphal formation, or response to host) or strongly respond to infection or infection-like conditions (i.e., temperature stress, nutrient limitation, or iron regulation). For *M. musculus*, we prioritized genes with phenotypes relating to the immune defense or response, or susceptibility to pathogens in a systemic mouse model of infection.

A number of recent studies have shown the reverse engineering approach is greatly improved by the integration of different data sources (Werhli and Husmeier, 2007; Gustafsson et al., 2008; Hecker et al., 2009a,b). We therefore collected putative regulatory interactions based on additional data obtained from literature, referred to as prior knowledge, for each gene. Based on the confidence of the prior knowledge source, a score is attributed to each interaction (see Materials and Methods). Since different data sources might be contradictory, prior knowledge was softly integrated so that if a proposed interaction contradicts the measured data to a great extent, it can be removed from the resulting network (see Materials and Methods). Genes with no known or predicted function were therefore excluded from the analysis. Based on these criteria, we narrowed our gene lists to five from *C. albicans* and six from *M. musculus*. Prior knowledge scores (Figure 1C) and expression kinetics (Figure 1D) for the candidate genes were combined in NetGenerator to generate the final network inference (Figure 1E).

To verify the fit of the model to the actual expression kinetics of the candidate genes, we first used NetGenerator to interpolate and simulate gene expression for the measured data points of each gene (Figure 1D). The closer expression profiles for the individual

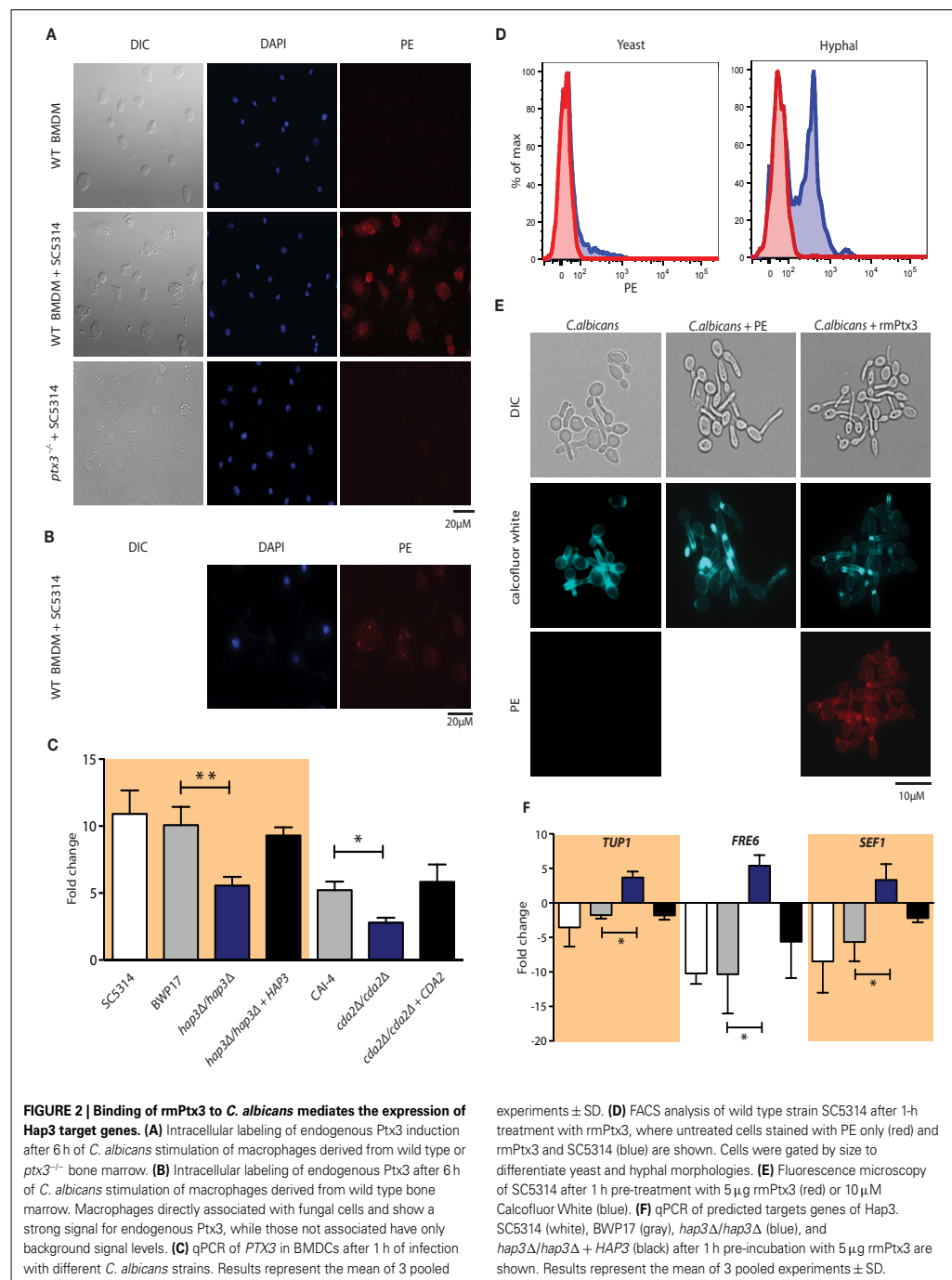
genes fit to the measured data points, the better the inference prediction is. We found a close relationship between the simulated and measured data points, showing that the NetGenerator model is representative of the measured data. The final interspecies network was based on these predictions (Figure 1E). Only edges that were robust against the addition of Gaussian noise and partial skipping of prior knowledge were used in the construction of the model. The final network predicts 21 putative edges, including 4 interspecies edges.

To specifically test the robustness of the interspecies edges experimentally, we focused on a sub-network composed of a single *C. albicans* transcription factor Hap3 that was predicted in the inference model to contain two interactions with *M. musculus* genes. These interactions include a predicted blunt or repressing edge between fungal Hap3 by murine Ptx3, and a predicted blunt edge of murine Mta2 by fungal Hap3 itself.

THE BINDING OF MURINE Ptx3 REGULATES Hap3 TARGET GENES

Pentraxin 3 (Ptx3) is a soluble pattern recognition receptor that has been previously shown to function as an opsonin to facilitate pathogen uptake by phagocytic cells in a lectin-1 dependent manner (Diniz et al., 2004). To determine if *M. musculus* Ptx3 blocked fungal Hap3 function or the expression of Hap3-regulated genes in *C. albicans* as suggested by the inference model, we first asked whether Ptx3 was induced upon infection with *Candida* cells. Using intracellular staining for Ptx3, we detected a strong fluorescence signal in BMDMs infected with *C. albicans*, whereas no signal was detected in BMDMs alone or *ptx3*^{−/−} macrophages (Figures 2A,B). Furthermore, we found an increase in *PTX3* mRNA levels in BMDCs (Figure 2C) and BMDMs (data not shown) after 1 h of *C. albicans* infection, verifying that Ptx3 is indeed induced in our experimental setup. Interestingly, the amount of *PTX3* induced significantly decreased in the absence of Hap3 (Figure 2C). We detected a similarly significant decrease in *PTX3* induction in the gene containing the Hap3 binding box, *Cda2*, a predicted chitin deacetylase in *C. albicans*.

Ptx3 has been previously shown to bind to numerous fungi, including *Aspergillus fumigatus* (Jaillon et al., 2007) as well as zymosan-coated particles (Diniz et al., 2004). Therefore, we asked whether recombinant mouse Ptx3 (rmPtx3) could also bind to the *C. albicans* cell wall. We assessed rmPtx3 binding using fluorescence microscopy and flow cytometry. Fungal cells pre-incubated with rmPtx3 for 1 h at 37°C showed surface localization of Ptx3 to the cell wall. No signal was visible on cells treated with the PE-conjugated secondary alone or the untreated control (Figure 2E). Notably, the labeling pattern of Ptx3 coincided with areas of expected chitin exposure. Hence, we also stained *C. albicans* cells with Calcofluor White, a fluorescent dye that binds to exposed chitin (Bulawa et al., 1995). Interestingly, the localization of rmPtx3 on the *C. albicans* overlapped with the signal detected for Calcofluor White alone (Figure 2E), suggesting that Ptx3 might bind to accessible cell surface chitin. We confirmed and quantified the amount of Ptx3 binding to the cell wall using flow cytometry. Based on the binding observed in the fluorescence microscopy, we analyzed our data using both the complete pool of *C. albicans* cells as well as discriminating yeast and hyphal forms (Figure 2D; Figure A1 in Appendix). We detected a high level of



binding in fungal hyphae compared to yeast form cells. This is consistent with our fluorescence microscopy analysis, where we

detected much stronger signals on the hyphal cell walls compared to the bud scars on yeast form cells.

Given that rmPtx3 binds to the fungal cell surface, we assessed if rmPtx3 binding influenced the expression of predicted fungal Hap3 target genes as predicted by the inferred network using qPCR (Figure 2F). There are 10 predicted target genes of Hap3 that were recently identified in a network inference study using microarray data from *C. albicans* during *in vitro* epithelial infection, where iron is assumed to be limited (Linde et al., 2010). Out of the 10 putative Hap3 target genes, we found three, *TUP1*, *FRE6*, and *SEF1*, whose expressions were significantly decreased in *C. albicans* after rmPtx3 binding, verifying their functionality as Hap3 target genes. Interestingly, the levels of these genes increased in the Hap3 knock-out. These data strongly suggest that their down-modulation upon binding of rmPtx3 is Hap3-dependent.

CANDIDA ALBICANS BOUND BY Ptx3 ATTENUATES THE IMMUNE RESPONSE IN BMDCs

Recently it was shown that the binding of recombinant human Ptx3 increases *A. fumigatus* conidia phagocytosis and influences cytokine production. Those mice lacking Ptx3 were additionally found to be more susceptible to *A. fumigatus* infection (Moalli et al., 2011). To determine if the binding of Ptx3 to *C. albicans* changed the cytokine production of murine immune cells *in vitro*, we first investigated the gross immune response using a cytokine array after 24 h of co-culture with BMDCs (Figure A2 in Appendix). RmPtx3-bound *C. albicans* induced multiple cytokines compared to untreated *C. albicans* including IL-2, a cytokine regulated by Mta2, as well as the inflammatory cytokines KC, JE, and TNF α (Figures A2 in Appendix). Interestingly, when we then compared Hap3 knock-out cells pre-incubated with rmPtx3, we detected a general increase in these cytokines in addition to IL-23, IL-17, IL-16, and IL-10 that were not detected in using wild type *C. albicans* (Figure A2 in Appendix).

Since the cytokine array is a qualitative assessment of cytokine production with a relatively high detection threshold, we verified the changes in cytokine levels for the cytokines most relevant to the inference model, namely, IL-2 and IL-4, both target cytokines of Mta2, and TNF α , a pleiotropic inflammatory cytokine, by ELISA. Mta2 is a member of the NuRD (nucleosome remodeling and histone deacetylase) complex in *M. musculus* (Manavathi et al., 2007) and predicted in our network model as repressed by Hap3. The cytokines IL-2 and IL-4 produced during the host immune response were both recently identified as targets of the Mta2/NuRD complex (Lu et al., 2008). We found that in BMDCs, *MTA2* increased in the absence of Hap3, suggesting that Hap3 might indirectly regulate expression of Mta2 (Figure 3A). We quantified cellular cytokine release using wild type BMDCs with *C. albicans*, wild type BMDCs with rmPtx3-bound *C. albicans*, and *ptx3*^{-/-} BMDCs. We found that the production of IL-2 and IL-4 significantly increased in the absence of Hap3 (Figures 3B,C). This increase was augmented by *C. albicans* cells pre-incubated with rmPtx3, confirming our observation from the cytokine array that there is a general increase in the production of these cytokines when there is an increase in Ptx3. Interestingly, compared to wild type BMDCs, the basal level of cytokine production of IL-2 and IL-4 increased in *ptx3*^{-/-} BMDCs, corresponding to our inferred network prediction that the loss of its predicted negative regulators, Ptx3 and fungal Hap3, would increase the expression of

Mta2 and thereby increase the expression of its target cytokines. In Hap3 knock-out cells, we found both on the cytokine arrays and by ELISA a significant decrease in TNF α (Figure 3D). These data show that the binding of Ptx3 to fungal cells alters the cytokine production by immune cells in a Hap3-dependent manner, and the regulation of Mta2 target cytokines indirectly suggests an involvement of Mta2 as predicted by the network inference model.

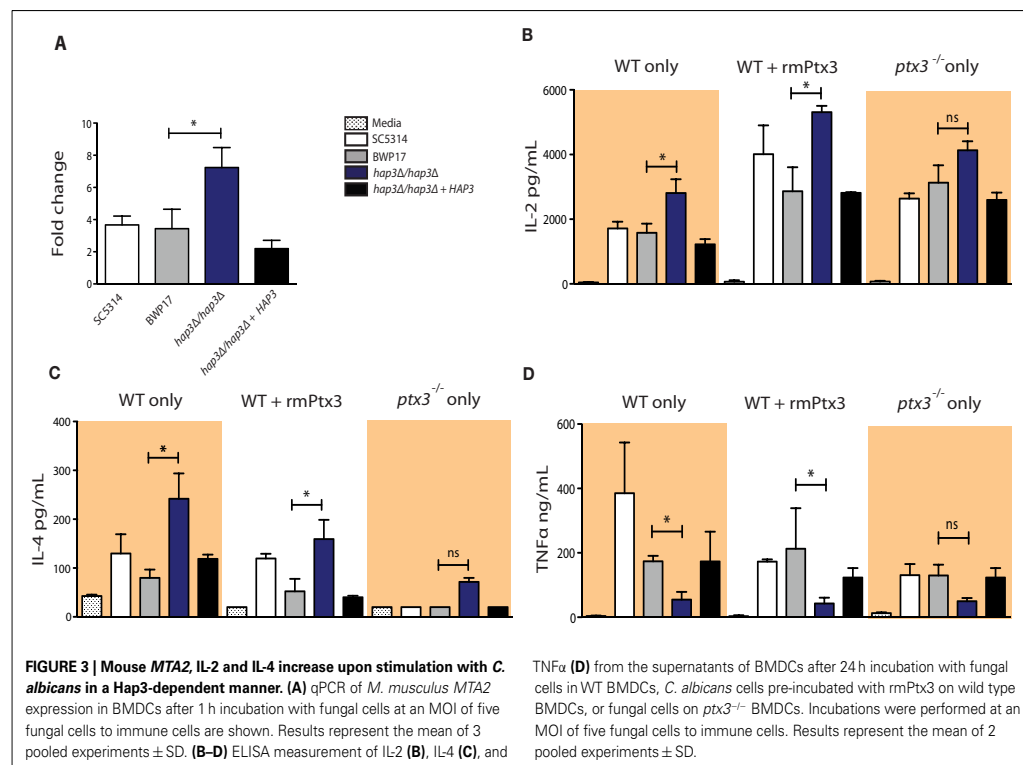
IDENTIFYING CELL SURFACE Hap3 TARGET GENES

To identify how immune cells could detect the regulation of the transcription factor Hap3 in *C. albicans*, we searched for putative Hap3 target genes that could have more direct contact with immune cells, including: cell wall, plasma membrane or secretory proteins. We focused on *C. albicans* genes of cluster 2, since their expression strongly increased expression over the time course of invasion (Figure 1B). Within this cluster, we scanned for genes harboring the binding site of the Hap-complex in their upstream regulatory regions (Baek et al., 2008). We further narrowed down the candidate list by removing genes that did not have a predicted cellular localization or function in the *C. albicans* database (Skrzypek et al., 2010). Following these selection criteria, nine candidate genes were left that we used to infer an additional network in combination with Ptx3, Hap3, and Mta2 to determine if an interaction could be inferred with a protein that could come in direct contact with immune cells (Figure A3 in Appendix). To increase the reliability of the putative Hap3 interactions within the new interaction network, we included the validated interactions from our experiments within this study (repression of *HAP3* by Ptx3 and *MTA2* by Hap3), as additional prior knowledge. Of all of the candidate genes, only the activation of *CDA2* (a putative chitin deacetylase in *C. albicans*) by Hap3, was robust against Gaussian noise and partial skipping of prior knowledge.

DISCUSSION

In this study, we aimed to infer a network that predicts interactions between host and pathogenic species under infection settings. To our knowledge, this is the first network inference approach predicting host–pathogen interactions. This approach allowed for the prediction, identification, and experimental verification of interdependent sub-networks composed of a single *C. albicans* transcription factor Hap3, and the *M. musculus* genes Ptx3 and Mta2. The experimental validation suggests a putative mechanism to explain how these interactions could be regulated during infection of immune cells by fungal pathogens.

Our modeling approach was fundamentally based on differential equations, which have been previously used to infer regulatory network models (Toepfer et al., 2007; Linde et al., 2010). This approach is generally suitable for time series data. Nevertheless, this approach is inappropriate for large-scale modeling, because a large number of genes incorporated into a differential equation based model leaves open a number of parameters to be identified. This may result in an over-fitting of the data. The modeling approach uses four attempts to prevent over-fitting. First, we restrict the number of genes within the model such that a smaller number of parameters need to be identified. Second, it aims at inferring a sparse network where many parameters are zero. Thirdly, it makes use of re-sampling techniques where the data are



perturbed in a random manner. Finally, we make use of prior knowledge guiding the inferred structure to a knowledge-based solution. Thus skipping incorrect network structures.

Gene expression levels, as well as available prior biological knowledge, were used to aid in the narrowing of genes that we chose to incorporate into the model. For this reason, genes where no biological knowledge was available were excluded from further analysis. However, we cannot exclude the possibility that additional genes of unknown function might also play a role in our inference model. This remains a limitation of the modeling approach, in so far as predictions can only be made for genes where a reasonable amount of prior knowledge is available. The genes incorporated into the model represent only one possible scenario of interactions and we do not exclude the possibility that other genes may play a role under other conditions. We have already started to take first step for a full-genomic network modeling for *C. albicans* utilizing a compendium of all available expression data (Altwater et al., 2012). Moreover, we primarily focused on genes acting as putative network “hubs” in their organisms (Bulawa et al., 1995). Hubs are genes such as transcription factors that regulate many other downstream genes within a network either directly or indirectly. Hubs were chosen because they are less likely to have redundant roles. Therefore, we would expect a stronger phenotype than investigating genes that are sparsely connected. This also means that the interactions we are investigating are more likely to be indirect and should be interpreted with caution.

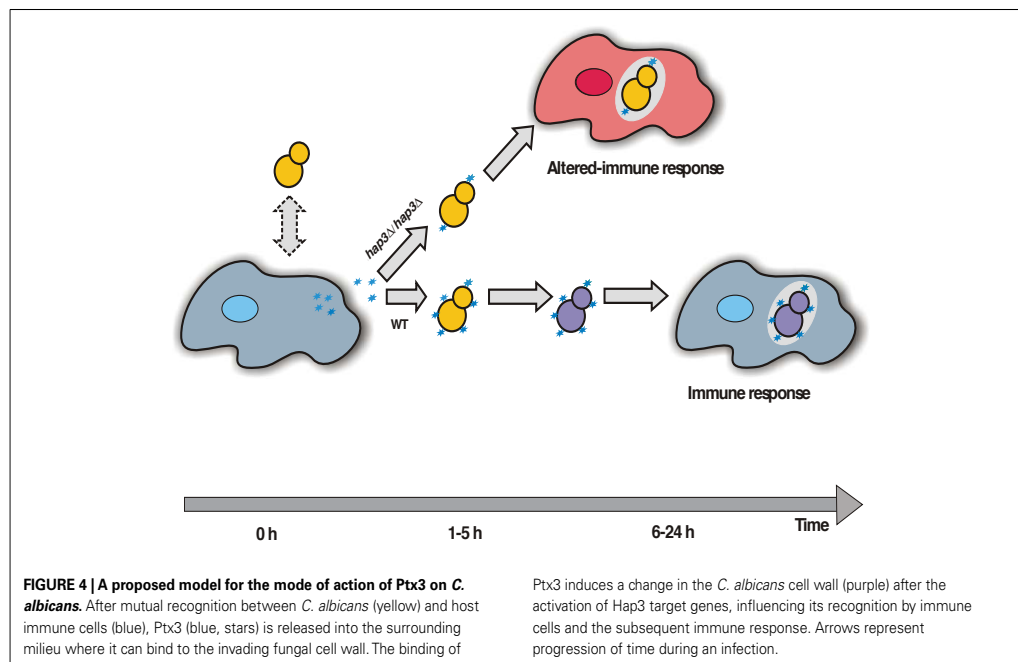
From our original candidate gene list, we inferred *HAP3* as a putative network hub targeted by innate immune cells. Interestingly, several putative target genes of Hap3 identified in this study are predicted to localize to the plasma membrane, cell wall, or are involved with cell wall reorganization in *C. albicans*. The fungal cell wall is a dynamic structure, which undergoes significant structural and molecular composition remodeling throughout its life cycle, as well as in response to a variety of external stimuli (Chaffin, 2008). As Hap3 in *C. albicans* is a transcription factor up-regulated under iron-limiting conditions (Linde et al., 2010), it is likely that its function during fungal recognition or phagocytosis by immune cells is indirect. Of all of the candidate cell surface Hap3 targets, only Cda2, a putative chitin deacetylase forms a robust interaction with Hap3 within the second network (Figure A3 in Appendix). Chitin deacetylase enzymes exist in both intracellular and secreted forms in different fungi, where they hydrolyzes the acetamido group in the *N*-acetylglucosamine units of chitin and chitosan, leaving glucosamine units and acetic acid form as byproducts (Zhao et al., 2010). Chitin deacetylases exist in both *Saccharomyces cerevisiae* (Martinou et al., 2002) and in the opportunistic fungal pathogen *Cryptococcus neoformans*, where they have been suggested as an antifungal target due to their severe effect on cell wall integrity (Baker et al., 2007). Notably, chitin deacetylases are secreted during different developmental stages of some other fungi (Zhao et al., 2010). For example, in *Colletotrichum lindemuthianum*, a plant fungal pathogen, chitin deacetylases are

exclusively secreted during hyphal penetration into plant tissue (Tokuyasu et al., 1996). We find that Ptx3 induction is decreased in the *CDA2* knock-out (Figure 2C), further suggesting a possible connection to the inferred network model. These data are consistent with the overlap of Ptx3 staining and that of Calcofluor White (Figure 2E), which binds to exposed chitin. Therefore, it is tempting to speculate that the recognition of *C. albicans* by immune cells triggers the production of this enzyme to induce cell wall remodeling as an evasion strategy. However, further work beyond the scope of this study is needed to decipher the specific function of Cda2 in *C. albicans* and its connection to Hap3.

We observed that upon binding of rmPtx3 to fungal cells, the *C. albicans* virulence genes *TUP1*, *FRE6*, and *SEF1* mRNA levels significantly decreased in a Hap3-dependent manner (Figure 2F). Tup1 has a well-characterized role as a key regulator in *C. albicans* morphogenesis (Braun and Johnson, 1997). We cannot exclude the possibility that Hap3 and Tup1 may have similar or even complementary functions during interaction with immune cells. Interestingly, both Tup1 and Fre6 are either directly or indirectly involved in the *C. albicans* cell wall homeostasis. Tup1 is a multifunctional transcriptional co-repressor of filamentous growth in *C. albicans* whose lack leads to constitutive filamentous growth (Braun and Johnson, 1997; Park and Morschhäuser, 2005). Fre6 is an uncharacterized protein, for which *in silico* predictions suggest it to reside in the plasma membrane with a putative functional similarity to the ferric reductase Fre10, an important protein in iron acquisition (Knight et al., 2005). Therefore, their regulation upon binding or phagocytosis might play an additional role in cell wall remodeling during infection. Since fungal cells experience severe iron-limiting condition within phagosomes of host

cells, Hap3 and Fre6 appear as logical candidates involved in this reciprocal interaction. Likewise, Sef1 regulates iron uptake, and has recently been shown to promote virulence in a mouse model of bloodstream infections (Chen et al., 2011). Interestingly, it was shown that knock-out mice lacking Ptx3 are hyper-susceptible to *A. fumigatus* (Moalli et al., 2010). However, no *in vivo* work has been performed to date using *ptx3*^{-/-} mice and *C. albicans*. A recent study has shown that the activation of the complement system via the lectin pathway can be triggered via a complex of Ptx3 and mannose binding lectin (MBL) on *C. albicans* mannan *in vitro* (Ma et al., 2011). They showed the MBL–Ptx3 complex could enhance the deposition of the complement components C3 and C4 and thereby increase phagocytosis of *C. albicans* by polymorphonuclear leukocytes. It has previously been shown that C3 knock-out mice are additionally more susceptible to *C. albicans* infections (Han et al., 2001). Therefore, it is possible that the absence of Ptx3 could result in reduced activation of the complement pathway and reduced fungal killing. *In vivo* studies using *ptx3*^{-/-} mice would be needed to investigate this hypothesis.

We found that the expression of *MTA2* and the regulation of its downstream targets such as cytokines IL-2 and IL-4, are increased during immune cell invasion by *C. albicans* in a Hap3-dependent manner (Figures 3B,C). Moreover, we found that an altered-immune response is one consequence of rmPtx3 binding. Mta2 knock-out mice display partial embryonic lethality, while the surviving mice develop lupus-like autoimmune symptoms, including severe developmental phenotypes (Lu et al., 2008). Mta2 is uniquely associated with the NuRD chromatin complex, which has both nucleosome remodeling and histone deacetylase activity (Feng and Zhang, 2003). Although there has been no data



to date in fungi indicating a role for host chromatin in pathogenesis, recent work in bacteria and viruses (Hamon and Cossart, 2008; Rohde, 2011) shows that chromatin remodeling is induced in host cells during invasion. Consistent with these observations, our data suggests that the regulation of *MTA2* may affect chromatin remodeling in immune cells in the response to fungal pathogens. The resultant altered-immune response may be disadvantageous to the pathogen because it would promote fungal clearance.

We propose that Hap3 constitutes a target hub of *C. albicans*, which actively regulates immune responses through the reorganization of the *C. albicans* cell wall during invasion of innate immune cells (Figure 4). Specifically, we propose a model in which the binding of Ptx3 released from immune cells to *C. albicans* cell wall triggering the reorganization of the *C. albicans* cell wall and plasma membrane via the activation of Hap3 target genes. This reorganization in turn changes the recognition of the fungus by immune cells and attenuates the host immune response. This work demonstrates the possibility to experimentally verify predicted host–pathogen relationships based on an interspecies model of network inference, showing that inference modeling can be used in the investigation of microbial pathogenesis. We propose that this method could be useful for the identification of antifungal target genes.

CONTRIBUTION

Lanay Tierney designed research, performed experiments, analyzed data and co-wrote the manuscript. Jörg Linde generated

network inference maps, analyzed data, and co-wrote the manuscript. Sebastian Müller performed the sequence alignment and normalization. Sascha Brunke, Juan Camilo Molina, and Bernhard Hube provided *C. albicans* Hap3 deletion strains. Reinhard Guthke and Karl Kuchler designed the research and co-wrote the manuscript.

ACKNOWLEDGMENTS

We would like to thank Alberto Mantovani and Cecilia Garlanda (Milan, Italy) for kindly providing *ptx3*^{−/−} bone marrow. We are also indebted to Neil Gow (Aberdeen, UK) for kindly providing the Cda2 knock-out *C. albicans* strains. This work was supported by a grant from the Christian Doppler Research Society to Karl Kuchler, and in part by the FWF-DACH grant of the Austrian Science Foundation (FWF-Proj.: I-746-B11) to Karl Kuchler and Bernhard Hube. Jörg Linde was supported by the excellence graduate school “Jena School for Microbial Communication (JSMC).” We would like to thank laboratory members for critical reading and helpful comments on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/Microbial_Immunology/10.3389/fmicb.2012.00085/abstract

Table S1 | RPKM values for *C. albicans* genes over the infection time course.

Table S2 | RPKM values for *M. musculus* genes over the infection time course.

REFERENCES

- Almeida, R. S., Brunke, S., Albrecht, A., Thewes, S., Laue, M., Edwards, J. E., Filler, S. G., and Hube, B. (2008). The hyphal-associated adhesin and invasin Als3 of *Candida albicans* mediates iron acquisition from host ferritin. *PLoS Pathog.* 4, e1000217. doi:10.1371/journal.ppat.1000217
- Altwasser, R., Linde, J., Buyko, E., Hahn, U., and Guthke, R. (2012). Genome-wide scale-free network inference for *Candida albicans*. *Front. Microbiol.* 3:51. doi:10.3389/fmicb.2012.00051
- Aranda, B., Achuthan, P., Alam-Faruke, Y., Armean, I., Bridge, A., Derow, C., Feuermann, M., Ghanbarian, A. T., Kerrien, S., Khadake, J., Leroy, C., Menden, M., Michaut, M., Montecchi-Palazzi, L., Neuhauser, S. N., Orchard, S., Perreau, V., Roechert, B., van Eijk, K., and Hermjakob, H. (2010). The IntAct molecular interaction database in 2010. *Nucleic Acids Res.* 38, D525–D531.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* 25, 25–29.
- Baek, Y. U., Li, M., and Davis, D. A. (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.
- Baker, L. G., Specht, C. A., Donlin, M. J., and Lodge, J. K. (2007). Chitosan, the deacetylated form of chitin, is necessary for cell wall integrity in *Cryptococcus neoformans*. *Eukaryot. Cell* 6, 855–867.
- Barabasi, A. L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113.
- Bezdek, J. (1992). *Fuzzy Models for Pattern Recognition: Methods that Search for Structures in Data*. New York: Institute of Electrical and Electronics Engineers (IEEE) Press.
- Blake, J. A., Bult, C. J., Kadin, J. A., Richardson, J. E., and Eppig, J. T. (2011). The mouse genome database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res.* 39, D842–D848.
- Bourgeois, C., Majer, O., Frohner, I., and Kuchler, K. (2009). *In vitro* systems for studying the interaction of fungal pathogens with primary cells from the mammalian innate immune system. *Methods Mol. Biol.* 470, 125–139.
- Bourgeois, C., Majer, O., Frohner, I. E., Tierney, L., and Kuchler, K. (2010). Fungal attacks on mammalian hosts: pathogen elimination requires sensing and tasting. *Curr. Opin. Microbiol.* 13, 401–408.
- Braun, B. R., and Johnson, A. D. (1997). Control of filament formation in *Candida albicans* by the transcriptional repressor TUP1. *Science* 277, 105–109.
- Bulawa, C. E., Miller, D. W., Henry, L. K., and Becker, J. M. (1995). Attenuated virulence of chitin-deficient mutants of *Candida albicans*. *Proc. Natl. Acad. Sci. U.S.A.* 92, 10570–10574.
- Butte, A. J., and Kohane, I. S. (2000). Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac. Symp. Biocomput.* 5, 418–429.
- Chaffin, W. L. (2008). *Candida albicans* cell wall proteins. *Microbiol. Mol. Biol. Rev.* 72, 495–544.
- Chen, C., Pande, K., French, S. D., Tuch, B. B., and Noble, S. M. (2011). An iron homeostasis regulatory circuit with reciprocal roles in *Candida albicans* commensalism and pathogenesis. *Cell Host Microbe* 10, 118–135.
- D'Haeseleer, P., Wen, X., Fuhrman, S., and Somogyi, R. (1999). Linear modelling of mRNA expression levels during CNS development and injury. *Pac. Symp. Biocomput.* 4, 41–52.
- Diniz, S. N., Nomizo, R., Cisalpino, P. S., Teixeira, M. M., Brown, G. D., Mantovani, A., Gordon, S., Reis, L. F., and Dias, A. A. (2004). PTX3 function as an opsonin for the dectin-1-dependent internalization of zymosan by macrophages. *J. Leukoc. Biol.* 75, 649–656.
- Doedt, T., Krishnamurthy, S., Bockmuhl, D. P., Tebarth, B., Stempel, C., Russell, C. L., Brown, A. J., and Ernst, J. E. (2004). APSES proteins regulate morphogenesis and metabolism in *Candida albicans*. *Mol. Biol. Cell* 15, 3167–3180.
- Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J. J., and Gardner, T. S. (2007). Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* 5, e8. doi:10.1371/journal.pbio.0050008
- Feng, Q., and Zhang, Y. (2003). The NuRD complex: linking histone modification to nucleosome remodeling. *Curr. Top. Microbiol. Immunol.* 274, 269–290.

- Flicek, P., Amode, M. R., Barrell, D., Beal, K., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., Gil, L., Gordon, L., Hendrix, M., Hourlier, T., Johnson, N., Kähäri, A. K., Keefe, D., Keenan, S., Kinsella, R., Komorowska, M., Koscielny, G., Kulesha, E., Larsson, P., Longden, I., McLaren, W., Muffato, M., Overduin, B., Pignatelli, M., Pritchard, B., Riat, H. S., Ritchie, G. R., Ruffier, M., Schuster, M., Sobral, D., Tang, Y. A., Taylor, K., Trevanion, S., Vandrovcova, J., White, S., Wilson, M., Wilder, S. P., Aken, B. L., Birney, E., Cunningham, P., Dunham, I., Durbin, R., Fernández-Suarez, X. M., Harrow, J., Herrero, J., Hubbard, T. J., Parker, A., Proctor, G., Spudich, G., Vogel, J., Yates, A., Zadissa, A., and Searle, S. M. (2011). Ensembl 2011. *Nucleic Acids Res.* 39, D800–D806.
- Fradin, C., De Groot, P., MacCallum, D., Schaller, M., Klis, F., Odds, F. C., and Hube, B. (2005). Granulocytes govern the transcriptional response, morphology and proliferation of *Candida albicans* in human blood. *Mol. Microbiol.* 56, 397–415.
- Friedman, S. R., Kottiri, B. J., Neaigus, A., Curtis, R., Vermund, S. H., and Des Jarlais, D. C. (2000). Network-related mechanisms may help explain long-term HIV-1 seroprevalence levels that remain high but do not approach population-group saturation. *Am. J. Epidemiol.* 152, 913–922.
- Frohner, I. E., Bourgeois, C., Yatsyk, K., Majer, O., and Kuchler, K. (2009). *Candida albicans* cell surface superoxide dismutases degrade host-derived reactive oxygen species to escape innate immune surveillance. *Mol. Microbiol.* 71, 240–252.
- Gillum, A. M., Tsay, E. Y., and Kirsch, D. R. (1984). Isolation of the *Candida albicans* gene for orotidine-5'-phosphate decarboxylase by complementation of *S. cerevisiae* ura3 and *E. coli* pyrF mutations. *Mol. Gen. Genet.* 198, 179–182.
- Gola, S., Martin, R., Walther, A., Dunkler, A., and Wendland, J. (2003). New modules for PCR-based gene targeting in *Candida albicans*: rapid and efficient gene targeting using 100 bp of flanking homology region. *Yeast* 20, 1339–1347.
- Griffin, T. A., Barnes, M. G., Ilowite, N. T., Olson, J. C., Sherry, D. D., Gottlieb, B. S., Aronow, B. J., Pavlidis, P., Hinze, C. H., Thornton, S., Thompson, S. D., Grom, A. A., Colbert, R. A., and Glass, D. N. (2009). Gene expression signatures in polyarticular juvenile idiopathic arthritis demonstrate disease heterogeneity and offer a molecular classification of disease subsets. *Arthritis Rheum.* 60, 2113–2123.
- Gudlaugsson, O., Gillespie, S., Lee, K., Vande Berg, J., Hu, J., Messer, S., Herwaldt, L., Pfaller, M., and Diekema, D. (2003). Attributable mortality of nosocomial candidemia, revisited. *Clin. Infect. Dis.* 37, 1172–1177.
- Gustafsson, M., Björkengren, J., and Tegne, J. (2008). "Soft integration of data for reverse engineering," in *International Conference on Systems Biology*, Göteborg, 127–127.
- Guthke, R., Albrecht, D., Brackhage, A. A., and Möller, U. (2007). Discovery of gene regulatory networks in *Aspergillus fumigatus*. *Lect. Notes Bioinform.* 4366, 22–41.
- Guthke, R., Möller, U., Hoffmann, M., Thies, F., and Topfer, S. (2005). Dynamic network reconstruction from gene expression data applied to immune response during bacterial infection. *Bioinformatics* 21, 1626–1634.
- Hamon, M. A., and Cossart, P. (2008). Histone modifications and chromatin remodelling during bacterial infections. *Cell Host Microbe* 4, 100–109.
- Han, Y., Kozel, T. R., Zhang, M. X., MacGill, R. S., Carroll, M. C., and Cutler, J. E. (2001). Complement is essential for protection by an IgM and an IgG3 monoclonal antibody against experimental, hematogenously disseminated candidiasis. *J. Immunol.* 167, 1550–1557.
- Hardcastle, T. J., and Kelly, K. A. (2010). Bayseq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* 11, 422. doi:10.1186/1471-2105-11-422
- Hecker, M., Goertsches, R. H., Engelmann, R., Thiesen, H. J., and Guthke, R. (2009a). Integrative modelling of transcriptional regulation in response to antirheumatic therapy. *BMC Bioinformatics* 10, 262. doi:10.1186/1471-2105-10-262
- Hecker, M., Lambeck, S., Toepfer, S., van Someren, E., and Guthke, R. (2009b). Gene regulatory network inference: data integration in dynamic models—a review. *BioSystems* 96, 86–103.
- Hinze, C. H., Fall, N., Thornton, S., Mo, J. Q., Aronow, B. J., Layh-Schmitt, G., Griffin, T. A., Thompson, S. D., Colbert, R. A., Glass, D. N., Barnes, M. G., and Grom, A. A. (2010). Immature cell populations and an erythropoiesis gene-expression signature in systemic juvenile idiopathic arthritis: implications for pathogenesis. *Arthritis Res. Ther.* 12, R123.
- Holter, N. S., Maritan, A., Cieplak, M., Fedoroff, N. V., and Banavar, J. R. (2001). Dynamic modelling of gene expression data. *Proc. Natl. Acad. Sci. U.S.A.* 98, 1693–1698.
- Hou, J., Aerts, J., den Hamer, B., van Ijcken, W., den Bakker, M., Riegman, P., van der Leest, C., van der Spek, P., Foekens, J. A., Hoogsteden, H. C., Grosveld, F., and Philipsen, S. (2010). Gene expression-based classification of non-small cell lung carcinomas and survival prediction. *PLoS ONE* 5, e10312. doi:10.1371/journal.pone.0010312
- Jaillon, S., Peri, G., Delneste, Y., Fremaux, I., Doni, A., Moalli, F., Garlanda, C., Romani, L., Gascan, H., Bellocchio, S., Bozza, S., Cassatella, M. A., Jeannin, P., and Mantovani, A. (2007). The humoral pattern recognition receptor PTX3 is stored in neutrophil granules and localizes in extracellular traps. *J. Exp. Med.* 204, 793–804.
- Knight, S. A., Vailaire, G., Lesuisse, E., and Dancis, A. (2005). Iron acquisition from transferrin by *Candida albicans* depends on the reductive pathway. *Infect. Immun.* 73, 5482–5492.
- Lane, S., Birse, C., Zhou, S., Matson, R., and Liu, H. (2001). DNA array studies demonstrate convergent regulation of virulence factors by Cph1, Cph2, and Efg1 in *Candida albicans*. *J. Biol. Chem.* 276, 48988–48996.
- Le Novère, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., Demir, E., Wegner, K., Aladjem, M. I., Wimalaratne, S. M., Bergman, F. T., Gauges, R., Ghazal, P., Kawaji, H., Li, L., Matsuoka, Y., Villéger, A., Boyd, S. E., Calzone, L., Courtot, M., Dogrusoz, U., Freeman, T. C., Funahashi, A., Ghosh, S., Jouraku, A., Kim, S., Kolpakov, F., Luna, A., Sahle, S., Schmidt, E., Watterson, S., Wu, G., Goryanin, I., Kell, D. B., Sander, C., Sauro, H., Snoep, J. L., Kohn, K., and Kitano, H. (2009). The systems biology graphical notation. *Nat. Biotechnol.* 27, 735–741.
- Linde, J., Wilson, D., Hube, B., and Guthke, R. (2010). Regulatory network modelling of iron acquisition by a fungal pathogen in contact with epithelial cells. *BMC Syst. Biol.* 4, 148. doi:10.1186/1752-0509-4-148
- Lu, X., Kovalev, G. I., Chang, H., Kallin, E., Knudsen, G., Xia, L., Mishra, N., Ruiz, P., Li, E., Su, L., and Zhang, Y. (2008). Inactivation of NuRD component Mta2 causes abnormal T cell activation and lupus-like autoimmune disease in mice. *J. Biol. Chem.* 283, 13825–13833.
- Ma, Y. J., Doni, A., Skjoedt, M. O., Honore, C., Arendrup, M., Mantovani, A., and Garred, P. (2011). Heterocomplexes of mannose-binding lectin and the pentraxins PTX3 or serum amyloid P component trigger cross-activation of the complement system. *J. Biol. Chem.* 286, 3405–3417.
- Manavathi, B., Singh, K., and Kumar, R. (2007). MTA family of coregulators in nuclear receptor biology and pathology. *Nucl. Recept. Signal* 5, e010.
- Marbach, D., Prill, R. J., Schaffter, T., Mattiussi, C., Floreano, D., and Stolovitzky, G. (2010). Revealing strengths and weaknesses of methods for gene network inference. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6286–6291.
- Martchenko, M., Alarco, A. M., Harcus, D., and Whiteway, M. (2004). Superoxide dismutases in *Candida albicans*: transcriptional regulation and functional characterization of the hyphal-induced SOD5 gene. *Mol. Biol. Cell* 15, 456–467.
- Martinou, A., Koutsoulis, D., and Bouriotis, V. (2002). Expression, purification, and characterization of a cobalt-activated chitin deacetylase (Cda2p) from *Saccharomyces cerevisiae*. *Protein Expr. Purif.* 24, 111–116.
- Moalli, F., Doni, A., Deban, L., Zelante, T., Zagarella, S., Bottazzi, B., Romani, L., Mantovani, A., and Garlanda, C. (2010). Role of complement and Fcγ receptors in the protective activity of the long pentraxin PTX3 against *Aspergillus fumigatus*. *Blood* 116, 5170–5180.
- Moalli, F., Paroni, M., Veliz Rodriguez, T., Riva, F., Polentarutti, N., Bottazzi, B., Valentino, S., Mantero, S., Nebuloni, M., Mantovani, A., Bragonzi, A., and Garlanda, C. (2011). The therapeutic potential of the humoral pattern recognition molecule PTX3 in chronic lung infection caused by *Pseudomonas aeruginosa*. *J. Immunol.* 186, 5425–5434.
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628.
- Murad, A. M., Lee, P. R., Broadbent, I. D., Barelle, C. J., and Brown, A. J. (2000). Clp10, an efficient and convenient integrating vector for *Candida albicans*. *Yeast* 16, 325–327.
- Netea, M. G., Brown, G. D., Kullberg, B. J., and Gow, N. A. (2008). An integrated model of the recognition

- of *Candida albicans* by the innate immune system. *Nat. Rev. Microbiol.* 6, 67–78.
- Nobile, C. J., Solis, N., Myers, C. L., Fay, A. J., Deneault, J. S., Nantel, A., Mitchell, A. P., and Filler, S. G. (2008). *Candida albicans* transcription factor Rim101 mediates pathogenic interactions through cell wall functions. *Cell. Microbiol.* 10, 2180–2196.
- Oberholzer, U., Nantel, A., Berman, J., and Whiteway, M. (2006). Transcript profiles of *Candida albicans* cortical actin patch mutants reflect their cellular defects: contribution of the Hog1p and Mkc1p signaling pathways. *Eukaryot. Cell* 5, 1252–1265.
- Park, Y. N., and Morschhäuser, J. (2005). *Candida albicans* MTL α tup1 Δ mutants can reversibly switch to mating-competent, filamentous growth forms. *Mol. Microbiol.* 58, 1288–1302.
- Priebe, S., Linde, J., Albrecht, D., Guthke, R., and Brakhage, A. A. (2011). FungiFun: a web-based application for functional categorization of fungal genes and proteins. *Fungal Genet. Biol.* 48, 353–358.
- Raman, T., O'Connor, T. P., Hackett, N. R., Wang, W., Harvey, B. G., Attiyeh, M. A., Dang, D. T., Teater, M., and Crystal, R. G. (2009). Quality control in microarray assessment of gene expression in human airway epithelium. *BMC Genomics* 10, 493. doi:10.1186/1471-2164-10-493
- Rizzetto, L., and Cavalieri, D. (2011). Friend or foe: using systems biology to elucidate interactions between fungi and their hosts. *Trends Microbiol.* 19, 509–515.
- Rohde, J. R. (2011). Microbiology. *Listeria* unwinds host DNA. *Science* 331, 1271–1272.
- Ruepp, A., Zollner, A., Maier, D., Albermann, K., Hani, J., Mokrejs, M., Tetko, I., Guldener, U., Mannhaupt, G., Munsterkotter, M., and Mewes, H. W. (2004). The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Res.* 32, 5539–5545.
- Sellam, A., Al-Niemi, T., McInerney, K., Brumfield, S., Nantel, A., and Suci, P. A. (2009). A *Candida albicans* early stage biofilm detachment event in rich medium. *BMC Microbiol.* 9, 25. doi:10.1186/1471-2180-9-25
- Skrzypek, M. S., Arnaud, M. B., Costanzo, M. C., Inglis, D. O., Shah, P., Binkley, G., Miyasato, S. R., and Sherlock, G. (2010). New tools at the *Candida* genome database: biochemical pathways and full-text literature search. *Nucleic Acids Res.* 38, D428–D432.
- Smith, J. J., Deane, N. G., Wu, F., Merchant, N. B., Zhang, B., Jiang, A., Lu, P., Johnson, J. C., Schmidt, C., Bailey, C. E., Eschrich, S., Kis, C., Levy, S., Washington, M. K., Heslin, M. J., Coffey, R. J., Yeatman, T. J., Shyr, Y., and Beauchamp, R. D. (2010). Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. *Gastroenterology* 138, 958–968.
- Spira, A., Beane, J. E., Shah, V., Steiling, K., Liu, G., Schembri, E., Gilman, S., Dumas, Y. M., Calner, P., Sebastiani, P., Sridhar, S., Beamis, J., Lamb, C., Anderson, T., Gerry, N., Keane, J., Lenburg, M. E., and Brody, J. S. (2007). Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat. Med.* 13, 361–366.
- Stark, C., Breitkreutz, B. J., Chatri-Aryamontri, A., Boucher, L., Oughtred, R., Livstone, M. S., Nixon, J., Van Aukun, K., Wang, X., Shi, X., Regul, T., Rust, J. M., Winter, A., Dolinski, K., and Tyers, M. (2011). The BioGRID interaction database: 2011 update. *Nucleic Acids Res.* 39, D698–D704.
- Thewes, S., Kretschmar, M., Park, H., Schaller, M., Filler, S. G., and Hube, B. (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.
- Toepfer, S., Guthke, R., Driesch, D., Woetzel, D., and Pfaff, M. (2007). The NetGenerator algorithm: reconstruction of gene regulatory networks. *Lect. Notes Bioinform.* 4366, 119–130.
- Tokuyasu, K., Ohnishi-Kameyama, M., and Hayashi, K. (1996). Purification and characterization of extracellular chitin deacetylase from *Colletotrichum lindemuthianum*. *Biosci. Biotechnol. Biochem.* 60, 1598–1603.
- Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111.
- Wachtler, B., Wilson, D., Haedicke, K., Dalle, F., and 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. doi:10.1371/journal.pone.0017046
- Wang, Q., Diskin, S., Rappaport, E., Attiyeh, E., Mosse, Y., Shue, D., Seiser, E., Jagannathan, J., Shusterman, S., Bansal, M., Khazi, D., Winter, C., Okawa, E., Grant, G., Cnaan, A., Zhao, H., Cheung, N. K., Gerald, W., London, W., Matthay, K. K., Brodeur, G. M., and Maris, J. M. (2006). Integrative genomics identifies distinct molecular classes of neuroblastoma and shows that multiple genes are targeted by regional alterations in DNA copy number. *Cancer Res.* 66, 6050–6062.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63.
- Werhli, A. V., and Husmeier, D. (2007). Reconstructing gene regulatory networks with Bayesian networks by combining expression data with multiple sources of prior knowledge. *Stat. Appl. Genet. Mol. Biol.* 6, 15.
- Zakikhany, K., Naglik, J. R., Schmidt-Westhausen, A., Holland, G., Schaller, M., and Hube, B. (2007). *In vivo* transcript profiling of *Candida albicans* identifies a gene essential for interepithelial dissemination. *Cell. Microbiol.* 9, 2938–2954.
- Zhang, J., Chiodini, R., Badr, A., and Zhang, G. (2011). The impact of next-generation sequencing on genomics. *J. Genet. Genomics* 38, 95–109.
- Zhao, Y., Oh, S. H., Cheng, G., Green, C. B., Nuessen, J. A., Yeater, K., Leng, R. P., Brown, A. J., and Hoyer, L. L. (2004). ALS3 and ALS8 represent a single locus that encodes a *Candida albicans* adhesin; functional comparisons between Als3p and Als1p. *Microbiology* 150, 2415–2428.
- Zhao, Y., Park, R. D., and Muzzarelli, R. A. (2010). Chitin deacetylases: properties and applications. *Mar. Drugs* 8, 24–46.

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

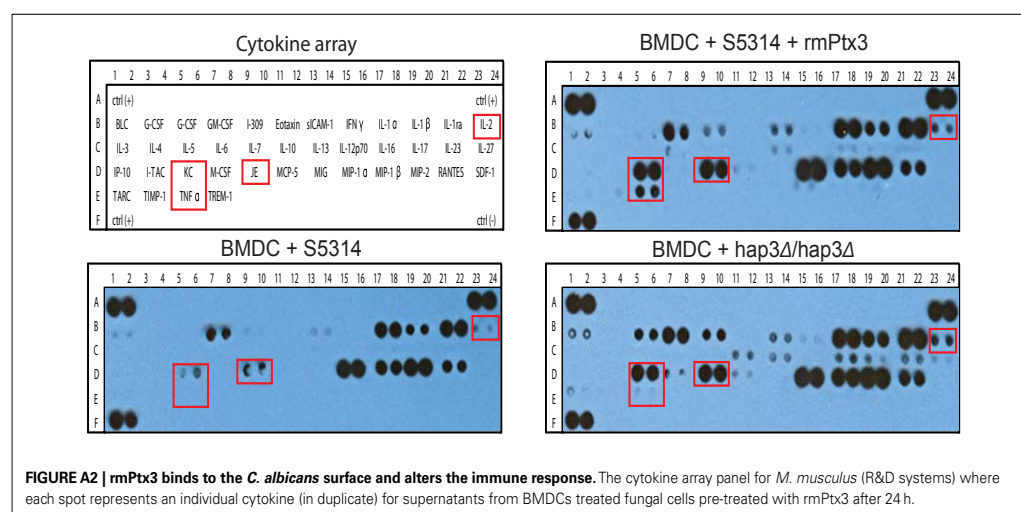
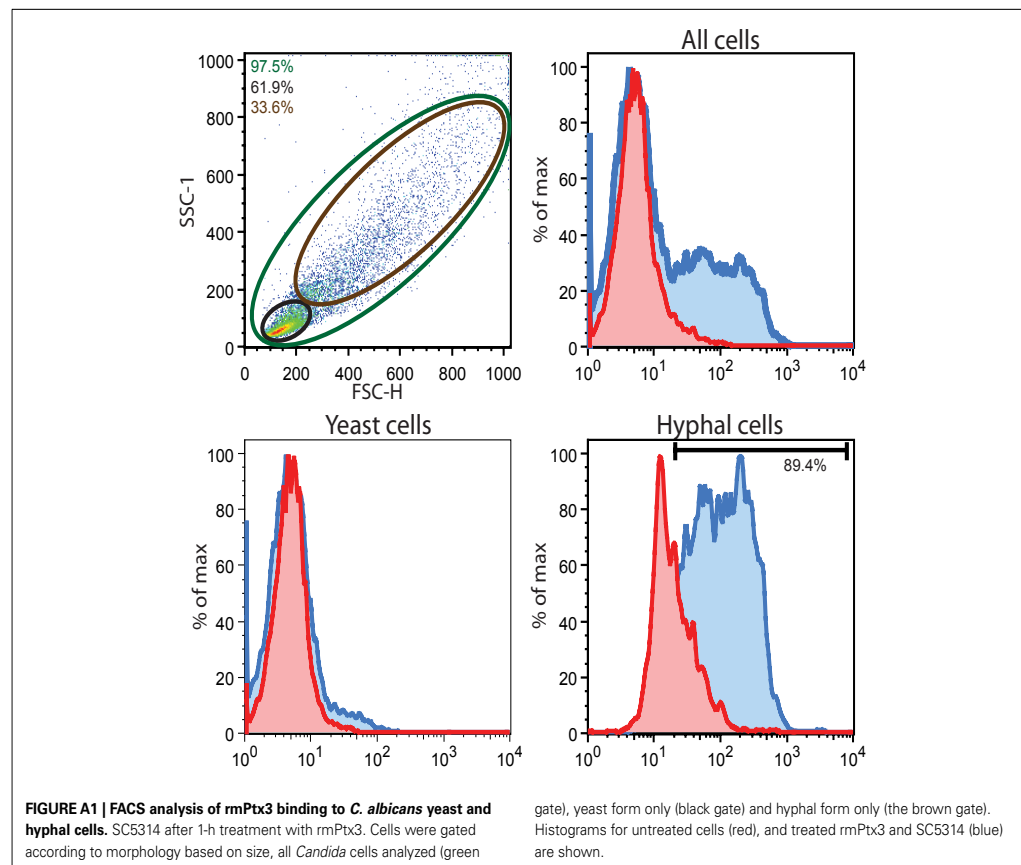
Received: 08 December 2011; accepted: 20 February 2012; published online: 12 March 2012.

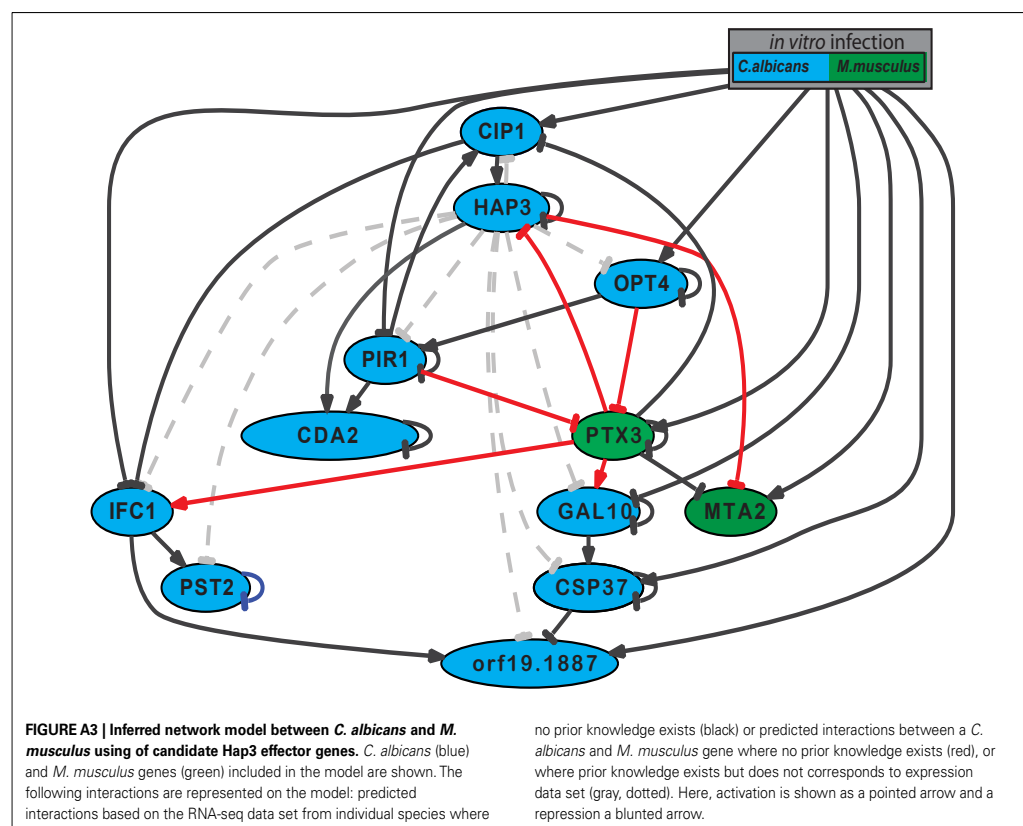
Citation: Tierney L, Linde J, Müller S, Brunke S, Molina JC, Hube B, Schöck U, Guthke R and Kuchler K (2012) An interspecies regulatory network inferred from simultaneous RNA-seq of *Candida albicans* invading innate immune cells. *Front. Microbio.* 3:85. doi: 10.3389/fmicb.2012.00085

This article was submitted to *Frontiers in Microbial Immunology*, a specialty of *Frontiers in Microbiology*.

Copyright © 2012 Tierney, Linde, Müller, Brunke, Molina, Hube, Schöck, Guthke and Kuchler. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.

APPENDIX





Systems Biology of Host-Fungus Interactions: Turning Complexity Into Simplicity

Lanay Tierney¹, Karl Kuchler¹, Lisa Rizzetto², Duccio Cavalieri^{2,3}

This review addresses recent advances in understanding the complexity of host-pathogen interactions using systems biology approaches among the major opportunistic and pathogenic fungi. The manuscript addresses the mounting complexity in this field from three different perspectives including that of the host, the pathogen, and the host-pathogen interface. We describe recent work on the evaluation and identification of the transition from the commensal to pathogenic state in fungi using bioinformatics and functional genomics tools. Additionally, the review focuses on novel methods that aid in the integration of genome-wide-datasets from both fungal and host immune responses including inference modeling (similar to what is described in Paper 1 of this dissertation), web based resources, and game theory mathematics. Moreover, we propose strategies for parsing genome-wide-datasets into testable hypotheses in order to more easily validate current systems biology based methods.

For this review, I conceived the outline for the paper, prepared the figures, and wrote the majority of the manuscript.



ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in
Microbiology

Systems biology of host–fungus interactions: turning complexity into simplicity

Lanay Tierney¹ and Karl Kuchler¹
Lisa Rizzetto² and Duccio Cavalieri^{2,3}

Modeling interactions between fungi and their hosts at the systems level requires a molecular understanding both of how the host orchestrates immune surveillance and tolerance, and how this activation, in turn, affects fungal adaptation and survival. The transition from the commensal to pathogenic state, and the co-evolution of fungal strains within their hosts, necessitates the molecular dissection of fungal traits responsible for these interactions. There has been a dramatic increase in publically available genome-wide resources addressing fungal pathophysiology and host–fungal immunology. The integration of these existing data and emerging large-scale technologies addressing host–pathogen interactions requires novel tools to connect genome-wide data sets and theoretical approaches with experimental validation so as to identify inherent and emerging properties of host–pathogen relationships and to obtain a holistic view of infectious processes. If successful, a better understanding of the immune response in health and microbial diseases will eventually emerge and pave the way for improved therapies

Addresses

¹ Medical University of Vienna, Christian Doppler Laboratory Infection Biology, Max F. Perutz Laboratories, A-1030 Vienna, Austria

² Department of Preclinical and Clinical Pharmacology, University of Florence, 50139 Firenze, Italy

³ Research and Innovation Centre, Edmund Mach Foundation, San Michele all'Adige, 38010, Trento, Italy

Corresponding authors: Kuchler, Karl (karl.kuchler@meduniwien.ac.at)

Current Opinion in Microbiology 2012, 15:1–7

This review comes from a themed issue on
Host-microbe interactions: Fungi
Edited by Mihai G Netea and Gordon D Brown

1369-5274/\$ – see front matter

© 2012 Elsevier Ltd. All rights reserved.

<http://dx.doi.org/10.1016/j.mib.2012.05.001>

Introduction

In ecology and immunology, tolerance usually refers to host mitigation of the fitness costs of an infection [1]. This is distinct from resistance, whereby the host reduces the microorganism burden. These costs may tip the balance of an immune response towards tolerance of environmental microorganisms, including fungi. Modern pressures on the immune system and the natural composition human microbiome have partially resulted from the expansion of

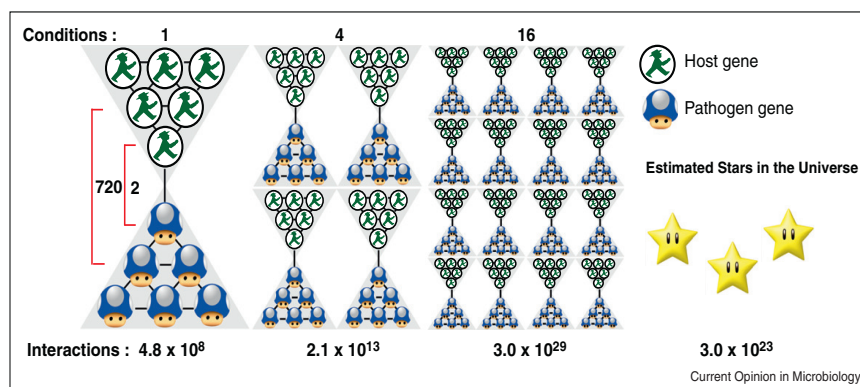
fungi in fermented foods, including opportunistic pathogens colonizing humans. This is particularly important for intestinal tissues, where mucosal immunity faces life-long challenges by beneficial and detrimental microbes [2,3]. These microbes, including pathogenic fungi, possess a molecular arsenal to escape diverse defense mechanisms of immunocompetent hosts. It is thought that the co-evolution of opportunistic pathogens with their healthy host may aid in their ability to exploit host defenses and remain tolerated [4].

The history of host and fungal interactions will strongly influence resistance against and tolerance to microorganisms. Cross-talk mechanisms during host–pathogen interactions will impact the outcome of infections and further influence subsequent pathogen exposure. As a result, genome-wide studies have gained in popularity to investigate global response patterns to infections from both the host and pathogen side. However, biological interpretations of genome-wide studies are limited to only a fraction of the theoretically possible interactions between genes, environmental conditions, and life cycles taking part in a host–pathogen setting (Figure 1). The enormous complexity underlying the host–pathogen interplay when considering the theoretically possible genetic interactions of even a few genes, necessitates the simplification of systems to cellular or pathway levels. A systems biology approach at different levels — genomic, proteomic, and metabolomic — is an emerging strategy to better understand the pathophysiology of infectious processes and their underlying mechanisms during host–pathogen interactions [5,6].

Systems biology is a rapidly evolving integrative approach that connects many disciplines and aims to create a quantitative and predictive understanding of biological processes. Systems biology has evolved by two parallel approaches: ‘top-down’ network inference, reconstruction, and modeling based on functional genomics data, and the ‘bottom-up’ approach of modeling well-defined circuits based on their functional conservation with other systems. Systems biology approaches follow iterative cycles of modeling and data generation, based on a given biological and testable hypothesis [7]. Recent seminal reviews highlight the power of these different approaches in the dissection of mammalian innate immunity [8[•],9–12], the reconstruction of immune signaling, transcriptional networks [13[•]], and host–pathogen interactions [14,5,6].

2 Host-microbe interactions: Fungi

Figure 1



The factorial complexity of host-pathogen interactions. Schematic representation of the number of possible interactions between limited numbers of host-pathogen genes, where interactions are represented as a line between objects. The numbers of potential interactions between genes show a factorial increase based on the number of environmental conditions applicable. The number of putative interactions is calculated by taking the factorial $n!$ of the total number of genes and environmental conditions in the system. This astronomical number for a limited number of genes emphasizes the necessity to simplify host-pathogen interaction studies.

This review will address recent work aimed at investigating the transition of opportunistic fungi, with a focus on *Candida* spp., from the commensal to the pathogenic state, emphasizing fungal mechanisms to escape host immune surveillance. We will discuss new approaches in functional genomics that facilitate modeling, and those which are aimed at understanding the fungal response in the host environment. Furthermore, we discuss the advantages of combining different approaches to gain a better understanding of how the cross-talk between fungal pathogens and their hosts shapes the progress and outcome of invasive infections.

Host perspectives

Innate and adaptive immune responses are responsible for recognizing, responding, and adapting to opportunistic microbial pathogens, including fungi [15*,16]. These responses determine whether microbes require the activation of pathogen-specific defense or attack mechanisms [17,18]. Recognition of fungal pathogens by innate immune cells elicits immune responses by engaging multiple cell-bound, soluble, or intracellular receptors, in a stage-specific and cell-specific manner [19**]. To date, hundreds of proteins and genes have been implicated in the innate immune response [20]. The transcriptional response to a microbial stimulus is further tailored to both the stimulus and the responsive immune cell [21]. The analysis of the transcriptome of human dendritic cells (DCs) to *Aspergillus fumigatus*, *C. albicans*, and *S. cerevisiae* showed how the expression of immune-relevant genes increases depending on the morphology, life-stage, and incubation period with the fungus [22*,23,24]. Models of downstream signal transduction networks using gene expression data have been generated based

on similarities in expression profiles in related species, the prediction of shared regulatory motifs, and their integration at the pathway level [25]. Recently, a combination of a forward-genomics and reverse-genomics approach enabled the reconstruction of transcriptional and regulatory networks driving the immune response in DCs to a viral infection [26**]. The resultant network model investigated how pathogen-sensing pathways achieve specificity and the influence of a single regulator on mediating inflammatory genes and viral responses depending on the timing of the regulator activation. A regulatory network of potential interactions between microRNAs and mRNAs is an additional level of complexity of how pathogens could manipulate host cell responses [27].

The extent to which early transcriptional regulatory events determine the decision-making process in immune cells responding to different pathogenic fungi is still an open question. However, an increasing number of databases are collecting and annotating functional information. For example, the InnateDB, curates the innate immunity interactome [28], and ImmGen collects immunological microarray data (www.immgen.org). The further development of cell-specific bioinformatic tools to analyze the response in macrophages [29] or DCs [30*] will allow for the classification of stimuli by their species-specific transcriptional programs governing fungal recognition (Rizzetto *et al.*, unpublished observation).

While the analysis of gene expression is commonly used to study the activation of immune cells, proteomics constitute a complementary approach providing a direct view on protein levels as well as their activities. Proteomics

however poses additional challenges, including cost and the technical limitations to make the process quantitative [31^{*}]. Moreover, mRNA expression levels are not necessarily correlated with protein production, hampering the comparative analysis of these data sets. A recent study combined a comprehensive quantitative proteome and transcriptome analysis on immature and cytokine-matured human DCs [32]. Although the overall correlation between differential mRNA and protein expression was low, the correlation between components of DC relevant pathways was significantly higher, underscoring that the integration of related data sets at the pathway level can significantly increase the predictive power of multiple -omics analyses. Recently, a global investigation of the macrophage phosphoproteome and its dynamic changes upon TLR activation has been identified [33]. Functional bioinformatic analyses confirmed already known players of the TLR-mediated signaling and identified new transcriptional regulators previously not implicated in TLR-induced gene expression.

Pathogen perspectives

Fungal adaptation to host immune surveillance

Fungal pathogens have developed sophisticated means to evade or persist in the host, despite normal immune surveillance [34]. The use of genome-wide technologies to study global transcriptional changes has revealed the complexity of fungal adaptation to various host niches. Recent studies provide insights into the mechanisms of adaptation during infection, which include: the expression of anti-phagocytic functions and specific nutrient acquisition systems, the remodeling of central carbon metabolism, and the hypoxia response [35,36]. Virulence factor expression is, to a large extent, embedded in the regulation of functions needed for growth in the mammalian hosts. Pioneering early work on the differential gene expression of fungi phagocytosed by immune cells including macrophages, neutrophils, and granulocytes, revealed, among others, a dynamic response to nutrient starvation, oxidative stress, and iron limitation. Attempts by fungi and especially *Candida* spp. to adapt to the damaging effects of the environment via the activation of genes encoding antioxidant and detoxifying enzymes, and iron uptake proteins were shown [37]. A physiological role for cell surface superoxide dismutases in detoxifying reactive oxygen species (ROS) in innate immune cells and facilitating immune evasion was found [38]. In addition, autophagy and pexophagy mechanisms are important virulence traits of fungi to enable persistence and survival [39,40]. Notably, a global model of iron homeostasis in *A. fumigatus* has integrated data from Northern blot analysis, microarray expression, transcription factor knock-out mutants, and the occurrence of transcription factor binding motifs in regulatory regions of the genes to predict new transcription factor to target interactions [41].

Fungi may also evade the immune system by changing virulence gene expression at different infection stages upon encountering host-conditions. For example, a novel flow cytometry-based technique showed how changes in fungal gene expression profiles occurring over time influenced patient outcomes with clinical strains of *Cryptococcus neoformans* [42,43]. Using an *in vitro* oral candidiasis model, *C. albicans* mutants defecting in regulators of hyphal formation were attenuated in their ability to invade and damage epithelial cells [44]. The further use of microarray and RNA-seq technology in conjunction with *in vitro* infection models could be used to further investigate the role of stage-specific virulence gene expression.

Genome dynamics of fungal pathogens

Many fungal clinical isolates display a large degree of genetic and genomic heterogeneity. Segmental or whole-chromosome aneuploidy can be a source of selectable phenotypic variation in fungal species [45], conferring a selective advantage in a host setting [46]. For example, exposure to specific antifungal drugs increases the frequency of adaptive events, promoting drug resistance in independent lineages of *C. albicans* cells [47]. Additionally, loss of heterozygosity events is elevated in *C. albicans* in response to oxidative, heat, and antifungal drug stress *in vitro* [48^{*}]. Although rare, even *S. cerevisiae* may become an opportunistic pathogen under very specific conditions or genetic alterations [49]. Hence, cell population dynamics and evolutionary forces imposed by host stress and other factors may represent the driving force of genomic plasticity in fungal pathogens that enable colonization of various host niches. Strain variability and surface alterations could also explain differences in the host immune response [50], providing new opportunities to model host immune system interactions. Pathogenicity itself could reflect adaptive advantages conferred by the acquisition of virulence traits in different strains, thereby increasing pathogen fitness.

Contrary to *S. cerevisiae*, *C. albicans* lacks a complete sexual cycle, impeding efficient genetic analyses and limiting systems biology approaches with this obligatory diploid fungus. Under certain environmental conditions, *C. albicans* can switch from to a mating-competent state [51]. This transition modulates metabolic preferences, antifungal drug resistance, niche distribution, and host immune cell-specific interactions among many others, and is therefore an important consideration in the investigation of fungal fitness within host niches. Comparative genomics studies have the potential to identify new virulence-associated gene networks [5,52]. The number of sequenced fungal genomes publically available has significantly expanded in recent years [53]. In addition, the *Candida* Genome Database (CGD) and the *Aspergillus* database, among others now offer multiple species, facilitating these comparisons. The availability of genomic datasets studying specifically host–fungi inter-

4 Host-microbe interactions: Fungi

Table 1

Databases integrating host-pathogen data. All databases are publicly available.

Database	URL	Description	Pathogen
BiologicalNetworks	http://flu.sdsc.edu/index.jsp	Interactive networks, proteomics, transcriptomics, and metabolomics	Viruses
GPS-Prot	http://www.gpsprot.org	HIV-human protein-protein interactions	Viruses
HPIDB	http://agbase.msstate.edu/hpi	Protein interaction database	Bacteria, viruses, fungi
PATRIC	http://www.patricbrc.org	Bacterial and human proteins	Bacteria
PHI-Base	http://www.phibase.org	Data-mining and gene expression data	Fungi, oomycetes, bacteria
PHIDIAS	http://www.phidias.us	Data-mining, genomic sequences and gene expression data	Bacteria, viruses, parasites, fungi
PhylomeDB	http://phylomedb.org	Gene phylogenies for phylogenetic and comparative analysis	Bacteria, fungi
PIG	http://molvis.vbi.vt.edu/pig	Protein interaction database	Bacteria, viruses, fungi
vHoT	http://dna.korea.ac.kr/vhot	Prediction of mRNA targets of viral microRNAs	Viruses

actions have also expanded (Table 1), along with the number of software platforms available for the analysis and integration of genome-wide data sets [54]. Exploring commonalities and differences among fungi could be used to further understand the genetic basis for pathogenic phenotypes.

Infection modeling and microbial arcades

Spatio-temporal modeling of infection dynamics is an emerging field to incorporate the dynamics of pathogenesis [55,6]. One approach is evolutionary game theory, an application of game theory mathematics based on the relationship between the behavior of an organism and its evolution, or co-evolution, with other species. These studies formulate a simplified infection *in silico* and predict pathogen fitness by identifying game rules, often from genome-wide expression data. Most recently, it has been used to describe infections including: mixed viral infections of *Arabidopsis thaliana* [56], persistent bacterial infections [57], a simulated multi-species biofilm [58], and the mechanisms enabling survival of *C. albicans* inside macrophages [59•]. For *C. albicans*, the outcome was analyzed based on the mean evolutionary cost of a cell population to obtain a positive fitness and the infection strategy employed by *C. albicans* to enable proliferation in the host was hypothesized to be responsive to this cost. These studies emphasize the importance of analyzing microbes as adaptive social components of biological systems, because of their ability to sense and respond to the requirements of their own population, and that of their environment [60•].

Computational modeling has been used to reconstruct the complex network between the immune cells and the bacterial pathogen *Mycobacterium tuberculosis*. On the basis of known interactions of the bacteria during infection, they estimated the influence of specific factors, such as an increase in specific cytokines or vaccination, on bacterial clearance and thereby identified the overall propensity for

the bacteria to persist in the host under a wide range of conditions [61].

Most modeling approaches use genome-wide microarray expression data or RNA-seq. RNA-seq provides the advantage of simultaneous expression profiling of genes of the pathogens and their hosts, reducing concerns about platform-dependent effects. In addition, RNA-seq can potentially be used to investigate allelic variants of a transcript, and the evolution of microorganisms within its host. Small-scale network inference from the simultaneous analysis of *C. albicans* and DCs from *M. musculus* has predicted novel host-pathogen genetic interactions [62•]. Furthermore, a genome-wide inference network using *C. albicans* has identified a number of candidate antifungal target genes [63]. These studies emphasize the advantages of simplifying genome-wide expression data using modeling and inference techniques to identify novel interactions and strategies utilized by the host and pathogen during infection.

Significant hurdles remain in order to use infection modeling on a large scale. One major limitation is that the experimental data is generated at different time scales. The transcriptional response of fungi takes place after minutes, proteomics from minutes to hours, and the subsequent immune response to the fungus from hours to days or even weeks. Choosing a mathematical approach to relate these time scales is not trivial. Moreover, the use of different units, strains, and animal models between laboratories can limit the ability to compare data sets. There been a push to standardize genome-wide data sets, including the Minimum Information for Biological and Biomedical Investigations (MIBBI, <http://mibbi.sourceforge.net/>), which will significantly aid in dataset comparison between laboratories. Relatedly, the maintenance and integration of new and existing fungal databases is needed to make the available information accessible and decrease the bottleneck for data analysis. Curation based on data models that incorporate pathway information [64] will make it easier to

integrate new types of data sets, such as metabolomics, proteomics or host–pathogen data sets, as they become available.

Conclusions and outlook

A frequent critique of systems biology is that the massive influx of data has led to a fundamental loss of perspective because data generation has outpaced our capacity and ability to analyze them. It is therefore easy to lose the scale in which -omics data is biologically meaningful. Taking a lesson from Schrödinger's philosophy, the understanding of inner workings of the eye does not bring one closer to the perception of color: the additional information is irrelevant to the question. In other words, the biological context and proper parameter estimation of biological data sets is the key to generate models of predictive power. An initial definition of the system and its potential impact on the interacting species it contains is therefore required for analysis, including responses that determine pathogen clearance or host killing. Understanding the evolution of fungal strategies to survive and infect the host requires simultaneous investigation of microorganism–host interactions in both pathogenic and commensal species. Lessons learned from modeling the cell cycle show the importance of obtaining time course information either at the whole genome, or at the single molecule level, including the identification of biologically meaningful parameters, to obtain identifiable models. Developing strategies for the integration of multiple and complementary — quantitative -omics data sets, in a dynamic manner, will also be essential to further our understanding of microbial infections by reducing available data sets into testable models.

The host immune response is a complex entity and its behavior cannot be investigated in isolation from the environment that is driving adaptive changes, such as host immune defense. Systems biology holds the promise of helping us to obtain holistic views on the extent of this environment, and to generate predictions of host–microbe behavior and disease outcome. Combining the major schools of thought of mathematical modeling and functional genomics is a promising solution to reach the goals of deciphering infectious processes and eventually improving therapeutic approaches to fungal infections.

Conflict of interest

The authors have declared that no conflicts of interest.

Acknowledgements

We would like to thank Sara Tierney for contributing to the artwork. We apologize to all colleagues whose work we could not cite because of constraints regarding length and publication year.

Grant support: We would like to thank EU Framework Programme 7 Collaborative Project SYBARIS, Grant Agreement Number 242220 for supporting our work in this field. This work was supported by a grant from the Christian Doppler Research Society to KK, by the SysMO MOSES project to KK, and in part by the FWF-DACH grant of the Austrian Science Foundation (FWF-Project I-746-B11) to KK.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Sears BF, Rohr JR, Allen JE, Martin LB: **The economy of inflammation: when is less more?** *Trends Parasitol* 2011, **27**:382-387.
2. DuPont AW, DuPont HL: **The intestinal microbiota and chronic disorders of the gut.** *Nat Rev Gastroenterol Hepatol* 2011, **8**:523-531.
3. Mason KL, Erb Downward JR, Falkowski NR, Young VB, Kao JY, Huffnagle GB: **Interplay between the gastric bacterial microbiota and *Candida albicans* during postantibiotic recolonization and gastritis.** *Infect Immun* 2012, **80**:150-158.
4. Collette JR, Lorenz MC: **Mechanisms of immune evasion in fungal pathogens.** *Curr Opin Microbiol* 2011, **14**:668-675.
5. Sturdevant DE, Virtaneva K, Martens C, Bozinov D, Ogundare O, Castro N, Kanakabandi K, Beare PA, Omsland A, Carlson JH *et al.*: **Host-microbe interaction systems biology: lifecycle transcriptomics and comparative genomics.** *Future Microbiol* 2010, **5**:205-219.
6. Cottier F, Pavelka N: **Complexity and dynamics of host-fungal interactions.** *Immunol Res* 2012. Epub Mar 14.
7. Westerhoff HV: **Systems biology left and right.** *Methods Enzymol* 2011, **500**:3-11.
8. Chuang HY, Hofree M, Ideker T: **A decade of systems biology.** *Annu Rev Cell Dev Biol* 2010, **26**:721-744.
- A comprehensive review of systems biology approaches with a focus on pathway analysis, predictive maps, as well as integrative software to investigate human disease.
9. Shapira SD, Hacohen N: **Systems biology approaches to dissect mammalian innate immunity.** *Curr Opin Immunol* 2011, **23**:71-77.
10. Schubert C: **Systems immunology: complexity captured.** *Nature* 2011, **473**:113-114.
11. Santamaria R, Rizzetto L, Bromley M, Zelante T, Lee W, Cavalieri D, Romani L, Miller B, Gut I, Santos M *et al.*: **Systems biology of infectious diseases: a focus on fungal infections.** *Immunobiology* 2011, **216**:1212-1227.
12. Germain RN, Meier-Schellersheim M, Nita-Lazar A, Fraser ID: **Systems biology in immunology: a computational modeling perspective.** *Annu Rev Immunol* 2011, **29**:527-585.
13. Amit I, Regev A, Hacohen N: **Strategies to discover regulatory circuits of the mammalian immune system.** *Nat Rev Immunol* 2011, **11**:873-880.
- A clear and concise review on recent tools developed for network reconstruction with a strong focus on the mammalian innate immune system.
14. Yan Q: **Immunoinformatics and systems biology methods for personalized medicine.** *Methods Mol Biol* 2010, **662**:203-220.
15. Gow NA, van de Veerdonk FL, Brown AJ, Netea MG: ***Candida albicans* morphogenesis and host defence: discriminating invasion from colonization.** *Nat Rev Microbiol* 2012, **10**:112-122.
- A comprehensive summary on the mechanisms driving host–pathogen interactions that emphasizes the importance of signaling at the interface of host invasion.
16. Bourgeois C, Majer O, Frohner IE, Tierney L, Kuchler K: **Fungal attacks on mammalian hosts: pathogen elimination requires sensing and tasting.** *Curr Opin Microbiol* 2010, **13**:401-408.
17. Zak DE, Aderem A: **Systems biology of innate immunity.** *Immunol Rev* 2009, **227**:264-282.
18. Gardy JL, Lynn DJ, Brinkman FS, Hancock RE: **Enabling a systems biology approach to immunology: focus on innate immunity.** *Trends Immunol* 2009, **30**:249-262.

6 Host-microbe interactions: Fungi

19. Romani L: **Immunity to fungal infections.** *Nat Rev Immunol* 2011, **11**:275-288.
An exhaustive overview on the immune responses activated upon fungal infections with a particular attention on the integration of appropriate signaling to better cope with the invading pathogen.
20. Li S, Wang L, Berman M, Kong YY, Dorf ME: **Mapping a dynamic innate immunity protein interaction network regulating type I interferon production.** *Immunity* 2011, **35**:426-440.
21. Smale ST: **Transcriptional regulation in the innate immune system.** *Curr Opin Immunol* 2012, **24**:51-57.
22. Loeffler J, Haddad Z, Bonin M, Romeike N, Mezger M, Schumacher U, Kapp M, Gebhardt F, Grigoleit GU, Stevanovic S *et al.*: **Interaction analyses of human monocytes co-cultured with different forms of *Aspergillus fumigatus*.** *J Med Microbiol* 2009, **58**:49-58.
This study uses microarrays for both *S. cerevisiae* and dendritic cells to investigate the transcriptional response of both species during *in vitro* co-culture.
23. Rizzetto L, Cavalieri D: **A systems biology approach to the mutual interaction between yeast and the immune system.** *Immunobiology* 2010, **215**:762-769.
24. Rizzetto L, Kuka M, De Filippo C, Cambi A, Netea MG, Beltrame L, Napolitani G, Torcia MG, D'Oro U, Cavalieri D: **Differential IL-17 production and mannan recognition contribute to fungal pathogenicity and commensalism.** *J Immunol* 2010, **184**:4258-4268.
25. Ramsey SA, Klemm SL, Zak DE, Kennedy KA, Thorsson V, Li B, Gilchrist M, Gold ES, Johnson CD, Litvak V *et al.*: **Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics.** *PLoS Comput Biol* 2008, **4**:e1000021.
26. Amit I, Garber M, Chevrier N, Leite AP, Donner Y, Eisenhaure T, Guttman M, Grenier JK, Li W, Zuk O *et al.*: **Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses.** *Science* 2009, **326**:257-263.
This study is of exceptional interest to the community, as it is presenting a strategy to perturb transcriptional regulators in response to pathogen challenge.
27. Sharbati S, Sharbati J, Hoeke L, Bohmer M, Einspanier R: **Quantification and accurate normalisation of small RNAs through new custom RT-qPCR arrays demonstrates *Salmonella*-induced microRNAs in human monocytes.** *BMC Genom* 2012, **13**:23.
28. Lynn DJ, Chan C, Naseer M, Yau M, Lo R, Sribnaia A, Ring G, Que J, Wee K, Winsor GL *et al.*: **Curating the innate immunity interactome.** *BMC Syst Biol* 2010, **4**:117.
29. Raza S, McDerment N, Lacaze PA, Robertson K, Watterson S, Chen Y, Chisholm M, Eleftheriadis G, Monk S, O'Sullivan M *et al.*: **Construction of a large scale integrated map of macrophage pathogen recognition and effector systems.** *BMC Syst Biol* 2010, **4**:63.
30. Cavalieri D, Rivero D, Beltrame L, Buschow SI, Calura E, Rizzetto L, Gessani S, Gauzzi MC, Reith W, Baur A *et al.*: **DC-ATLAS: a systems biology resource to dissect receptor specific signal transduction in dendritic cells.** *Immunome Res* 2010, **6**:10.
This work describes a new interactive tool available for dendritic cell signaling networks to infer information starting from high-throughput data.
31. Schmidt F, Volker U: **Proteome analysis of host-pathogen interactions: investigation of pathogen responses to the host cell environment.** *Proteomics* 2011, **11**:3203-3211.
This review addresses new proteomics tools, with a strong focus on cell culture or *in vivo* infection settings.
32. Buschow SI, Lasonder E, van Deutekom HW, Oud MM, Beltrame L, Huynen MA, de Vries IJ, Figdor CG, Cavalieri D: **Dominant processes during human dendritic cell maturation revealed by integration of proteome and transcriptome at the pathway level.** *J Proteome Res* 2010, **9**:1727-1737.
33. Weintz G, Olsen JV, Fruhauf K, Niedzielska M, Amit I, Jantsch J, Mages J, Frech C, Dolken L, Mann M *et al.*: **The phosphoproteome of toll-like receptor-activated macrophages.** *Mol Syst Biol* 2010, **6**:371.
34. Nish S, Medzhitov R: **Host defense pathways: role of redundancy and compensation in infectious disease phenotypes.** *Immunity* 2011, **34**:629-636.
35. Kronstad J, Saikia S, Nielson ED, Kretschmer M, Jung W, Hu G, Geddes JM, Griffiths EJ, Choi J, Cadieux B *et al.*: **Adaptation of *Cryptococcus neoformans* to mammalian hosts: integrated regulation of metabolism and virulence.** *Eukaryot Cell* 2011, **11**:109-118.
36. Morton CO, Varga JJ, Hornbach A, Mezger M, Sennfelder H, Kneitz S, Kurzai O, Krappmann S, Einsele H, Niernan WC *et al.*: **The temporal dynamics of differential gene expression in *Aspergillus fumigatus* interacting with human immature dendritic cells *in vitro*.** *PLoS ONE* 2011, **6**:e16016.
37. Almeida RS, Wilson D, Hube B: ***Candida albicans* iron acquisition within the host.** *FEMS Yeast Res* 2009, **9**:1000-1012.
38. Frohner IE, Bourgeois C, Yatsyk K, Majer O, Kuchler K: ***Candida albicans* cell surface superoxide dismutases degrade host-derived reactive oxygen species to escape innate immune surveillance.** *Mol Microbiol* 2009, **71**:240-252.
39. Hu G, Gibbons J, Williamson PR: **Analysis of autophagy during infections of *Cryptococcus neoformans*.** *Methods Enzymol* 2008, **451**:323-342.
40. Roetzer A, Gratz N, Kovarik P, Schuller C: **Autophagy supports *Candida glabrata* survival during phagocytosis.** *Cell Microbiol* 2010, **12**:199-216.
41. Linde J, Hortschansky P, Fazius E, Brakhage AA, Guthke R, Haas H: **Regulatory interactions for iron homeostasis in *Aspergillus fumigatus* inferred by a Systems Biology approach.** *BMC Syst Biol* 2012, **6**:6.
42. Alanio A, Desnos-Ollivier M, Dromer F: **Dynamics of *Cryptococcus neoformans*-macrophage interactions reveal that fungal background influences outcome during cryptococcal meningoencephalitis in humans.** *MBio* 2011, **2**:e00158.
43. Mansour MK, Vyas JM, Levitz SM: **Dynamic virulence: real-time assessment of intracellular pathogenesis links *Cryptococcus neoformans* phenotype with clinical outcome.** *MBio* 2011, **2**:e00217.
44. Martin R, Wachtler B, Schaller M, Wilson D, Hube B: **Host-pathogen interactions and virulence-associated genes during *Candida albicans* oral infections.** *Int J Med Microbiol* 2011, **301**:417-422.
45. Hu G, Wang J, Choi J, Jung WH, Liu I, Litvinseva AP, Bicanic T, Aurora R, Mitchell TG, Perfect JR *et al.*: **Variation in chromosome copy number influences the virulence of *Cryptococcus neoformans* and occurs in isolates from AIDS patients.** *BMC Genom* 2011, **12**:526.
46. Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, Sanderson BW, Hattam GL, Li R: **Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast.** *Nature* 2010, **468**:321-325.
47. Huang M, McClellan M, Berman J, Kao KC: **Evolutionary dynamics of *Candida albicans* during *in vitro* evolution.** *Eukaryot Cell* 2011, **10**:1413-1421.
48. Forche A, Abbey D, Pisithkul T, Weinzierl MA, Ringstrom T, Bruck D, Petersen K, Berman J: **Stress alters rates and types of loss of heterozygosity in *Candida albicans*.** *mBio* 2011, **2**:e00129.
This work generated a high-resolution map of the *C. albicans* transcriptome in response to numerous environmental conditions, identifying novel transcripts involved in stress responses.
49. de Llanos R, Llopis S, Molero G, Querol A, Gil C, Fernandez-Espinar MT: ***In vivo* virulence of commercial *Saccharomyces cerevisiae* strains with pathogenicity-associated phenotypic traits.** *Int J Food Microbiol* 2011, **144**:393-399.
50. Foligne B, Dewulf J, Vandekerckhove P, Pignede G, Pot B: **Probiotic yeasts: anti-inflammatory potential of various non-pathogenic strains in experimental colitis in mice.** *World J Gastroenterol* 2010, **16**:2134-2145.

51. Alby K, Bennett RJ: **Interspecies pheromone signaling promotes biofilm formation and same-sex mating in *Candida albicans*.** *Proc Natl Acad Sci U S A* 2011, **108**:2510–2515.
52. Bruno VM, Wang Z, Marjani SL, Euskirchen GM, Martin J, Sherlock G, Snyder M: **Comprehensive annotation of the transcriptome of the human fungal pathogen *Candida albicans* using RNA-seq.** *Genome Res* 2010, **20**:1451–1458.
53. Marcet-Houben M, Gabaldon T: **The tree versus the forest: the fungal tree of life and the topological diversity within the yeast phylome.** *PLoS ONE* 2009, **4**:e4357.
54. Ghosh S, Matsuoka Y, Asai Y, Hsin KY, Kitano H: **Software for systems biology: from tools to integrated platforms.** *Nat Rev Genet* 2011, **12**:821–832.
55. Horn F, Heinekamp T, Kniemeyer O, Pollmacher J, Valiante V, Brakhage AA: **Systems biology of fungal infection.** *Front Microbiol* 2012, **3**:108.
56. Martin S, Elena SF: **Application of game theory to the interaction between plant viruses during mixed infections.** *J Gen Virol* 2009, **90**:2815–2820.
57. Eswarappa SM: **Location of pathogenic bacteria during persistent infections: insights from an analysis using game theory.** *PLoS ONE* 2009, **4**:e5383.
58. Mitri S, Xavier JB, Foster KR: **Social evolution in multispecies biofilms.** *Proc Natl Acad Sci U S A* 2011, **108**:10839–10846.
59. Hummert S, Hummert C, Schroter A, Hube B, Schuster S: **Game •• theoretical modelling of survival strategies of *Candida albicans* inside macrophages.** *J Theor Biol* 2010, **264**:312–318.
This study is the first to use game theory modeling to investigate *C. albicans* infection strategies based on genome-wide transcriptional data sets.
60. Xavier JB: **Social interaction in synthetic and natural microbial •• communities.** *Mol Syst Biol* 2011, **7**:483.
An extensive and clear review on strategies and techniques available for integrating genome-wide data sets, including workflow analysis examples.
61. Raman K, Bhat AG, Chandra N: **A systems perspective of host–pathogen interactions: predicting disease outcome in tuberculosis.** *Mol Biosyst* 2010, **6**:516–530.
62. Tierney L, Linde J, Müller S, Brunke S, Molina JC, Hube B, Schöck U, •• Guthke R, Kuchler K: **An interspecies regulatory network inferred from simultaneous RNA-seq of *Candida albicans* invading innate immune cells.** *Front Microbiol* 2012, **3**:85.
This is the first study to both generate and experimentally verify a host–pathogen network inference map using simultaneous RNA-seq of both the host and pathogen during infection.
63. Altwasser R, Linde J, Buyko E, Hahn U, Guthke R: **Genome-wide scale-free network inference for *Candida albicans*.** *Front Microbiol* 2012, **3**:51.
64. Beltrame L, Calura E, Popovici RR, Rizzetto L, Guedez DR, Donato M, Romualdi C, Draghici S, Cavalieri D: **The Biological Connection Markup Language: a SBGN-compliant format for visualization, filtering and analysis of biological pathways.** *Bioinformatics* 2011, **27**:2127–2133.

3.3 Paper 3: (Tierney *et al.*, submitted)

Dendritic Cells Promote Stage-specific Alternative Splicing in *Candida albicans* Infections

Lanay Tierney^{1*}, Denes Hnisz^{1*}, Fritz Joachim Sedlazeck², Stefanie Tauber², Jörg Linde³, Reinhard Guthke³, Arndt von Haeseler² & Karl Kuchler¹

Due to heavy intron loss over evolutionary history, splicing and alternative splicing are rare events in both *S. cerevisiae* and *C. albicans*. Nevertheless, several recent studies have shown that the remaining introns can be spliced in response to conditions mimicking the environment present in the host cells including immune acid starvation and a temperature shift to 37°C, among others. Combined, these data suggest that splicing as a process may play a greater role on the adaptation of these species to the host environment than previous thought. In this study, we identify and verify a subset of genes in that splice specifically in response to murine dendritic cells, including novel alternative spliced transcripts in that undergo splicing during discrete infection stages. In addition, we developed a mathematical model to quantify splicing efficiency. Based on these data, we propose that splicing could be a general adaptive mechanism in microbial pathogenesis. We further propose that the interaction between host and microbe is a potential driving force of intron evolution.

For this study, I conceived and designed all wet lab experiments, performed, analyzed, and interpreted the data, and wrote the manuscript.

Dendritic cells promote stage-specific alternative splicing in *Candida albicans* infections

Lanay Tierney^{1*}, Denes Hnisz^{1*}, Fritz Joachim Sedlazeck², Stefanie Tauber²,
Jörg Linde³, Reinhard Guthke³, Arndt von Haeseler² & Karl Kuchler¹

¹Medical University of Vienna, Christian Doppler Laboratory Infection Biology, Max F Perutz Laboratories, Campus Vienna Biocenter, Dr.-Bohr-Gasse 3-9, A-1030 Vienna, Austria. ²Center for Integrative Bioinformatics Vienna, Max F Perutz Laboratories, University of Vienna, Medical University of Vienna, University of Veterinary Medicine, Dr.-Bohr-Gasse 3-9, A-1030 Vienna, Austria. ³Research Group Systems Biology & Bioinformatics, Leibniz-Institute for Natural Product Research and Infection Biology-Hans-Knoell-Institute, Jena, Germany.

* These authors contributed equally to the manuscript.

KEYWORDS: host-pathogen interactions, alternative splicing, RNA-seq, *C. albicans*

Abstract

Environmental stimuli regulate splicing in eukaryotes. However, the global regulation of splicing during pathogenesis is unknown. Here, we use simultaneous RNA-seq of both host and pathogen transcriptomes to define a time-resolved set of genes in the human fungal pathogen *Candida albicans* that undergo infection specific splicing and alternative splicing during discrete infection stages with murine dendritic cells. We identify multiple novel alternative splice junctions in *C. albicans*, including non-canonical splice junction sequences, suggesting a higher level of complexity in the *C. albicans* spliceosomal machinery. We developed a mathematical model to quantify

splicing events and find that splicing during infection is generally an all-or-nothing effect, supporting the idea that splicing is co-transcriptional. Our data identifies infection specific splicing as an adaptive mechanism of *Candida albicans* to the host. The high conservation of the splicing machineries among other eukaryotes suggests the regulation of splicing may play a general role in virulence and intron evolution of pathogens.

Introduction

In eukaryotes, splicing is a universal mechanism of gene regulation. Splicing removes introns from precursor mRNAs (pre-mRNAs) via the nuclear spliceosome to produce mature messenger RNAs (mRNA) by catalyzing two sequential transesterification reactions.¹ This ribonuclearprotein complex is highly conserved across species and is composed of over 100 proteins and five small nuclear RNAs.² Splicing and alternative splicing (AS) are essential processes for regulating gene expression among higher eukaryotes during development and in response to stress.³ AS utilizes combinations of exons to generate diverse mature mRNAs from a single pre-mRNA transcript. It is the major mechanism responsible for bridging the gap between a comparatively small genome size in higher eukaryotes and their genetic diversity.

Introns themselves play an important role in genome adaptation in response to environmental pressures. Both intron gain⁴ and intron loss⁵ have been implicated as adaptive strategies over evolutionary history for a variety of organisms.^{6,7} The non-pathogenic *Saccharomyces cerevisiae* and the opportunistic fungal pathogen in humans *Candida albicans*, represent two extreme examples of intron loss in fungi, as over 90% of genes lack introns.⁸ Interestingly, the remaining introns in both species

display a non-random distribution and this enrichment has been postulated to confer selective advantages for these organisms.^{9, 10}

Although splicing and AS are rare events in both *S. cerevisiae* and *C. albicans* due to the low number of introns in the genome, recent work suggests that these processes may play a much greater role in the adaptation of these species to the environment than previously thought. For instance, amino acid starvation impairs splicing of a subset of *S. cerevisiae* genes,¹¹ and the AS of the *C. albicans* gene *RPL30* (orf19.3788.1) is temperature-dependent.¹⁰ Most recently, the localization of the *Gnd1* was shown to be directed to either the cytosol or the peroxisomes based on the AS of its protein.¹²

C. albicans is a natural colonizer of the human gut. In immunocompromised patients, it can disseminate systemically, inflicting severe diseases with a high mortality rate.¹³ Phagocytosis of *C. albicans* by innate immune cells, such as macrophages or dendritic cells, exposes fungal cells to a hostile and nutrient poor environment. Although the transcriptional response of *C. albicans* has been investigated using several *in vitro* infection models^{14, 15} a contribution of post-transcriptional processing mechanisms, such as splicing, to pathogenesis, remains entirely unexplored.

Here, we examine the global prevalence of splicing and during *C. albicans* pathogenesis using simultaneous RNA-seq of both host and pathogen transcriptomes. We identify a subset of genes that splice specifically in response to murine dendritic cells, including novel AS transcripts in *C. albicans*, that undergo splicing during discrete infection stages. Our data identifies infection-specific splicing as a novel adaptive mechanism of *C. albicans* to the host. The high conservation of the molecular splicing machinery among other microbial pathogens suggests that splicing control could be a

general adaptive mechanism in microbial pathogenesis to counteract host immune surveillance.

Results

We performed massively parallel RNA sequencing of cDNA (RNA-seq) obtained from co-cultures of *C. albicans* and *M. musculus* bone marrow-derived dendritic cells (BMDCs) over a time course of infection. BMDCs were chosen as the model host because they are required for the activation of adaptive immune responses and are an important cell type in pathogen recognition.^{16, 17} The selected time points encompass four discrete *C. albicans* infection stages including: fungal adhesion, phagocytosis, hyphae formation, and phagosomal escape (Figure 1A) as quantified in the co-cultures by fluorescence microscopy (Figure 1B). RNA mixtures were purified from 3 independent biological repeats for each time point and subjected to deep-sequencing. The resulting ~120 million 36-nucleotide long reads were simultaneously mapped to both the *C. albicans* and *M. musculus* transcriptomes for each time point. Over 80% of the sequence reads were successfully mapped (Supplemental Table 2).

We first determined the differential expression values of all genes for each species over the time course of the infection. Only transcripts that changed their expression levels at least two-fold relative to the pre-infection state were considered (Figure 2A). In total, approximately 600 *C. albicans* and 200 *M. musculus* genes were identified as differentially expressed over the infection time course. Strikingly, more than 16 times as many *C. albicans* genes decreased (n=507) in transcription than were up-regulated (n=31) in response to the initial exposure to BMDCs, and 5 times as many remained repressed at the point of phagocytosis (n=89, n=18), suggesting a general transcriptional repression on the side of *C. albicans* during the onset of infection.

Out of the differentially expressed *C. albicans* genes, 75 contain experimentally verified introns.¹⁸ We found that this intron expression was time dependent. We detected significant intron enrichment at 120 min above the genomic average of approximately 6% within the genome to over 25% (p value 2.2×10^{-16} , chi squared test) (Figure 2B), suggesting an importance for intron regulation during the infection process.

To see if genes associated with splicing were additionally regulated during the time course, we used the GO term finder from the *Candida* genome database (CGD) (<http://www.candidagenome.org/>) on the differentially expressed gene set. At 30, 60, and 90 min, we find that genes associated with RNA binding are enriched as a category (Figure 2C). Furthermore, at 120 min, when we see the highest percent of intron containing genes, we find rRNA processing (p value 0.062), and several terms corresponding to ribosome biogenesis (p value 2.77×10^{-14}) and mRNA transport are enriched (Figure 2D). These data combined suggest that splicing in *C. albicans* as a process is generally enriched during the infection time course.

Based on sequencing of the *Schizosaccharomyces pombe* transcriptome and high-density tiling arrays with *S. cerevisiae*, it was recently proposed that transcriptional and splicing efficiencies could be mechanistically linked.^{19, 20} Therefore, we investigated how the spliced genes were regulated at the transcriptional level. First, the expression values (RPKM) for all annotated *C. albicans* genes for all time points were collected from the TopHat read mapping. We separated those genes in which splice junctions were detected during all three sequencing repeats during at least one infection time point, which we refer to as our spliced gene set (Figure 3A) from those which were expressed, intron containing, but where either no splice junction was detected, or it was not consistently detected in the sequencing repeats. When all data

were pooled, the mean expression values of intron-containing genes (Figure 2E, blue line) and spliced genes (Figure 2E, red line) were around three-fold (p value 1×10^{-6} , permutation test) and six-fold (p value 1.4×10^{-4} , permutation test) higher, respectively, than the mean value of all genes (Figure 2E, green line). When the mean expression values for individual infection stages are resolved, we find a similar correlation for both subsets at time point 0 (spliced and intron-containing p value 1×10^{-6} , permutation test) (Figure 2F). This indicates that spliced genes in general are higher expressed than the genomic average, which was similarly observed for *S. pombe*.^{19, 20}

For genes consistently detected as spliced, we verified the splicing of the sequenced cDNA using intron-spanning primers designed around the annotated junctions on CGD for each gene all time points over the 3 repeats (Figure 3A). Using this PCR based approach, we found that of our set of 50 junctions investigated 33 showed only the spliced product at all time points with BMDs, however the remaining 18 junctions varied in their splicing pattern during the infection time course. We found 10 junctions in which partial splicing took place, in which a mixture of both spliced and non-spliced product were identifiable, 6 junctions belonging to genes with multiple introns where only one of the two introns were spliced (*HHO1*, *MMD1*, and *ATP1*), a form of AS referred to as intron skipping (Figure 4A),²¹ and two in which multiple novel alternative splice forms were detected, in which there was a difference in 5' donor site selection (Orf19.4149.1 and *PRE8*) (Figure 4B-C). Interestingly, these AS forms were detected only at specific time points, suggesting that the infection stage was important for the generation of these different forms.

Intron removal is catalyzed by the spliceosome in sequence specific manner. The intron boundaries maintain highly conserved motifs that identify them.²² To see if there was any selectivity for the spliced gene set on the sequence level, we looked for

motif variations among the donor, acceptor, and branch site motif sequences for the spliced gene set compared to the total set of intron containing genes in *C. albicans*. We found a common sequence motif variation among the spliced gene set, with a significant enrichment for an “A” 5 bp before the donor site (p-value 0.028, Chi squared test), and 6bp after the acceptor site (p value 0.043, Chi squared test) (Supplemental Table 4). No differences were found in the branch point site for the spliced gene set (data not shown). Differences in the sequences flanking the donor and acceptor site may aid in the recognition of the spliced gene set within the pool of intron containing genes by the spliceosome complex.

In order to see if the splicing or AS was dependent on the presence of BMDCs, we performed the same PCR based splicing analysis for all 53 junctions with *C. albicans* in a variety of media conditions alone including complete dendritic cell media (DCM), DMEM, YPD, and minimal media (MM) (Figure 3A). We found significant variation in the mRNA species present under media conditions alone compared the interaction time course. We saw an increase in the number of genes in which partial splicing was observed, and several genes in which splicing was dependent on the presence of the BMDCs, and not observed under any of the tested media conditions (Figure 3F). Importantly, under media conditions, we did not detect any difference in the mRNA species present throughout the time course (Figure 4A-C), whereas in the presence of BMDCs, especially for the AS genes, we detected some mRNA species only during specific time points, further suggesting the importance of association of *C. albicans* with BMDCs for the generation of these products.

In order to see if the spliced gene set functionally grouped, we checked again for any GO term enrichment of the spliced gene set. Although intron-containing genes in *C. albicans* are known to be enriched in ribosomal components ¹⁰ we found that,

although the ribosome was significantly enriched (p value 0.0028), it was one of the least enriched categories overall (Figure 3G), suggesting a greater importance for the uncharacterized genes spliced during the infection time course.

As we observed many cases in which both spliced and unspliced products were detected, we wanted to quantify the extent of splicing within a single *C. albicans* locus. To do this, we developed a mathematical model to calculate the splicing efficiency. In brief, we used the expression values (RPKM, i.e. Reads mapped per Kilobase in a million mapped reads) of the individual genes obtained from the sequencing reads mapped onto the coding and genomic *C. albicans* assemblies. As a control, we verified with 50 random intron lacking genes that the expression values for the CDS and GDS are identical (Figure 3B) and only for intron containing genes is there a shift away from the $y=x$ line (Figure 3C). With those conditions being satisfied, the ratio of the two expression values (CDS and GDS) can then be expressed as a function of the intron length relative to the total gene length (z), and the ratio of spliced and total mRNA originating from the locus (Figure 3D). This approach was used to calculate the splicing efficiencies for all *C. albicans* intron-containing genes for all time points (Figure 3E). On the whole, 85% of the calculated splicing efficiency values were over 90% or below 10% indicating that under these conditions, splicing is generally an all-or-nothing effect, and supporting the idea that this process is regulated co-transcriptionally.

During the verification of our spliced gene set, we detected several additional bands in our PCR assay, corresponding to novel AS junctions for the *C. albicans* genes *HHO1*, orf19.4149.1 and *PRE8* (Figure 4A-C) which were then sequenced to verify the shift in the junction compared to the annotated junction on CGD. We checked the expression of these genes by qPCR to verify that they were indeed expressed at the time in which the alternative junctions were detected (Figure 4D-F). Interestingly, we

found that the regulation of the alternative forms in a time dependent manner took place only upon interaction with BMDCs, and not in response to media only conditions including DCM (Figure 4A-C), DMEM, MM, or YPD (data not shown) suggesting the AS of these genes is induced by dendritic cells.

To investigate the functional role of the AS products, we looked for differences in protein domains of the AS products using the InterProScan.²³ We did not detect any changes in the domains for the AS products of orf19.4149.1, or *HHO1* compared to the normally spliced version of the genes. In Pre8, the normally spliced version contains two functional domains: InterPro entry IPR001353 (<http://pfam.sanger.ac.uk/family/PF00227>), a member proteasome family with homology to the alpha and beta subunits of the eukaryotic proteasome, and InterPro entry IPR000426 (<http://pfam.sanger.ac.uk/family/PF10584>), which contains the proteasome subunit A N-terminal signature. Interestingly, in both alternative splice products identified for Pre8, the N-terminal signature domain is absent. In *C. albicans*, 4 proteins are currently annotated as members of the proteasome complex, including: Pre5, Pre10, Pre9 and Pup2.²⁴ All of the members contain both domains, as does the putative member of the complex Pre8, which might suggest that both domains are needed to be functional components of the proteasome complex.

Discussion

In this study, we identify infection-specific splicing in *C. albicans* as a hitherto unrecognized time-dependent response mechanism during host invasion by fungal pathogens. Using simultaneous RNA-seq of host-pathogen transcriptomes, we identified a subset of spliced and AS genes in *C. albicans*, including several novel AS transcripts in response to BMDCs. We further developed a mathematical model to

quantify the extent of splicing during the interaction. This is the first study to identify a contribution of regulating post-transcriptional processing mechanisms to *C. albicans* pathogenesis.

Phagocytosis of *C. albicans* by immune cells exposes it to starvation-like conditions, altering nutrient availability, pH, and attack by reactive oxygen species, among numerous others tissue and cell specific defense mechanisms by the host.^{14, 25} The infection stages used in this study were chosen to maximize number of environmental stresses on *C. albicans* by profiling its transcriptional response before, during, and after phagocytosis by BMDCs. By using BMDCs, which have significantly reduced killing capacity compared to other host immune cells, such as macrophages and neutrophils, we extend the time for *C. albicans* to transcriptionally respond and adapt to environment within the phagosome. Therefore, the splicing response we detected may not be present in immune cells where different environmental stresses need to be considered.

A strong net down-regulation of spliced and non-spliced intron containing genes is observed immediately after infection is initiated. These data demonstrate that the infection provokes a global transcriptional repression of *C. albicans* that is not specific to the spliced genes. Since a correlation between splicing and elevated transcription was observable for the pre-infection state only, we conclude that splicing during infection constitutes a gene regulatory mechanism that is at least partially independent of the transcriptional response.

Out of the identified spliced gene set, over 50% are uncharacterized genes. Of those that have been characterized so far, previous studies indicate that the deletion of several of them leads to numerous phenotypes implicated in fungal pathogenesis, including: haploinsufficient growth defects (*HHO1*, *ATP2*)²⁶, decreased invasive growth

(Orf19.6220.4, *LTP1*, *IDH1*, *ATP1*²⁶, *ASC1*²⁷) decreased or abnormal filamentous growth (*CDC14*²⁸, *ASC1*²⁹), increased resistance to antifungal drugs (*FCA1*³⁰), and several putative essential genes (*CDC12*³¹ and *INO4*³²). Notably, these include genes where we identified splicing or alternative splicing as infection specific, including *CDC14* and *HHO1* respectively. Based on our *in vitro* data, these defects combined, suggests a general decrease in the virulence capacity among the spliced gene set and may play a role in the overall virulence of *C. albicans* *in vivo*.

C. albicans is an intron poor species in which splicing and AS are rare events. Nevertheless, that the remaining introns are highly enriched in specific processes, suggest a functional significance for their retention,¹⁰ however what this function may be is not yet understood. As *C. albicans* is a natural human commensal and its evolution is coupled to its host, it is not surprising that some responses of the pathogen would only be detectable upon interaction with host cells. The phagocytosis of *C. albicans* by immune cells is an environment in which precise transcriptional regulation is required to counteract host challenge. The regulation of splicing during this process supports the idea that the remaining introns contribute to the fitness of the organism, and that environmental pressures exerted by the host may function as a driving force in intron evolution. The tight regulation of the AS products at specific infection stages is consistent with this hypothesis. Additionally, the overall splicing response itself is on par to what would be expected among higher eukaryotes, suggesting a higher level of transcriptional complexity than has previously been attributed to *C. albicans*.

In general, introns maintain the canonical dinucleotides GT-AG for the donor-acceptor site for over 98% of introns.³³ However, variability in the splice site sequences is a known regulatory mechanism for selection by the spliceosome and non-canonical splice sites are more common for alternatively spliced genes.³⁴ Of the

novel AS transcripts identified in this study, we find that one AS form of *PRE8*, contains a non-canonical splice site: AT-AG. It has been previously shown in eukaryotes, that the splicing signals and the splicing machinery have coevolved.³⁵ Non-canonical splice signals are indeed processed by different spliceosomal components and schemes have been devised to classify introns based on the type of spliceosomal machinery that recognizes them.³⁶ From an evolutionary perspective, combinations of these two signals have been suggested by Burge *et al.* to identify a transitional stage between the major spliceosomal machineries. Therefore, the ability of *C. albicans* to splice this non-canonical site may suggest the presence of additional spliceosomal components than have yet to be identified in the species.

We propose that the infection specific spliced and AS genes described in this work may represent a new set of *C. albicans* virulence genes, whose transcripts are differentially regulated upon interaction with host immune cells. The high conservation of the splicing machinery among other eukaryotic organisms suggests that changing splicing patterns could be a general adaptive mechanism for pathogenic species to counteract host immune surveillance.

Materials and Methods

***Candida* growth conditions**

The *C. albicans* clinical isolate SC5314³⁷ was used in all *in vitro* experiments. The GFP-labeled *C. albicans* strain derived from the CAF2-1³⁸ was used for fluorescence microscopy experiments. *C. albicans* strains were routinely grown on YPD plates (1% yeast extract, 2% peptone, 2% glucose, 2% agar) and in standard rich media YPD (1% yeast extract, 2% peptone, 2% glucose) for liquid culture at 30°C. Fungal cells were collected in the logarithmic growth phase by a brief centrifugation,

washed in sterile PBS, and diluted for co-culture interaction studies in dendritic cell media (DCM) or standard minimal yeast media, YPD, or DMEM (PAA Laboratories GmbH, Pasching) for *in vitro* assays as previously described.³⁹

Differentiation of bone marrow derived dendritic cells (BMDCs)

BMDCs were differentiated from the femurs of 7-9 week old wild type C57BL/6 mice as previously described.³⁹ Briefly, bone marrow was differentiated for 7 days in DMEM, supplemented with penicillin and streptomycin (Life Technologies BRL), 10% heat-inactivated FCS (Invitrogen, Carlsbad, CA; USA) and 15 % GM-CSF-containing x-conditioned media. BMDCs were analyzed for typical dendritic cell markers by flow cytometry using a panel of marker antibodies. BMDCs were F4/80⁻, CD11b⁺, where at least 70% of the cells were CD11c⁺ representing dendritic cell lineages.

Fungal-mammalian cell co-culture

Fungal-mammalian cell co-culture was performed as previously described.³⁹ BMDCs were plated at a density of 1.0×10^5 cells/cm² in sterile cell culture dishes and incubated with fungal cells at a target ratio of 1:1 *C. albicans* cell to BMDCs. Samples were incubated at 37°C 5% CO₂, 95% humidity.

***C. albicans* internalization assay**

Fungal-mammalian cell co-culture was performed with GFP-*C. albicans* on sterile microscopy cover slips. Cells were fixed in 2% para-formaldehyde for 3 min in PBS at room temperature. Blocking was performed in 4% BSA in sterile PBS at 4°C. Slides were then treated with a *C. albicans* cell wall-specific primary polyclonal antibodies (1:500 in 4% BSA in PBS) overnight at 4°C. Slides were then washed in sterile PBS, and treated with anti-rabbit CY5-conjugated antibodies (Sigma-Aldrich) in glycerol (1:350 in 4% BSA in PBS) for two hours. Finally, slides were inverted in mounting media (Sigma-Aldrich) containing 1µg/ml DAPI (Sigma-Aldrich). Imaging was

performed on an Olympus Cell-R Live Imaging Unit (Olympus, Essex, UK) for all experiments. We scored *C. albicans* cells based on their morphology and association with BMDCs in three independent biological repeats and 100 cells were counted per time point.

Sample preparation for RNA-seq

Total RNA was isolated from BMDCs and *C. albicans* co-culture samples and prepared for RNA-seq as previously described.⁴⁰ All samples were sequenced on the Genome Analyzer IIx at GATC (Konstanz, Germany) using 36bp, single run, indexed read mode (36+7bp). All primary sequence data are MIAME-compliant and available at the ArrayExpress database (<http://www.ebi.ac.uk/arrayexpress/>) under accession number E-MTAB-595.

Sequence read mapping and raw data pre-processing

We first generated a reference set to include both *C. albicans* and *M. musculus* genomic and coding assemblies by joining all coding sequences from *M. musculus* assembly 37^{41, 42} and the orf-21 *C. albicans* assembly freely available from the CGD <http://www.candidagenome.org/>.⁴³ We used Bowtie⁴⁴ to map sequencing reads for all samples and repeats against our reference set using the default parameters with an allowance of two mismatches per read. We included the reporting of all valid alignments per read (-a), and singleton alignments were reported as best fit based on the quality and number of mismatched positions (-best). We extracted read alignments using NextGenTool (NGT) to avoid a potential bias to any specific reference region. NGT counts the number of mapped reads per coding sequence. If a read showed more than one valid equally scoring region, then the corresponding read was counted for each region respectively. However, the reads contribution was counted only once and was divided by the number of regions to which it aligned.

Sequence normalization for differential gene expression analysis

We normalized read counts and analyzed for differential gene expression as previously described.⁴⁵ The Benjamini-Hochberg⁴⁶ procedure was used to control the false discovery rate. In general, RNA sequencing can be modeled as a random sampling process, in which each read is sampled independently and uniformly from the RNA pool in the sample. Under this assumption, the read-counts follow a multinomial distribution, which can be approximated by the Poisson distribution. The Poisson distribution considers the property of equality of mean and variance. However, it does not account for frequently observed over-dispersion or bias in biological data. For this reason, we decided to use the negative binomial distribution, which produces a suitable underlying distribution. The negative binomial distribution provides one additional parameter, which can be used to adjust the variance independently of the mean. As the number of replicates of our sequencing datasets is relatively low, we decided to use a method that suggests the use of local regression on the graph of sample means and sample variance to estimate the per-gene variance.⁴⁵ The final variance estimate for the individual read counts is then decomposed into an inherent Poisson shot noise and the per-gene raw variance term.

Splice junction identification

To identify splice junctions in *C. albicans*, we used TopHat⁴⁷ via the GALAXY server^{48, 49} using the default parameters on the read alignments obtained from Bowtie using the SC5314 21-genomic and 21-coding assemblies. We set the constraints for intron detection between 20-1000bp and visualized the detected splice junctions on the Integrated Genome Viewer (IGV).⁵⁰ The existing intron annotations available on CGD were used to verify the positions of detected splice junctions. Only junctions that were detected in all 3 repeats for at least one time point were subjected to further analysis.

Splice junctions were verified using a PCR based protocol of treated samples. We used gDNA from *C. albicans* alone as the unspliced control for all samples. Primers were designed to flank existing intron annotations available on CGD for the spliced gene set (Supplemental Table 1). PCR amplifications were performed samples either from sequenced cDNA or of *C. albicans* treated in media conditions alone were for 32 cycles at 94°C 3 min, 94°C 30s, 55°C 30s, 72°C for 1 min. Samples which showed putative alternative splice products were sequenced to determine to the position of the splice junctions.

qPCR

RNA sample preparation, reverse transcription, and real-time PCR were performed previously described ³⁹ using the following primers: *PRE8* forward 5'-TGGGACCAGATTTCCGAGTA-3' reverse 5'-TTCACCGGTGGGTATTCATT-3', *HHO1* forward 5'-AAAGAAAGCTGCCACCAAAA-3', reverse 5'-TGGAGCAGCTTTGGTTTCTT-3', Orf19.4149.1 forward 5'-ACCACACATT-TTCACCAACC-3', and reverse 5'-TCTTGGTTTTCTTGTTGGCTTT-3'. Results were calculated using the $\Delta\Delta ct$ method and are expressed as the fold of the gene expression of interest versus the expression of the *C. albicans* housekeeping gene, *RIP1*, forward 5'-TGCTGACAGAGTCAAGA-AACC-3' reverse 5'-GAACCAACCACCGAAATCAC-3', in treated versus untreated conditions.

Frequency distributions of spliced genes

Frequency distributions were graphed using the mean of the CDS (RPKM_{CDS}) pooled from all time points for all detected *C. albicans* genes with and without introns, intron-containing genes, and the spliced gene set. Distributions were graphed as relative frequencies as a percent (bin-width 100) and fitted with a Lorentizan function. The significance of the mean values for the spliced and intron-containing genes relative

to the total gene set was determined using a Permutation test for all time points separately and combined. For the permutation test, 1 million gene sets were drawn and compared the resulting difference-of-means distribution to the value obtained by our gene sets.

Calculating *C. albicans* splicing efficiency

For all spliced genes, we estimated the transcript abundance using the default parameters for Cufflinks⁵¹ on both the coding and genomic *C. albicans* assemblies. Bias correction was included. We used quartile normalization to estimate the RPKM values (reads per kilobase of exon model per million mapped reads) for each sample, and then used the mean of the RPKM values for the three repeats of the CDS (RPKM_{CDS}) and GDS (RPKM_{GDS}) assemblies, respectively for further analysis.

We based our splicing calculation on the quantification of transcript abundances retrieved from mapping the sequenced reads to the coding and genomic orf-21 *C. albicans* assemblies using Cufflinks. The RPKM retrieved from Cufflinks can be expressed as the following functions when mapping the reads to the CDS and GDS reference sequences. The RPKM formula was adapted from⁴²):

$$RPKM_{CDS} = \frac{10^9(E_1 + E_2)}{l_e N_{CDS}} \quad \text{and} \quad RPKM_{GDS} = \frac{10^9(E_1 + E_2 + I)}{(l_e + l_i) N_{GDS}} \quad \text{Where reads mapped to}$$

the exonic sequence originating from (spliced) mRNA (E_1), reads mapped to the exonic sequence originating from (unspliced) pre-mRNA (E_2), reads mapped to the intron (I), length of the exonic sequence (l_e), length of the intron (l_i), and the total number of mapped reads (N) are represented. For genes lacking introns ($E_2=0$, $I=0$, $l_i=0$), the RPKM_{CDS} and RPKM_{GDS} values should be identical, if the total number of mapped reads for both the coding and genomic assemblies do not differ significantly

($N_{CDS} \approx N_{GDS}$). We validated this approximation using the CDS and GDS assemblies with Bowtie for all samples (Supplemental Table 3).

We therefore expect values of the $RPKM_{CDS}$ and $RPKM_{GDS}$ to differ only for spliced genes, which can be expressed by their ratio as:

$$\frac{RPKM_{CDS}}{RPKM_{GDS}} = \frac{\frac{10^9(E_1 + E_2)}{l_e N_{CDS}}}{\frac{10^9(E_1 + E_2 + I)}{(l_e + l_i)N_{GDS}}}$$

We base our further calculations on the assumption that the extent by which this ratio differs from one is proportional to the length of the intron relative to the length of the gene, and the ratio of spliced and un-spliced RNA originating from the locus (Supplemental Figure 1).

As the next step, the N_{CDS}/N_{GDS} is again approximated as one. Therefore, the expression simplifies to:

$$\frac{RPKM_{CDS}}{RPKM_{GDS}} = \frac{\frac{10^9(E_1 + E_2)}{l_e N_{CDS}}}{\frac{10^9(E_1 + E_2 + I)}{(l_e + l_i)N_{GDS}}} = \frac{\frac{E_1 + E_2}{l_e}}{\frac{E_1 + E_2 + I}{l_e + l_i}}$$

In the numerator, we substitute in as the next step:

$$\frac{10^9 E_2}{l_e N_{CDS}} \approx \frac{10^9 E_2}{l_e N_{GDS}} = \frac{10^9 (E_2 + I)}{(l_e + l_i)N_{GDS}} \text{ which is simplified as: } \frac{E_2}{l_e} = \frac{(E_2 + I)}{(l_e + l_i)}$$

We base this approximation on the assumption that the average coverage of exonic segments of the unspliced pre-mRNA is approximately equal to the average coverage over the total gene length. We note, that although this approximation step does introduce an error margin into the calculation (which assumes an ideal gene in which the transcript coverage is approximately equal across the entire length), it is required for the downstream arithmetic simplification.

The expression is thus converted as:

$$\frac{RPKM_{CDS}}{RPKM_{GDS}} = \frac{\frac{E_1 + E_2}{l_e}}{\frac{E_1 + E_2 + I}{l_e + l_i}} = \frac{\frac{E_1}{l_e} + \frac{E_2 + I}{l_e + l_i}}{\frac{E_1 + E_2 + I}{l_e + l_i}}$$

The expression can arithmetically be converted as:

$$\begin{aligned} \frac{RPKM_{CDS}}{RPKM_{GDS}} &= \frac{\frac{E_1 + E_2}{l_e}}{\frac{E_1 + E_2 + I}{l_e + l_i}} = \frac{\frac{E_1}{l_e} + \frac{E_2 + I}{l_e + l_i}}{\frac{E_1 + E_2 + I}{l_e + l_i}} = \frac{\frac{E_1 l_i + E_1 l_e + E_2 l_e + I l_e}{l_e(l_e + l_i)}}{\frac{E_1 + E_2 + I}{l_e + l_i}} = \\ &= \frac{\frac{E_1 l_i}{l_e} + E_1 + E_2 + I}{\frac{E_1 l_i}{l_e} + E_1 + E_2 + I - \frac{E_1 l_i}{l_e}} = \frac{\frac{E_1(l_i + l_e)}{l_e} + E_2 + I}{\frac{E_1(l_i + l_e)}{l_e} + E_2 + I - \frac{E_1 l_i}{l_e}} = \\ &= \frac{1}{\frac{\frac{E_1(l_i + l_e)}{l_e} + E_2 + I - \frac{E_1 l_i}{l_e}}{\frac{E_1(l_i + l_e)}{l_e} + E_2 + I}} = \frac{1}{1 - \frac{\frac{E_1 l_i}{l_e}}{\frac{E_1(l_i + l_e)}{l_e} + E_2 + I}} = \frac{1}{1 - \frac{\frac{E_1 l_i}{l_e}}{\frac{E_1(l_i + l_e)}{l_e} + \frac{(E_2 + I)(l_i + l_e)}{l_i + l_e}}} = \\ &= \frac{1}{1 - \left(\frac{l_i}{l_i + l_e} \right) \left(\frac{\frac{E_1}{l_e}}{\frac{E_1}{l_e} + \frac{E_2 + I}{l_i + l_e}} \right)} \end{aligned}$$

This equation cannot be solved directly, since E_1 and E_2 cannot be determined. However, by introducing two derived variables: the relative length of the intron to the total gene length (z) and the splicing efficiency (y), it is possible to determine the ratio of exonic reads derived from spliced mRNA and the total transcript derived from the locus.

$$z = \frac{l_i}{l_i + l_e} \quad \text{and} \quad y = \frac{\frac{E_1}{l_e}}{\frac{E_1}{l_e} + \frac{E_2 + I}{l_i + l_e}} \quad \left(= \frac{mRNA}{mRNA + pre-mRNA} \right)$$

Therefore, we can calculate the splicing efficiency of the locus using the final expression:

$$\frac{RPKM_{CDS}}{RPKM_{GDS}} = \frac{1}{1 - z \cdot y}$$

Consequently, after substituting in the relative intron length for each intron-containing gene (which is determined from the annotated genomic coordinates), the splicing efficiency is retrieved representing the fraction of spliced mRNA in the total amount of transcript, spliced mRNA and unspliced pre-mRNA, originating from the locus. We used this final equation to calculate the splicing efficiencies for all intron-containing genes.

Figure Legends

Figure 1.

A. *C. albicans* - BMDC infection stages (30 min, attachment; 60 min, phagocytosis; 90 min, hyphal formation; and 120 min, phagosomal escape). *C. albicans* expressing GFP are shown in green, extracellular *C. albicans* are stained red with antibody-conjugated FITC, and nuclei are stained blue with DAPI. The contours of the membrane of BMDCs are overlaid as manually traced from bright-field images.

B. Homogeneity of the infection stages. Mean percentage values and standard deviations are shown.

Figure 2.

A. Dot plots of the complete set of transcript abundances for *C. albicans* (black) and *M. musculus* (orange) genes over time, normalized against the expression values measured for the pre-infection stage. Each transcript is represented by one dot. The axes represent the base mean, i.e. the normalized number of reads per transcript and

the exponents of the $\log_2$ scale. Gray lines show the cutoff value used to define significance.

B. Percentage of intron containing genes among the total set of differentially expressed genes for each time point. The percent of introns present in the *C. albicans* genome, approximately 6%, is overlaid of the graph as a blue dotted line. P-values were calculated using the Chi-squared test.⁵²

C. *C. albicans* GO term slim categories enriched during the infection time course where intron enrichment was not significant. Categories related to splicing are shown in red, all other categories are shown in gray scale.

D. *C. albicans* GO term categories enriched at 120 mins. Categories related to splicing are shown in red, all other categories are shown in gray.

E. Frequency distributions of the expression values (RPKM_{CDS}) measured for all annotated *C. albicans* genes. Values are shown as a percent of the total input for all detected *C. albicans* genes (blue), intron-containing genes only (red), and spliced genes (green).

F. Average CDS over the time course of infection for all detected *C. albicans* genes (blue), intron-containing genes only (red), and spliced genes (green).

Figure 3.

A. Verification of *C. albicans* splice junctions during BMDC infection and under various media conditions: dendritic cell media alone (DCM), DMEM (D), YPD, and minimal media (MM) over the time course of infection. Each gene is scored on the following criteria: spliced product alone (+), unspliced product alone (-), a mixture of spliced and unspliced product (M), Alternative 5' or 3' selection (AS, blue), alternative splicing via

intron skipping (green). Genes with previously annotated alternative splice junctions are marked (*).

B. RPKM CDS and GDS values of 50 randomly selected intron-lacking genes. The red line shows the $y = x$ line to which the dots regress with a linear regression coefficient of $R^2 = 1$.

C. RPKM CDS and GDS values of 50 randomly selected intron-containing genes where the regression line (red) for the intron-lacking genes is overlaid on the plot. The regression line (red) for the intron-containing genes is overlaid on the plot.

D. A mathematical model to calculate *C. albicans* splicing efficiency. The dependence of the RPKM ratio of the CDS/GDS values on the relative length of the intron to the total gene (z) is shown for five theoretical splicing efficiency values.

E. Overlay of experimental data on the mathematical model of splicing efficiency. For all *C. albicans* intron-containing genes (gray dots) and spliced genes (blue dots) the RPKM ratio of CDS/GDS ratio is plotted against the relative intron length (z) for all time points. Functions for 100% (red) and 0% (blue) splicing efficiencies from panel B are overlaid for orientation. Green box denotes the empirically defined threshold of the function ($z > 0.25$).

F. Representative splice gel for *CDC14* distinguishing spliced versus unspliced products in different media only conditions.

G. GO term categories enriched in the splice gene set for component. The ribosome category is shown in red because it has been previously shown to be an enriched category for intron containing genes. All other categories are shown in gray.

Figure 4.

A – C. PCR Splice gels of novel alternative splice junctions for in co-culture with BMDCs or in media alone (DCM) for **(A)** *HHO1*, **(B)** orf19.4149.1, and **(C)** *PRE8*. Here, gDNA serves as the unspliced (US) control; alternative spliced (AS) junctions and annotated spliced (S) are shown for each gene. For each alternative splice junction, the verified sequence position for the novel and annotated introns are shown for each gene.

D – F. qPCR of novel alternative spliced genes **(D)** *HHO1*, **(E)** orf19.4149.1, and **(F)** *PRE8*.

Supplemental Tables

Supplemental Table 1. Intron spanning primers used in this study.

Supplemental Table 2. Summary statistics for sequencing of *C. albicans* and *M. musculus* coding assemblies.

Supplemental Table 3. Summary statistics for CDS and GDS Bowtie read assignments to orf-21 *C. albicans* assembly.

Supplemental Table 4. Sequence motif variation surrounding annotated donor and acceptor sites for spliced, alternative spliced, and unspliced gene sets.

Supplemental Figure Legends

Figure S1.

A. Representation of parameters of pre-mRNA (unspliced) and mRNA (spliced) used to calculate the splicing efficiency for intron-containing genes. For simplicity, untranslated regions and the poly A tail are omitted from the figures.

B. Graphical representation of “ideal” coverage maps, where the reads are mapped onto either the CDS or the GDS sequence assemblies shown for an example gene. The maps are shown for different theoretical degrees of splicing, where “y” splicing represents all splicing between 1-99%. The dotted line depicts the contour of the gene coverage. Sequence reads that map against the pre-mRNA (black) and mRNA (blue) are shown. Note that it is not possible to distinguish which exonic reads originate from pre-mRNA and mRNA, however their ratio can be calculated.

C. The average coverage of an “ideal” pre-mRNA. When no splicing takes place, the average coverage over the intron is approximately equal to the average coverage over the exon sequences (i.e. the average coverage measured in the two green boxes should be approximately identical).

Acknowledgements

We would like to thank Kai Sohn and Andrea Barta for helpful comments on the manuscript. This work was supported by grants from the Christian Doppler Research Society and in part by the FFG ETB CanVac project. These funding agencies had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. S.T., F.J.S., and A.vH acknowledge financial support from the Wiener Wissenschafts- Forschungs- und Technologiefonds. A.vH and F.J.S. are also supported by the DFG priority program SPP 1174 Deep Metazoan Phylogeny.

References

1. Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell* 2009; 136:701-18.

2. Jurica MS, Moore MJ. Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* 2003; 12:5-14.
3. Kalsotra A, Cooper TA. Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet* 2011; 12:715-29.
4. Rogozin IB, Wolf YI, Sorokin AV, Mirkin BG, Koonin EV. Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Curr Biol* 2003; 13:1512-7.
5. Lynch M. Intron evolution as a population-genetic process. *Proc Natl Acad Sci U S A* 2002; 99:6118-23.
6. Roy SW, Gilbert W. Rates of intron loss and gain: implications for early eukaryotic evolution. *Proc Natl Acad Sci U S A* 2005; 102:5773-8.
7. Jeffares DC, Mourier T, Penny D. The biology of intron gain and loss. *Trends Genet* 2006; 22:16-22.
8. Davis CA, Grate L, Spingola M, Ares M, Jr. Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res* 2000; 28:1700-6.
9. Mitrovich QM, Tuch BB, De La Vega FM, Guthrie C, Johnson AD. Evolution of yeast noncoding RNAs reveals an alternative mechanism for widespread intron loss. *Science* 2010; 330:838-41.
10. Mitrovich QM, Tuch BB, Guthrie C, Johnson AD. Computational and experimental approaches double the number of known introns in the pathogenic yeast *Candida albicans*. *Genome Res* 2007; 17:492-502.
11. Pleiss JA, Whitworth GB, Bergkessel M, Guthrie C. Rapid, transcript-specific changes in splicing in response to environmental stress. *Mol Cell* 2007; 27:928-37.

12. Strijbis K, van den Burg J, Visser WF, van den Berg M, Distel B. Alternative splicing directs dual localization of *Candida albicans* 6-phosphogluconate dehydrogenase to cytosol and peroxisomes. *FEMS Yeast Res* 2012; 12:61-8.
13. Gudlaugsson O, Gillespie S, Lee K, Vande Berg J, Hu J, Messer S, et al. Attributable mortality of nosocomial candidemia, revisited. *Clin Infect Dis* 2003; 37:1172-7.
14. Lorenz MC, Bender JA, Fink GR. Transcriptional response of *Candida albicans* upon internalization by macrophages. *Eukaryot Cell* 2004; 3:1076-87.
15. Rubin-Bejerano I, Fraser I, Grisafi P, Fink GR. Phagocytosis by neutrophils induces an amino acid deprivation response in *Saccharomyces cerevisiae* and *Candida albicans*. *Proc Natl Acad Sci U S A* 2003; 100:11007-12.
16. Bourgeois C, Majer O, Frohner IE, Tierney L, Kuchler K. Fungal attacks on mammalian hosts: pathogen elimination requires sensing and tasting. *Curr Opin Microbiol* 2010; 13:401-8.
17. Netea MG, Brown GD, Kullberg BJ, Gow NA. An integrated model of the recognition of *Candida albicans* by the innate immune system. *Nat Rev Microbiol* 2008; 6:67-78.
18. Bruno VM, Wang Z, Marjani SL, Euskirchen GM, Martin J, Sherlock G, et al. Comprehensive annotation of the transcriptome of the human fungal pathogen *Candida albicans* using RNA-seq. *Genome Res* 2010; 20:1451-8.
19. Carrillo Oesterreich F, Preibisch S, Neugebauer KM. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol Cell* 2010; 40:571-81.
20. Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, et al. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 2008; 453:1239-43.

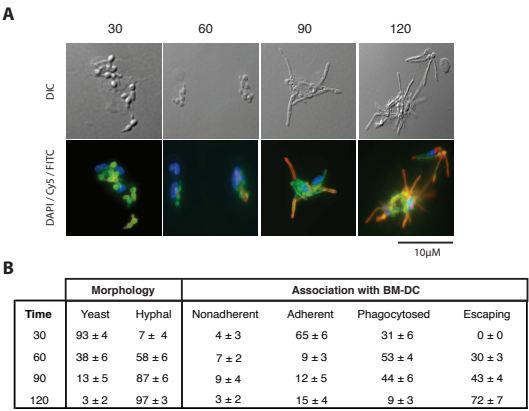
21. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. *Nature* 2010; 463:457-63.
22. Neuveglise C, Marck C, Gaillardin C. The intronome of budding yeasts. *C R Biol* 2011; 334:662-70.
23. Zdobnov EM, Apweiler R. InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 2001; 17:847-8.
24. Fernandez Murray P, Biscoglio MJ, Passeron S. Purification and characterization of *Candida albicans* 20S proteasome: identification of four proteasomal subunits. *Arch Biochem Biophys* 2000; 375:211-9.
25. Hube B. From commensal to pathogen: stage- and tissue-specific gene expression of *Candida albicans*. *Curr Opin Microbiol* 2004; 7:336-41.
26. Oh J, Fung E, Schlecht U, Davis RW, Giaever G, St Onge RP, et al. Gene annotation and drug target discovery in *Candida albicans* with a tagged transposon mutant collection. *PLoS Pathog* 2010; 6:e1001140.
27. Kim SW, Joo YJ, Kim J. Asc1p, a ribosomal protein, plays a pivotal role in cellular adhesion and virulence in *Candida albicans*. *J Microbiol* 2010; 48:842-8.
28. Clemente-Blanco A, Gonzalez-Novo A, Machin F, Caballero-Lima D, Aragon L, Sanchez M, et al. The Cdc14p phosphatase affects late cell-cycle events and morphogenesis in *Candida albicans*. *J Cell Sci* 2006; 119:1130-43.
29. Liu X, Nie X, Ding Y, Chen J. Asc1, a WD-repeat protein, is required for hyphal development and virulence in *Candida albicans*. *Acta Biochim Biophys Sin (Shanghai)* 2010; 42:793-800.
30. Hope WW, Tabernero L, Denning DW, Anderson MJ. Molecular mechanisms of primary resistance to flucytosine in *Candida albicans*. *Antimicrob Agents Chemother* 2004; 48:4377-86.

31. Warena AJ, Konopka JB. Septin function in *Candida albicans* morphogenesis. *Mol Biol Cell* 2002; 13:2732-46.
32. Hoppen J, Dietz M, Warsow G, Rohde R, Schuller HJ. Ribosomal protein genes in the yeast *Candida albicans* may be activated by a heterodimeric transcription factor related to Ino2 and Ino4 from *S. cerevisiae*. *Mol Genet Genomics* 2007; 278:317-30.
33. Burset M, Seledtsov IA, Solovyev VV. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res* 2000; 28:4364-75.
34. Stamm S, Zhang MQ, Marr TG, Helfman DM. A sequence compilation and comparison of exons that are alternatively spliced in neurons. *Nucleic Acids Res* 1994; 22:1515-26.
35. Schwartz SH, Silva J, Burstein D, Pupko T, Eyras E, Ast G. Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res* 2008; 18:88-103.
36. Burge CB, Padgett RA, Sharp PA. Evolutionary fates and origins of U12-type introns. *Mol Cell* 1998; 2:773-85.
37. Gillum AM, Tsay EY, Kirsch DR. Isolation of the *Candida albicans* gene for orotidine-5'-phosphate decarboxylase by complementation of *S. cerevisiae* *ura3* and *E. coli* *pyrF* mutations. *Mol Gen Genet* 1984; 198:179-82.
38. Hull CM, Johnson AD. Identification of a mating type-like locus in the asexual pathogenic yeast *Candida albicans*. *Science* 1999; 285:1271-5.
39. Bourgeois C, Majer O, Frohner I, Kuchler K. *In vitro* systems for studying the interaction of fungal pathogens with primary cells from the mammalian innate immune system. *Methods Mol Biol* 2009; 470:125-39.

40. Tierney L, Linde J, Muller S, Brunke S, Molina JC, Hube B, et al. An Interspecies Regulatory Network Inferred from Simultaneous RNA-seq of *Candida albicans* Invading Innate Immune Cells. *Front Microbiol* 2012; 3:85.
41. Hubbard TJ, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, et al. Ensembl 2007. *Nucleic Acids Res* 2007; 35:D610-7.
42. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002; 420:520-62.
43. Skrzypek MS, Arnaud MB, Costanzo MC, Inglis DO, Shah P, Binkley G, et al. New tools at the *Candida* Genome Database: biochemical pathways and full-text literature search. *Nucleic Acids Res* 2010; 38:D428-32.
44. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009; 10:R25.
45. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010; 11:R106.
46. Benjamini Y, Hochberg Y. Controlling the false discovery rate-A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* 1995; 57:289-300.
47. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; 25:1105-11.
48. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, et al. Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol* 2010; Chapter 19:Unit 19 0 1-21.
49. Goecks J, Nekrutenko A, Taylor J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 2010; 11:R86.

50. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol* 2011; 29:24-6.
51. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 2010; 28:511-5.
52. Kielbasa SM, Gonze D, Herzel H. Measuring similarities between transcription factor binding sites. *BMC Bioinformatics* 2005; 6:237.

Figure 1



A

30 min 60 min 90 min 120 min

Fold Change (Log2)

Base mean

B

$p = 2.2 \times 10^{-16}$

%

Time (min)

C

30 min 60 min 90 min

phosphatase activity, ion binding, oxidoreductase activity, transmembrane transporter activity, DNA binding, RNA binding, nucleic acid binding, ATPase activity, hydrolyase activity, transferase activity, peptidase activity, RNA binding, DNA binding, hydrolyase activity, phosphatase activity, transmembrane transporter activity, oxidoreductase activity

D

2.77×10^{-14}

0.0062

$-\log_{10} p\text{-value}$

unannotated, Ribosome biogenesis, RNA nuclear export, rRNA transport, Ribosome assembly, Cellular A-to-I editing, Organic A-to-I editing, Catalytic A-to-I editing, Cellular component biogenesis, small molecule biosynthesis, Translational regulation, rRNA processing, Ribosomal small subunit

E

RPKM

F

RPKM

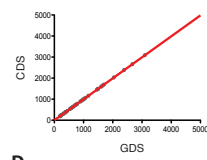
Time (min)

Figure 3

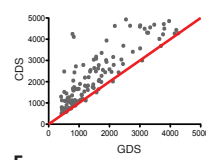
A

Gene	Function	DC + DCM				DCM				YPD			
		0	30	60	90	120	0-120	0-120	0-120	0-120	0-120	0-120	0-120
Orf19.1114	Unknown	+	+	+	+	+	M	M	M	M	M	M	M
ARO3	AA biosynthesis	+	+	+	+	+	M	M	M	M	M	M	M
LSC2	ATP binding	+	+	+	+	+	+	+	+	+	+	+	+
Orf19.2257	Unknown	+	+	+	+	+	M	M	M	M	M	M	M
Orf19.2322.3	Unknown	+	+	+	+	+	+	+	+	+	+	+	+
HRT1	Zinc binding	+	+	+	+	+	+	+	+	+	+	+	+
PHA2	AA binding	+	+	+	+	+	+	+	+	+	+	+	+
AUT7*	Unknown	M	M	M	M	M	M	M	M	M	M	M	M
CDC12	GTP binding	+	+	+	+	+	M	M	M	M	M	M	M
Orf19.3273	Ribosome binding	+	+	+	+	+	M	M	M	M	M	M	M
RHO3	GTP binding	+	+	+	+	+	+	+	+	+	+	+	+
Orf19.4090.1	Unknown	M	M	M	M	M	+	+	+	+	+	+	+
Orf19.4132	Unknown	M	M	M	M	M	+	+	+	+	+	+	+
Orf19.4149.1	Translation	+	AS	AS	AS	+	M	M	M	M	M	M	M
CDC14	Protein P activity	+	+	+	+	+	-	-	-	-	-	-	-
FCA1	Zinc binding	+	+	+	+	+	M	M	M	M	M	M	M
CTT1	Citrate synthesis	+	+	+	+	+	+	+	+	+	+	+	+
Orf19.446.1	Unknown	+	+	+	+	+	M	M	M	M	M	M	M
IDH1	NAD binding	M	M	M	M	M	-	-	-	-	-	-	-
PET100	Protein binding	+	+	+	+	+	M	M	M	M	M	M	M
LTP1	Protein P activity	+	+	+	+	+	M	M	M	M	M	M	M
HHO1	DNA binding	+	+	+	+	+	+	+	+	+	+	+	+
RPL27A	Translation	+	+	+	+	+	+	+	+	+	+	+	+
Orf19.5628	Membrane transport	M	M	M	M	M	-	-	-	-	-	-	-
SLB2	ATP binding	+	+	+	+	+	+	+	+	+	+	+	+
ATP2	ATP binding	+	+	+	+	+	+	+	+	+	+	+	+
RIP1	Electron transport	+	+	+	+	+	+	+	+	+	+	+	+
Orf19.6062.3	Unknown	+	+	+	+	+	+	+	+	+	+	+	+
MWD1	Unknown	+	+	+	+	+	M	M	M	M	M	M	M
Orf19.6220.4	Translation	+	+	+	+	+	+	+	+	+	+	+	+
Orf19.6264.3	Unknown	M	M	M	M	M	-	-	-	-	-	-	-
Orf19.6462	Unknown	+	+	+	+	+	M	M	M	M	M	M	M
Orf19.6563.1	Unknown	+	+	+	+	+	+	+	+	+	+	+	+
Orf19.6718	DNA binding	M	M	M	M	M	M	M	M	M	M	M	M
FMA1	Nucleotide binding	+	+	+	+	+	M	M	M	M	M	M	M
ATP1*	ATP binding	+	+	+	+	+	+	+	+	+	+	+	+
ASC1	Ribosome binding	-	-	-	-	-	+	+	+	+	+	+	+
GTT11	Unknown	+	+	+	+	+	M	M	M	M	M	M	M
RAD6	Transcription	+	+	+	+	+	M	M	M	M	M	M	M
PRE8	Protein catabolism	-	AS	AS	M	M	AS	AS	AS	AS	AS	AS	AS
ECM15	Unknown	+	+	+	+	+	M	M	M	M	M	M	M
FDH3	Zinc binding	M	M	M	M	M	-	-	-	-	-	-	-
JRC4	DNA binding	M	M	M	M	M	M	M	M	M	M	M	M
Orf19.909.1	Unknown	M	M	M	M	M	M	M	M	M	M	M	M
Orf19.913.2	Electron transport	+	+	+	+	+	+	+	+	+	+	+	+

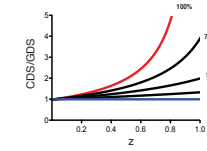
B



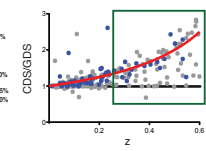
C



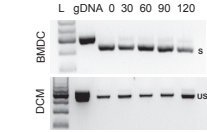
D



E



F



G

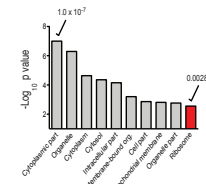
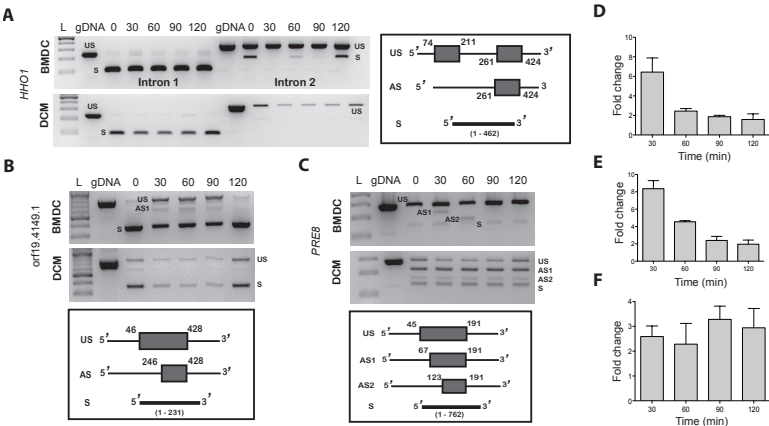
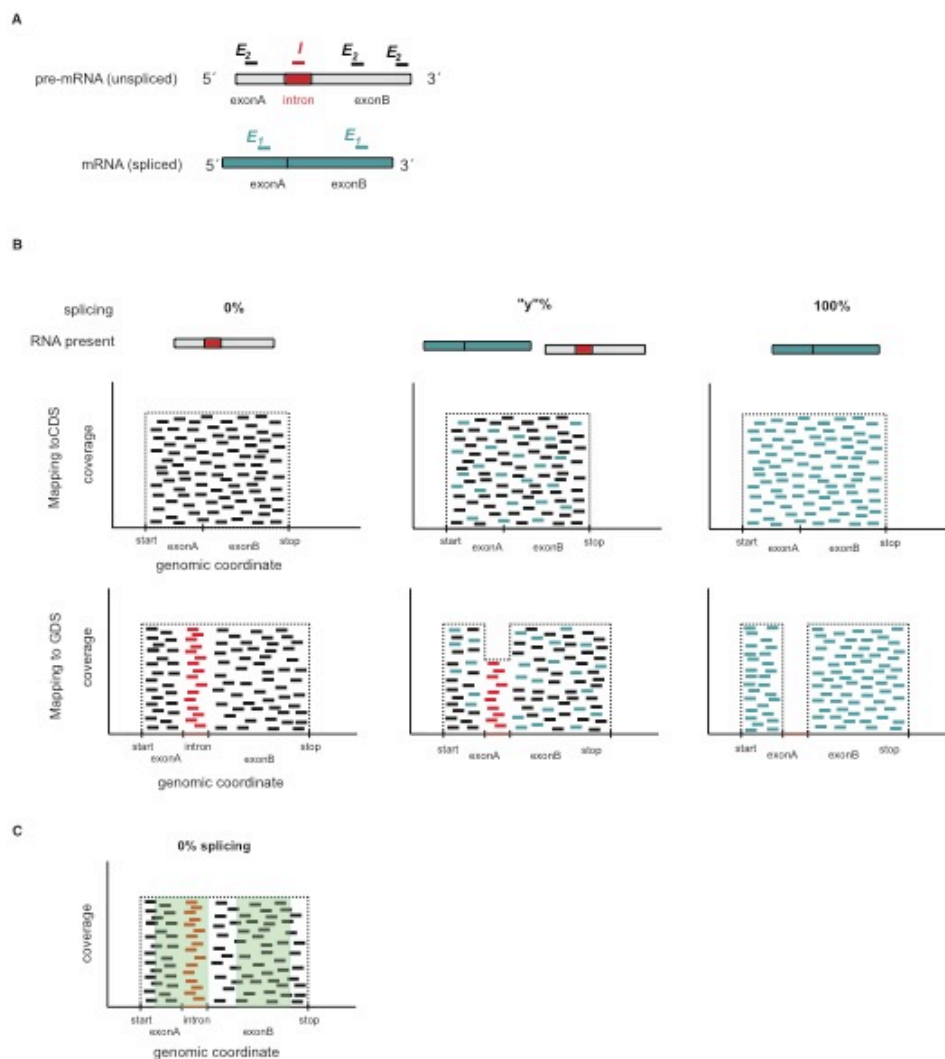


Figure 4



Supplemental Figure 1



Supplemental Table 1

Gene name	Forward primer (5'-3')	Reverse primer (5'-3')
Orf19.1114	CCGTTAGACTTCCATCATCCA	GTTGTGACGTTTCCGTACCA
ARO3	CAACGAACACGTTGGTGAC	ACCTTTCCAACCACTGTGG
LSC2	GATCATTTGCCCGTATTTCAA	AAAGCTTGGGCTTTGATGAC
Orf19.2257	CCAATAGTTGCCCAACAAGG	ACCAGTGGCGTAGGATGAAT
Orf19.2322.3	TAAGACCGGGGAAACCAAGT	ATGTTCCCCAGAGTCCAATG
HRT1	ACGATCAAGATCGCATGGAC	TGCATATCCCATGACCAAAA
PHA2	TTTCTAGGACCAGAGGGAACA	TGTCTTCAATTTTGAAGCATT
AUT7	AAAGACGAGCATCCTTTTGG	CGGTTGGGGTAATATGTCA
CDC12	CCTGAACCAAGTTGGTATTGC	GTGTCTGTGGCTTGGCTCTT
Orf19.3273	CTGCCGCTGGAGGTAGATTA	CTTCCCTTGCAGCAGAATTT
RHO3	TTTGTGGTGGTGGTCAAAAA	TTTTCAAGCGAGTCTGGTGA
Orf19.4090.1	CCTGCTCCAGATATCCCAAA	TCAAAACAACATTGGCTGCT
Orf19.4132	GACGGTCACGATTGGTTTCT	AGCCGCATAAGTTGCTCTGT
Orf19.4149.1	TCTGAAAGAGCTTTCCAAAAGG	ATTGACCAACGGTGACAACA
CDC14	TCACTGGTTCATGTGCCTCT	CTCCACATGGCATAACGAC
FCA1	AAAGGTTTACAAATCGCTCTTGA	TGACATTTTCTCCCATCACAA
CIT1	AACGTTCAACCAACGTAGCC	CTTCACCACCTGGTGCTTTT
Orf19.446.1	GGCTCACACGGAGAATTATGA	CAGAAGGAAGTACCTTTATCGAGAC
IDH1	CCGTTATTGCTCAAAAGGTATG	TTCGACAGCTTCGTCAACAC
PET100	TGGGCTTCATACGTGCATT	TTAATACGGGCGATTTCCTGC
LTP1	TGGCTCCTAAATTTGATCCTAAA	TGAGCATGGTGAGAAACAGG
HHO1	CGCTGCAACTGCTAAATCAA	TTCTTTAAGGTAAGGATAGCAGA
HHO1(I2)	AAGGAAAGAAATGGTCTTCG	GAAGCCGGTTTTTCTTCTT
RPL27A	TGGCTAAGTTCATCAAATCTGG	ATGTGGGAATGGGTGAGATT
Orf19.5628	TCCGTATGTAAATCACCGGATA	TTGTTCTGTGCCTCCAGTTG
SUB2	TCGTGTTTCAACCACTCA	TCGTTACGGATTGTTGGTAGGC
ATP2	GTGCTGCATTGAGAGCAGTC	CGAATTGGACATCAACAACG
RIP1	TGCTTCTCTTGCTTTTCAAGACA	CCATAGCCAAAACATCAGCA
Orf19.6062.3	CCAATCGACTTTAGTGCCACA	CGCAACTGAAAGTTTCTGATCTT
MMD1	TGTTGCGCAACAAATTGAGT	TGGTGCCTGTGAGGTTCTAA
MMD1(I2)	AAGCAATTAAAGTCAACGGTACG	AACGTTCTGAATGGCTTGCT
Orf19.6220.4	GGCTCAACGTGTACCTACA	CAGCCTTCTTGACGTGTTGA
Orf19.6264.3	TTTAGCAGTATGTTTCAAGAACTCA	CCATAAATACCCCAAGAACG
Orf19.6462	TCTGTTGCAATGGTATTGGTG	AATTACCTTTTGCCATTGTCAT
Orf19.6563.1	ACCCAGATTATTACTAGATCCTCA	TTTTCACGGAAACATCTAGCA
Orf19.6718	TCACTCGAACAGTTTTTCAAGACA	TGCTGCTGATCAATTGGTTT
FMA1	CGTCTCAATTGTAAGTGGTGCT	TCCAACACACCAGCATTAGC
ATP1	TGTTGTGAGCTCGTCTGTT	GGCAGAGGCAATCTTTGAG
ATP1(I2)	CGTCGGGTATGTATCAATTCAA	CAGAACCGAACAACAACAACC
ASC1	TCTGCTTCTTGGGATAAGACTG	TGTAACCGGTGTGACCAATG
GTT11	GACACCAAAATTATTGTTCACTGG	CGATTGGCAGGAGTCAAGAT
RAD6	TGTCACACCTGCTAGAAGAAGAC	CACCTTGGTGGATCTTGTTC
RAD6(I2)	ATCAGCATCTCCATTACCAGA	AAACATAATTCTCCACTGGCATA
PRE8	TGACCACATTTTCAACAAGG	GGGAATTGGCTTTACGTTCA
ECM15	ACATTGCATTGTTTGGCTGA	GTCCAGCAATTGTGGTTCCT
FDH3	TGTTGAAAGGTAAGTTGAAACAA	GACCAATGGCTTAGCAGCTT
INO4	CAAAGTCGTCCGTTTTATCAGA	TCTCTGCTCTCAACTCGTCAA
Orf19.909.1	GGCGCTTATACGATTTTGG	TTGGACAATAGGAGCAACTGG
Orf19.909.1(I2)	CAGGTATTGCCATTCTAAGTTTT	TTTCTCTTTTGGTCAAATCACT
Orf19.913.2	TGTTCCAACGTCTATGCTGA	CCGTTAGACTTCCATCATCCA

Supplemental Table 2

	#	
Transcripts		46937
MM		40732
Ca		6205
	#	
Transcripts with at least one hit in one sample		
MM		32846
Ca		6176
	#	
Number of Transcripts with adjusted p value < 0.1		
TP 30 vs TP 0		620
out of these		
MM		72
Ca		548
TP 60 vs TP 0		197
out of these		
MM		82
Ca		115
TP 90 vs TP 0		220
out of these		
MM		109
Ca		111
TP 120 vs TP 0		480
out of these		
MM		167
Ca		313
	#	
Number of Transcripts significant at		
30_0 AND 60_0		146
30_0 AND 90_0		94
30_0 AND 120_0		164
60_0 AND 90_0		115
60_0 AND 120_0		125
60_90 AND 120_0		123
all comparisons		71

Supplemental Table 3

Repeat	Time	hits GDS	hits CDS	% Difference
1	0	119657	119532	0.01
	30	610529	609908	0.01
	60	211591	211625	0.02
	90	450532	450475	0.01
	120	785477	785371	0.01
2	0	3018850	3009552	0.31
	30	2095337	2094621	0.03
	60	181437	1814001	0.02
	90	968138	967334	0.08
	120	930053	930171	0.01
3	0	1713732	1710442	0.19
	30	1326965	1326476	0.04
	60	1727113	1725144	0.11
	90	1149141	1149361	0.02
	120	352407	352460	0.02

Supplemental Table 4

Donor site																
unspliced																
positions	-6	-5	-4	-3	-2	-1	G T		2	3	4	5	6	7	8	9
A	112	117	137	149	124	104	64	90	196	17	4	137	104	66	76	141
G	58	45	50	51	65	125	189	16	55	156	174	3	16	138	32	31
T	61	67	56	44	40	31	21	15	16	11	1	2	26	22	32	28
C	90	92	78	77	92	61	47	200	54	137	142	179	175	95	181	121
spliced																
positions	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9
A	29	17*	29	34	38	26	21	28	50	4	0	35	25	14	20	33
G	12	11	11	16	13	24	42	5	12	44	44	0	5	36	10	10
T	17	18	23	19	12	12	7	6	4	4	0	0	7	7	5	8
C	23	35	18	12	18	19	11	42	15	29	37	46	44	24	46	30
alternatively spliced																
positions	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9
A	4	0	3	4	4	2	3	3	5	0	0	5	1	3	1	5
G	1	0	0	0	2	2	5	1	4	6	4	0	1	6	0	2
T	0	4	5	5	3	2	2	0	0	1	0	0	0	0	0	0
C	5	6	2	1	1	4	0	6	1	3	6	5	8	1	9	3

*Position -5 donor p value 0.028 (spliced vs unspliced, Chi squared test)

Acceptor site																	
unspliced																	
positions	-6	-5	-4	-3	-2	-1	A		G	2	3	4	5	6	7	8	9
A	101	85	68	116	82	192	60	85	171	71	110	93	95	113	112	115	
G	21	24	23	22	14	34	173	44	30	148	78	67	62	51	48	70	
T	45	38	37	39	34	14	29	49	57	39	46	63	69	66	52	48	
C	154	174	193	144	191	81	59	143	63	63	87	98	95	91	109	88	
spliced																	
positions	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	
A	23	23	15	25	22	42	20	18	39	21	25	22	32*	23	19	24	
G	5	8	4	4	2	6	40	9	3	42	22	18	6	14	16	16	
T	10	10	12	10	8	4	8	10	16	4	8	19	21	17	20	16	
C	43	40	50	42	49	29	13	44	23	14	26	22	22	27	26	25	
alternatively spliced																	
positions	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	
A	3	3	3	5	1	5	2	1	6	1	1	1	8	2	1	3	
G	1	1	0	0	0	0	4	2	0	7	2	5	1	0	3	1	
T	2	0	2	2	2	0	1	0	2	0	1	2	0	2	4	2	
C	4	6	5	3	7	5	3	7	2	2	6	2	1	6	2	4	

*Position 6 acceptor p value 0.043 (spliced vs unspliced, Chi squared test)

4. Unpublished Results

4.1 Establishing simultaneous RNA-seq with fungi

One clear advantage to using a sequencing approach is that it is not organism specific, which theoretically allows for the investigation of multiple species in parallel. Recently, Mimivirus (100) and Vaccinia virus (101) have been sequenced during host-cell invasion showing that the sequencing of mixed species samples is technically feasible. The metabolic dependence of viruses on the host cell however, hinders the identification of transcriptional changes that are the cause of, or the result of, host cell invasion. The independent life cycles of fungal, bacterial, and protozoan pathogens in principle, should allow for the separation of host-pathogen crosstalk and therefore is an attractive method to establish for pathogenesis research.

In order to adapt simultaneous RNA-seq for fungal pathogens, we first determined if it would be possible to separate the sequencing reads with a high confidence for our species of interest. To investigate the fidelity of our read assignment to the *C. albicans* or *M. musculus* genome, we used the sequence identity of all orthogous genes between the two species as measure of the potential for reads to be assigned to the wrong genome (Figure 10A). We found that the majority of the transcripts maintain an identity around 20-40%, with less than 5% having an identity of over 90%. These data shown that the vast majority of sequence reads would not be falsely assigned to one or both species as a positive match because they are distinct at the sequence level. Moreover, because of the evolution distance between host and pathogen pairs in general, this method should also be applicable to other species so long as they have distinct genomic sequences.

The rigid cell wall of *C. albicans* hinders RNA extraction using most mammalian cell lysis buffers alone. In order to extract RNA from both *C. albicans* and mammalian cells, we adapted an RNA extraction protocol for mammalian cells and added glass beads in order to break the *C. albicans* cell wall (92). To ensure that RNA was efficiently extracted from both species at a high quality, all total RNA samples were measured for their integrity on a Bioanalyzer 2100 (Figure 10B). The main bands present on the gel correspond to the large and small ribosomal subunits due to the high proportion of rRNA in total RNA samples. Due to the size difference in the large ribosomal subunit between yeast and mammals, it is possible to distinguish them on an

RNA gel using mixed samples. Additionally, An RNA integrity number (RIN) is given on a scale of 1-10 for each sample, where 10 represents the highest measurable sample integrity. Only samples with a RIN number of 9 or higher were used for sequencing. Using these conditions, we were able to verify that the samples collected were not only of high quality, but also were of a similar quality for both species.

After sequencing, several factors needed to be taken into consideration during analysis to decrease possible biases between the samples at the species level, the sample time point, and the sequencing repeat. First, to determine if there could be a bias in the read assignment due to significant differences in the average transcript length between *C. albicans* and *M. musculus*, we plotted the gene density versus the transcript length for both species (Figure 10C). Since the two curves strongly overlap for their gene density distribution, we did not need to take into additional consideration a potential bias in overall transcript length between the two species during normalization.

Next, we looked for a way to assign reads that would that would provide the most random read distribution compared to the transcript length to decrease a read assignment bias within the total data set. When assigning reads along to the transcript length, we observed a clear length bias in which longer transcripts had higher read counts using both the mean and median read values (data not shown). We overcame this bias by using the median value for the base coverage, as this value provided the most randomize read assignment (Figure 10D).

Finally, we wanted to determine the best way to normalize the total data between samples and repeats. High-dimensionality data such as microarray or RNA-seq samples are a significant hurdle for data analysis because a much higher number of variables are measured compared to the number of measured samples. One method to visualize this dimensionality and clearly identify batch effects is by using principle component analysis (PCA) (102). Briefly, PCA identifies data trends within samples, called principal components, so that very large data sets can be graphically represented using individual data points. We performed PCA on all RNA-seq samples and repeats before and after normalization (Figure 10E). We found a clear batch effect in the raw data between the samples that were sequenced together (repeats 2 and 3) compared to repeat 1 which was sequenced on a different day (Figure 10F). These effects were then taken into consideration during normalization and were able to remove the batch effect (Figure 10F, normalized) so the data could be used for

downstream differential gene expression analysis (see Tierney *et al.*, submitted, this dissertation: Paper 3 methods section for a complete description of the normalization procedure).

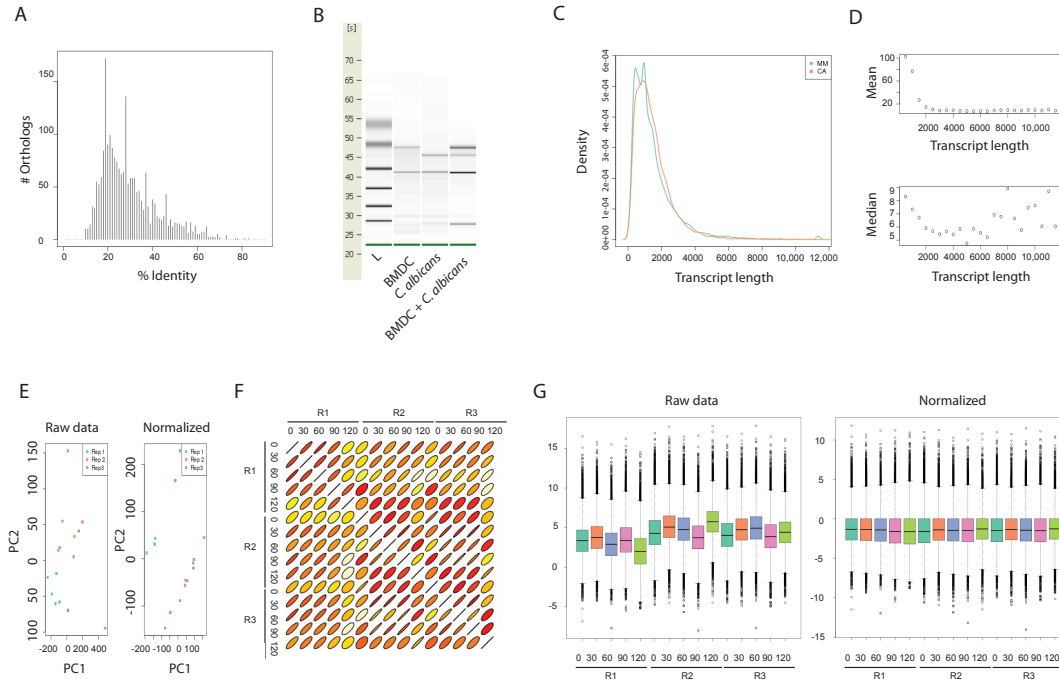


Figure 10. Simultaneous RNA-seq sequence normalization

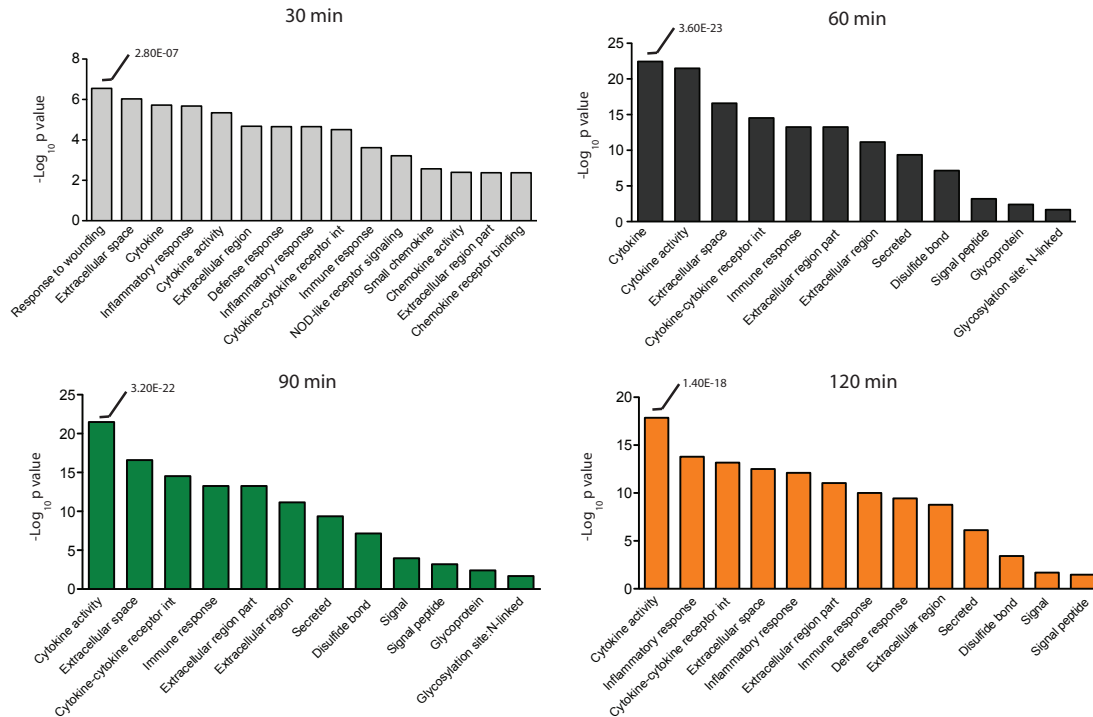
(A) Representative Bioanalyzer gel of total RNA extracted from BMDcs, *C. albicans*, and co-culture samples. (B) Density plot of the length distribution of *M. musculus* and transcripts. (C) The percent identity of *M. musculus* and transcripts for all orthologous genes between the two species. (D) Typical base coverage for sequence reads for the combined data set using the mean and median transcript coverage. (E) A principle component (PC) analysis of the raw and normalized total data set over the three sequencing repeats. (F) A plot of the pairwise correlation of the raw $\log_2$ transformed data of all samples from all time points against one another. A line corresponds to a correlation of one. The size of the circle corresponds to a decrease in correlation, where the larger the circle is, the worse the correlation between the samples. The intensity of the overall read count is shown as a gradient from white to red, from low to high intensities respectively. (G) Boxplot of the raw and quantile-normalized $\log_2$ transformed data. The sum of reads per quartile for total sequencing reads from *M. musculus* and is shown for each sample.

4.2 Putative genes are among the highest enriched during infection

From the normalized data set, we were left with a list of approximately 200 and 600 differentially expressed genes for *C. albicans* and *M. musculus* respectively at the

investigated time points. With these data, we first used GO term enrichment to identify the overlap of the pathways induced in our system in comparison to previously published work. For *C. albicans*, we used the CGD Gene Ontology Term Finder (<http://www.candidagenome.org/cgi-bin/GO/goTermFinder>) from the Candida genome database (103) for each time point. We found that biofilm formation (p value 2.8×10^{-4}) and interspecies interactions (p value 2.8×10^{-4}) were the most highly induced categories, and translation (p value 2.63×10^{-36}) was the most repressed category at all time points investigated (data not shown). These data correspond well to previous *in vitro* microarray studies with *C. albicans* invading innate immune cells including macrophages (104), monocytes and neutrophils (105) showing that the our data set, as a whole, is in concordance with the state of knowledge. However, when we looked for which genes among those differentially expressed underwent the greatest transcriptional change, we found a high percentage of putative genes and those with unknown function, especially at later time points including: *HSP21* (120min, fold change + 258.2), orf19.1363 (120min, fold change + 142.3) and orf19.3938 (120min, fold change + 53) whereas the widely recognized and highly induced adhesion *ALS3* (106, 107) was induced to a much lower extent (60 min, fold change +25.82). Interestingly, *HSP21* was recently shown to be an important component in the adaptation response stress (108), however its mode of action is still unknown. To our knowledge, these putative genes were not identified in previous microarray or RNA-seq studies with. Based on the transcriptional regulation of the putative genes, we believe that some of them may represent unidentified virulence components of *C. albicans* that are targets for future work.

We then performed GO term enrichment on the mouse data set (Figure 11). We found again that the pathways induced corresponded well to those previously identified for a variety of pathogens infecting dendritic cells (109, 110). Unexpectedly, when we looked for which genes displayed the highest transcriptional variation in the data set, we found that the predicted mouse lincRNA Gm16869 is among the highest differentially expressed genes after 30 min co-culture with *C. albicans*, with an over 20-fold increase relative to the zero time point. LincRNAs are long non-coding RNA species which have been shown to play a major role in the functional developmental of mammals (111). In additional, lincRNAs have been shown to be responsive to LPS stimulation in murine macrophages (112). Our data suggests that lincRNAs may play a



role in the immune response to *C. albicans* challenge, however further work would be needed to verify this hypothesis.

Figure 11. *M. musculus* GO term enrichment during an infection time course with *C. albicans*

Statistically significant GO terms enriched for differentially expressed *M. musculus* genes over the time course of the infection using the Database for Annotation, Visualization, and Integrated Discovery (DAVID), version 6.7.

4.3 A putative *C. albicans* splicing component is regulated during infection

In our RNA-seq analysis, we identified splicing and alternatively splicing as a novel adaptive response to *in vitro* infection (Tierney *et al.*, submitted). In order to see specifically what splicing machinery was involved in this process, we reanalyzed the RNA-seq data set to see if any direct components of the *C. albicans* spliceosome, or any of its accessory proteins, were additionally regulated during the infection process which may provide mechanistic clues for the splicing regulation of this gene set. Interestingly, we found that orf19.4964 increased over the infection time course, and peaked at 90mins after the onset of infection (Figure 12A). Although its molecular function in *C. albicans* is unknown, orf19.4964 is homologous to the *S. cerevisiae* gene Bud13, which is linked to splicing and mRNA-retention (113). In *S. cerevisiae*, Bud13 is one of 3 members of the Retention and Splicing (RES) complex (114) and the

disruption of some of the complex's components has been linked with increased mRNA retention of some transcripts (115).

To see if the ortholog orf19.4964 also played a role in splicing, we attempted to generate a knockout of the gene. We were unable to generate a homozygous knockout of orf19.4964, suggesting that it may be an essential gene in. To see if orf19.4964 had a haploinsufficient-splicing defect, we looked for an increase in mRNA retention for a subset of spliced, which were consistently detected as spliced during the infection process with the heterozygous orf19.4964 knockout (see Tierney *et al.*, submitted for gene list). Using intron-spanning primers for a subset of intron containing genes, we did not detect any consistently significant changes in the splicing pattern for those specific genes and therefore could not link the lack of orf19.4964 with a splicing response (data not shown). However, as only a small number of intron containing genes were investigated using the heterozygous knockout, it would be necessary to use a genome-wide approach to determine if orf19.4964 does not play a role in splicing, either directly or indirectly, of any intron containing gene in using a conditional knockout. Notably, we find that the heterozygous orf19.4964 knockout has a significantly decreased survival against neutrophil challenge (Figure 12B), suggesting that orf19.4964 may represent a new virulence gene in, however further work would be needed to uncover how the partial loss of orf19.4964 affects survival, especially in the context of host challenge.

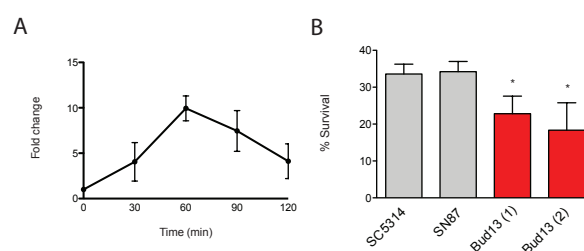


Figure 12. Bud13 (orf19.4964) is a putative splicing factor in *C. albicans*

(A) Sequencing verification of the orf19.4964 fold-change with qPCR over the time course of infection using the cDNA sent for sequencing using all three repeats. **(B)** Neutrophil killing assay for orf19.4964 heterozygous knockout strains (two independently transformed strains are shown) compared to the background strain for the knockout SN87, and the clinical isolate SC5314. A p value < 0.05 was considered significant.

4.4 *C. pintolopesii* as a model to investigate *Candida* pathogenesis

C. pintolopesii (also known as *Torulopsis pintolopesii* and *Arxiozyma terruris*) is the only indigenous *Candida* spp. in the GI tract of mice (116, 117). Unlike current *in vivo* models of *Candida* infection, which often require the use of immunocompromised mice or invasive techniques to maintain infection, the colonization of mice by *C. pintolopesii* does not need to be induced, as it is naturally highly abundant in the GI tract. For this reason, *C. pintolopesii* represents a unique model to study the transition of a natural commensal to the pathogenic state.

In order to see which *Candida* spp. *C. pintolopesii* most resembled, we first checked to see whether it was able to change its morphology to a hyphal or pseudo-hyphal state similar to that of. Using several hyphal inducing conditions, no true hyphal cells were detected after 3 (Figure 13A) or 6 hours (data not shown). Some partially elongated cells, resembling pseudo-hyphae were detected among the population in response to Lee's medium, however they represented, in general, the minority of the cells.

In collaboration with Toni Gabaldon (CRG, Center for Genomic Regulation), we performed the annotation of the genomic assembly for *C. pintolopesii* and investigated its gene homology to other fungi and yeast (Figure 13B). shows a similar homology profile for other well studied *Candida* spp., including *C. glabrata* (data not shown). Based on these data, should be more comparable to *C. glabrata* infection for *in vivo* colonization studies.

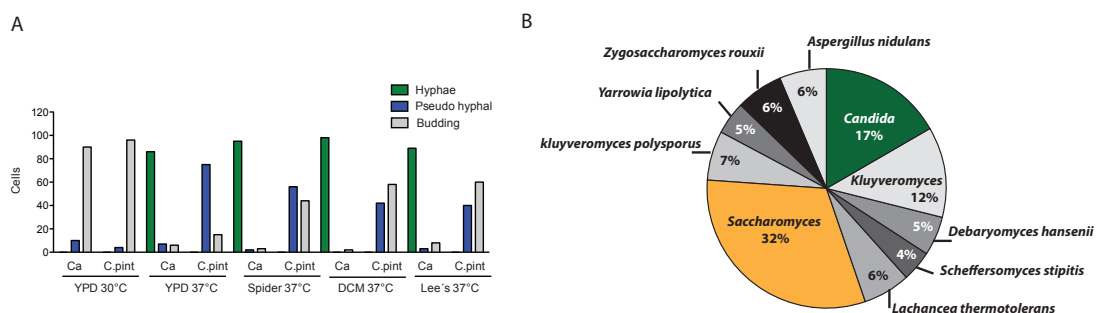


Figure 13. *C. pintolopesii* is homologous to *C. glabrata*

(A) Morphology of *C. pintolopesii* (*C. pint*) and *C. albicans* (*Ca*) cells under different hyphal inducing media conditions after 3 hours incubation. **(B)** Gene homology prediction of *C. pintolopesii* to related fungi.

In order to investigate to commensal to pathogen transition, we attempted to establish the GI tract colonization model of Koh *et al.* (118) using BL6 mice. We were able to deplete the neutrophils over 50% and maintain the depletion for at least 4 days using at least 100µg of the neutrophil depletion antibody RB6-8C5. We obtained the best efficiency using 200µg from Biolegend (Figure 14A). We also verified GI flora depletion using several treatments of antibiotics and antifungals in the drinking water (Figure 14B). However we were not able to detect significant dissemination to the major body organs including kidney, liver, and brain and had high variability between mice (data not shown). We found that and *C. glabrata* remained in the gut at high levels of colonization even with 5% DSS treatment (Figure 14C) showing that for BL6 mice, in our hands, the model was not efficient to study disseminated infection. This could be a mouse strain-specific effect as the model was previously applied to C3H/HeN mice in the original publication. Therefore, it is more applicable as a colonization model for BL6 mice. Notably, after *Candida* feeding was halted, the colonization levels among *Candida spp.* varied dramatically. The levels of *C. albicans* and in stool peaked in colonization on week two, and decreased for the remainder of the experiment. However, actually proliferated in the gut, reaching levels of over 6.0×10^5 CFUs/g stool after 6 weeks (Figure 14D), emphasizing its applicability for a long-term colonization model.

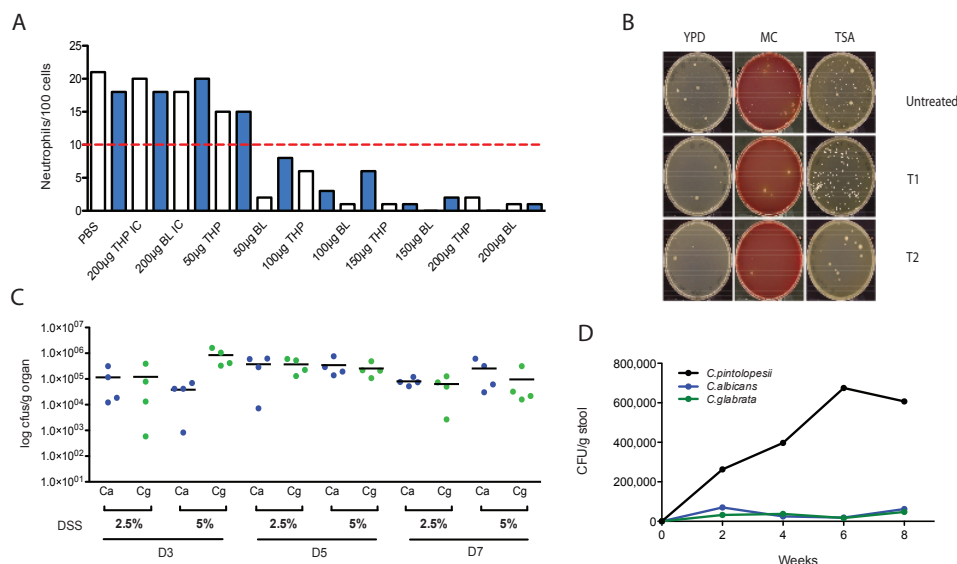


Figure 14. GI tract colonization model for *Candida spp.*

(A) Neutrophil depletion efficiency 2 days (white bar) and 4 days (blue bar) after depletion antibody RB6-8C5 injection. The red dotted line corresponds to a 50% reduction compared to the PBS injected control

mice (IC = isotype control, BL = BioLegend, THP = THP Medical Products). **(B)** Validation of the GI tract depletion with homogenized stool pellets plated before the experiment, after the first treatment (T1, 2mg/mL streptomycin, 1mg/mL penicillin G and 0.250mg/mL fluconazole, 3 days) and after the 2nd treatment (T2, 2mg/mL streptomycin, 1mg/mL penicillin G) on YPD, Mac Conkey Agar (MC), or Tryptic Soy Agar (TSA) plates.**(C)** GI tract CFUs after flora depletion and DDS treatment for *C. albicans* (Ca) or *C. glabrata* (Cg), 3, 5, and 7 days after the onset of *Candida* feeding. Each dot corresponds to one mouse. **(D)** Persistence of different *Candida spp.* in stool samples after *Candida* feeding was stopped.

5. Additional Materials and Methods

Ethics statement – All animal experiments were discussed and approved through the University of Veterinary Medicine Vienna's institutional ethics committee and carried out in accordance with animal experimentation protocols approved by the Austrian law (GZ 680 250/67-BrGt/2003, GZ-BMWF-68.250/0233-II/10b/2009).

5.1 Fungal strains and growth conditions

The *C. albicans* SC5314 (119) and *C. glabrata* ATCC2001 (120) clinical isolates were used in this study. SN87 (*ura3::imm434::URA3/ura3::imm434 iro1::IRO1/iro1::imm434 his1::hisG/his1::hisG leu2/leu2*) derived from the SC5314 (121) strain was used for all knockout experiments. *Candida pintolopesii* var. *pintolopesii* (NCYC 1562) was obtained from the National Collection of Yeast Cultures (NCYC, UK). All strains were routinely grown on YPD plates (1% yeast extract, 2% peptone, 2% glucose, 2% agar) and in standard rich media YPD (1% yeast extract, 2% peptone, 2% glucose) for liquid culture at 30°C. Fungal cells were collected in the logarithmic growth phase by a brief centrifugation, washed in sterile PBS, and diluted for all experiments as described.

5.2 *Candida pintolopesii* de novo sequencing

C. pintolopesii cells were grown overnight in 5mL YPD. Cells were collected by brief centrifugation and resuspended in 250µL lysis buffer (2% Triton x-100, 1% SDS, 100mM NaCl, 10mM Tris-HCL pH8, 1mM EDTA). Cells were homogenized in a Fast Prep-24 (MP Biomedicals, Santa Ana, CA) with acid washed glass beads (Sigma Aldrich, St. Louis, MO) and 250µL Phenol chloroform (PCI) (Sigma Aldrich) at level 6 for 30s. Samples were then centrifuged at 12,000 RPM for 5 min, the aqueous phase transferred to a tube containing 1mL ice cold isopropanol, and precipitated at -80°C for 30 min. After precipitation, samples were centrifuged at 12,000 RPM for 30 min and resuspended in 400µL 2x TE buffer containing 10mg/mL RNase (Sigma Aldrich). Samples were incubated for 1hr at 37°C. DNA was re-extracted with PCI as described above and resuspended in 100µL sterile TE for sequencing. The resulting DNA was sequenced using Roche/454 (mate-pair reads to a depth of 5x, insert size 8kb) and the

Illumina sequencing (paired-end reads, insert size 350bp sequenced in mode 2 x 100bp) platforms at GATC (Konstanz, Germany) with an estimated coverage of 20x. The genome annotation was performed by Toni Gabaldon at the Center for Genomic Regulation in Barcelona, Spain.

5.3 *Candida* hyphal induction

Hyphae were induced in 1×10^7 *Candida* cells diluted in 5mL of the following media: YPD at 30°C or 37°C, Spider media (122), or Lee's media (123) for 3hrs with shaking. Cells for each strain and condition were evaluated as being in one of the following stages: budding, hyphal or pseudo-hyphal by counting 100 cells on an Olympus Cell-R Live Imaging Unit (Olympus, Essex, UK).

5.4 Strain construction

The heterozygous gene deletion of orf19.4964 was created in the SN87 strain using the C.d.HIS1 marker cassette (124) and fusion PCR strategy as previously described (121) using the following sequence primers: (sequences 5'-3') 55_Ca4964 TTGGTGGTCCTATAATGAGTG, 53_Ca4964 cacggcgcgcctagcagcggCATTGTA-TGTTATGGTGTAAATGT, 35_Ca4964 gtcagcgccgcacccctgcTGAGAGGGGTTG-ATATATTGTT, and 33_Ca4964 GATACCTCGATACTGATTGTG.

5.5 Mice

Wild type C57/BL6 mice purchased from Charles Rivers were used in all experiments using gender and age matched 7-9 week old mice.

5.6 Neutrophil killing assay

Neutrophils were isolated from bone marrow collected from the femurs and tibias of mice using a percoll (Sigma Aldrich) gradient. Briefly, percoll solutions of different densities (78%, 69%, and 52%) were layered in a 15mL falcon using a needle. Red blood cells were lysed from the bone marrow using red blood cell lysis buffer (8.3g/L NH₄Cl in 0.01 M Tris-HCL) for 1 min. The lysis reaction was halted using 20mL RPMI

(PAA, Pasching, Austria). Samples were briefly centrifuged and resuspended in 2mL 10mM HEPES buffer (115mM NaCl, 1.2mM CaCl₂, 1.2mM MgCl₂, 2.4mM K₂HPO₄, 20mM HEPES, pH 7.4) and layered on the percoll gradient. Samples were centrifuged at 1500 x g at 4°C for 30 min. After centrifugation, neutrophils were isolated from between the 78% and 69% percoll layers, diluted in 5mL RPMI, and 1 x 10⁵ cells were distributed per well in 96 well format in sterile cell culture dishes. Fungal cells were added (1:10) in RPMI and incubated for 1hr at 37°C. After incubation, neutrophils were lysed in the well with 50µL of 4% Triton-X 100 in PBS. Cell suspensions were collected and the well washed with 400µL sterile PBS 3 times to collect any remaining cells. Cell suspensions were diluted 1:10 and plated on YPD plates. Colonies were counted after 3 days incubation at 30°C. *Candida* survival was calculated from the number of cells collected from the neutrophil treated versus untreated wells.

5.7 Gastrointestinal tract colonization with *Candida* species

The GI tract colonization model was adapted from Koh *et al.* (118). Briefly, mice were depleted of their GI flora using an antibiotic regime in their drinking water including 2mg/mL streptomycin (Sigma Aldrich), 1mg/mL penicillin G (Sigma Aldrich) and 0.250mg/mL fluconazole (Sigma Aldrich) for 3 days. GI tract depletion was confirmed by plating homogenized stool pellets diluted 1:1000 in 1% protease peptone (Sigma Aldrich). Homogenates were plated on YPD plates, Mac Conkey Agar No 1 (Fluka Analytical, Buchs, Switzerland) (2% glucose, 2% agar), and Tryptic Soy Agar Vegitone (Fluka Analytical) (2% glucose, 2% agar) in triplicate and left to grow for 2 days at 30°C. Plates were then visually inspected for reduced colony numbers after each treatment. After depletion, mice maintained the same concentrations of streptomycin and penicillin G in the drinking water for the remainder of the treatment and additionally including *Candida* 1 x 10⁷ CFUs/mL. Colonization was confirmed by the presence of *Candida* colonies in the stool sample plates as described above. Neutrophils from the mice were depleted after 4 days of antibiotic treatment by injecting interperitoneal (IP) 50-200µg of the anti-granulocyte receptor-1 (Gr-1) mAb, RB6-8C5 from either THP Medical (Vienna, Austria) or Biolegend (Franklin Lakes, New Jersey). Neutrophil depletion was confirmed by GIMSA stain (Sigma Aldrich) following manufactures instructions. We used 1 drop of blood from treated mice and 100 cells were counted from each smear to estimate the neutrophil depletion efficiency in the blood by the

antibody on the Olympus Cell-R Live Imaging Unit (Olympus). PBS was injected for all untreated mice. Dissemination of *Candida* to the organs was induced one week after the start of antibiotic treatment by including 2.5% or 5% Dextran sulfate sodium (DSS) (Sigma Aldrich) with the antibiotic treatment in the drinking water. To determine dissemination efficiency, the liver, spleen, GI tract, brain, and kidney were removed aseptically, rinsed in sterile PBS, weighed, and homogenized in 1.5mL tissue lysis buffer (200 mM NaCl, 5mM EDTA, 10mM Tris, 10% glycerol) on ice. Samples were homogenized using an Ika T10 basic Ultra-Turrax homogenizer (Ika, Staufen, Germany). Colony forming units (CFUs) were determined for each organ by plating serial dilutions of homogenates on YPD plates. Colonies were counted after 2 days of incubation at 30°C and fungal burden is calculated for CFUs/gram tissue.

5.8 qPCR

RNA sample preparation, reverse transcription, and real-time PCR were performed as previously described (125) using the following primers: *RIP1*, forward 5'-TGCTGACAGAGTCAAGA-AACC-3' reverse 5'-GAACCAACCACCGAAATCAC-3', Orf19.4964 forward 5'-TATCTCCATTATCTATTCGTTTG-3, reverse 5'-CTCCCAAACCATAGGTGTAC-3'. Results were calculated using the $\Delta\Delta ct$ method using the expression of the *C. albicans* housekeeping gene, *RIP1*, versus orf19.4964 in treated versus untreated conditions.

6. General Discussion

6.1 Major results presented in this thesis

This study aimed to investigate host-pathogen cross talk between *Candida spp.* and their host, by resolving transcriptional changes of both *C. albicans* and *M. musculus* under infection conditions using simultaneous RNA-seq of the host and pathogen transcriptome. We decipher novel adaptive responses of *C. albicans* to the host environment, including previously unidentified genes involved in host-pathogen cross talk on both the host and pathogen side using network inference. Additionally, we identify post-transcriptional modifications, including infection specific alternative splicing as a new adaptive response of *C. albicans* to counteract host challenge. These results are presented in two major publications and one review.

In summary, the work presented in this thesis established the following methods:

- ***In silico***: Network Inference as robust strategy to model host-pathogen interactions using NetGenerator
- ***In vitro***: Simultaneous RNA-seq as a method to study pathogenesis using *C. albicans* and BMDCs in a genome-wide, time-resolved fashion
- ***In vivo***: A long-term GI tract colonization model for *Candida spp.* to investigate the commensal to pathogen transition

Using these methods, this thesis discovered the following major points:

- Hap3 is a fungal hub in *C. albicans* infections, which interacts with Ptx3 and Mta2 in *M. musculus*, where the loss of Hap3 alters the immune response to the fungus.
- *C. albicans* undergoes infection specific splicing and alternative splicing during discrete infection stages with murine dendritic cells, including novel alternative splice junctions.
- The natural *Candida* mouse commensal *C. pintolopesii* (genome annotated as part of this thesis) is a reasonable organism to investigate the commensal state for *Candida spp.* during *in vivo* infection.

6.1.1 Using simultaneous RNA-seq to investigate host-pathogen interactions

Our work characterized the complex niche adaptation, in molecular terms, of *C. albicans* to its mammalian host *in vitro* using simultaneous RNA-seq of the host-pathogen transcriptomes over a time course of infection. By using deep sequencing instead of other genome-wide technologies such as microarray, we were able to overcome several significant hurdles in our data analysis. Specifically, we had the flexibility of investigating multiple species from the same sample, which eliminated platform dependent effects and other biases as a result of additional sample handling.

RNA-seq as a technology is still being developed as a standard laboratory technique and standards for sample preparation, documentation, pipelines for analysis, are still being established (126-128). Especially in regards to the experimental design, it has been clearly argued that standard design principles of are grounded in low throughput data, and the application of such approaches to high-throughput technologies can create significant problems in data interpretation (129). Therefore, it is essential in RNA-seq experiments for particular attention to be paid to the statistical analysis, the number of biological replicates, as well as comparatively low-throughout biological validation of the data set. For our simultaneous RNA-seq experiment, it was necessary for us to verify the contribution of potential biases specifically resulting from our mixed data set, including transcript length differences between *M. musculus* and *C. albicans*, the possibility of sequence overlap, and the normalization technique used for the biological repeats, the results of which we were able to take into consideration for downstream analysis. We found that evolution distance between host and pathogens in general, should allow for the applicability of simultaneous RNA-seq technology to a wide-variety of pathogenic scenarios both *in vitro* and *in vivo*. However, based on the need for the sequence conservation to be so great between those investigated, the method would not be suitable for the majority of meta-genomics or super-infection studies, because of the increased probability that reads would be miss assigned to the incorrect species and is best suited for co-culture experiments.

Based on our data analysis, the resolution of the *C. albicans* genome, in general, was much higher than that of *M. musculus*, in terms of the number of differentially expressed genes detected and pathways enriched, throughout the time course of the infection. The mouse genome, based on length alone, is 189.7 times larger than that of *C. albicans* using the estimates from the CGD and the mouse

genome database (MGI) (<http://www.informatics.jax.org/>) (130). Although these numbers do not necessarily reflex the transcriptional variation for either genome at the time that the sequencing was performed, they nevertheless suggest that the read depth will be greater for the smaller genome. A recent study has designed a model to estimate the read depth required in order to reasonably detect low and high abundance transcripts with over 95% efficiency (131), and suggests that pilot studies in general should have a minimum of 1 million reads per sample. As our reads were divided among multiple organisms, the coverage between the two was not equal. This may explain why the resolution of the *C. albicans* genome, in terms of novel adaptive responses and mechanistic insight over our time course was much greater than that of the host response. The *M. musculus* data may have benefited overall with a greater sequence depth. These data emphasize the necessity for the biological validation of the observed RNA-seq phenomena especially in transcripts with low coverage.

6.1.2 Predicting host-pathogen interactions using network inference

Once we were able to validate the use of simultaneous RNA-seq as an appropriate method, we were interested in using our data to model processes related to the pathogenesis of *C. albicans*, specifically in regards to host-pathogen cross talk. Since the majority of players involved in host-pathogen interactions were identified using low-throughput approaches, we wanted to try to apply an *in silico* modeling strategy in order to increase the throughput of identifying putative components involved, using our genome-wide expression data as a test set.

To this end, we used network inference to model a subset of genes taken from both the host and pathogen side and were able to present the first use of inferred network analysis to host-pathogen interaction (92). A sub-network from this analysis was experimentally verified, validating the use of network inference for the analysis of multiple species data sets.

6.1.3 Novel spliceosome proteins may account for infection specific splicing

During our gene selection for the network model, we used a clustering approach to identify co-regulated genes and categories to be included in our downstream analysis. We discovered several enriched categories that were not previously

associated with host-pathogen interactions including a high enrichment for ribosomal genes and post-transcriptional processing mechanisms. The importance of quick transcriptional responses of *C. albicans* to counteract host challenge has been investigated in a variety of *in vitro* models (104, 105). Based on the high number of intron containing genes identified as differentially expressed for *C. albicans* in our simultaneous RNA-seq data set, we asked whether there may be a contribution of post-transcriptional processing mechanisms, such as splicing, in the regulation in the transcriptional responses of *C. albicans* upon host challenge; a field which had been almost entirely unexplored for this species.

Among eukaryotes, splicing is a ubiquitous form of gene regulation throughout development and in response to stress within its environment (132). Compared to higher eukaryotes with an average of 6-8 introns per gene, yeast species, such as *S. cerevisiae* as well as *C. albicans* have a significantly lower intron density (133). This genomic simplicity coupled with the high conservation of spliceosomal components among eukaryotes, has made them historically the model organism of choice for studying splicing (134). Interestingly, although the overall intron density is low, over 70% of the pool of actively translated RNA in yeast originates from genes that are intron containing (135), emphasizing their overall importance. Similarly, the proportion of intron containing genes in our data set was much higher than would be expected based on the intron density for alone.

After further investigations, we found that infection-specific splicing and alternative splicing were a hitherto unrecognized time-dependent response mechanism of during pathogenesis (Tierney *et al.*, submitted). Importantly, we found that splicing was limited to a subset of intron containing genes. This made us interested to investigate what could be the selective mechanism for this specific gene set.

Pre-mRNA splicing is catalyzed by the spliceosome, a ribonuclearprotein complex that excises introns through a highly orchestrated process involving RNA-RNA, RNA-protein, and protein-protein interactions (136). The spliceosome is comprised of five small nuclear RNAs, U1, U2, U4, U5 and U6, and over 100 protein factors involved in pre-mRNA splicing. It has been shown to recognize a number of consensus motifs on the pre-mRNA, including the donor, acceptor site, and the branch point (137) and the strength of these different signals combined is important for splice site selection, especially in cases of alternative splicing (138). Moreover, recent work has suggested that splicing signals and their corresponding splicing factors have co-

evolved (139), highlighting the interdependence of the components of the spliceosomal machinery and the introns spliced.

Essential as well as non-essential splicing factors have been shown to be important in eukaryotic development. To date, numerous splicing protein complexes have been linked with nuclear pre-mRNA retention and splicing. Importantly, errors in the expression, concentration, or composition, of splicing factors underline many human diseases (140). Due to the highly dynamic nature of the environment that *C. albicans* is exposed to during phagocytosis, it is possible that one or more of these environmental conditions either triggers or represses some of the splicing factors in *C. albicans*, which would influence splicing, and possibly the selection of the spliced gene set. As splicing in *C. albicans* has been poorly studied, it is not surprising that the identity of many of the splicing factors in are either unknown or based entirely on homology prediction to *S. cerevisiae*. Based on the data that are available, the regulation in our data set of a homolog of Bud13 in *C. albicans*, orf19.4964, a member of the RES complex in yeast (114), implies that the regulation of splicing components is taking place under infection conditions for *C. albicans*; this could explain for the selection of our intron containing gene set among the pool of intron containing genes. However, further work be needed specifically on the regulation of splicing factors themselves under different conditions in *C. albicans* to show that this is indeed the case.

6.1.4 Studying the commensal to pathogen transition using *C. pintolopesii*

In order to see if infection specific splicing and other adaptive strategies identified using the simultaneous RNA-seq approach also takes place under *in vivo* infection conditions for *Candida spp.*, we decided to adapt the GI tract colonization model of Koh *et al.*, (118) to be used for the mouse species from which the BMDCs cells were collected, namely C57/BL6 mice.

For *Candida spp.* the border between the commensal and pathogenic state remains poorly defined. Since this transition is a key component of the virulence of the species as well as a major topic for *Candida* research, and the fine-tuned immune response the fungus will be influenced by the route of infection (25), we decided to choose a model that most resembled a natural infection situation. Therefore, we chose to establish a mucosal inoculation model. Popular systemic models of *C. albicans* such

as intravenous or interperitoneal injection represent atypical route of infection for the fungus (141). Mucosal inoculation however mimics the common entry point for systemic dissemination of the fungus. On the other hand, in order to force *C. albicans* systemic dissemination in mucosal models, they still require antibiotic treatment and some form of immune suppression. These models therefore more closely mimic patients who have undergone a similar treatment strategy.

A yeast causing widespread infections in laboratory mice and the only natural *Candida* commensal in mice *C. pintolopesii* (116, 117, 142) is a useful model to study to commensal to pathogen transition. To date, only a limited number of studies have used *C. pintolopesii* in *in vivo* experiments using immuno-compromised mice (143). We found that *C. pintolopesii* quickly colonizes and proliferates within the GI tract of mice without neutropenia in BL6 mice for over 6 weeks. However, we only tested the model after first treating the mice with antibiotic and antifungal treatment in order to be certain that the colonizing fungus we detected was on the plates belonged to *C. pintolopesii* and not to other indigenous fungi or contaminating bacteria. It still needs to be seen how well the colonization work in completely untreated mice. Overall, *C. pintolopesii* represents a unique control species to model to study the transition of a natural commensal to the pathogenic state. Based on *C. pintolopesii* homology profile from the genome annotation, it should be more comparable to *C. glabrata* or *C. albicans* in the yeast state for *in vivo* colonization studies.

In total, we believe that the combination of the *in vivo* colonization method with simultaneous RNA-seq will provide an enhanced view of the colonization process of *C. albicans* in response to the host, its complex niche adaptation strategies, in order to devise strategies to limit dissemination in, or colonization of, mammalian hosts by *Candida* spp.

7. References

1. Sobel JD (1988) Pathogenesis and epidemiology of vulvovaginal candidiasis. *Ann N Y Acad Sci* 544:547-557 .
2. Pfaller MA & Diekema DJ (2007) Epidemiology of invasive candidiasis: a persistent public health problem. *Clin Microbiol Rev* 20(1):133-163 .
3. Bassetti M, Mikulska M, & Viscoli C (2010) Bench-to-bedside review: therapeutic management of invasive candidiasis in the intensive care unit. *Crit Care* 14(6):244 .
4. Casadevall A & Pirofski LA (2003) The damage-response framework of microbial pathogenesis. *Nat Rev Microbiol* 1(1):17-24 .
5. Morgan J, *et al.* (2005) Excess mortality, hospital stay, and cost due to candidemia: a case-control study using data from population-based candidemia surveillance. *Infect Control Hosp Epidemiol* 26(6):540-547 .
6. James TY, *et al.* (2006) Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 443(7113):818-822 .
7. Berman J & Sudbery PE (2002) *Candida Albicans*: a molecular revolution built on lessons from budding yeast. *Nat Rev Genet* 3(12):918-930 .
8. Santos MA & Tuite MF (1995) The CUG codon is decoded in vivo as serine and not leucine in *Candida albicans*. *Nucleic Acids Res* 23(9):1481-1486 .
9. Hube B (2004) From commensal to pathogen: stage- and tissue-specific gene expression of *Candida albicans*. *Curr Opin Microbiol* 7(4):336-341 .
10. Butler G, *et al.* (2009) Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature* 459(7247):657-662 .
11. Almeida RS, *et al.* (2008) the hyphal-associated adhesin and invasin Als3 of *Candida albicans* mediates iron acquisition from host ferritin. *PLoS Pathog* 4(11):e1000217 .
12. Lo HJ, *et al.* (1997) Nonfilamentous *C. albicans* mutants are avirulent. *Cell* 90(5):939-949 .
13. Braun BR, Kadosh D, & Johnson AD (2001) NRG1, a repressor of filamentous growth in *C.albicans*, is down-regulated during filament induction. *Embo J* 20(17):4753-4761 .

14. Brown AJ & Gow NA (1999) Regulatory networks controlling *Candida albicans* morphogenesis. *Trends Microbiol* 7(8):333-338 .
15. Gow NA, van de Veerdonk FL, Brown AJ, & Netea MG (2012) *Candida albicans* morphogenesis and host defence: discriminating invasion from colonization. *Nat Rev Microbiol* 10(2):112-122 .
16. Naglik JR, Moyes DL, Wachtler B, & Hube B (2011) *Candida albicans* interactions with epithelial cells and mucosal immunity. *Microbes Infect* 13(12-13):963-976 .
17. Phan QT, *et al.* (2007) Als3 is a *Candida albicans* invasin that binds to cadherins and induces endocytosis by host cells. *PLoS Biol* 5(3):e64 .
18. Argimon S, *et al.* (2007) Developmental regulation of an adhesin gene during cellular morphogenesis in the fungal pathogen *Candida albicans*. *Eukaryot Cell* 6(4):682-692 .
19. Staab JF, Bradway SD, Fidel PL, & Sundstrom P (1999) Adhesive and mammalian transglutaminase substrate properties of *Candida albicans* Hwp1. *Science* 283(5407):1535-1538 .
20. Moreno-Ruiz E, *et al.* (2009) *Candida albicans* internalization by host cells is mediated by a clathrin-dependent mechanism. *Cell Microbiol* 11(8):1179-1189 .
21. Dalle F, *et al.* (2010) Cellular interactions of *Candida albicans* with human oral epithelial cells and enterocytes. *Cell Microbiol* 12(2):248-271 .
22. Wisplinghoff H, Seifert H, Wenzel RP, & Edmond MB (2006) Inflammatory response and clinical course of adult patients with nosocomial bloodstream infections caused by *Candida* spp. *Clin Microbiol Infect* 12(2):170-177 .
23. Romani L, Bistoni F, & Puccetti P (2002) Fungi, dendritic cells and receptors: a host perspective of fungal virulence. *Trends Microbiol* 10(11):508-514 .
24. Netea MG, Brown GD, Kullberg BJ, & Gow NA (2008) An integrated model of the recognition of *Candida albicans* by the innate immune system. *Nat Rev Microbiol* 6(1):67-78 .
25. Bourgeois C, Majer O, Frohner IE, Tierney L, & Kuchler K (2010) Fungal attacks on mammalian hosts: pathogen elimination requires sensing and tasting. *Curr Opin Microbiol* 13(4):401-408 .
26. Chaffin WL (2008) *Candida albicans* cell wall proteins. *Microbiol Mol Biol Rev* 72(3):495-544 .

27. Fuchs BB & Mylonakis E (2009) Our paths might cross: the role of the fungal cell wall integrity pathway in stress response and cross talk with other stress response pathways. *Eukaryot Cell* 8(11):1616-1625 .
28. Moyes DL, *et al.* (2010) A biphasic innate immune MAPK response discriminates between the yeast and hyphal forms of *Candida albicans* in epithelial cells. *Cell Host Microbe* 8(3):225-235 .
29. Gow NA & Hube B (2012) Importance of the *Candida albicans* cell wall during commensalism and infection. *Curr Opin Microbiol* .
30. Wheeler RT, Kombe D, Agarwala SD, & Fink GR (2008) Dynamic, morphotype-specific *Candida albicans* beta-glucan exposure during infection and drug treatment. *PLoS Pathog* 4(12):e1000227 .
31. van der Graaf CA, Netea MG, Verschueren I, van der Meer JW, & Kullberg BJ (2005) Differential cytokine production and Toll-like receptor signaling pathways by *Candida albicans* blastoconidia and hyphae. *Infect Immun* 73(11):7458-7464 .
32. Lenardon MD, Munro CA, & Gow NA (2010) Chitin synthesis and fungal pathogenesis. *Curr Opin Microbiol* 13(4):416-423 .
33. Chai LY, Netea MG, Vonk AG, & Kullberg BJ (2009) Fungal strategies for overcoming host innate immune response. *Med Mycol* 47(3):227-236 .
34. Heinsbroek SE, Brown GD, & Gordon S (2005) Dectin-1 escape by fungal dimorphism. *Trends Immunol* 26(7):352-354 .
35. Gantner BN, Simmons RM, & Underhill DM (2005) Dectin-1 mediates macrophage recognition of *Candida albicans* yeast but not filaments. *Embo J* 24(6):1277-1286 .
36. Tavanti A, *et al.* (2006) *Candida albicans* isolates with different genomic backgrounds display a differential response to macrophage infection. *Microbes Infect* 8(3):791-800 .
37. Seider K, Heyken A, Luttich A, Miramon P, & Hube B (2010) Interaction of pathogenic yeasts with phagocytes: survival, persistence and escape. *Curr Opin Microbiol* 13(4):392-400 .
38. Tobudic S, *et al.* (2011) Biofilm formation of *Candida* spp. isolates from patients at a cardiothoracic intensive care unit. *Int J Artif Organs* 34(9):818-823 .
39. Frohner IE, Bourgeois C, Yatsyk K, Majer O, & Kuchler K (2009) *Candida albicans* cell surface superoxide dismutases degrade host-derived reactive

- oxygen species to escape innate immune surveillance. *Mol Microbiol* 71(1):240-252 .
40. Ullmann BD, *et al.* (2004) Inducible defense mechanism against nitric oxide in *Candida albicans*. *Eukaryot Cell* 3(3):715-723 .
 41. Hocine S, Singer RH, & Grunwald D (2010) RNA processing and export. *Cold Spring Harb Perspect Biol* 2(12):a000752 .
 42. Pan Q, Shai O, Lee LJ, Frey BJ, & Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40(12):1413-1415 .
 43. Kupfer DM, *et al.* (2004) Introns and splicing elements of five diverse fungi. *Eukaryot Cell* 3(5):1088-1100 .
 44. Mitrovich QM, Tuch BB, Guthrie C, & Johnson AD (2007) Computational and experimental approaches double the number of known introns in the pathogenic yeast *Candida albicans*. *Genome Res* 17(4):492-502 .
 45. Liu N, *et al.* (2009) SnoRNAs from the filamentous fungus *Neurospora crassa*: structural, functional and evolutionary insights. *BMC Genomics* 10:515 .
 46. Bernstein DA, *et al.* (2012) *Candida albicans* Dicer (CaDcr1) is required for efficient ribosomal and spliceosomal RNA maturation. *Proc Natl Acad Sci U S A* 109(2):523-528 .
 47. Webster KA (2003) Evolution of the coordinate regulation of glycolytic enzyme genes by hypoxia. *J Exp Biol* 206(Pt 17):2911-2922 .
 48. Barker BM, *et al.* (2012) Transcriptomic and proteomic analyses of the *Aspergillus fumigatus* hypoxia response using an oxygen-controlled fermenter. *BMC Genomics* 13:62 .
 49. de Groot MJ, *et al.* (2007) Quantitative proteomics and transcriptomics of anaerobic and aerobic yeast cultures reveals post-transcriptional regulation of key cellular processes. *Microbiology* 153(Pt 11):3864-3878 .
 50. Chen M & Manley JL (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* 10(11):741-754 .
 51. Graveley BR (2005) Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* 123(1):65-73 .
 52. Hiller M, Zhang Z, Backofen R, & Stamm S (2007) Pre-mRNA secondary structures influence exon recognition. *PLoS Genet* 3(11):e204 .

53. Grover A, *et al.* (1999) 5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. *J Biol Chem* 274(21):15134-15143 .
54. Cartegni L, Chew SL, & Krainer AR (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 3(4):285-298 .
55. Graveley BR (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet* 17(2):100-107 .
56. Schmucker D, *et al.* (2000) Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101(6):671-684 .
57. Blencowe BJ (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci* 25(3):106-110 .
58. Roy SW & Gilbert W (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat Rev Genet* 7(3):211-221 .
59. Dujon B (2006) Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends Genet* 22(7):375-387 .
60. Derr LK (1998) The involvement of cellular recombination and repair genes in RNA-mediated recombination in *Saccharomyces cerevisiae*. *Genetics* 148(3):937-945 .
61. Kent WJ & Zahler AM (2000) Conservation, regulation, synteny, and introns in a large-scale *C. briggsae*-*C. elegans* genomic alignment. *Genome Res* 10(8):1115-1125 .
62. Mitrovich QM, Tuch BB, De La Vega FM, Guthrie C, & Johnson AD (2010) Evolution of yeast noncoding RNAs reveals an alternative mechanism for widespread intron loss. *Science* 330(6005):838-841 .
63. Engebrecht JA, Voelkel-Meiman K, & Roeder GS (1991) Meiosis-specific RNA splicing in yeast. *Cell* 66(6):1257-1268 .
64. Li Z, Paulovich AG, & Woolford JL, Jr. (1995) Feedback inhibition of the yeast ribosomal protein gene CRY2 is mediated by the nucleotide sequence and secondary structure of CRY2 pre-mRNA. *Mol Cell Biol* 15(11):6454-6464 .
65. Vilardell J, Chartrand P, Singer RH, & Warner JR (2000) The odyssey of a regulated transcript. *Rna* 6(12):1773-1780 .

66. Preker PJ, Kim KS, & Guthrie C (2002) Expression of the essential mRNA export factor Yra1p is autoregulated by a splicing-dependent mechanism. *Rna* 8(8):969-980 .
67. Westerhoff HV & Palsson BO (2004) The evolution of molecular biology into systems biology. *Nat Biotechnol* 22(10):1249-1252 .
68. Ideker T, Galitski T, & Hood L (2001) A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet* 2:343-372 .
69. Zhang W, Li F, & Nie L (2010) Integrating multiple 'omics' analysis for microbial biology: application and methodologies. *Microbiology* 156(Pt 2):287-301 .
70. Chuang HY, Hofree M, & Ideker T (2010) A decade of systems biology. *Annu Rev Cell Dev Biol* 26:721-744 .
71. Fontana JM, Alexander E, & Salvatore M (2012) Translational research in infectious disease: current paradigms and challenges ahead. *Transl Res* 159(6):430-453 .
72. Watson JD & Cook-Deegan RM (1991) Origins of the Human Genome Project. *Faseb J* 5(1):8-11 .
73. Chin CS, *et al.* (2011) The origin of the Haitian cholera outbreak strain. *N Engl J Med* 364(1):33-42 .
74. Garten RJ, *et al.* (2009) Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science* 325(5937):197-201 .
75. Rappuoli R (2001) Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine* 19(17-19):2688-2691 .
76. Emonet SE, Urata S, & de la Torre JC (2011) Arenavirus reverse genetics: new approaches for the investigation of arenavirus biology and development of antiviral strategies. *Virology* 411(2):416-425 .
77. Cuomo CA & Birren BW (2010) The fungal genome initiative and lessons learned from genome sequencing. *Methods Enzymol* 470:833-855 .
78. Frohlich T & Arnold GJ (2006) Proteome research based on modern liquid chromatography--tandem mass spectrometry: separation, identification and quantification. *J Neural Transm* 113(8):973-994 .
79. Lasonder E, *et al.* (2002) Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* 419(6906):537-542 .

80. Lindon JC, Holmes E, & Nicholson JK (2007) Metabonomics in pharmaceutical R&D. *Febs J* 274(5):1140-1151 .
81. Hameed S, Dhamgaye S, Singh A, Goswami SK, & Prasad R (2011) Calcineurin signaling and membrane lipid homeostasis regulates iron mediated multidrug resistance mechanisms in *Candida albicans*. *PLoS One* 6(4):e18684.
82. Munger J, *et al.* (2008) Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. *Nat Biotechnol* 26(10):1179-1186 .
83. Xu D, *et al.* (2007) Genome-wide fitness test and mechanism-of-action studies of inhibitory compounds in *Candida albicans*. *PLoS Pathog* 3(6):e92 .
84. Lum PY, *et al.* (2004) Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. *Cell* 116(1):121-137 .
85. Oh J, *et al.* (2010) Gene annotation and drug target discovery in *Candida albicans* with a tagged transposon mutant collection. *PLoS Pathog* 6(10):e1001140 .
86. Ghazal P, Watterson S, Robertson K, & Kluth DC (2011) The *in silico* macrophage: toward a better understanding of inflammatory disease. *Genome Med* 3(1):4 .
87. Tierney L, Kuchler K, Rizzetto L, & Cavalieri D (2012) Systems biology of host-fungus interactions: turning complexity into simplicity. *Curr Opin Microbiol* .
88. Santamaria R, *et al.* (2011) Systems biology of infectious diseases: a focus on fungal infections. *Immunobiology* 216(11):1212-1227 .
89. Rizzetto L & Cavalieri D (2011) Friend or foe: using systems biology to elucidate interactions between fungi and their hosts. *Trends Microbiol* 19(10):509-515 .
90. Horn F, *et al.* (2012) Systems biology of fungal infection. *Front Microbiol* 3:108.
91. Linde J, Wilson D, Hube B, & Guthke R (2010) Regulatory network modelling of iron acquisition by a fungal pathogen in contact with epithelial cells. *BMC Syst Biol* 4:148 .
92. Tierney L, *et al.* (2012) An Interspecies Regulatory Network Inferred from Simultaneous RNA-seq of *Candida albicans* Invading Innate Immune Cells. *Front Microbiol* 3:85 .
93. Loeffler J, *et al.* (2009) Interaction analyses of human monocytes co-cultured with different forms of *Aspergillus fumigatus*. *J Med Microbiol* 58(Pt 1):49-58 .

94. Rizzetto L, *et al.* (2010) Differential IL-17 production and mannan recognition contribute to fungal pathogenicity and commensalism. *J Immunol* 184(8):4258-4268 .
95. Morton CO, *et al.* (2011) The temporal dynamics of differential gene expression in *Aspergillus fumigatus* interacting with human immature dendritic cells *in vitro*. *PLoS One* 6(1):e16016 .
96. Kamthan M, Mukhopadhyay G, Chakraborty N, Chakraborty S, & Datta A (2012) Quantitative proteomics and metabolomics approaches to demonstrate N-acetyl-D-glucosamine inducible amino acid deprivation response as morphological switch in *Candida albicans*. *Fungal Genet Biol* 49(5):369-378 .
97. Dyer MD, Murali TM, & Sobral BW (2008) The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathog* 4(2):e32 .
98. Sogro AG, *et al.* (2011) Effects of fluconazole on the secretome, the wall proteome, and wall integrity of the clinical fungus *Candida albicans*. *Eukaryot Cell* 10(8):1071-1081 .
99. Tuckwell D, Denning DW, & Bowyer P (2011) A public resource for metabolic pathway mapping of *Aspergillus fumigatus* Af293. *Med Mycol* 49 Suppl 1:S114-119 .
100. Legendre M, *et al.* (2010) mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. *Genome Res* 20(5):664-674 .
101. Yang Z, Bruno DP, Martens CA, Porcella SF, & Moss B (2010) Simultaneous high-resolution analysis of vaccinia virus and host cell transcriptomes by deep RNA sequencing. *Proc Natl Acad Sci U S A* 107(25):11513-11518 .
102. Ringner M (2008) What is principal component analysis? *Nat Biotechnol* 26(3):303-304 .
103. Inglis DO, *et al.* (2012) The *Candida* genome database incorporates multiple *Candida* species: multispecies search and analysis tools with curated gene and protein information for *Candida albicans* and *Candida glabrata*. *Nucleic Acids Res* 40(Database issue):D667-674 .
104. Lorenz MC, Bender JA, & Fink GR (2004) Transcriptional response of *Candida albicans* upon internalization by macrophages. *Eukaryot Cell* 3(5):1076-1087 .
105. 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(19):11007-11012 .
106. Bensen ES, Martin SJ, Li M, Berman J, & Davis DA (2004) Transcriptional profiling in *Candida albicans* reveals new adaptive responses to extracellular pH and functions for Rim101p. *Mol Microbiol* 54(5):1335-1351 .
 107. Fernandez-Arenas E, *et al.* (2007) Integrated proteomics and genomics strategies bring new insight into *Candida albicans* response upon macrophage interaction. *Mol Cell Proteomics* 6(3):460-478 .
 108. Mayer FL, *et al.* (2012) Small but crucial: the novel small heat shock protein Hsp21 mediates stress adaptation and virulence in *Candida albicans*. *PLoS One* 7(6):e38584 .
 109. Tavares AH, *et al.* (2012) Murine dendritic cells transcriptional modulation upon *Paracoccidioides brasiliensis* infection. *PLoS Negl Trop Dis* 6(1):e1459 .
 110. Genini S, *et al.* (2011) Strengthening insights into host responses to mastitis infection in ruminants by combining heterogeneous microarray data sources. *BMC Genomics* 12(1):225 .
 111. Mercer TR, Dinger ME, & Mattick JS (2009) Long non-coding RNAs: insights into functions. *Nat Rev Genet* 10(3):155-159 .
 112. Garmire LX, *et al.* (2011) A global clustering algorithm to identify long intergenic non-coding RNA--with applications in mouse macrophages. *PLoS One* 6(9):e24051 .
 113. Dziembowski A, *et al.* (2004) Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. *Embo J* 23(24):4847-4856 .
 114. Brooks MA, *et al.* (2009) Structure of the yeast Pml1 splicing factor and its integration into the RES complex. *Nucleic Acids Res* 37(1):129-143 .
 115. Scherrer FW, Jr. & Spingola M (2006) A subset of Mer1p-dependent introns requires Bud13p for splicing activation and nuclear retention. *Rna* 12(7):1361-1372 .
 116. Suegara N, Siegel JE, & Savage DC (1979) Ecological determinants in microbial colonization of the murine gastrointestinal tract: adherence of *Torulopsis pintolopesii* to epithelial surfaces. *Infect Immun* 25(1):139-145 .

117. Artwohl J, McClain A, & Cera L (1988) Population changes of indigenous murine *Candida pintolopesii* under various experimental conditions and routes of inoculation. *Appl Environ Microbiol* 54(10):2371-2374 .
118. Koh AY, Kohler JR, Coggshall KT, Van Rooijen N, & Pier GB (2008) Mucosal damage and neutropenia are required for *Candida albicans* dissemination. *PLoS Pathog* 4(2):e35 .
119. Gillum AM, Tsay EY, & Kirsch DR (1984) Isolation of the *Candida albicans* gene for orotidine-5'-phosphate decarboxylase by complementation of *S. cerevisiae* *ura3* and *E. coli* *pyrF* mutations. *Mol Gen Genet* 198(1):179-182 .
120. Espinel-Ingroff A, *et al.* (1998) Comparison of RapID yeast plus system with API 20C system for identification of common, new, and emerging yeast pathogens. *J Clin Microbiol* 36(4):883-886 .
121. Noble SM & Johnson AD (2005) Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen *Candida albicans*. *Eukaryot Cell* 4(2):298-309 .
122. Liu H, Kohler J, & Fink GR (1994) Suppression of hyphal formation in *Candida albicans* by mutation of a STE12 homolog. *Science* 266(5191):1723-1726 .
123. Lee KL, Buckley HR, & Campbell CC (1975) An amino acid liquid synthetic medium for the development of mycelial and yeast forms of *Candida Albicans*. *Sabouraudia* 13(2):148-153 .
124. Hnisz D, Schwarzmuller T, & Kuchler K (2009) Transcriptional loops meet chromatin: a dual-layer network controls white-opaque switching in *Candida albicans*. *Mol Microbiol* 74(1):1-15 .
125. Bourgeois C, Majer O, Frohner I, & Kuchler K (2009) *In vitro* systems for studying the interaction of fungal pathogens with primary cells from the mammalian innate immune system. *Methods Mol Biol* 470:125-139 .
126. Jiang H & Wong WH (2009) Statistical inferences for isoform expression in RNA-Seq. *Bioinformatics* 25(8):1026-1032 .
127. Li B, Ruotti V, Stewart RM, Thomson JA, & Dewey CN (2010) RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* 26(4):493-500 .
128. Oshlack A, Robinson MD, & Young MD (2010) From RNA-seq reads to differential expression results. *Genome Biol* 11(12):220 .

129. Fang Z & Cui X (2011) Design and validation issues in RNA-seq experiments. *Brief Bioinform* 12(3):280-287 .
130. Waterston RH, *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420(6915):520-562 .
131. Bashir A, Bansal V, & Bafna V (2010) Designing deep sequencing experiments: detecting structural variation and estimating transcript abundance. *BMC Genomics* 11:385 .
132. Kalsotra A & Cooper TA (2011) Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet* 12(10):715-729 .
133. Jeffares DC, Mourier T, & Penny D (2006) The biology of intron gain and loss. *Trends Genet* 22(1):16-22 .
134. Meyer M & Vilardell J (2009) The quest for a message: budding yeast, a model organism to study the control of pre-mRNA splicing. *Brief Funct Genomic Proteomic* 8(1):60-67 .
135. Ares M, Jr., Grate L, & Pauling MH (1999) A handful of intron-containing genes produces the lion's share of yeast mRNA. *Rna* 5(9):1138-1139 .
136. Wahl MC, Will CL, & Luhrmann R (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell* 136(4):701-718 .
137. Burset M, Seledtsov IA, & Solovyev VV (2000) Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res* 28(21):4364-4375 .
138. Thanaraj TA & Clark F (2001) Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. *Nucleic Acids Res* 29(12):2581-2593 .
139. Schwartz SH, *et al.* (2008) Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res* 18(1):88-103 .
140. Cooper TA, Wan L, & Dreyfuss G (2009) RNA and disease. *Cell* 136(4):777-793 .
141. Naglik JR, Fidel PL, Jr., & Odds FC (2008) Animal models of mucosal *Candida* infection. *FEMS Microbiol Lett* 283(2):129-139 .
142. Savage DC & Dubos RJ (1967) Localization of indigenous yeast in the murine stomach. *J Bacteriol* 94(6):1811-1816 .

143. Kurtzman CP, *et al.* (2005) Multigene phylogenetic analysis of pathogenic candida species in the *Kazachstania* (*Arxiozyma*) *telluris* complex and description of their ascosporic states as *Kazachstania bovina* sp. nov., *K. heterogenica* sp. nov., *K. pintolopesii* sp. nov., and *K. slooffiae* sp. nov. *J Clin Microbiol* 43(1):101-111 .

8. Appendix

8.1 Additional publications

Current Opinion in Microbiology. 2010 Aug;13(4):401-8.

Fungal Attacks on Mammalian Hosts: Pathogen Elimination Requires Sensing and Tasting

Bourgeois C, Majer O, Frohner IE, **Tierney L**, Kuchler K.

Abstract: Recognition of *Candida spp.* by immune cells is mediated by dedicated pattern recognition receptors (PRRs), including Toll-like receptors (TLRs) and lectins expressed on innate immune cells (e.g., macrophages, neutrophils and dendritic cells (DCs)). PRRs recognize *Candida*-specific pathogen-associated molecular patterns (PAMPs). Binding of fungal PAMPs (e.g., cell wall sugar polymers and proteins, fungal nucleic acids) to PRRs triggers the activation of innate effector cells. Recent findings underscore the role of DCs in relaying PAMP information through their PRRs to stimulate the adaptive response. In agreement, deficiencies in certain PRRs strongly impair survival to *Candida* infections in mice and is associated with enhanced susceptibility to mucocutaneous fungal infections in humans. Understanding the complex signaling networks protecting the host against fungal pathogens remains a challenge in the field.

Contribution: For this study, I was involved in the writing and preparation of the manuscript.

8.2 Summary

Candida spp. are a major cause of systemic fungal infections in immunocompromised individuals, inflicting severe diseases of high mortality. Their success as a pathogen is strongly correlated with their ability to adapt to diverse micro-environmental challenges encountered within their host. Their inherent genomic plasticity poses further challenges to diagnosis and treatment. My thesis aimed to study adaptive strategies employed by *Candida albicans* during *in vitro* pathogenesis from a transcriptional perspective. I developed a method to perform and analyze transcriptional data simultaneously from host-pathogen interactions using deep sequencing technology. I quantified *C. albicans* and *M. musculus* gene expression dynamics during phagocytosis by murine dendritic cells in a genome-wide, time-resolved fashion. Using this approach, I have identified novel strategies of *C. albicans* to counteract host challenge and divided my work in the verification of these different adaptive mechanisms.

The first part of this thesis characterizes several components previously unknown to be involved in host-pathogen crosstalk using the first inferred network of a mixed host-pathogen data set. I identified a role for the interactions between the iron regulatory transcription factor Hap3 in *C. albicans*, and the soluble receptor Ptx3 in *M. musculus*, in the global immune response to the fungus. My data suggest that this interaction takes place via cell wall remodeling of the pathogen. Moreover, the experimental verification of the predicted interspecies interactions is proof of principal that network inference can be used to investigate pathogenesis.

The second part of this thesis focuses on post-transcriptional processing mechanisms of. I identified numerous cases of infection specific alternative splicing in addition to novel alternative splice junctions. Furthermore, I developed a mathematical model to quantify splicing for and these data support the hypothesis that splicing is co-transcriptional. The conservation of the splicing machineries among higher eukaryotes suggests that splicing regulation by the pathogen may affect its virulence and intron evolution.

Taken together, this thesis proposes novel methods to analyze host-pathogen interactions and new putative targets for *Candida* treatment.

8.3 Zusammenfassung

Pilzarten der Gattung *Candida* führen in immunsupprimierten Patienten zu schweren invasiven Erkrankungen. Der Erfolg des Pathogens ist dabei abhängig von dessen Fähigkeit sich auf die verschiedensten Bedingungen im Wirt anzupassen. Zusätzlich erschwert die hohe genetische Flexibilität von *Candida* die Diagnose und Behandlung von Erkrankungen. Ziel dieser Dissertation war es daher, genau diese adaptiven Prozesse von *Candida* aus der transkriptionellen Sicht zu studieren.

Dazu entwickelte ich eine auf RNA-Seq basierende Methode, welche ermöglicht simultan sowohl das Pathogen als auch das Wirt Transkriptom während der Infektion zu untersuchen. Mittels dieser Methode wurden dynamische Genexpressionsprofile von Maus dendritischen Zellen und *C. albicans* unter *in vitro* Infektionsbedingungen quantifiziert. Die Arbeit entschlüsselte neuartige Abwehrstrategien von *C. albicans* um sich der Immunantwort des Wirt zu entziehen. Im Folgenden lässt sich meine Dissertation in folgende zwei Teilprojekte unterteilen.

Der erste Teil beschreibt die Charakterisierung neuer Faktoren, die zuvor nicht mit Wirt-Pathogen Wechselwirkungen assoziiert waren und mittels einem „Inferred Network“ des gemischten Wirt-Pathogen Datensets sichtbar wurden. Das Projekt entschlüsselte eine neuartige Interaktion zwischen dem Eisen-regulierendem Transkriptionsfaktor Hap3 in *C. albicans* und dem löslichen Maus Immunrezeptor Ptx3. Die Daten deuten an, dass diese Interaktion zum Teil durch Zellwand-„Remodeling“ im Pathogen vermittelt wird. Außerdem repräsentiert meine Arbeit ein „proof of principle“, dass „Network Inference“ erfolgreich als Model zur Studie von Wirt-Pathogen Interaktionen eingesetzt werden kann.

Der zweite Teil der Dissertation beschäftigt sich mit Mechanismen der post-transkriptionalen Modifizierung in *C. albicans*. Diese Arbeit entschlüsselte die Existenz von infektionsspezifischen alternativen „splicing“, sowie zuvor unbekannte alternative „splice junctions“ in *C. albicans*. Zudem wurde ein mathematisches Modell entwickelt um „splicing“ zu quantifizieren, dessen Daten ergaben, dass „splicing“ sehr wahrscheinlich co-transcriptional stattfindet. Die starke Konservierung des „splicing“ Apparats in höheren Eukaryoten lässt vermuten, dass „splicing“ Regulierung zur Virulenz und Intron Evolution von *Candida* beiträgt.

Zusammenfassend, schlägt meine Dissertation neue Methoden zur Untersuchung von Wirt-Pathogen Interaktionen vor und entschlüsselt neue potentielle „targets“ für die Behandlung von *Candida* Infektionen.

8.4 Abbreviations

<i>A. fumigatus</i>	<i>Aspergillus fumigatus</i>
AS	alternative splicing
BMDC	bone marrow derived dendritic cells
<i>C. albicans</i>	<i>Candida albicans</i>
<i>C. glabrata</i>	<i>Candida glabrata</i>
<i>C. krusei</i>	<i>Candida krusei</i>
<i>C. parapsilosis</i>	<i>Candida parapsilosis</i>
<i>C. pintolopesii</i>	<i>Candida pintolopesii</i>
<i>C. tropicalis</i>	<i>Candida tropicalis</i>
CaDcr1	<i>Candida albicans</i> gene Dicer
<i>Candida spp.</i>	<i>Candida</i> species
CWPs	cell wall proteins
DCs	dendritic cells
Dscam	down syndrome cell-adhesion molecule
DSS	dextran sulfate sodium
GI tract	gastrointestinal tract
GO	gene ontology
GPI	glycosylphosphatidylinositol
GR-1	granulocyte receptor-1
HPLC	high performance liquid chromatography
IC	isotype control
IP	interperitoneal
lincRNA	long non-coding RNA
<i>M. musculus</i>	<i>Mus musculus</i>
MC	Mac Conkey agar
mRNA	messenger RNA
MS	mass spectrometry
MyD88	myeloid differentiation primary response gene 88
NMR	nuclear magnetic resonance
NO	nitric oxide
NuRD	nucleosome remodeling and histone deacetylase complex
PAMPs	pathogen associated molecular patterns

PBS	phosphate buffered saline
PCA	principle component analysis
PRR	pathogen recognition receptors
RES	retention and splicing
RNA-seq	RNA sequencing
RNase	ribonuclease
ROS	reactive oxygen species
rRNAs	ribosomal RNA
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
snoRNAs	small nucleolar RNA
snRNPs	small nuclear ribonucleoprotein particles
TLRs	toll-like receptors
TSA	tryptic soy agar
YPD	yeast extract peptone dextrose medium

8.5 Curriculum Vitae

Date of birth: 30.05.1986	Address: Rabensteig 1/16 1010
Place of birth: Springfield, MA, USA	Vienna, Austria
E-mail: Lanay.Tierney@gmail.com	Nationality: American, Residency: Austrian

Education

2008 - 12.2012 (Expected)	Ph.D., Molecular Biology University of Vienna – Vienna Biocenter International Ph.D. program	<i>Vienna, Austria</i>
2004 - 2008	B.Sc., Biology Boston College	<i>Chestnut Hill, USA</i>

Professional Experience

2007 - 2008	Young Ambassador German Academic Exchange Service (DAAD) – Promoted DAAD programs and study abroad opportunities – Representative of DAAD at Boston College and within Boston	<i>Boston, USA</i>
2006 - 2008	Research Assistant Massachusetts State Science and Engineering Fair – Assisted the Executive Director in alumni relations and publicity	<i>Boston, USA</i>
Jun - Aug 2007	Laboratory Research Internship Heinrich Heine University – Technical training at the Institute for Cardiovascular Physiology	<i>Düsseldorf, Germany</i>
Jun - Aug 2006	English Teacher Learning Enterprises Organization – Volunteer teacher to children and adults in rural village locations	<i>Hungary & Slovakia</i>

Honours

2011 May	Poster Presentation Award – Best presentation at a FEBS course on fungal pathogens	<i>Nice, France</i>
2010 June	Poster Presentation Award – Ph.D. Symposium Young Scientists Association	<i>Vienna, Austria</i>
2009 May	Youth Travel Fund Fellowship – To attend the 2009 FEBS course on fungal pathogens	<i>Nice, France</i>
2004 Jul	International Science and Engineering Camp – One of four chosen to represent the USA – Awarded by the American Junior Academy of Sciences	<i>Seoul, South Korea</i>
2004 Feb	Massachusetts Student Representative	<i>Seattle, USA</i>

- Chosen for a poster presenter at AJAS annual meeting

Publications

- Tierney *et al.* 2012 Systems biology of host–fungus interactions: turning complexity into simplicity. *Current Opinion in Microbiology*
- Tierney *et al.* 2012 An interspecies regulatory network inferred from simultaneous RNA seq of *Candida albicans* invading innate immune cells. *Frontiers in Microbiology*

Training Courses

- 2010 FINSysB Training workshop on microarray analysis, *Paris, France*
next generation sequencing and bioinformatics
- 2009 Certificate course on Laboratory Animal Science *Berlin, Germany*
from Berliner Fortbildungen

Selected Oral Presentations

- 2012 International Society for Human and Animal Mycology *Berlin, Germany*
- 2011 Medical Mycology MedMyc2011 *Vienna, Austria*
- 2011 FEBS Human Fungal Pathogen Course *Nice, France*
- 2009 Wissenschaftlichen Tagung der Deutschsprachigen *Vienna, Austria*
Mykologischen Gesellschaft

Leadership

- 2009 - 2010 Vienna Biocenter Ph.D. Student Representative
- 2010 Organizing committee member for the Vienna Biocenter symposium
- 2010 Organizing committee member for the 1st annual MFPL student retreat

Languages English (native speaker), German (professional working proficiency)

Skills Data analysis, Statistics, Research design, Communication

8.6 Acknowledgements

I would like to personally thank all of those who supported me throughout my work. First and foremost, I would like to thank Karl for giving me the opportunity to work on this very interesting topic in his lab, who gave me motivation when experiments did not always work according to plan, and who was always was enthusiastic when they did. I would of course like to thank the entire Kuchler lab for their support in all facets of the dissertation, experimentally, professionally, and personally. I would like to especially thank Olivia, Denes, Christelle, and Walter for their endless patience, helpful feedback, and for providing a fun and positive working environment in the lab.

I would like to thank all of our collaborative partners both in Vienna and abroad, especially Jörg, Sebastian, Reinhard, Sascha, and Bernhard, in Jena and Arndt, Fritz, and Stefanie in Vienna, for their helpful technical support and feedback on the project throughout its development.

Finally, I would like to thank my parents, Laura and Brad, and sister Sara, who supported me throughout my education, to follow my aspirations, regardless of what they were, or where in the world they would take me.