Host-fungal pathogen interactions: A study of *Candida albicans* and mammalian macrophage and epithelial cells at the transcriptional level

by

Toni Marie Delorey

A Dissertation

Submitted to the Faculty of

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Biology and Biotechnology

by

_____

May 28th, 2019

Approved by: _____

Dr. Reeta Prusty Rao

_____	_____	_____
Dr. Elizabeth Ryder	Dr. Scarlet Shell	Dr. Patrick Flaherty

Abstract

It is estimated that fungal infections kill greater than 1.6 million people annually, a number that is comparable to the number of deaths associated with tuberculosis. *Candida* species are the fourth leading cause of hospital-acquired blood infections and *Candida albicans* is the most common cause of these fungal blood infections (known as candidemia). Immunocompromised individuals, such as those who have HIV/AIDS, those undergoing chemotherapy treatments, or those on broad-spectrum antibiotics, are most likely to develop candidemia. Candidemia is associated with a 20-40% mortality rate. However, when patient treatment for candidemia is delayed for over 48 hours, associated mortality rates increase to 78%. Blood infections can disseminate *Candida albicans* throughout the body, eventually leading to infection in vital organs like the liver, kidney and brain. Optimal patient outcomes are achieved if antifungal therapy is given within 12 hours after a blood sample is obtained for culture and testing. However, current blood tests cannot reliably detect *Candida* this early and thus antifungals are not routinely given to patients in this time frame.

Counterintuitively, it is believed that some fungi, like many bacteria, are non-harmful residents in small intestines of most adults and this hypothesis is supported by the fact that the most common fungal species in the human gut is *Candida albicans*. However, intestinal overgrowth of *C. albicans* is linked to Crohn's disease, and disease-causing forms of *C. albicans* can arise from commensal strains that once resided in the patient's gastrointestinal tract. The specific molecular mechanisms by which *C. albicans* interacts with host immune cells versus intestinal cells, and those that trigger *Candida* pathogenicity remain unknown. Many strains of *Candida albicans* have developed resistance to azoles, the major class of drugs used to treat both superficial and systemic infections. In order to develop new treatments, we must better understand host-fungal pathogen

biology to determine novel antifungal targets or therapeutics to fight fungal infections at the early stages of infection. In this work, we developed a novel tool that allows us to measure which genes are important to both the host and *Candida albicans* simultaneously, in specific infection states. We have applied this tool to measure gene expression in *Candida albicans* interacting with mammalian macrophages and small intestine epithelial cells– at both the population and single cell levels.

When examining populations of sorted infection samples, we found that host immune cells both exposed to and infected with fungal cells exhibit similar expression patterns. In contrast, phagocytosed *C. albicans* exhibit unique expression patterns compared to those merely exposed to macrophages. We found that immune response genes in single, *Candida* infected macrophages exhibited bimodal expression patterns for some immune response genes. We also observed examples of expression bimodality in live *Candida* inside of single macrophages. Both *Candida albicans* and host small intestine epithelial cells demonstrate distinct patterns of expression when exposed to each other at the population level, compared to unexposed controls. However, the magnitude of these differences is dependent on the multiplicity of infection. Some expression programs overlapped with those observed in populations of *Candida* cells interacting with macrophages, with key differences. We also observed expression bimodality among epithelial cells infected with *C. albicans*. We believe the information obtained using this technique could be used when considering new antifungal or therapeutics targets; if uniform and high expression of particular in genes in *Candida* populations phagocytosed by macrophages or invading epithelial cells leads to high and early protein production, these proteins may be effective antifungal targets. Similarly, if some host immune response genes are not expressed in a population of *Candida*

infected macrophages as uniformly and highly as expected, these genes or proteins could be target of a therapeutic for patients with *Candida* infections that are resistant to azoles.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

°C: degrees Celsius

3’: three prime end of a nucleic acid strand

3-D: three-dimensional

5’: five prime of a nucleic acid strand

BSA: Bovine serum albumin

C. albicans: *Candida albicans*

CO_2: Carbon dioxide

CBC cells: Crypt Based Columnar cells

cDNA: complementary DNA

DAPI: 4',6-diamidino-2-phenylindole; binds to adenine–thymine rich regions in DNA but can’t breach intact cell membranes

DNA: deoxyribonucleic acid

dNTP(s): deoxyribonucleotide triphosphate

E: enterocyte

EC: enteroendocrine

EP: enterocyte progenitor

FACS: fluorescent activated cell sorting

FBS: fetal bovine serum

FC: fold change

FCS: fetal calf serum

FDR: false discovery rate

FPKM: fragments per kilobase million

FU: fluorescent units

G1/S: The cell cycle check point between the Gap 1 phase (cell growth) and the Synthesis phase (DNA replication) of the cell cycle, where the cell assess DNA integrity and can pause for repair before proceeding.

gDNA: genomic DNA

GEM: gel bead in emulsion

GO term: Gene Ontology term

GFP: green fluorescent protein

gRNA: Guide RNA

GUT cells: Gastrointestinally induced transition cells

H. polygyrus: *Heligmosomoides polygyrus*

IBD: irritable bowel disease

ISC: base intestinal stem cell

LGR5: Leucine rich repeat-containing G-protein coupled receptor 5

LGR5+ cells: Leucine rich repeat-containing G-protein coupled receptor 5 containing cells

log2FC: Log 2 fold change

LOF: loss of function mutation

ISPCR primer: in silico PCR primer

LNA: locked nucleic acid

Macrophage(s): MØ

mCherry: monomeric red fluorescent protein

MIC: Minimum Inhibitory Concentration

MOI: multiplicity of infection

ml: milliliter

M-MLV: Moloney Murine Leukemia Virus

MW: molecular weight marker

ng/ul: nanograms

oligo(s): oligonucleotide(s)

oligo-dT(s): single-stranded sequence(s) of deoxythymine (dT), used for priming reverse transcriptase reactions with mRNA molecules containing polyadenylated 3' tails.

Orf: open reading frame

PAMPs: pathogen-associated molecular patterns

PBS: phosphate buffered saline

PCA: Principle component analysis

PCR: Polymerase Chain Reaction

PMA: phorbol 12-myristate 13-acetate

QSC: quiescent stem cell

RNA: ribonucleic acid

poly(A): polyadenylated

ROS: reactive oxygen species

PRRs: pathogen recognition receptors

rcf: relative centrifugal force

RPKM: reads per kilobase million

rpm: revolutions per minute

RT: reverse transcription

Salmonella: *Salmonella enterica*

SAP/Sap/sap: secreted aspartyl protease

scRNA-Seq: single-cell RNA-sequencing

SEM: standard error of the mean

SNP: single nucleotide polymorphism

TA: transit amplifying

TPM: transcripts per kilobase million

t-SNE: t-distributed stochastic neighbor embedding

TSO: template switching oligo

UC: ulcerative colitis

ug/ml: micrograms per milliliter

ul: microliter

μm: micron

UMI(s): unique molecular identifier(s)

Acknowledgements

First and foremost, I thank Dr. Reeta Prusty Rao for continual support and encouragement throughout this processes. She is an excellent mentor who pushed me to complete tasks that I thought were impossible.

I thank Dr. Elizabeth Ryder for chairing my dissertation committee, providing moral support, encouragement and building my confidence as a scientific communicator. Her excellent questions about my data and project improved the quality of this work.

I thank Dr. Scarlet Shell for always asking both detailed and wide-ranging questions about this work and for teaching me to think more critically and deeply about all aspects of experimental design, analysis and scientific communication.

I thank Dr. Patrick Flaherty for traveling to Worcester to attend my qualifying exam and dissertation defense. I also thank Dr. Flaherty for excellent questions which taught me to think both more broadly and mathematically about my scientific work.

I offer my enduring gratitude to the faculty, staff and my fellow students at WPI-especially in the BBT department, who have inspired me to continue my work in this field. I owe a special thanks to Dr. Jill Rulfs for coordinating Teaching Assistant assignments, as well as Dr. JoAnn Whitefleet-Smith and Dr. Michael Buckholt who both advised me on how to best teach at the undergraduate level. I am eternally grateful for the Boynton Fellowship and Teaching Assistantships from WPI

that have financially supported me throughout graduate school. I thank Dean Terri Camesano and Joan Murphey for their support and kindness while coordinating my non-traditional graduate student funding paradigms and always assisting with financial matters. I am also thankful to the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under award N°: U19AI110818, for financial support of this work and myself through Research Assistantships.

I would like to thank you Rao lab members and classmates, past and present. Thank you to Jefferey Letourneau for his work on *Candida albicans* survival assays. Thank you to Amanda Sullivan for lab preparation work. Special thanks to Cat Harwood, Dr. Melissa Mobley, Giulia Galotto, Diego Vargas Blanco, Dr. Christopher Chute, Matthew Desrosiers, Michelle McKee and Doug Reilly for moral support and scientific advice.

Many thanks to Dr. Jose Muñoz and Dr. Christina Cuomo at the Broad Institute for data analysis in the macrophage and *Candida* albicans project and for support throughout this work.

I thank Dr. Aviv Regev and all members of her lab at the Broad Institute for providing generous support for the experimental work in this dissertation. Special thanks for Jenna Pfiffner-Borges, Jane Lee and Dr. Orit Rozenblatt-Rosen for administrative support and encouragement during my work in the Regev lab. I thank Dr. Moshe Biton, Grace Burgin and Malika Sud for lending their experience and assistance with organoids experiments. I think Dr. Christopher Ford, Biyu Li and Dr. Dawn Thompson for study design, *Candida albicans* reporter strain construction and

experimental help. I also thank Dr. Raktima Raychowdhury for help with preparation of the primary BMDM cells.

I thank Dr. Valmik Vyas and Dr. Rebecca Shapiro for *Candida albicans* CRISPR reagents and assistance with experimental design.

Enormous thanks are owed to my family and partner, whose have supported me throughout my years of education.

Dedication

I dedicate this work to the women who have supported me throughout my life and academic journey: Mallory Coppenrath, Colleen Fitzgerald, Valery Young, Laura Draleau, Shirley Delorey, Delia Delorey, Jenna Pfiffner-Borges, Dawn Thompson, Danielle Dionne, Elizabeth Upsall Aderhold, Laura Lambiase and Jillian Nolan.

Chapter 1: Introduction

1.1 *Candida albicans* infections

It is estimated that fungal infections kill greater than 1.6 million people each year; this is comparable to the number of deaths caused by tuberculosis and approximately three times more deaths than caused by malaria, annually[1]. Billions of people have superficial fungal infections on their skin or nails, and most of these infections are not life threatening. However, approximately 150 million people have a fungal infection that causes major medical problems or threatens their life[1]. *Candida albicans* is the fourth leading cause of hospital-acquired infections[2,3] and blood infections caused by *Candida* species (candidemia) lead to increased care cost and length of hospital stay[4–8]. *C*andidemia is the most common type of blood infection in the United states and is associated with an estimated 30-40% crude mortality rate[8,9]. Blood infections can disseminate *C. albicans* throughout the body, eventually leading to infection in vital organs such as the brain, liver, and kidneys. The majority of adults have *Candida albicans* on their skin, gastrointestinal or urogenital tract, but most people will not develop deadly and invasive fungal infections. Immunocompromised individuals, such as those who have HIV/AIDs, those undergoing chemotherapy treatments, or those on broad-spectrum antibiotics are the most likely to develop invasive *Candida albicans* infections. When patient treatment for *C. albicans* blood infections is delayed for over 48 hours, associated mortality rates increases to 78%[10]. Many strains of *C. albicans* have developed resistance to azoles[11,12], the major class of drugs used to treat both superficial and systemic infections. In 2013, 7% of the *Candida* blood strains tested at the Centers for Disease Control and Prevention were resistant to fluconazole, a commonly used antifungal in the azole family of drugs[13].

C. albicans can also evolve to become more drug resistant within a patient over the course of antifungal treatment; in a recent study, Ford et al. performed phenotypic and genomic analyses on *C. albicans* strains that were collected between 1992 and 1997 from the mouths of HIV/AIDs patients with thrush, a *Candida* infection of the oral epithelial cells[14]. In a partial table from this study shown below, *C. albicans* showed rapid drug resistance evolution within a single patient, Patient 1 (PT1), after only one week of fluconazole treatment **(Table 1.1)**. Each strain was tested for the minimum inhibitory concentration (MIC) of fluconazole needed to inhibit *C. albicans* growth. Azole treatment over two years led to an over 300 fold increase in fluconazole MIC for *C. albicans* strains isolated from PT1.

Table 1.1: *Candida albicans* strains isolated from patients rapidly evolve to become drug resistant over the course of fluconazole treatment
Adapted from Ford et al., 2015.

PT	Strain	Entry date	Drug treatment	Dose (mg/day)	E-test MIC (ug/mL)
1	1	9/10/90	Fluconazole	100	0.25
	2	12/14/90	Fluconazole	100	1
	3	12/21/90	Fluconazole	100	4
	4	12/31/90	Fluconazole	100	3
	5	2/8/91	Fluconazole	100	4
	6	2/22/91	Fluconazole	100	4
	7	3/25/91	Fluconazole	100	4
	8	4/8/91	Fluconazole	100	4
	9	6/4/91	Fluconazole	100	4
	11	7/15/91	Fluconazole	100	4
	12	11/26/91	Fluconazole	200	4
	13	12/13/91	Fluconazole	400	32
	14	1/28/92	Fluconazole	400	24
	15	2/21/92	Clotriminazole	50	24
	16	4/1/92	Fluconazole	400	96
	17	8/25/92	Fluconazole	800	96

C. albicans is a dynamic organism, that adapts to the host environment by changing morphology[15], transcriptional programs[16–20] and the composition of its genome[14,21]. *Candida* species are present in the guts of 70% of adults in the United States[22] and are generally commensal, non-pathogenic residents. *Candida albicans* over-growth is usually inhibited by the host immune system and other microflora. However, if a patient is on broad-spectrum antibiotics, for example, resident bacterial population dynamics can be altered and/or drastically reduced, resulting in the establishment of *C. albicans* infections. The commensal *C. albicans* phenotype has been typically characterized as white-phase budding yeast cells while the pathogenic *C. albicans* forms pseudo-hyphae and hyphae[17,23]. A special type of commensal *C. albicans* in the gut is characterized by over-expression

of a transcription factor Wor1 and a darker colony, containing vacuoles with an elongated shape[17]. While many proteins have been linked to virulence[24], there is no classically hyper-virulent strain of *C. albicans*.

C. albicans can interact with the host epithelium by expressing adhesion proteins that allow cell-to-cell contact (adhesins) and promote biofilm formation[24]. This biofilm, a conglomerate of microorganisms, can also shield the *Candida albicans* pathogen associated molecular pattern (PAMP) molecules from host pathogen recognition receptors (PRRs), protecting the *Candida albicans* from opsonization. Additionally, when *Candida albicans* switch from the yeast to the hyphal form, the beta glucan molecules on the external cell wall are shielded from host PRRs, preventing host immune system detection[25]. The hyphal form of *C. albicans* is capable of invading host tissue and organs. Not only can this result in organ damage and failure, it can also lead to dissemination of the fungus into the bloodstream, leading to additional organ infections. While *C. albicans* generally exists as a commensal organism in the small intestine of a majority of adults[22], the genes important for maintaining commensalism and the host cues that may signal *C. albicans* to switch to an invasive/pathogenic form are not well understood.

At the genomic level, chromosomal aneuploidies and point mutations can also occur upon exposure to host and treatment with azoles[14]. For example, a specific aneuploidy involving the two left arms of chromosome 5 leads to increased expression of genes involved in drug efflux pump activity and ergosterol biosynthesis[26], a pathway important for cell wall integrity that is commonly targeted by antifungal drugs[27]. Additionally, the transcriptional flexibility of *C. albicans* allows

the fungus to respond to a dynamic host environment. Switching from traditional, circular yeast form to hyphal form, a process known as phenotypic switching, requires dynamic regulation of several types of genes deemed virulence factors. Some examples of virulence factors include secreted aspartyl proteases (Saps), which are fungal proteins can break down host adhesion proteins that aid in pathogen recognition and binding[24].

C. albicans is able to not only survive and grow in commensal fashion inside of the host but is able to cause infections in diverse host niche environments like the oral cavity, vagina or gut. *C. albicans* can also survive engulfment by phagocytic host macrophages, requiring rapid adjustment to changes in pH, oxygen levels, iron concentration, and other host molecules like immune cells. Macrophages are phagocytic monocytes that are among the first immune cell responders to pathogens in the tissue and blood[28]. Macrophages contain cell surface receptors (such as Dectin-1) that recognize fungal cell wall components[28]. Exactly how or why some individual *C. albicans* cells are able to survive in these diverse host environments, like that of a macrophage, while others die is still unclear. *C. albicans* is a dynamic fungus that is able to overcome host defense and anti-fungal drugs, motivating the need for new, potential fungal drug targets.

1.2 *Candida albicans* morphology

Candida albicans is commonly thought of as a dimorphic organism and an obligate commensal. However, it is now known that *C. albicans* can form at least six different distinct cell types, many of which correspond to the specific host niche from which *C. albicans* is occupying **(Figure 1.1)**.

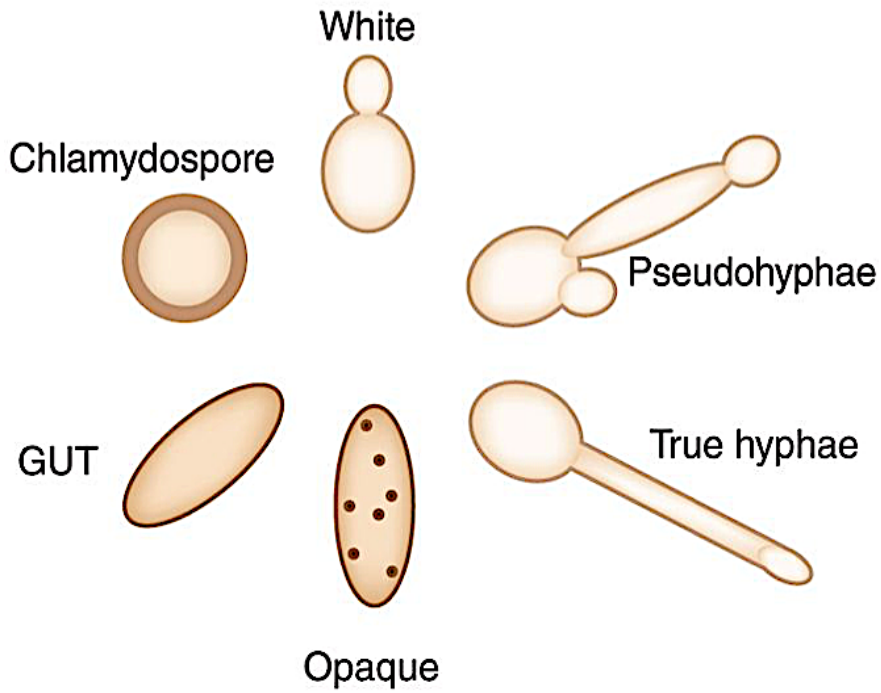

Figure 1.1: *Candida albicans* has at least 6 defined cell morphologies
Adapted from Gow, 2013. The white cell form is a budding yeast cell form and has been traditionally considered non-pathogenic. *C. albicans* can also form pseudohyphae and hyphae, both of which can invade host cells and tissue; hyphae are considered the pathogenic *C. albicans* cell form. Opaque cells are mating-competent, elongated and contain dimples on the cell surface. GUT cells were isolated after *C. albicans* intestine colonization in the. mouse; they are transcriptionally distinct from other cell types and are considered commensal and adapted to this specific host niche. Chlamydospores have thick cell walls and have been observed on hyphae and pseudohyphae; the biological role of these cells has not yet been determined.

Additionally, there is now evidence that *C. albicans* may be able to grow outside of the host, as it has been collected from ancient oak trees but was likely not deposited from host fecal matter[29]. The white *C. albicans* cell form was originally considered non-pathogenic. However, it's been discovered that white cell type can easily disseminate into the blood stream of the host, where it can then travel to organs like the brain, liver, and most commonly, the kidneys. In these organs, *C. albicans* is usually found in a filamentous form (hyphae or pseudohyphae), where the fungus is then able to penetrate the tissue **(Figure 1.2).**

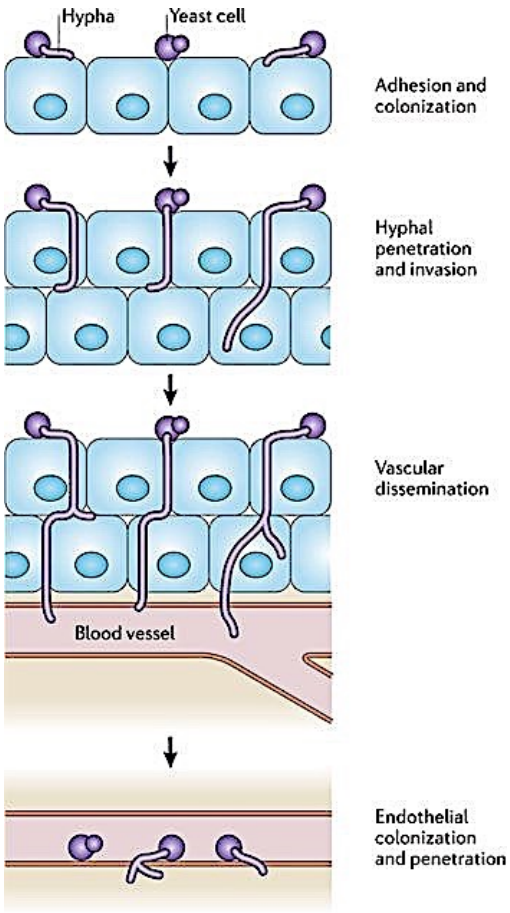

Figure 1.2: *Candida albicans* can disseminate through the host body through cell invasion
Adapted from Gow et al., 2012. *Candida albicans* can reversibly switch from budding yeast cell morphology that allows for adhesion and colonization of cell surfaces to hyphae formation, which allows for cell penetration and invasion. This enables *C. albicans* to reach blood vessels, where it can colonize endothelial cells and also disseminate through the body. *C. albicans* is most commonly found in yeast form in the blood stream.

White cells usually range between 5-10 μm in diameter and an oval shape capable of forming buds[25,30–32]. Pseudohyphae are more elongated than white cells and have borders at the septa[32],a and thus are composed of multiple cells. Hyphae are generally thinner than pseudohyphae, have parallel sides and no constrictions at the septa[30–32] and are technically one cell. Yeast and pseudohyphal cell growth is synchronized at the beginning of the cell cycle, whereas hyphal cell germ tube formation occurs prior the G1/S transition[32–35]. Both pseudohyphal and hyphal cells can are able to invade agar in laboratory settings[31]. The specific structural and chemical composition

difference between the pseudohyphal and hyphal forms, and whether or not the host immune system can discriminate between them, is not known[36]. *C. albicans* cells usually transition from yeast form, to pseudohyphal, to hyphal form. However, this transition is reversible and usually coincides with signals from the environment and/or host cells.

Opaque cells are different from white *C. albicans* cells, phenotypically and physiologically; they are larger, with about three time more volume and twice the mass of white cells. Despite this size difference, opaque cells contain a single nucleus and about the same amount of DNA as white cells. Compared to white cells, opaque cells flatter and darker in color with dimples on the cell surface. Opaque cells are considered the mating competent cell type, as they mate 10^6 times more efficiently than white cells[37]. Interestingly, white cells are more frequently phagocytosed than opaque cells; opaque cells do not filament and thus are less likely to activate neutrophils, despite physically interacting with them[38]. However, this difference is abrogated when the cells were incubated in gel matrix and dendritic cells phagocytosed and both cell types at a similar rate[38]. Macrophages from mice and *Drosophila* have been reported to preferentially phagocytose white cells over opaque cells[39]. The white to opaque switch is epigenetic, heritable and controlled via a transcriptional feedback loop[40–45]. It can be effected by environmental cues like temperature, carbon source and carbon dioxide levels[38,40,46,47].

Gastrointestinally induced transition (GUT) cells are a newly discovered *C. albicans* cell type[17]. Using a model of candidiasis in the gastrointestinal tract of mice that were otherwise healthy but contained high titers of yeast in the gut for three months, a strain of *C. albicans* was isolated that constitutively overexpressed *WOR1*, a master switch required for the white opaque transition[44]. In

a competition experiment, GUT cells outcompeted a wild-type strain in the gut of mice beginning at 10 days after inoculation[17]. GUT cells are phenotypically distinct from other *C. albicans* cells types; they are white in cell type when propagated *in vitro*, but switch to a darker color when grown inside of the host gut. GUT cells also do not respond to mating pheromone, as white cells do. GUT cells are also transcriptionally distinct when grown at room temperature in liquid yeast media compared to white cells; GUT cells upregulated genes involved catabolism of N-acetylglucosamine and fatty acids, while downregulating genes involved in iron acquisition and glucose catabolism. This expression program makes sense, because iron bio-availability is low in the gastrointestinal tract and glucose availability differs depending on intestinal location. However, gut microbiota can release glucans into the intestine lumen, including N-acetylglucosamine, which is found in the small and large intestine[48–51]. This metabolic shift is likely key to the ability of the GUT strain to outcompete the wild type strain of *Candida albicans* in the mammalian gut[17].

1.3 *Candida albicans* and host cell interactions

Interactions between fungal pathogens and the host innate immune system are critical to determining the course of infection. Phagocytic cells, including macrophages and dendritic cells, are key players in the recognition of and response to fungal infections[52]. *C. albicans* is considered an obligate commensal organism, as it has not been isolated outside of the host, until recently[29]. It can be found as a commensal resident of the skin, gastrointestinal system, and urogenital tract[53]. Therefore, *C. albicans* interacts with a variety of host cell types prior to and during pathogenesis. Host cells recognize pathogen associated molecular patterns (PAMPs) on the cell wall surface of *C. albicans* and other microbes. Host cells contain specific receptors for mannans, glucans and chitin, examples of PAMPs that are found on the *C. albicans* cell wall (**Figure 1.3**).

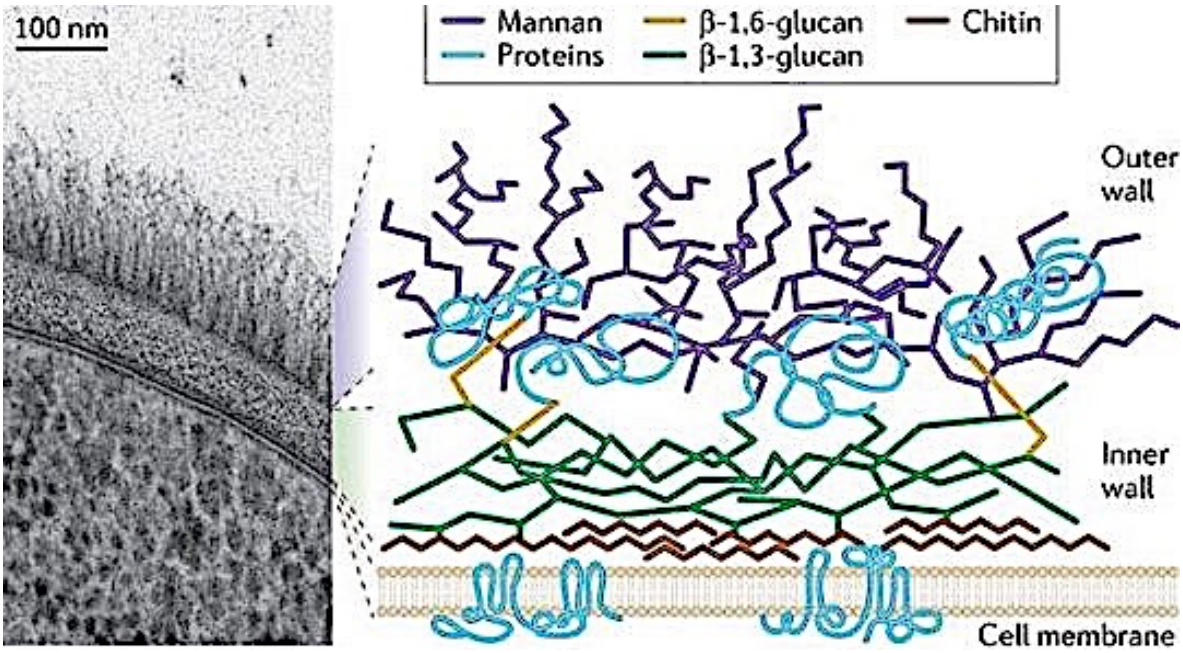

Figure 1.3: The *Candida albicans* cell wall is composed of two layers
Adapted from Gow et al., 2012. The left image is high resolution microscopy image showing the two main cell wall layers of *C. albicans*. The right illustration is a depiction of both *C. albicans* cell wall layer components. The outer layer contains mannans covalently linked with other proteins to form glycoproteins. The inner layer contains the polysaccharides chitin and glucans, which retain the cell wall rigidity and cell shape.

Candida albicans cell shape can determine which pattern recognition receptors (PRRs) on host cells are able to directly detect the fungus[32]. The main PRRs in mammalian cells include C-type lectin receptors (CLRs), Toll-like receptors (TLRs), NOD-like receptors (NLRs) and RIG I-like receptors (RLRs), each of which recognize and bind specific *C. albicans* cell well components (**Figure 1.4**). Specific host cell receptors can also recognize fungal cell presence indirectly, through binding of antibodies to fungal cells or receptor binding to fungal DNA or RNA[32]. *C. albicans* can withstand harsh host environments, including the macrophage phagosome, by regulating metabolic and cell morphology pathways[16,54]. Macrophages are some of the first responders to the site of fungal infections and the most adept at phagocytosing *C. albicans*.

Macrophages contain many cell surface PRRs, including: mannose receptor (MR), Dectin-1, Dectin-2, Minicle, DC-SIGN, Galectin-3, TLRs and NLRs[32].

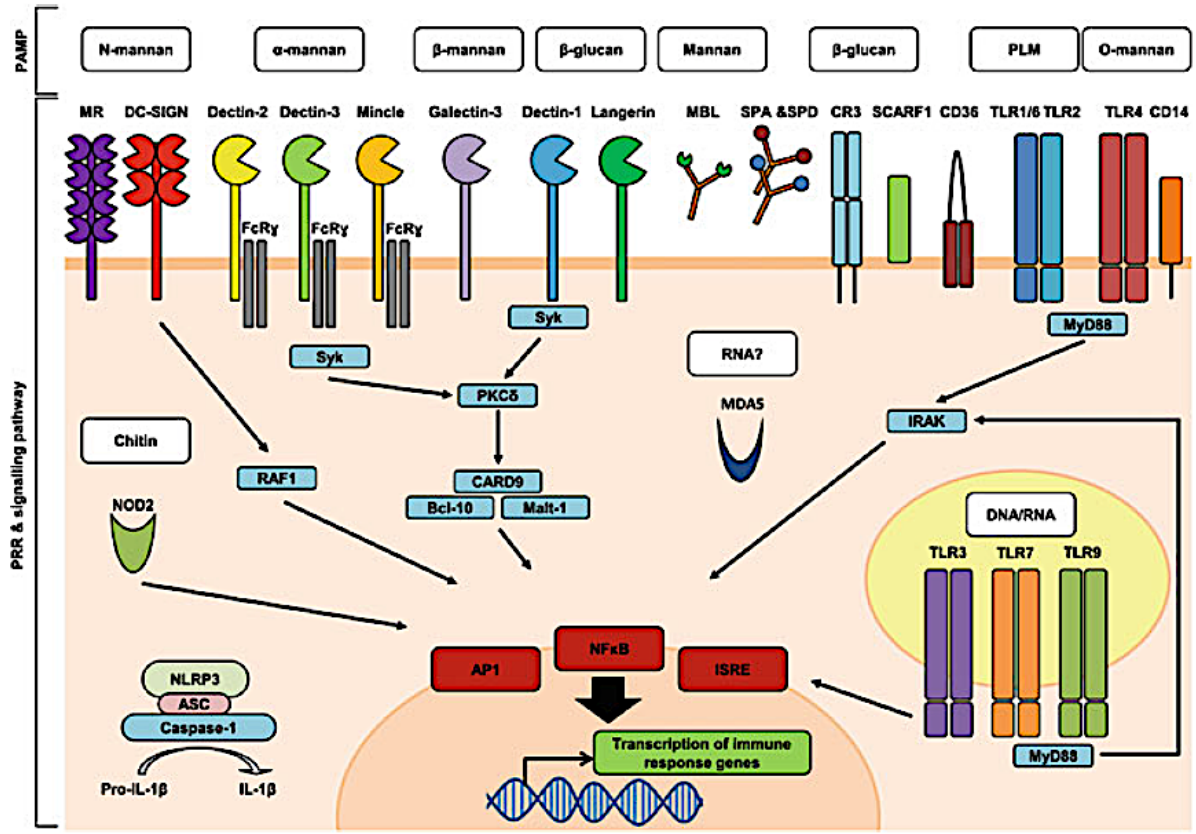

Figure 1.4: *Candida albicans* pathogen associated molecular patterns (PAMPs) and mammalian host cell pathogen recognition receptors (PRRs)
Adapted from Prasad et al., 2017. Depiction of PAMPs on *C. albicans* and the host PRRs which recognize them. The binding of PAMPs to PRRs can lead to signaling cascade in the host cell that elicits both pro- and anti-inflammatory immune responses.

Dectin-1 (encoded by *CLEC7A*) is expressed on the surface of macrophages, dendritic cells and neutrophils[32,55]and primarily recognizes the ß-glucan components of the *C. albicans* cell wall[55]. The cell wall glucans are usually masked when *C. albicans* cell morphology takes on the white, non-hyphal form. However, when filamentation occurs, the ß-glucans become unmasked. Dectin-1 binding of *C. albicans* leads to induction of TH17 cell differentiation, as well as reactive oxygen species (ROS) and pro-inflammatory cytokine production. In humans, a single nucleotide

polymorphism (SNP) in *CLEC7A* led to decreased production of pro-inflammatory cytokines during *C. albicans* infection but did not lead to candidemia because phagocytosis and *C. albicans* killing was not decreased in these patients[32,56]. Similarly, bone-marrow transplant patients with this *CLEC7A* SNP had higher oral and mucosal *C. albicans* burden and vulvovaginal candidiasis; while these patients also did not suffer from candidemia, the higher fungal burden did increase the number and dosage of fluconazole treatments needed to control fungal growth[32,57,58].

Like Dectin-1, Dectin-2 (produced by the gene *CLEC4n*) is expressed on macrophages and other immune cells, such as dendric cells and neutrophils[32]. Dectin-2 specifically recognizes *Candida* hyphae and the outer layer of the *Candida* cell wall which is high in mannose[32,59,60]. Similar to Dectin-1, Dectin-2 binding of *C. albicans* induces pro-inflammatory cytokine production, recruitment of other immune cells, Th17 cell responses and phagocytosis of *C. albicans*[32,61–63]. While Dectin-2 is relatively less studied than Dectin-1, it's been demonstrated that mice lacking Dectin-2 are more susceptible to candidemia[32,61,62].

While macrophages directly control fungal proliferation and coordinate the response of other immune cells, the outcomes of these interactions are heterogeneous; some *C. albicans* cells are effectively killed by macrophage engulfment whereas others evade or survive macrophage interactions and escape from the host cells[64] (**Figure 1.5**).

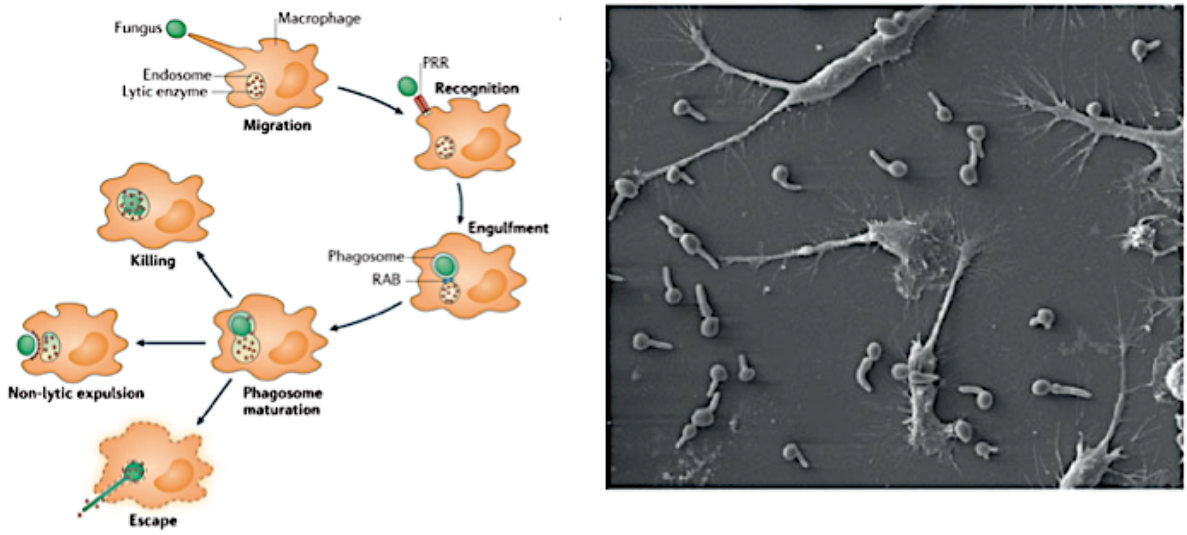

Figure 1.5: *Candida albicans* and macrophage interaction outcomes vary
Adapted from Erwig and Gow, 2016. On the left, a schematic illustrates the way that fungus, including *C. albicans*, interacts with macrophages, leading to fungal death, host death or pathogen escape. On the right, a scanning electron microscope image of macrophages interacting with filamenting *C. albicans*.

1.3.1 Secreted Aspartyl Proteinases (*SAP*) genes in *Candida albicans*

There are 21 predicted Secreted Aspartyl Proteinases (*SAP*) genes in *Candida albicans,* ten of which produce protein products that have been studied to varying degrees. These proteins contain conserved aspartate residues that bind water molecules to hydrolyze target substrates. Increased Sap protein production and activity has been implicated in oral candidiasis[65,66], vaginal candidiasis[67–70] and pathogenicity in animal models[71–74]. Saps1-8 are secreted from *Candida albicans* into the environment, while Sap9 and Sap10 contain glycophosphatidylinositol (GPI) groups attached to the C-terminus of the proteins, indicating that they are likely anchored to the

cell membrane[75]. Sap proteinases 1-10 appear to be expressed differentially, depending on the host environment. These proteases can act to break down host proteins involved in immune recognition, destroy tissue barriers and use parts of the degraded proteins for growth[76].

There are conflicting reports in the literature about the roles of Sap protein functions. For example, it has been demonstrated that individual sap1, sap2 and sap3 disruption mutants cause less tissue damage[77], are less adherent to host cells[78] and are less virulent than wild type *C. albicans* in mucosal models of infection[79]. However, in systemic infection models, it has been demonstrated that sap1, sap2 and sap3 mutants showed no difference in adherence to host cells, and that only sap2 mutants caused less tissue damage that wild type[80]. In peritoneum models of infection, a sap4-6 mutant caused decreased host tissue damage and invasion while sap1, sap2 and sap3 individual disruption mutants did not[81]. However, when sap1, sap2 and sap3 individual mutants and the sap4-6 mutant were injected into live animals, all strains led to attenuated virulence[82,83]. **Table 1.2** summarizes existing *C. albicans* sap mutant and host cell interaction studies.

Table 1.2 A summary of *Candida albicans* sap mutant and host cell interaction findings
Adapted from Naglik et al., 2003.

Infection Type	Main Findings
Mucosal	sap1, sap2 and sap3individual mutant strains cause less tissue damage than wild type A sap4-6 triple mutant strain causes the same level of damage as wild type.
Systemic	sap1, sap2 and sap3 individual mutant strains show no difference in tissue damage and invasion. A sap4-6 triple mutant strain causes less damage and invasion.
Systemic	sap1, sap2 and sap3 individual mutant strains are less lethal. A sap4-6 triple mutant strain shows the most attenuation.
Macrophages	A sap4-6 triple mutant strain is eliminated 53% more *in vitro.*

Sap proteins can function at various pH levels. For example, Sap2 and 3 are able to function at low pH levels (4.0 and 2.0, respectively), while Saps 4, 5 and 6 function optimally at pH 7.4[75,84]. This may explain some of the differential phenotypes measured by various research groups, when performing studies in a variety of complex host environments. Work has been done to determine where these proteinases are expressed within the host. Previous work has shown that Saps 4-6 are mostly expressed by *Candida albicans* in phagosomes of murine peritoneal macrophages and neutrophils, and it was postulated that these proteinases may contribute to virulence by resisting phagocytic attack[75,84]. It has also been demonstrated that sap4-6 mutants are hypersensitive to phagocytic killing by macrophages[84]. However, it has been noted that mutants lacking individual *SAP* genes rarely exhibit a full avirulent phenotype in infection models[75]. While it is certainly possible that other virulence factors, proteinases, host environmental factors or even other microflora can contribute to the pathogenicity of *Candida albicans*, it is believed that the dynamic behavior of these proteins warrants further investigation, especially during the early stages of immune defense.

1.4 *Candida albicans* and host cell RNA-sequencing

Previous studies of *C. albicans* and immune cell interactions in bulk populations have identified key pathways by characterization of either the fungal or host transcriptional response during these interactions[16,85,86]. From this work, it was found that phagocytosis of *C. albicans* by macrophages leads to a shift in *C. albicans* metabolic programming, from expression of genes involved in glucose catabolism, to upregulation of genes involved in catabolism of fatty acid and those

involved in the glyoxylate cycle[16] (**Figure 1.6**). Additionally, human macrophage precursor cells upregulate genes involved in the Type-I IFN response specifically after *C. albicans* exposure[85].

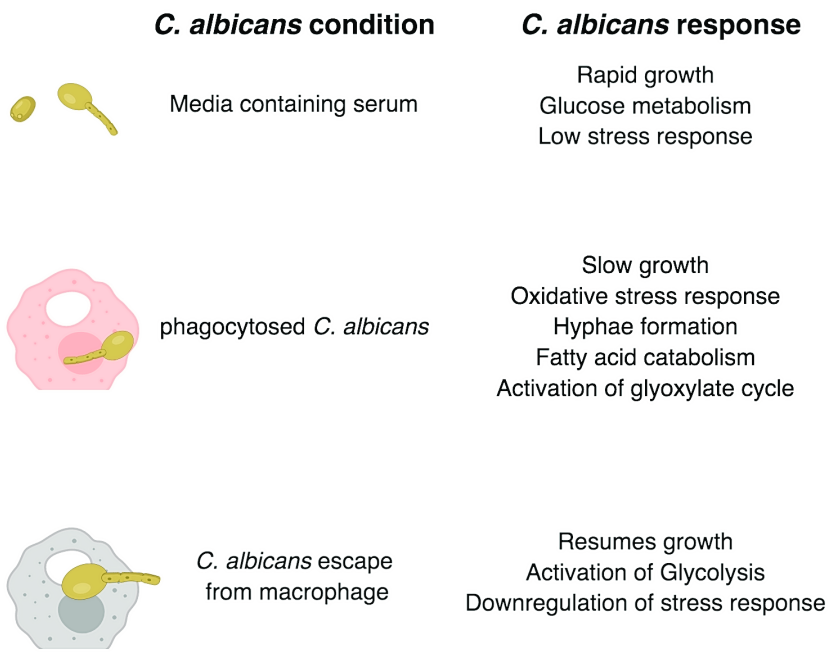

Figure 1.6: Summary of *Candida albicans* transcriptional response to macrophages
Adapted from Lorenz et al., 2004. *C. albicans* growing alone in mammalian cell culture media containing serum leads to rapid growth, glucose metabolism and downregulation of genes involved in the stress response. Both white and hyphal cell yeast forms can be found in media containing serum. Phagocytosed *C. albicans* slow growth, shift their metabolic program to be able to break down nutrients available in the host environment and begin hyphal morphogenesis. When *C. albicans* escapes from micrographs by piercing the host cell wall, growth and glycolysis resume and stress response genes are downregulated.

More recently, dual transcriptional profiling of host-fungal pathogen interactions has also examined populations of interacting cells[87–90]. This work was carried out in *C. albicans* exposed to oral epithelial and endothelial cells[88], neutrophils[89] and dendritic cells[90]. Also, the transcriptomes of *C. albicans* and whole murine liver and kidney tissues were profiled simultaneously[87]. However, even in a clonal population of phagocytes, many immune cells do not engulf any fungal cells, while others can phagocytose up to ten fungal cells[25]. These studies did not parse out heterogenous infection fates prior to gene expression profiling. To date, bulk

approaches measure the average transcriptional signal of millions of cells, obscuring any potential gene expression differences in host or fungal pathogen that vary in specific infection fates.

1.5 The use of single-cell RNA-sequencing to characterize heterogeneity in sample populations

Single-cell RNA-sequencing (scRNA-Seq) has highlighted the substantial variation in gene expression between cells within stimulated or infected immune cell populations[91–93]. For example, scRNA-Seq revealed that a subset of macrophages exposed to bacterial stimuli displayed a strong interferon-response, which was associated with cell surface variation between different bacteria[92]. A recent study measured host and pathogen gene expression in single host cells infected with the bacteria *Salmonella typhimurium*; however, the low number of pathogen transcripts detected per cell was only sufficient for analysis of sets of co-regulated genes[94]. To date, parallel transcriptional profiling of single host cells and fungal pathogens has not been reported.

Bulk RNA-sequencing technologies provide an average expression level for any given transcription of all cells in the sample population. However, the maturation of single-cell RNA sequencing technologies has led to the discovery that substantial variation in levels of gene expression in a sample population exists for certain cell types. For example, primary, bone derived dendritic cells stimulated with the gram-negative bacterial outer-membrane component lipopolysaccharide (LPS) were studied at both the population and single cell levels. Shalek, et al. found that while the average expression of the population yielded an upregulation of genes involved in the immune response, a small fraction of cells from within the population were

expressing extremely high levels of the transcript while the majority of the cells in the population where expressing lower levels.[91] It was determined that signal traditionally classified as low *population-level* expression of some transcripts may truly reflect high transcript expression in only a *small subset* of the cells within the sample population[91]. Similar results were observed in dendritic cells stimulated with *in vitro* generated viral-like double stranded RNA and a synthetic molecule made to mimic bacterial lipopeptides[95]. Genes exhibiting bimodal expression patterns early during infection tended to be those involved in the immune response[91,95]. However, Shalek et al. note that this degree of variation had not reported in single-cell RNA-sequencing studies of unstimulated mouse embryonic stem cells, fibroblasts[96], dividing human cells[97] or in mid-log phase yeast cells[98].

There are two main single-cell RNA-sequencing strategies that are commonly utilized at this time. The first is the Smart-seq2 method[99] (**Figure 1.7)**, wherein single cells are lysed in a buffer that contains free dNTPs and oligo-dTs with a universal 5' sequence. Polyadenylated mRNA transcript tails (present in eukaryotes, including mammals and yeast) are captured by oligos with strings of various lengths of thymine bases and then are reverse-transcribed. The Moloney Murine Leukemia Virus (M-MLV) enzyme used for reverse transcription (RT) allows both template switching and terminal transferase activity[100]. When the enzyme arrives at the end of the mRNA strand during RT, the terminal transferase activity of the enzyme adds non-templated cytosine residues to the cDNA. The cytosines anchor the locked nucleic acid (LNA) containing template-switching oligo (TSO). The TSO has two riboguanosines and a modified guanosine which produce a LNA as the last base at the 3' end[100]. LNAs also have greater thermal stability, thus have a stronger bond with the complementary sequence and can bind strongly the 3' end of untemplated cDNA[100,101]. The

TSO guanine residues base pair with the M-MLV deposited cytosines and the reverse transcriptase then uses the TSO as a template to complete 1st strand synthesis. With this method, synthesis of fragments at the 5' end is promoted and this overcomes the 3' bias observed with other RNA sequencing protocols that rely only on polyadenylated mRNA sequencing capture[100]. Sequencing libraries are made from cDNA using Illumina's Nextera kit. This method uses a transposon to randomly fragment cDNA into roughly 300 base pair sequences and simultaneously add synthetic molecular sequences (adapters) that are complimentary to sequencing primers. PCR is the completed with primers that are complementary to the adapters and add unique molecular indexes to each cell, so that samples may be multiplexed prior to sequencing. This strategy can also be used on populations of cells, with minor adjustments to reagent volumes. Using this method, 96 to 384 single cells are processed at one time.

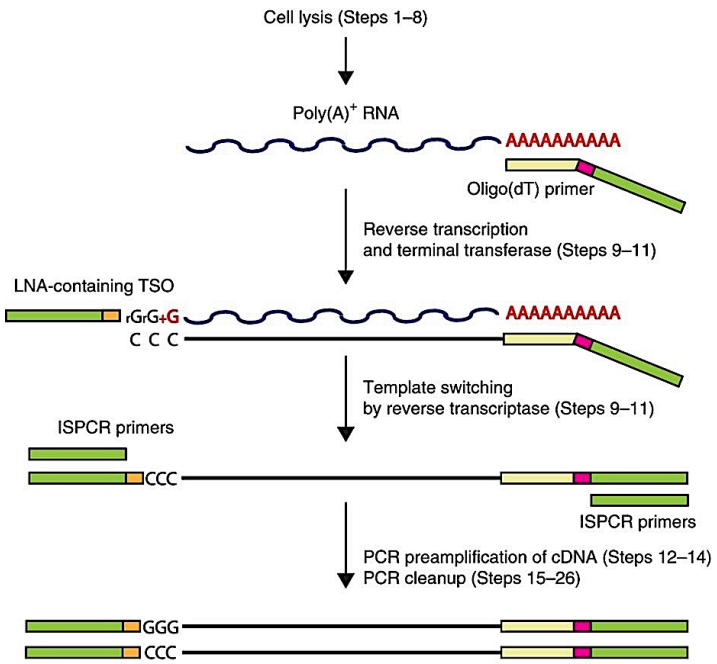

Figure 1.7: The Smart-Seq2 cDNA synthesis workflow
Adapted from Picelli et al., 2014. Single cells are lysed in a buffer that contains free dNTPs and oligo(dT)-tailed oligonucleotides with a universal 5' sequence. Reverse transcription is performed, which adds untemplated cytosines to the cDNA 3' end. A template-switching oligo (TSO) modified with locked nucleic acid (LNA) is added. After the first-strand reaction, the cDNA is amplified using ISPCR primers.

The second and more recently developed, common single-cell RNA-sequencing strategy is the Chromium Single-cell Gene Expression solution (10x genomics; **Figure 1.9**)[102]. This protocol enables users to load between 100-10,000 cells per sample on one of eight wells of a microfluidic chip. The chip contains reagents for cDNA synthesis, barcoded oligo-dT beads which also contain a unique molecular identifier (UMI), ensuring that each transcript receives a distinct synthetic molecular barcode. Cells and reagents are mixed together at an oil droplet interface, and the majority (~80%) of oil droplets contain one cell and one bead, or gel bead in emulsion (GEM). RT takes place inside of each GEM, resulting in a pool of GEMs that each contain uniquely barcoded cDNA. After the emulsion is broken, cDNA is amplified and libraries are constructed. Illumina sequencing "Read 1" contains the cDNA insert while "Read 2" captures the UMI. Sequencing

"Index reads" contain the sample indices and cell barcode information. This technique enables the parallel capture of up to ten thousand single cells per sample in up to eight samples.

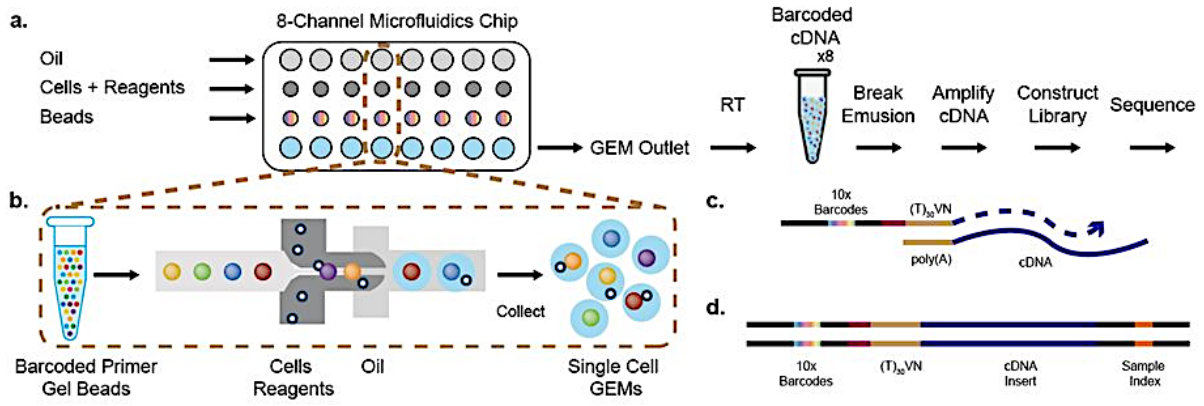

Figure 1.8: Single-cell RNA-sequencing using 10x GemCode™ technology
Adapted from Zheng et al., (2016). **(a)** Each microfluidic chip contains reagents and channels to process eight samples per chip. All cells from one sample are combined with reagents in one channel and mixed with oil and gel beads from additional channels to form gel beads in emulsions (GEMs). RT takes place inside of the GEMs and cDNA from each cell has a unique barcode. After breaking the emulsion, cDNA is amplified and libraries are created that are compatible with Illumina's sequencers. **(b)** Each gel bead contains barcoded primer and oligos. Beads are mixed with cells and reagents in oil interface and one bead and one cell are captured in an oil droplet, and these GEMs are collected in the microfluidic chip outlet. **(c)** Gel beads contain barcoded oligonucleotides consisting of Illumina sequencing adapters, 10x barcodes, UMIs to identify each transcript and oligo-dTs. (d.) Each cell from a sample contains also contains a unique sample index, allowing multiple samples to be pooled prior to sequencing. During analysis, transcripts are assigned back to the cell of origin based on this sample index.

However, *Candida albicans* cells do not lyse in traditional mammalian cell lysis buffers without the use of beads to break open the cell wall or some amount of freeze-thaw cycles. Currently, 10x Genomics does not support using this technology to measure gene expression in fungal cells.

Recently, single-cell RNA-sequencing in yeast has been reported[103]. However, this technique required ethanol fixation of yeast cells for efficient cell wall lysis. Ethanol has been reported to reduce the signal of some fluorescent reporters, such as GFP[104], and even low concentrations of ethanol can be toxic to mammalian cells[105]. Therefore, this method may only be appropriate for

experiments utilizing non-fluorescently labeled yeast or when host cell gene expression information is not required.

1.6 Gut organoids to study *Candida albicans* commensalism and pathogenicity

Yeasts are commensal resident of ~70% of the small intestines of all humans, and the majority of these are *Candida* species[22]. The most common fungal species in the human gut is *Candida albicans*[106,107]. Importantly, *C. albicans* infections can originate from strains that were once commensal residents of the patient gut, rather than from exposure to a new, more pathogenic strain[108]. Several studies have linked antibiotic usage to increased burden and lethality of *C. albicans* infections[109]. For example, in a study of antibiotic and non-antibiotic treated hamsters, *C. albicans* was found to be present in the gut and organs at higher levels in antibiotic treated animals, and it is presumed that the resident microflora of the non-antibiotic treated animals were protective[110]. However, patients on a long-term antibiotic regimen may not have the benefit of dynamic microflora in their gut.

It was recently discovered that exposure of a *C. albicans* lab strain to the mammalian gut caused increased expression of the Wor1 transcription factor in a small portion of the yeast cells recovered from the host[17]. These cells were phenotypically distinct from white cells and named Gastrointestinally induced transition cells (GUT) cells (**Figure1.1**). To investigate finding further, Pande et al. created a *C. albicans WOR1*-overexpression strain ($WOR1^{OE}$) to mimic the GUT cells. Using a mouse model of gastrointestinal candidiasis in which animals remain healthy, despite persistent infection with high titers of yeast[111] and an antibiotic regimen, these $WOR1^{OE}$ cells outcompete wild type cells after ten days. Microarray gene expression analysis revealed that

WOR1^{OE} cells upregulated genes involved in catabolism of fatty acids and N-acetylglucosamine (by-products of commensal bacteria found in the gut) and downregulated genes involved in biological adhesion and iron acquisition. However, prior to this microarray gene expression analysis both the wildtype and *WOR1*^{OE} strains were sampled in isolated yeast media culture, at room temperature sans host interaction[17]. While *WOR1*-overexpression was sufficient to maintain the new yeast cell form phenotype, these cells were less fit in traditional yeast culture conditions compared to the host. Growth conditions like temperature and media type have been shown to alter *C. albicans* morphology. For example, *C. albicans* grown in media supplemented with serum and grown at body temperature (37°C) will begin filamenting, even in the absence of host cells[112]. For this reason, it is important to recapitulate the host environment as much as possible when performing *in vitro* experiments, to capture biologically relevant morphological and transcriptional profiles.

The cell types in the small intestine vary greatly in terms of function and fate[113]. The small intestine is distinct from the large intestine, in that it contains villi to increase absorptive capabilities and crypts, which reside at the base of these villi. There are also several types of differentiated cells in the small intestine including absorptive enterocytes, goblet cells, which are responsible for secreting mucous, and paneth cells, which reside in the crypts and secrete cytokines and antimicrobial enzymes that defend the tissue from pathogens (**Figure 1.9**). The small intestine also contains M cells that reside in Peyer's patches (patches of lymphoid follicles) of the ileum of the small intestine and can phagocytose antigens[114]. The small intestine retains the capacity to harbor stem cells, which reside in the base of the crypts, because the small intestine epithelial cells turnover frequently; it's estimated that turnover occurs every 3-5 days in mice and humans[115].

Stem cells produce progenitor cells called transit-amplifying cells, which move up the crypt toward the villi. During transit, they differentiate and exit the crypt to the villi. Paneth cells remain at the base of the crypts and persist for six to eight weeks. Adjacent to Paneth cells are "Crypt Based Columnar" (CBC) cells, the stem cells at the base of the crypt. The gene for the Leucine rich repeat-containing G-protein coupled receptor 5 (Lgr5) is used as marker for these stem cells, as it encodes a component of the WNT signaling pathway, a pathway indicated in cell renewal and stem cell control[116].

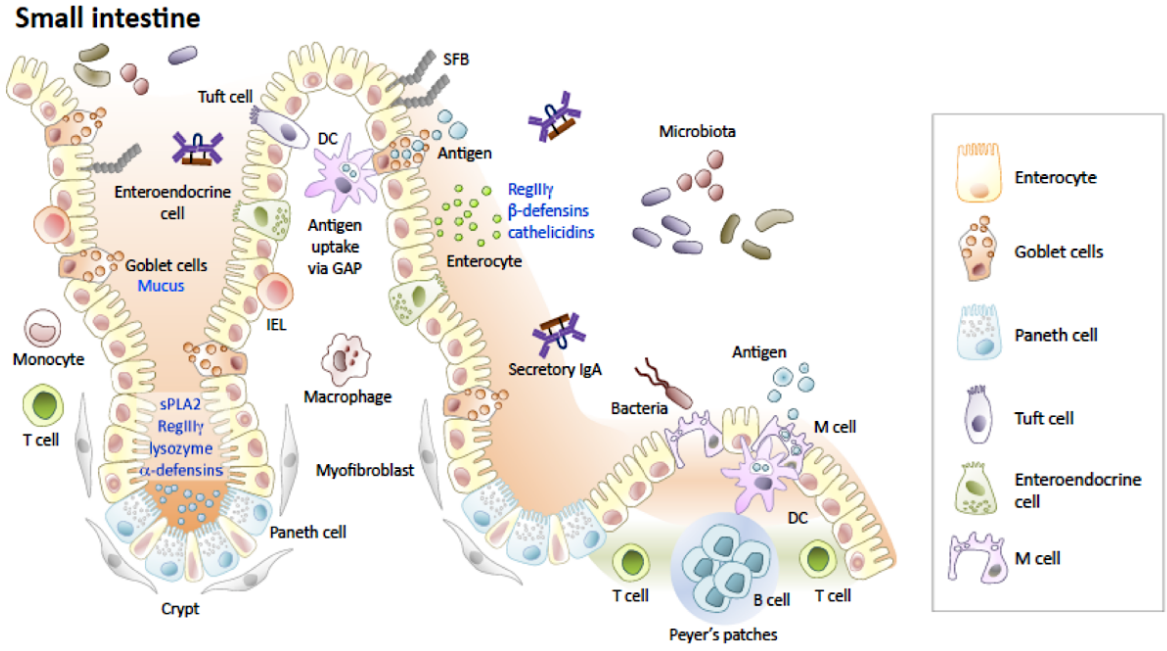

Figure 1.9: Diagram of small intestine cell types and cell type response to antigen
Adapted from Allaire et al., 2014. Schematic representation of the crypt and villus domains, as well as the cell types present and the relative location of these cells in the small intestine. Antigen uptake and the immune response are also depicted. Enterocytes are found in the crypt–villus axis secrete antimicrobial peptides (RegIIIg, b-defensins, cathelicidin). Paneth also can secrete antimicrobial peptides (lysozyme, a-defensins, sPLA2) and are located at the base of the crypt. M cells and Goblet cells (mucus producing cells) transport antigen from the lumen to dendritic cells (DC); this is known as goblet cell antigen passage (GAP). M cells are located in Peyer's patches, which are considered the immune sensors of the small intestine. Peyer's patches also contain pockets of virus killing T cells and antibody producing B cells, leading to the production of secreted IgA against specific pathogens. Intraepithelial lymphocytes (IELs) are a subpopulation of T cells located between epithelial cell. Segmented filamentous bacteria (SFB) can induce TH17 cells in the small intestine. The small intestine is the largest reservoir of macrophages in the human body and but also contains other phagocytes, such as monocytes. Myofibroblasts help repair epithelial stems cells after antigen damage at the crypts.

A recent single cell survey of the mouse small intestine epithelium was performed[117]. In addition to defining the cell types and proportion of cell types present in the small intestine during homeostasis, it was discovered that infection with pathogens can alter cell type proportions. Enterocytes are expanded upon infection in the small intestines of mice infected with *Salmonella enterica* and reduced upon *Heligmosomoides polygyrus* infection compared to uninfected controls, while tuft cells and goblet cells were expanded only upon infection with *H. polygyrus*[117] **(Figure 1.10).**

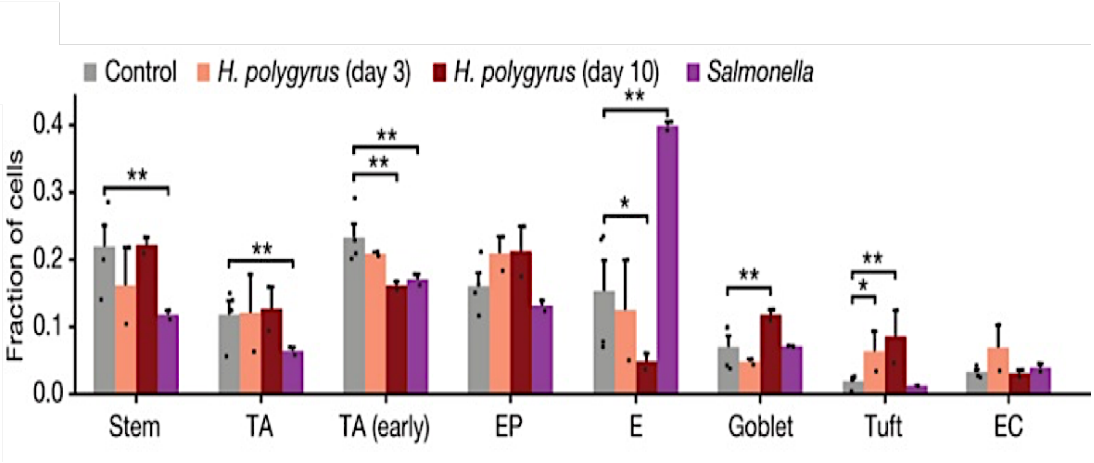

Figure 1.10: Small intestine epithelial cell type proportions shift in response to pathogen
Adapted from Haber et al., 2017. Mice were infected with either *Salmonella* (48 hours) or *H. polygyrus* (3 days and 10) and scRNA-seq was performed on infected cells and uninfected controls. The percentage of each cell type was measured based on gene expression and unsupervised clustering. TA, transit amplifying; EP, enterocyte progenitor; E, enterocyte; EC, enteroendocrine.

LGR5+ cells can be isolated from the small intestine crypt and used to initiate organoid culture with a Matrigel substrate and media containing R-spondin (a ligand for LGR5+ cells), and other cytokines[118,119]. Importantly, small intestine organoids produce villi and crypts containing paneth cells, goblet cells, M cells, enterocytes, Lgr5+ stem cells, tuft cells and other cell types found in the gut *in vivo* (**Figure 1.11**) Additionally, clonal organoid systems developed from one LGR5+ stem cell have been used as successful grafts in mice with damaged colon, leading to functional

organs for a minimum of 6 months after transplantation[113,120]. Because the organoids recapitulate some of the three-dimensional structure and cell types of the small intestine and are functional *in vivo* and dynamically respond to infection, they are ideal for use in gut studies when *in vivo* experimentation is not feasible.

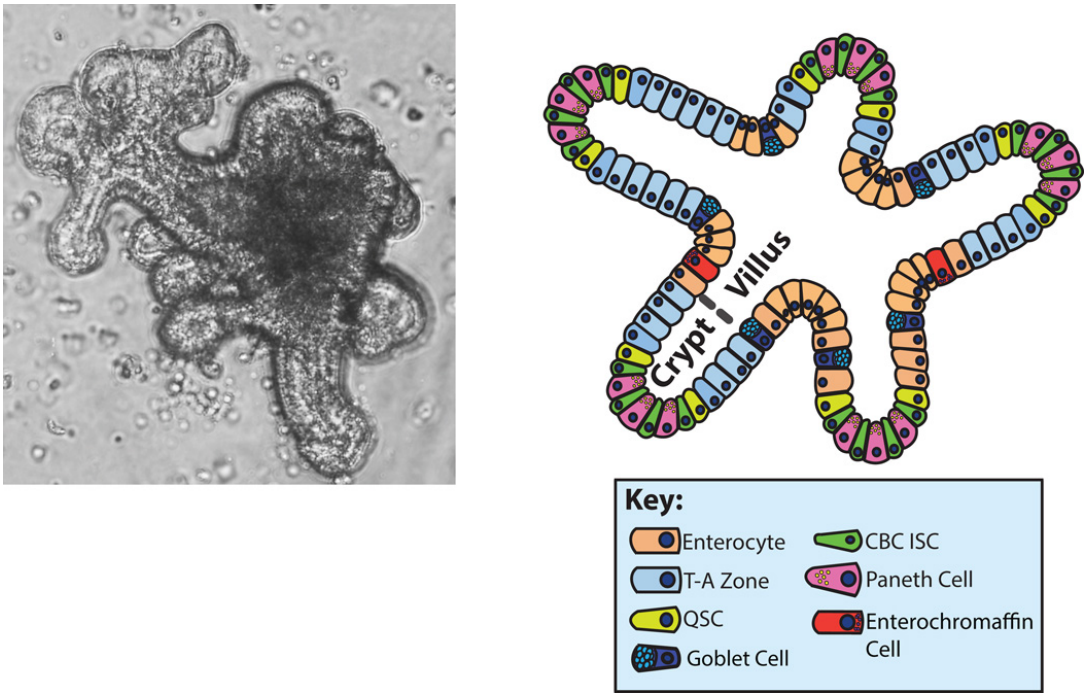

Figure 1.11: Murine small intestine epithelial cell organoids derived from stem cells
Adapted from Wallach et al., 2017. The left image is a microscopic picture of a mature small intestine organoid. The right image is a schematic depicting the crypt and villus domains of organoids, as well as the cell types present and the relative location of these cells in the organoid structure. T-A Zone, transit amplifying zone; QSC, quiescent stem cell; CBC ISC, crypt columnar base intestinal stem cell.

The ideal organoid tissue source for our studies would be human (mice have not been shown to be infected with *Candida albicans* in the wild), but patient-derived organoids are logistically difficult to obtain and are not as well characterized as murine-derived organoids. It's been demonstrated that that wild type *C. albicans* is capable of penetrating both human epidermal and intestinal

organoid models, while *C. albicans EFG1* and *CPH1* (transcription factors have previously shown to be important for filamentation and virulence in animal models[121,122]) single and double knockout mutants were unable to invade the matrices[123]. Thus, previous work supports the notion that intestinal organoids can be used in conjunction with *C. albicans* as a model system for studying infection[123,124].

Epithelial carcinoma and immortalized oral cell lines were examined to determine how the host discerns between pathogenic and commensal *C. albicans* forms[125,126]. Host cells demonstrated a "biphasic MAPK response" which was dependent on *C. albicans* morphology and levels of fungal burden. Activation of the NF-kB pathway and the first MAPK phase (c-Jun and mTOR activation) are likely triggered by fungal cell wall components (such as mannans, chitin and beta glucans), as these pathways were activated when host cells encountered non-pathogenic yeast *S. cerevisiae* and dead *C. albicans.* The second MAPK phase (MAPK phosphatase and c-Fos activation) is activated by hyphal *Candida albicans* and not by traditional yeast cell forms. In this work, a combination of techniques (western blotting, RNAi, qPCR) was used to elucidate the host genes involved in the discrimination between commensal versus pathogenic yeast forms.

However, to date, a transcriptome wide analysis and comparison of the host transcriptional signal after exposure to commensal and pathogenic *C. albicans* strains would provide a more complete picture of the host's response to this dimorphic fungus. Simultaneous capture of *Candida albicans* transcriptional signal may provide clues as to how the host and the fungus dynamically modulate transcription to respond to each other during both commensalism and infection.

Chapter 2: Coordinated host-pathogen transcriptional dynamics revealed using sorted subpopulations and single, *Candida albicans* infected macrophages

José F. Muñoz[1,*], Toni Delorey[1,2,*], Christopher B. Ford[1], Bi Yu Li[1], Dawn A. Thompson[1], Reeta P. Rao[1,2,#], Christina A. Cuomo[1,#]

[1]Broad Institute of MIT and Harvard, Cambridge, MA USA. [2]Worcester Polytechnic Institute, Worcester, MA USA.
*These authors contributed equally to this work
[#]Corresponding authors: Christina A. Cuomo cuomo@broadinstitute.org; Reeta P. Rao rpr@wpi.edu
https://doi.org/10.1038/s41467-019-09599-8

2.1 Abstract

The outcome of fungal infections depends on interactions with innate immune cells. Within a population of macrophages encountering *C. albicans*, there are distinct host-pathogen trajectories; however, little is known about the molecular heterogeneity that governs these fates. We developed an experimental system to separate interaction stages and single macrophage cells infected with *C. albicans* from uninfected cells and assessed transcriptional variability in the host and fungus. Macrophages displayed an initial up-regulation of pathways involved in phagocytosis and proinflammatory response after *C. albicans* exposure that declined during later time points. Phagocytosed *C. albicans* shifted expression programs to survive the nutrient poor phagosome and remodeled the cell wall. The transcriptomes of single infected macrophages and phagocytosed *C. albicans* revealed a tightly coordinated shift in gene expression co-stages and revealed expression bimodality and differential splicing that may drive infection outcome. This work establishes an approach for studying host-pathogen trajectories to resolve heterogeneity in dynamic populations.

2.2 Introduction

Interactions between microbial pathogens and the host innate immune system are critical to determining the course of infection. Phagocytic cells, including macrophages and dendritic cells, are key players in the recognition of and response to fungal infections[52]. *Candida albicans*, the most common fungal pathogen, can cause life threatening systemic infections in immunocompromised individuals; however, in healthy individuals, *C. albicans* can be found as a commensal resident of the skin, gastrointestinal system, and urogenital tract[53]. In addition, *C. albicans* can withstand harsh host environments, including the macrophage phagosome, by regulating metabolic and cell morphology pathways[16,54]. While macrophages directly control fungal proliferation and coordinate the response of other immune cells, the outcomes of these interactions are heterogeneous; some *C. albicans* cells are effectively killed by macrophage engulfment whereas others evade or survive macrophage interactions and persist in the host[64].

Previous studies of *C. albicans* and immune cell interactions in bulk populations have identified key pathways by characterization of either the fungal or host transcriptional response during these interactions[16,85,86]. More recently, dual transcriptional profiling of host-fungal pathogen interactions has also examined populations of cells[87–90]. Bulk approaches measure the average transcriptional signal of millions of cells, obscuring differences between infection fates. Even in a clonal population of phagocytes, many immune cells do not engulf any fungal cells, while others can phagocytose up to ten fungal cells[25]. Single-cell RNA sequencing (scRNA-Seq) has highlighted the substantial variation in gene expression between cells within stimulated or infected immune cell populations[91–93]. For example, scRNA-Seq revealed that a subset of macrophages exposed to bacterial stimuli displayed a strong interferon-response, which was associated with cell

surface variation between different bacteria[92]. A recent study measured host and pathogen gene expression in single host cells infected with the bacteria *Salmonella typhimurium*; however, the low number of pathogen transcripts detected per cell was only sufficient for analysis of sets of co-regulated genes[94]. To date, parallel transcriptional profiling of single host cells and fungal pathogens has not been reported.

To overcome these challenges, we developed an experimental system to isolate subpopulations of distinct infection outcomes and examined host and pathogen gene expression in sorted subpopulations and in single, infected macrophages. We focused on four distinct infection outcomes: (i) infected macrophages with live *C. albicans*, (ii) infected macrophages with dead, phagocytosed *C. albicans*, (iii) macrophages exposed to *C. albicans* that remained uninfected and (iv) *C. albicans* exposed to macrophages that remained un-engulfed. In addition to carrying out dual RNA-Seq on these subpopulations, we isolated single macrophages infected with *C. albicans* and adapted methods to measure gene expression of the host and pathogen to further resolve heterogeneity. By comparing the transcriptional profiles of *C. albicans* and primary, murine macrophages at both the subpopulation and single infected cell levels, we characterized the tightly coupled time-dependent transcriptional responses of the host and pathogen across distinct infection fates. We established that both host and pathogen gene expression can be measured from single cells; this revealed that genes involved in host immune response and in fungal morphology and adaptation show expression bimodality or changes in splicing patterns, variation that is important to consider in monitoring the dynamics of host-fungal pathogen interactions.

2.3 Results

2.3.1 Characterization of heterogeneous subpopulations *ex vivo* during macrophage and *Candida albicans* interactions

To capture infection subpopulations and more finely examine host and pathogen interactions, we developed a system for fluorescent sorting of *Candida albicans* with macrophages. We utilized a reporter to measure fungal cell status (live or dead) and infection status (engulfed or unengulfed). This construct, which constitutively expresses Green Fluorescent Protein (GFP) and mCherry, was integrated into *C. albicans* at the *NEUT5* locus (**2.5 Methods**); when *C. albicans* cells lyse in the acidic macrophage phagosome, GFP loses fluorescence upon the change in pH[93] whereas mCherry remains stable for up to 4 hours in this environment as visualized by microscopy. Primary, murine bone derived macrophages were stained with CellMask Deep Red plasma membrane stain. To study host-fungal pathogen infection stages at finer resolution, the *C. albicans* reporter strain was then co-incubated with primary bone-derived, stained macrophages and subpopulations were isolated using fluorescence-activated cell sorting (FACS) at time intervals (0,1, 2 and 4 hours; **2.5 Methods**; **Figure 2.1A**). These time points were selected to capture the early transcriptional changes of *C. albicans* in response to interactions with macrophages[16]. To examine gene expression, RNA of both host and fungal cells was extracted and adapted for Illumina sequencing using Smart-seq2 (**Methods**). Four major infection subpopulations were isolated by FACS: (i) macrophages infected with live *C. albicans* (GFP+, mCherry+, Deep red+), (ii) macrophages infected that phagocytosed and killed *C. albicans* (GFP-, mCherry+, Deep red+), (iii) macrophages exposed to *C. albicans* (GFP-, mCherry-, Deep red+) and (iv) *C. albicans* exposed to macrophages (GFP+, mCherry+, Deep red-; **Figure 2.1A**). The number of uninfected macrophages exposed to

C. albicans ranged from an average of 61% to 67%, while the number of uninternalized *C. albicans* ranged from 22% to 7% over the time course (**Figure 2.1B**). The number of infected macrophages varied from 11% to 29% over the time course; the largest increase in this population was observed between 0 and 1 hours post infection and then remained stable over 2 and 4 hours (**Figure 2.1B**). The number of macrophages infected with dead, phagocytosed *C. albicans* ranged between 0% to 3% over the time course (**Figure 2.1B**).

The number of RNA-Seq reads and transcripts detected for both host and fungal pathogen subpopulations was sufficient to profile parallel transcriptional responses (**Supplementary Note 1**). Based on alignments to a composite reference of both mouse and *C. albicans* transcriptomes (**Methods**), the fraction of mapped reads for host and pathogen was highly correlated with the size of the transcriptomes and percent of sorted cells for each subpopulation (*e.g.* 87% host and 13% fungus for macrophages infected with live *C. albicans*; **Figure 2.1C**, **S1B**; **Table S1**). In subpopulations of macrophages infected with live fungus, an average of 10,333 host and 4,567 *C. albicans* genes were detected (at least 1 fragment per replicate across all samples; **Figure S1A**; **Table S1**). We focused the differential expression analysis on subpopulations with high transcriptome coverage and highly correlated biological replicates (*e.g.* Pearson's $r > 0.85$, and > 6,000 and 2,000 transcripts detected in macrophages and *Candida*, respectively; **Figure S1B**). The detection of a high number of transcripts across sorted subpopulations supports that we have established a robust system for defining host and fungal pathogen gene expression profiles during phagocytosis.

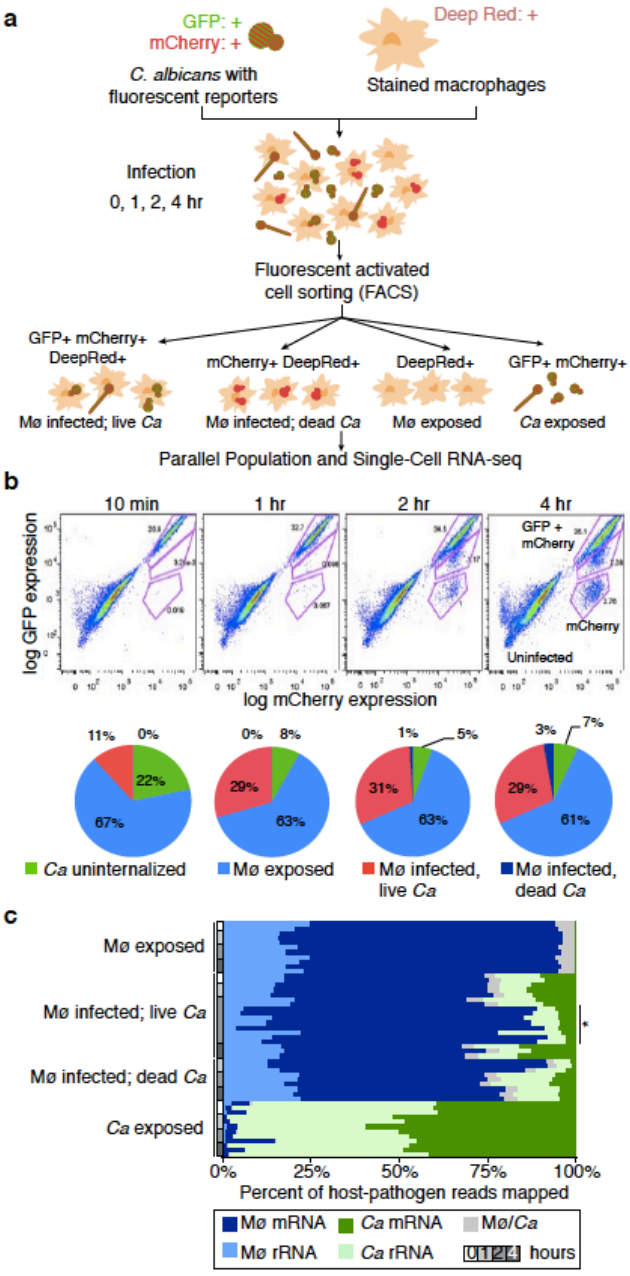

Figure 2.1: Heterogenous populations of macrophage-*Candida* encounters are captured by cell sorting and expression analysis

(a) Schematic representation of the experimental model, using primary, bone marrow-derived macrophages (BMDMs) incubated with *Candida albicans* reporter strain CAI4-F2-mCherry-GFP, sampling at time intervals, and sorting to separate subpopulations of interacting cells. **(b)** Bone marrow derived mouse macrophages (Mø) were incubated with the *Candida albicans* (*Ca*) reporter strain, then sorted at 10, 60, 120 and 240 minutes using fluorescence-activated cell sorting on the BDSORP FACSAria. Pie charts depicting the percent of cells sorted for each infection subpopulation. **(c)** Percent of host (*Mø*: macrophages) and pathogen (*Ca*: *C. albicans*) RNA-Seq reads mapped to the composite reference transcriptome, including mouse (GRCm38/mm10; mRNA and rRNA transcripts) and *Candida albicans* (CAI4-F2; mRNA and rRNA transcripts). Mø/*Ca* (gray color label) is the proportion of multiple mapped reads, *i.e.* reads that map to both the mouse and *C. albicans* transcriptomes, which were excluded from further analysis.

Next, we examined the major expression profiles in both the host and fungus. In *C. albicans* and macrophages, respectively, we identified 588 and 577 differentially expressed genes (DEGs; fold change (FC) > 4; false discovery rate (FDR) < 0.001) among all pairwise comparisons (**Methods**; **Figure 2.2**). To determine the major patterns of infection-fate specific or interaction-time specific, we used DEGs to perform Principal component analysis (PCA), then PC scores were clustered by *k-means* (**Methods**). PC analysis revealed that the transcriptional response in infected and exposed macrophages primarily varied over time and were highly similar at each time point between these subpopulations (**Figure 2.2C**). While exposed and phagocytosed *C. albicans* subpopulations also varied over time, these two populations appeared more distinct as described below (**Figure 2.2A**). These findings highlight how cell sorting can be used to separate different infection fates and distinguish gene expression signatures within populations of host and pathogen cells.

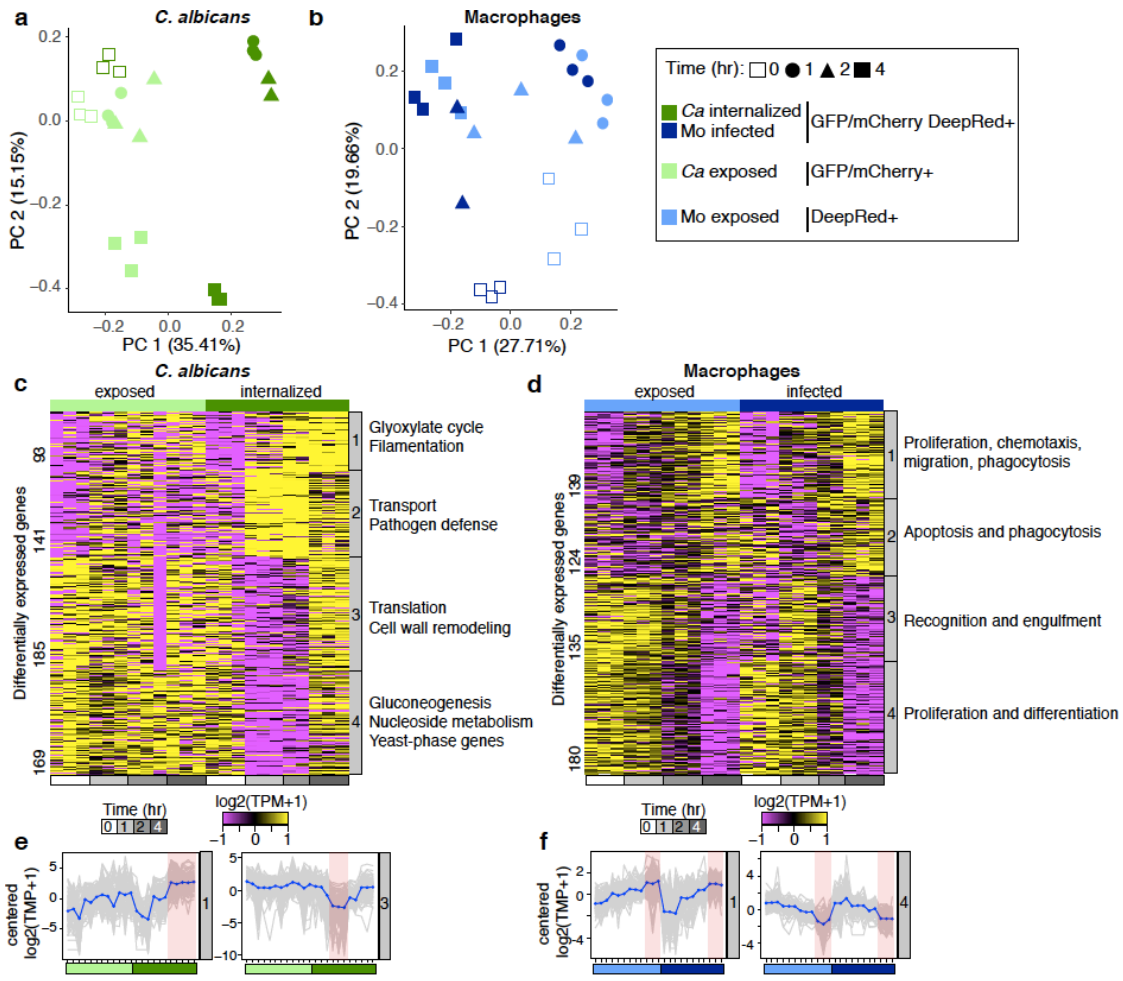

Figure 2.2: Subpopulations of infection outcomes reveal dynamic parallel host-pathogen orchestration of transcriptional response

Principal component analysis (PCA) using the transcript abundance for the pathogen (**a**; *C. albicans*) and the host (**c**; macrophages) using significantly differentially expressed genes (DEGs; FC >4, FDR < 0.001) in live *C. albicans*-phagocytized macrophages (GFP+ mCherry+ Deep Red+) compared to all other conditions. Contributions of each infection outcome replicate (points) to the first two principal components (PC1 and PC2) are depicted. The projection score (red: high; blue: low) for each gene (row) onto PC1 and PC2 revealed four clusters in both host and pathogen. For each projection score cluster, immune response genes (macrophages) and functional biological categories (*C. albicans*; deducted from GO term enrichment analysis) are shown. (**b**) Transcriptional response in subpopulations of *Candida albicans*. Heatmap depicts significantly differentially expressed genes for replicates of each *C. albicans* sorted populations (exposed and phagocytosed) at 0, 1, 2 and 4 hours post-infection, grouped by *k-means* (similar expression patterns) in five clusters. Each cluster includes synthesized functional biological relationships using Gene Ontology (GO) terms (corrected P < 0.05). **(d)** Transcriptional response in subpopulations of macrophages. Heatmap depicts significantly differentially expressed genes (DEGs; FC >4, FDR < 0.001) for replicates of each macrophage-sorted populations (exposed and infected) at 0, 1, 2 and 4 hours post-infection, clustered by *k-means* (similar expression patterns). Each cluster includes synthesized functional biological relationships using IPA terms (-log(p-value) > 1.3; z-score > 2). Expression patterns for **(f)** clusters 1 and 4 in *C. albicans* and **(g)** cluster 1 and 3 in macrophages. Each gene is plotted (gray) in addition to the mean expression profile for that cluster (blue).

2.3.2 Subpopulations of phagocytosed *Candida albicans* adapt to macrophages by switching metabolic pathways and regulating cell morphology

We next examined how *C. albicans* gene expression varied across exposed and phagocytosed cells over time. Using *k-means* clustering of 588 DEGs, we identified 4 clusters of genes with similar expression patterns; the major patterns of expression across time were either induced rapidly upon phagocytosis (clusters 1 and 2) or repressed at 1 and 2 hours upon phagocytosis (clusters 3 and 4; **Figure 2.2B,** S2A). Comparing the phagocytosed and un-engulfed *C. albicans* subpopulations at each time point, the highest number of DEGs was found at 1 hour (n = 165), highlighting a rapid and specific transcriptional response upon phagocytosis, which was maintained throughout the 4-hour infection time course (Table S2; **Figure 2.2B**).

Genes highly induced in phagocytosed *C. albicans* at 2 and 4 hours (cluster 1, n = 93) are involved in adaptation to the macrophage environment including changes in metabolic pathways (**Figure 2.2B**). These genes are involved in glucose and carbohydrate transport, carboxylic acid and organic acid metabolism, and fatty acid catabolic processes (enriched GO terms corrected-$P < 0.05$, hypergeometric distribution with Bonferroni correction; Table S3). Prior microarray analysis of bulk populations of *C. albicans* exposed to macrophages reported similar changes[3]; by sorting infection fates, our work suggests that this response is specific to engulfed *C. albicans*, allowing the pathogen to utilize the limited spectrum of nutrients available in the phagosome[16]. We also found that genes involved in glyoxylate metabolism, the beta-oxidation cycle, and transmembrane transport were significantly induced in phagocytosed *C. albicans* relative to exposed cells. Using sorted populations revealed that multiple classes of transporters were highly upregulated in both

engulfed and exposed *C. albicans* subpopulations (including oligopeptide transporters, several high affinity glucose transporters, and amino acid permeases), suggesting these changes are not in response to phagocytosis (**Figure 2.2B**; Table S2 and S3). Genes most strongly induced during at 4 hours upon phagocytosis (cluster 1) are involved in pathogenesis and associated with the formation of hyphae, including four of the eight genes involved in the core filamentation response[127] (*ALS3*, *ECE1*, *HWP1*, *orf19.2457*; Table S2). While media containing serum can also induce *C. albicans* filamentation, we found that filamentation genes were more highly induced in the phagocytosed *C. albicans* subpopulation (Table S2; **Figure S2B**). Genes induced most strongly at 1 and 2 hours upon phagocytosis (cluster 2) are involved in early fatty acid oxidation response and transmembrane transport (Opt and Hgt classes; Tables S2 and S3). These transporters differ from those in cluster 1 in that they show peak expression at 1 hour in phagocytosed cells and decrease in expression by 4 hours (**Figure 2.2B**). Clusters 1 and 2 contained 6 confirmed or putative transcription factors (*SUT1, STP4, TEA1, ADR1, ZCF38, TRY4*), all of which encode zinc finger containing proteins; zinc cluster transcription factors have been implicated in *C. albicans* virulence[128,129].

Two sets of genes were specifically downregulated in live, phagocytosed *C. albicans* at 1 and 2 hours. Expression levels of these genes largely did not change in exposed *C. albicans* over the time course (clusters 3 and 4; Figures **2.2B**, **E**). Notably, these repressed genes recovered their expression levels by 4 hours in phagocytosed *C. albicans*. Cluster 3 included genes related to the translation machinery and peptide biosynthesis, including ribosomal proteins, chaperones and transcription factors that regulate translation (enriched GO terms, corrected-P < 0.05, hypergeometric distribution with Bonferroni correction; Table S3; Figure S2B). Repression of the

translation machinery was previously noted in *C. albicans* during macrophage interaction[16]. Our results demonstrate that down-regulation of ribosomal proteins, chaperones and translation-regulator transcription factors are specific to phagocytosed *C. albicans* and that expression of these genes recovered at later time points (**Figures 2.2E**, S2B). Cluster 3 also encompassed genes involved in morphological and cell surface remodeling, including an essential negative regulator of filamentation, SSN6 (**Figures 2.2B**, S2B; Table S2). Repressed genes in Cluster 4 are largely involved in nucleoside metabolic processes, gluconeogenesis and host adaptation (enriched GO terms, corrected-P < 0.05, hypergeometric distribution with Bonferroni correction; **Figure 2.2B**; Table S3). A subset of these repressed genes are yeast-phase specific and included those involved in ergosterol biosynthesis, cell growth and cell wall synthesis (**Figure 2.2B**, S2). These results highlight that a large part of the observed transcriptional repression is specific to phagocytosed *C. albicans,* in contrast to the common sets of upregulated genes shared by exposed and phagocytosed cells.

2.3.3 Subpopulations of macrophages showed major pathogen recognition and pro-inflammatory response to *Candida albicans* and shift profiles at late time course

In parallel with the analysis of *C. albicans* gene expression, we also examined the transcriptional response of macrophages. Across all samples, we identified 577 DEGs (FC > 4; FDR < 0.001; which grouped into four clusters with similar expression patterns (**Figure 2.2C**, **D**). For both exposed and infected macrophages, a major difference was found between 1 and 4 hours along PC1, which highlights genes that were highly induced or repressed at 4 hours in these

subpopulations (clusters 1 and 4, respectively; **Figure 2.2C**, **D**). These 4 clusters together were significantly enriched for genes involved in activation of phagocytosis, migration of phagocytes, and triggering the innate immune response; this includes the induction of pathways such as IL-6, IL-8 and NF-κB signaling, Fcγ Receptor-mediated phagocytosis, production of nitric oxide and reactive oxygen species (ROS), Pattern Recognition Receptors (PRR), RhoA, ILK, and leukocyte extravasation signaling (p-value < 0.05 right-tailed Fisher's Exact Test; **Figure 2.2D**, **S3**). Activation of some of these pathways is consistent with previous analysis of phagocyte transcriptional responses to *C. albicans* infection (**Figure S4**)[87–89,130]; sorting distinct infection subpopulations during early time points of infection established that host cells regulate subsets of genes upon *C. albicans* exposure and phagocytosis (**Figure 2.2D**, **S5**).

In exposed and infected macrophages, many of the genes induced at 1 hour maintained this expression level at 2 and 4 hours (cluster 1; **Figure 2.2D**). These genes are related to defense mechanisms such as pro-inflammatory cytokine production and fungal recognition via transmembrane receptors. Upregulated genes related to pro-inflammatory cytokines included tumor necrosis factor (*Tnf*), interferon regulatory factor 1 (*Irf1*), and the chemokine receptor *Cx3cr1*, associated with an innate mechanism of fungal control in a model of systemic candidiasis[131] and colitis[132] (**Figure 2.2D**, **S5**; **Table S4**). A second set of genes was initially repressed at 1 hour and then upregulated in both exposed and infected macrophages at 2 and 4 hours (cluster 2; **Figure 2.2D**; **Table S4**). This set of genes is associated with pathogen recognition, opsonization, and activation of the engulfment (p-value < 0.05 right-tailed Fisher's Exact Test; **Table S5**), including the lectin-like receptor galectin 1 (*Lgals1*), transmembrane receptors (*Fcer1g*), chemokines (*Ccl3*, *Cxcl2*), extracellular complement protein (*C1qb*), and transcriptional

regulators that play a role in inflammation and programmed cell death (*Fos*, *Irf8*, *Cebpb* and *Card9*) (**Tables S4** and **S5**). Since the expression of these genes increased at 2 hours and maintained high expression in infected macrophages, they may also play an important role during phagocytosis or allow for uptake of additional *C. albicans* cells. The chemokines *Cxcl2, Ccl3* and *Cx3cr1* have also been previously shown to be induced during *C. albicans* interactions with other host cells, including neutrophils *in vitro*[89], in a murine kidney model[87], a murine vaginal model[130], and in mouse models of hematogenously disseminated candidiasis and of vulvovaginal candidiasis in humans[88], highlighting the role of these genes in host defense against *C. albicans* infection of different tissues (**Figure S4**).

We also examined subpopulations of macrophages that have phagocytosed and killed *C. albicans;* this data was analyzed separately, as the total number of cells sorted and therefore the transcriptome coverage were low (< 3,000 host transcripts detected) and had modestly correlated biological replicates (*e.g.* Pearson's $r < 0.56$). We found a small set of highly induced pro-inflammatory cytokines, including *Ccl3*, *Cxcl2*, *Il1rn* and *Tnf*, and transcription regulators such as *Cebpb*, *Irf8* and *Nfkbia* (**Figure S6**). These genes were also induced in macrophages infected with live *C. albicans* (clusters 1 and 2; **Figure 2.2D**), indicating that maintaining expression of these genes may be important for pathogen clearance and host cell survival after phagocytosis.

Another major shift in macrophage gene expression occurred at 4 hours, with sets of genes involved in the immune response highly repressed at this later time point (clusters 3 and 4, respectively; **Figure 2.2D**, **F**). Repressed genes at 4 hours (cluster 3) are enriched in cytokines (*Il1a, Cxcr4*) and transmembrane receptors, including intracellular toll-like receptor (*Tlr5*), C-type

lectin receptors (*Clec4a3*, *Clec10a, Olr1*), and interleukin receptors (*Il1r1*). This cluster was also enriched in categories associated with proliferation and immune cell differentiation (p-value < 0.05 right-tailed Fisher's Exact Test; **Table S5**). In addition, highly repressed host genes at 4 hours (cluster 4) were enriched for categories related to phagosome formation, phagocytosis signaling, and immune response signaling (p-value < 0.05 right-tailed Fisher's Exact Test; **Table S5**). Notably, these repressed genes included several interleukin receptors (*Il21r*, *Il4r*, *Il17ra*) and transmembrane receptors (*Tlr9*, *Mrc1*) that are typically highly expressed during the immune response to fungal infections[87,89]. This suggests that during phagocytosis of *C. albicans* there is a strong shift in macrophages toward a weaker pro-inflammatory transcriptional response by 4 hours.

2.3.4 Detection of host-pathogen transcriptional responses from single macrophages infected with *Candida albicans*

Even in sorted populations, individual cells may not have uniform expression patterns, as, even in a clonal population, cells can follow different trajectories over time. To address this, we next examined the level of single cell transcription variability during these infection time points. We collected sorted, single macrophages infected with live or with dead, phagocytosed *C. albicans* at 2 and 4 hours and adapted the RNA of both the host and pathogen for Illumina sequencing using Smart-seq2 (**Figure 1.8, 2.3A**; **2.5 Methods**). With this approach, each infected macrophage and the corresponding phagocytosed *C. albicans* received the same sample barcode, allowing us to pair transcriptional information for host and pathogen at the single, infected cell level (**Figure 2.3A**). While we successfully isolated single infected macrophages via FACS, we cannot control for the number of *C. albicans* cells inside of each macrophage with this approach, since macrophages can

phagocytose variable numbers of *C. albicans* cells[25]. We obtained 4.03 million paired-end reads per infected cell on average; a total of 449 single, infected macrophages had more than 1 million paired-end reads (**Table S1**). For macrophages with live *C. albicans*, we found an average 75% of reads mapped to host transcripts and 11% of reads mapped to *C. albicans* transcripts (**Figure 2.3B**). Although parallel sequencing of host and pathogen decreases the sensitivity to detect both transcriptomes from a single library, of the 224 single macrophages infected with live *C. albicans* with more than 0.5 million reads, 202 (90.2%) had at least 2,000 host-transcripts detected (> 1 Transcripts Per Million, TPM; 3,904 on average), and 162 (72.3%) had at least 600 *C. albicans* transcripts detected (> 1 TPM; 1,435 on average; **Figure 2.3B**, **S7**). The fact that we detected fewer fungal transcripts relative to the host was expected, as the fungal transcriptome is approximately four times smaller than the host transcriptome. Relative to the number of transcripts detected in the RNA-Seq of subpopulations of macrophages infected with live *C. albicans*, in single-infected-cells we detected 38% and 31% (on average) of the transcripts for host and *C. albicans* respectively, indicating that we obtained adequate sequencing coverage for both species. Additionally, we found that pooling single infected cell expression measurements could recapitulate the corresponding subpopulation expression levels. We found that the extensive cell-to-cell variation between single infected macrophages (average Pearson's from r 0.18 to 0.88; **Figure 2.3C**) was reduced when we aggregated the expression of 32 single cells (**Figure S8**). These results are consistent with previous single cell studies of immune cells[91,92,95] and indicate that we can accurately detect gene expression in single infected macrophages and phagocytosed *C. albicans*.

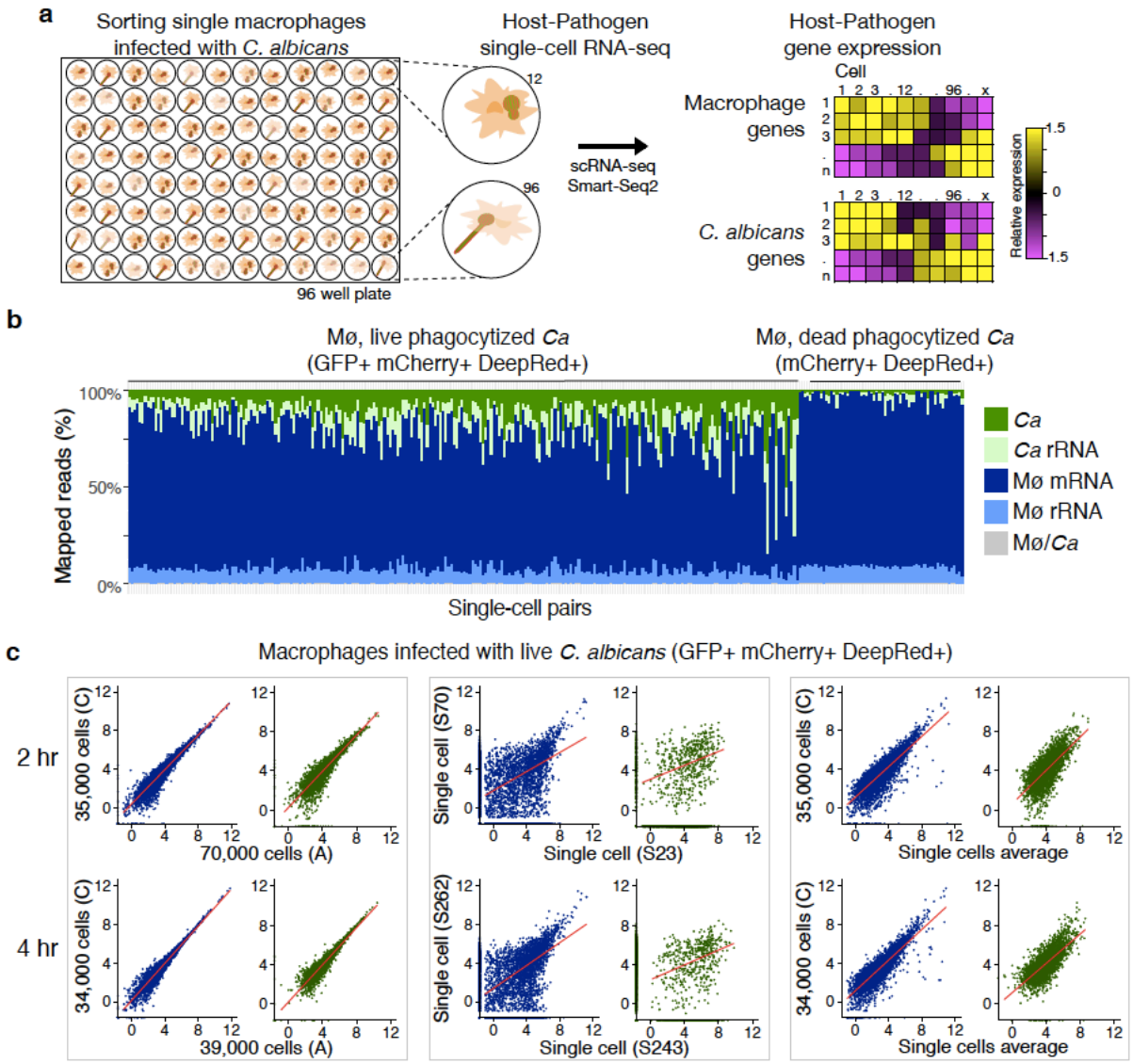

Figure 2.3: Processes and evaluation of parallel host-pathogen RNA-seq.
(a) Plot of the percent of mapped reads to the composite reference transcriptome (mouse messenger and ribosomal RNA + *Candida albicans* messenger and ribosomal RNA collections) of 314 single macrophages with live (mCherry + GFP+) or with dead, phagocytosed (mCherry + GFP-) phagocytosed *C. albicans* cells. **(b)** Plots of the gene expression correlation between (left) two replicates of the sorted subpopulation of macrophages (blue) with phagocytosed live *C. albicans* (green) at 2 and 4 hours post-infection; (middle) between two single macrophages and two phagocytosed *C. albicans*, and (right) between the single cells expression average and one replicate of the population.

2.3.5 Dynamic host-pathogen co-stages defined by analysis of single macrophages infected with live *Candida albicans*

To finely map the basis of heterogeneous responses during infection, we clustered cells by differential expressed genes and identified host-pathogen co-stages of infection in groups of infected macrophages and phagocytosed *C. albicans* pairs that showed similar gene expression profiles (**Figure 2.3A**, **Figure 2.4A**). Briefly, we used genes exhibiting high variability across the infected macrophages and live, phagocytosed *C. albicans* at 2 and 4 hours. We then reduced the dimensionality of the expression with principal components analysis (PCA), and clustered cells with the t-distributed stochastic neighbor embedding approach (t-SNE)[133] as implemented in Seurat[134] (**Methods**). The transcriptional response among single infected macrophages exhibited two time-dependent stages associated with expression shifts from 2 to 4 hours (stage 1^M and stage 2^M; **Figure 2.4A**; **Figure S9A**). While cells were largely separated into these two stages by time, a small subset of macrophages were assigned to the alternate cell stage by unsupervised clustering and appeared to be either early or delayed in the initiation of the transcriptional shift in genes involved in the immune response. Differentially expressed genes in stage 1^M (n = 88; likelihood-ratio test (LRT)[135], $P < 0.001$) are related to the pro-inflammatory response, and their expression significantly decreases in stage 2^M at 4 hours (DEGs n = 70; LRT[135], $P < 0.001$; **Figure 2.4A**, **top**; **Table S6**). This set comprises pro-inflammatory repertoire, such as cytokines (*Tnf, Ilf3*, *Ccl7*), transmembrane markers (*Cd83, Cd274*), interleukin receptors (*Il21r, Il4ra, Il6ra, Il17ra*) and the high affinity receptor for the Fc region (*Fcgr1*), and the transcriptional regulators *Cebpb* and *Cebpa* (**Figure 2.4A**, **4B**). Many genes variably expressed in these single, infected macrophages were also found to be upregulated in subpopulations of infected and exposed macrophages (*e.g. Tnf*, *Orl1*, **Figures S5**, **S10**); however, significant differences of these genes were not detected in

the RNA-Seq of subpopulations between 2 and 4 hours, which highlights cell-to-cell variability within each time point that can only be observed in single infected macrophages.

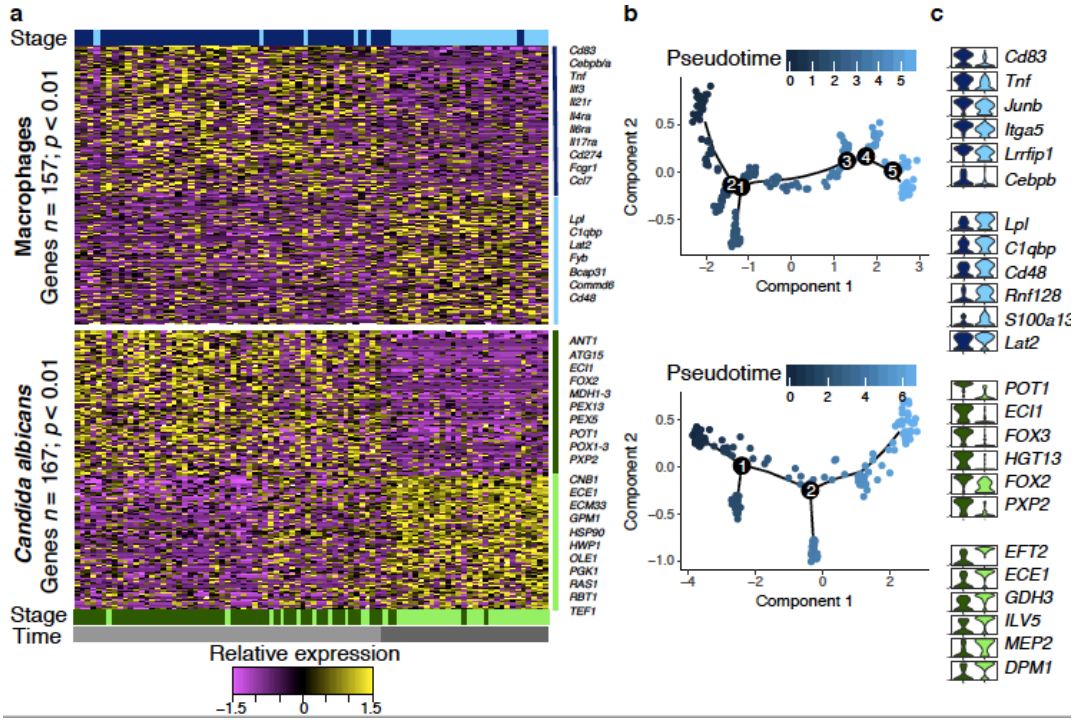

Figure 2.4: Parallel host-pathogen single-cell RNA-seq profiles reveal infection outcome shifts in pro-inflammatory response and pathogen immune evasion.
(a) Heat maps report parallel scaled expression [log TPM (transcripts per million) +1 values] of differentially expressed genes for each co-state of macrophages infected with live *C. albicans*. Each column represents transcriptional signal from a single, infected macrophage and the *C. albicans* inside of it. (Top: macrophage response; bottom: *C. albicans* response). Color scheme of gene expression is based on *z-score* distribution from -1.5 (purple) to 1.5 (yellow). Bottom and right margin color bars in each heat-map highlight co-state 1 (dark blue in macrophages, and dark green in *C. albicans*) and co-state 2 (blue in macrophages, and green in *C. albicans*), and time post-infection 2 hours (gray) and 4 hours (dark gray). **(b)** tSNE plot for macrophages (top) and *C. albicans* (bottom) colored by the pseudo-time. **(c)** Violin plots at right illustrate expression distribution of a subset six differentially expressed immune response or immune evasion genes for each co-state in macrophages and *C. albicans*, respectively.

As each single, infected macrophage received a unique sample barcode and host and fungal transcription were measured simultaneously (**Figure 2.3A**), the expression from each macrophage (**Figure 2.4A**, **top**) was matched with that of the live, phagocytosed fungus (**Figure 2.4A**,

bottom). Notably, independent analysis of the parallel fungal transcriptional response identified two pathogen stages that were also primarily distinguished by time. In *C. albicans*, genes significantly upregulated in stage 1^C (n = 80; LRT^{135}, $P < 0.001$; **Table S7**) were enriched in organic acid metabolism (P = 6.63e-11; enriched GO term, corrected-$P < 0.05$, hypergeometric distribution with Bonferroni correction; **Table S8**), including transporters (*HGT13*) and glyoxylate cycle genes, specifically those from beta-oxidation metabolism (*ECI1*, *FOX3*, *FOX2*, *PXP2*; **Figure 2.4A**, **4C**). Most macrophages infected by *C. albicans* in stage 1^M induced a strong pro-inflammatory response (co-stage of infection 1; **Figure 2.4**). At 4 hours in stage 2^C, expression of these transporters and glyoxylate cycle genes was reduced instead expression of genes (n = 86; LRT^{135}, $P < 0.001$; **Table S7**) enriched in carbon metabolism was increased of (P = 2.43e-05; enriched GO term, corrected-$P < 0.05$, hypergeometric distribution with Bonferroni correction; **Table S8**), including genes related to glycolysis and gluconeogenesis (*PGK1*), fatty acid biosynthesis (*FAS1*, *ACC1*), and genes associated with filamentation (*ECE1*, *HWP1*, *OLE1*, *RBT1*); at this stage, the majority of infected macrophages downregulated expression of pro-inflammatory cytokines (co-stage of infection 2; **Figure 2.4A**).

To more finely trace how cells shift their expression program between infection co-stages we performed pseudo-time analysis using Monocle[136] (**Methods**). We observed that the majority of infected macrophages and phagocytosed *C. albicans* pairs followed a linear expression trajectory between two major endpoints or clusters, recapitulating the two major infection co-stages; some cells exhibited alternative expression paths, suggesting the existence of minor cell trajectories (**Figure 2.4B**, **S11**). This ordering of cells revealed an early group of infected macrophages that expressed high levels of pro-inflammatory cytokines that decreased over time across intermediate

and late groups (**Figure 2.4B**, **top**; **S11**). Similarly, in phagocytosed *C. albicans*, an initial group of cells showed low expression levels of genes related to filamentation which increased over time (**Figure 2.4B**, **bottom**; **S11**). In this analysis, the fraction of phagocytosed *C. albicans* shifting to high expression of filamentation genes across pseudo-time increased slightly faster than the fraction of macrophages exhibiting decreased levels of expression of pro-inflammatory cytokines (10% more cells in the late pseudo-time range; **Figure 2.4B**). This suggests that the expression heterogeneity in macrophages could be driven by the expression heterogeneity in *C. albicans* (**Figure 2.4B**). This was also supported by unsupervised clustering, where some phagocytosed *C. albicans* collected at two hours were assigned to the second cell stage and appeared to initiate the transcriptional shift in immune response earlier than in the corresponding macrophage (**Figure 2.4A**). In summary, the major co-stages of infection largely correspond to time of infection; however, trajectory analysis highlights an asynchronous and linear transition associated with the induction of filamentation and metabolic adaptation in *C. albicans* which, correlates with a shift from a strong to a weak pro-inflammatory gene expression profile in the host.

2.3.6 Expression bimodality in host and fungal pathogen measured in single macrophages infected with live *Candida albicans*

To further examine heterogeneity in gene expression, we characterized modality of expression profiles across host-pathogen single cells, and then compared these distributions between 2 and 4 hours using a normal mixture model and Bayesian modeling framework as implemented in scDD[137] (**Methods**). Overall, an average of 84.5% of genes detected across single infected macrophages displayed unimodal gene expression distributions (**Figure 2.5A**; **Table S9**). Highly expressed unimodal genes (top 5%) with similar expression levels at 2 and 4 hours encompassed

genes involved in opsonization and *C. albicans* recognition, such as complement proteins (*C1qb*, *C1qc*) and galectin receptors that recognize beta-mannans (*Lgals1*, *Lgals3*). In phagocytosed *C. albicans*, an average of 76% had unimodal expression patterns (**Figure 2.5A**; **Table S10**). Highly expressed unimodal genes (top 5%) with similar expression levels in both co-stages were enriched in the oxidation-reduction process and defense against reactive oxygen species (enriched GO term, corrected-$P < 0.05$, hypergeometric distribution with Bonferroni correction; **Table S10**). These unimodal genes highlight the core genes involved in the host immune response to fungus and pathogen virulence, respectively.

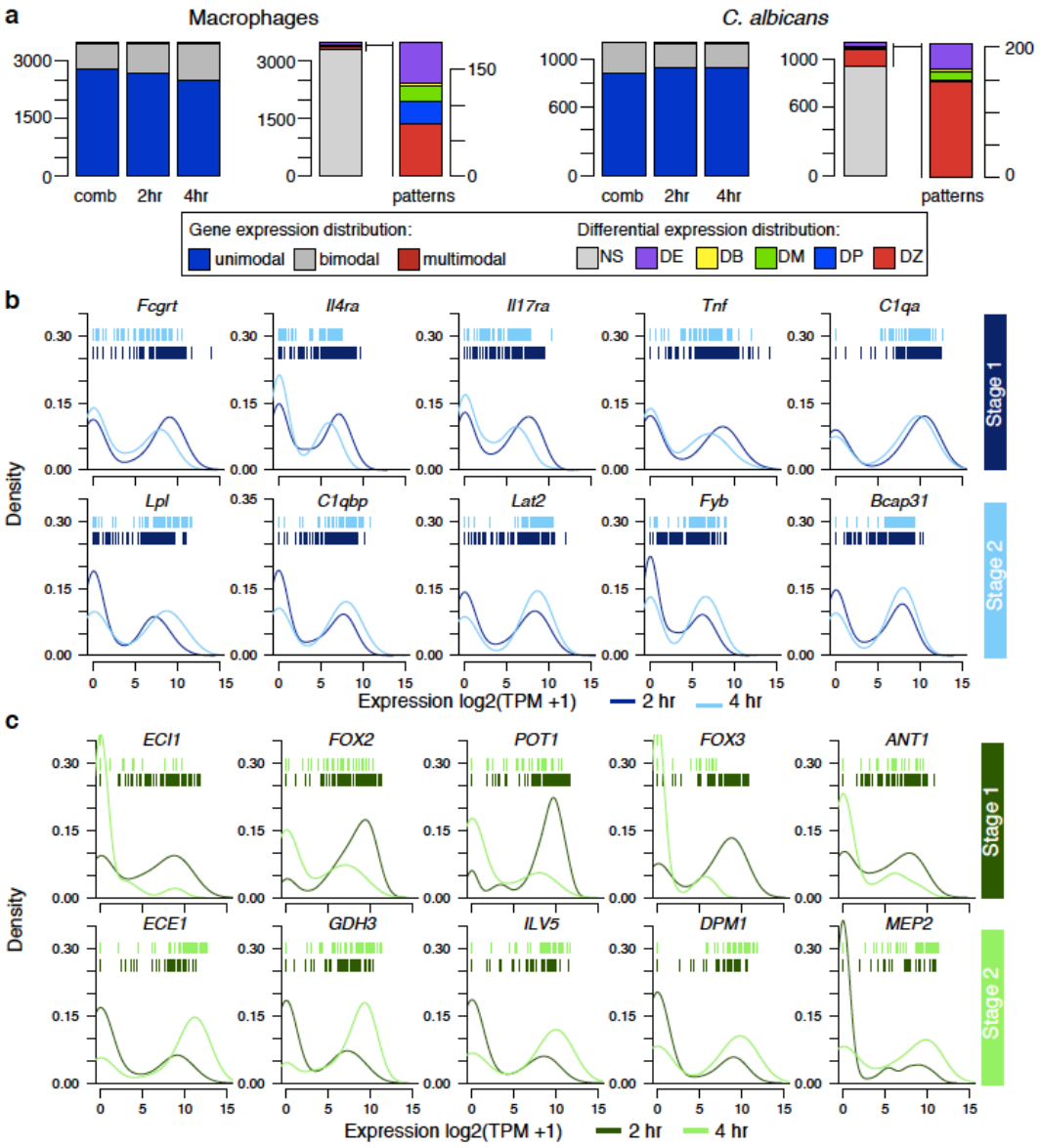

Figure 2.5: Expression variability and bimodality at the single cell level.
(a) Number of genes categorized as unimodal, bimodal, or multimodal (>2 components) according genes expression patterns in single, infected macrophages and corresponding phagocytized *C. albicans* across cells from 2 and 4 hours, and within each time point. Differential distributions were assigned as differential expression of unimodal genes (DE), differential modality and different component means (DB), differential modality (DM), differential proportion for bimodal genes (DP), differential proportion of zeroes (DZ), and not significant (NS). **(b)** Expression density distributions in parallel infected macrophages (n = 267; blue) and **(c)** phagocytized live *C. albicans* (n = 215; green) for 5 top marker genes in co-state1 and co-state2 across single macrophages infected with *C. albicans* at 2 and 4 hours post infection. Individual cells are plotted as bars for 2 hours (top row) and 4 hours (bottom row) for each distribution.

A subset of the genes highly expressed in co-stages of single infected macrophages and phagocytosed *C. albicans* showed bimodal expression distributions (**Figure 2.5B**). As bimodal transcriptional heterogeneity among single stimulated or infected cells can signify distinct immune cell expression programs[91,95], we next examined whether subgroups of macrophages could be defined by shared bimodality of genes involved in the immune response. An average of 15% host genes showed patterns of bimodal expression among or within 2 and 4 hours (exceeded Bimodality Index (BI) threshold; Dirichlet Process Mixture of normals model; **Table S9**, **Figure 2.5A**). In addition to expression bimodality, some of these genes showed patterns of differential distributions (*e.g.* shifts in mean(s) expression, modality, and proportions of cells) across and within 2 and 4 hours as implemented in scDD package[137]. This includes genes involved in pathogen intracellular recognition and pro-inflammatory response that had a bimodal expression distribution at 2 hours but not at 4 hours (*e.g. Olr1*, *Tnfrsf12a*), bimodal expression only at 4 hours (*Il4ra*), or bimodal expression at 2 and 4 hours with differential mean expression (*Il21r*, *Il17ra*; $P < 0.05$, Benjamini-Hochberg adjusted Fisher's combined test; **Figure 2.5B**; **Table S9**). As observed in macrophages infected with *Salmonella*[92] or stimulated with LPS[92], *Tnf* and *Il4ra* also exhibited bimodal expression patterns in single macrophages infected with *C. albicans* (**Figure 2.5B**, **top**). Additionally, we found unique subsets of genes displaying differential distributions in single macrophages infected with *C. albicans,* but not in response to bacterial stimuli, including *Il17ra* and other lectin-like receptors (*Olr1*; **Figure 2.5B** and **S12**). This suggests that variably expressed pathogen-specific receptors may play a role in these interactions, even in clonal populations of cells.

We next characterized variation in isoform usage between single macrophages during *C. albicans* infection, including immune response genes. Briefly, we calculated the frequency (percentage spliced in (PSI)) of previously annotated splicing events and identified differential isoform usage between single infected macrophages using BRIE[138] (**Methods**). We detected differential splicing between macrophages in 144 genes, including the immune response genes *Clec4n* (*Dectin-2*), *Il10rb* and *Ifi16* (cell pairs > 2000; Bayes factor > 200; **Table S10**). Notably, Dectin-2 had differential exon retention between macrophage stages, with the Dectin-2 α isoform (6 exons) predominantly in stage 1, and Dectin-2 β isoform (5 exons) predominately in stage 2 (**Figure 2.6**). The truncated isoform Dectin-2 β lacks part of the intracellular domain and most of the transmembrane domain of the receptor[139]. We found that Dectin-2 β has a lower posterior probability of transmembrane helix (p = 0.80; TMHMM2) as compared as Dectin-2 α (p = 0.99; TMHMM2; **Figure S13**). The lack of this transmembrane region has been proposed to encode a secreted protein, which may act as an antagonist to full-length Dectin-2[139]. Activation of Dectin-2 receptors on macrophages and dendritic cells by *C. albicans* leads to Th17 T cell differentiation to assist in the immune response[62,63]. These results suggest that splicing variation among single macrophages might indicate different potentials to respond to fungal infection.

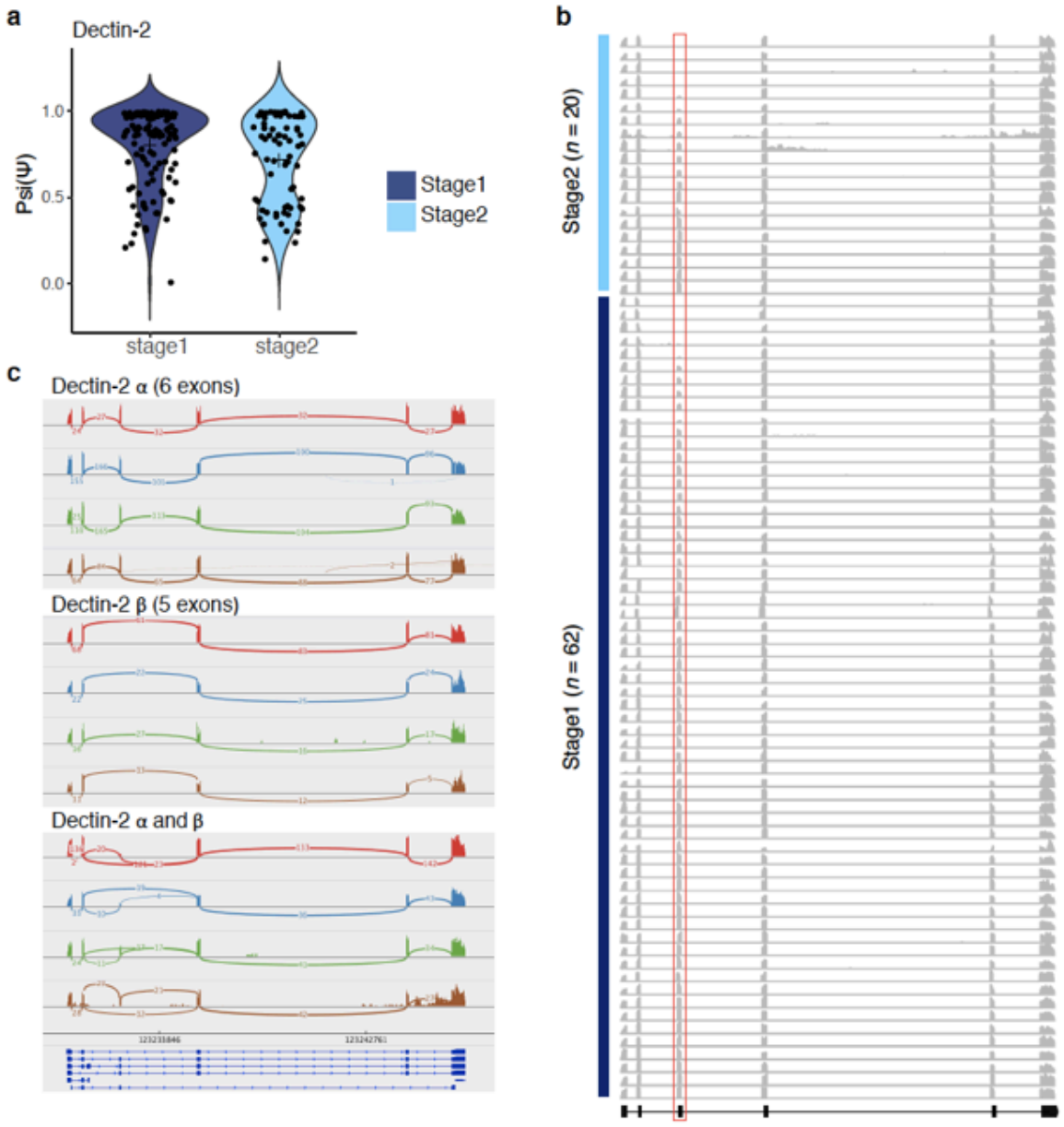

Figure 2.6: Differential splicing and isoform usage in single macrophages infected with *Candida albicans*
(a) Differential inclusion of exon 3 alternative splicing event in Clec4n (Dectin-2) in single macrophages infected with *C. albicans*. Violin plots depict the distribution of the percent spliced-in (Psi/Ψ) scores for single macrophages in stage 1 and 2 of infection. **(b)** Shown are scRNA-seq densities for all exons of Dectin-2 across 82 single macrophages infected with *C. albicans*; 62 in stage-1 and 20 in stage-2. Each row represents a single cell. Red box indicates the spliced exon 3 **(c)** Sashimi plots showing differential splicing between single macrophages, including the read densities and number of junction reads. Four cells were selected showing three scenarios: top: Dectin-2 α, middle Dectin-2 β, and bottom cell with both isoforms.

We next hypothesized that *C. albicans* may also demonstrate expression heterogeneity and bimodality that is linked with expression in the corresponding macrophage cell. We found an average of 23% *C. albicans* genes that showed patterns of bimodality at 2 and 4 hours, including a subset of virulence-associated genes that also showed shifts in mean expression, modality, and proportions of cells across and within 2 and 4 hours ($P < 0.05$, Benjamini-Hochberg adjusted Fisher's combined test; **Figure 2.5A; Table S10**). Differential expression of those genes explained most of the variation across infection co-stages (**Figure 2.4A**) and were more important for *C. albicans* fate decision and trajectory (**Figure 2.4B**). This set of genes were enriched in cell adhesion and filamentation, oxidation-reduction process and fatty acid oxidation (enriched GO term, corrected-$P < 0.05$, hypergeometric distribution with Bonferroni correction; **Table S10**), including genes involved in the core filamentation network (*ALS3*, *ECE1*, *HGT2*, *HWP1*, *IHD1, OLE1*), beta-oxidation and glyoxylate cycle (*ANT1*, *MDH1-3*, *FOX2*, *POX1-3*, *PEX5*, *POT1*), and response to oxidative stress (*DUR1,2*, *GLN1*, *PGK1*; **Table S10**; **Figure 2.5C**, **bottom**). Measuring dual species gene expression in sorted infection subpopulations and in single infected cells reveals that expression heterogeneity and bimodality of genes involved in fungal morphology and adaptation are tightly co-regulated. We observed shifts in the host response during host-fungal pathogen interactions, and, in some cases, this might result in different expression levels of host immune response genes and pathogen virulence genes. This approach can further enhance our understanding of distinct infection fates and the correlated gene regulation that governs host cells and fungal pathogen interaction outcomes.

2.4 Discussion

Host and fungal cell interactions are heterogeneous, even within clonal populations. One way to resolve this heterogeneity is to subdivide these populations by infection fate or stage to measure gene expression variability across subpopulations. Single-cell RNA-sequencing approaches offer the ultimate level of subdivision to study heterogeneity and may obviate the need for sorting when run at sufficient scale. Both these approaches rely on measuring host and fungal pathogen gene expression levels using dual RNA-sequencing to provide insight as to how both species respond in each infection stage. Here, we piloted both approaches to study fungal interactions with host cells. We developed a generalizable strategy to isolate distinct host and fungal pathogen infection fates over time, including single infected cells. In both sorted populations and single infected cells, we demonstrated that gene expression changes can be measured in both host and pathogen simultaneously. This approach allowed characterization of distinct infection fates within heterogeneous host and fungal pathogen interactions. This approach could be used to better further characterize the requirement for specific host and pathogen genes for these infection responses, and single cell analysis is well suited to characterize variability in both the host and pathogen during active infections.

This dual scRNA-Seq approach builds on prior transcriptional studies of interactions between microbial pathogens and immune cells. For fungi, RNA-Seq studies have largely measured gene expression profiles of either the host or the fungal pathogen[16,85,86] and have measured transcription profiles across infection outcomes[87–90]. To better examine heterogeneity of host-pathogen

interactions, recent approaches demonstrated the use of a similar GFP live/dead reporter to measure gene expression in phagocytes infected by bacterial pathogens using scRNA-Seq[92–94]. However, these studies mainly focused on the transcriptional response of the host, as bacterial transcriptomes can be difficult to measure due to their relatively low number of transcripts[92,94]. By contrast, we have demonstrated that both host and *C. albicans* gene expression can be measured in single infected cells; further studies will be needed to examine if the same or modified approaches can be extended to other microbial-host interactions.

By examining single infected macrophages, we showed that host and pathogen exhibit transcriptional co-stages that are tightly coupled during an infection time course, providing a high-resolution view of host-fungal interactions. This also revealed that expression heterogeneity of key genes in both infected macrophages and in phagocytosed *C. albicans* may contribute to infection outcomes. We identified two, time dependent linear co-stages of host-fungal pathogen interaction, with potential intermediate stages in some single cells suggested by pseudo-time analysis. The initial co-stage is characterized by induction of a pro-inflammatory host profile after 2 hours of interaction with *C. albicans* that then decreased by 4 hours. This is consistent with studies in human macrophages, where pro-inflammatory macrophages that interact with *C. albicans* for 8 hours or longer skew toward an anti-inflammatory proteomic profile[140]. Both commensal and invasive stages of *C. albicans* infection are impacted by the balance between pro-inflammatory and anti-inflammatory responses[141]. In single macrophages infected with *C. albicans*, the shift to an anti-inflammatory state, including upregulation of genes involved in the activation of inflammasomes, was coupled with the activation of filamentation and cell-wall remodeling in *C. albicans*. Previous work has shown that *C. albicans* can switch from yeast to hyphal growth within the nutrient-

deprived, acidic phagosome[142] and can escape by rupturing the macrophage membrane during intra-phagocytic hyphal growth[143]. Other mechanisms of escape include the activation of macrophage programmed cell death pathways, including the formation of inflammasomes and pyroptosis[144], or cell damage induced by a cytolytic peptide toxin (Candidalysin, produced by *ECE1*)[145]. Our analysis of single cells revealed bimodal expression in genes involved in these processes, and time analysis suggests that an initial shift in expression of filamentation and cell-wall remodeling programs in phagocytosed *C. albicans* rapidly result in down-regulation of the pro-inflammatory state of the host cells. This supports the hypothesis that an asynchronous *C. albicans* yeast-to-hyphae transition within the macrophage could drive expression heterogeneity in the host. While gene expression patterns are tightly correlated in host and fungal pathogen and the progression in co-stages of infection appears largely linear, greater time resolution across more diverse cell populations will be needed to fully resolve the expression heterogeneity among individual infected macrophages and phagocytosed *C. albicans*. Further work using this approach could also examine the role of specific genes in both *C. albicans* and the host to clarify the drivers of transcriptional responses and heterogeneity in host-fungal pathogen infection fates.

We found that both single infected macrophages and the corresponding phagocytosed *C. albicans* cells exhibit expression bimodality for a subset of genes. The expression bimodality observed for the host as well as the pathogen is consistent with the evolutionary concept of bet hedging[146]. Both cell types may rely on stochastic diversification of phenotypes to improve their survival rate in the event of an encounter with the other cell type. For instance, clonal populations of *C. albicans* that find themselves in the unpredictable, changing environment of a host phagocyte could increase the chance of survival by varying the expression of key genes involved in the response. A similar

scenario may provide an advantage for phagocytes that encounter pleomorphic *C. albicans*. These strategies are noted to occur within microbial populations, where a small fraction of "persister" cells might be capable of surviving exposure to lethal doses of antimicrobial drugs as a bet-hedging strategy[147]. Larger numbers of host-pathogen pairs collected over a longer time course could disambiguate how expression co-stages in *C. albicans* and macrophages results in distinct infection fates.

Heterogenous transcriptional responses are important to consider in the treatment of fungal infections. Genes expressed uniformly among fungal cells in a population may be more effective therapeutic targets than the products of genes expressed by only a subset of cells. A comparison of transcriptional signal of drug-treated single infected cells to non-treated cells would determine if all phagocytosed *C. albicans* cells co-regulate uniformly regulate genes involved in the response to drug treatment or if there is evidence of bimodal expression for genes involved in drug targets or efflux pumps, for example. Additionally, single cell analysis revealed bimodal expression and differential splicing of *Clec4n* (*Dectin-2*),a C-type lectin receptor that recognizes diverse fungal pathogens, as well as some bacteria and parasites[148]. The switch from producing a full length isoform to one missing an exon involved in signaling is related to the shift to an anti-inflammatory state. Parallel host-fungal pathogen expression profiling at single cell level could not only allow researchers to pinpoint changes in immune recognition and response pathways but also measure the dynamics of other stimuli, for example to characterize how new drug treatments affect both pathogen and host cells. While we used sorting to enrich for cell populations of interest, as scRNA-Seq microfluidic platforms continue to develop and profiling thousands of single cells becomes more cost effective, it will become tractable to interrogate more complex samples, including those

containing multiple host cell types and non-clonal pathogens. These data will provide fine mapping of diverse microbial immune responses and communication between immune cells. Importantly, this approach will allow further investigation into how fungal phenotypic and expression heterogeneity drives host responses and provide a systems view of these interactions.

2.5 Methods

2.5.1 *Candida albicans* reporter strain construction

The reporter construct used in this study was prepared by integrating the GFP and mCherry fluorescent tags driven by the bi-directional *ADH1* promoter and a nourseothricin resistance (NAT^R) cassette at the Neut5L locus of *Candida albicans* strain CAI4-F2[149]. CAI4-F2 harbors a homozygous *URA3* deletion which makes it less filamentous than other common laboratory strains of *C. albicans*, including SC5314 (**Figure S14**); this allowed more *C. albicans* cells to be sorted (**2.5.3 Methods; Macrophage and *Candida albicans* infection assay**). Briefly, the pUC57 vector containing mCherry driven by *ADH1* promoter (Bio Basic) was digested and this portion of the plasmid was ligated into a pDUP3 vector[150] containing GFP, also driven by *ADH1* promoter, a NAT^R marker, and homology to the Neut5L locus. The resulting plasmid was linearized and introduced via homologous recombination into a neutral genomic locus, Neut5L, using chemical transformation protocol with lithium acetate. Transformation was confirmed via colony PCR and whole-genome sequencing. A whole genome library was created from strain CAI4-F2-Neut5L-*NAT1*-mCherry-GFP using Nextera-XT library construction strategy(Illumina) and sequenced on an Illumina MiSeq (150x150 paired end sequencing; approximately 28 million reads). Sequencing reads were *de novo* assembled using dipSPAdes[151]. Scaffolds were queried back to the plasmid

sequence used to transform CAI4-F2 using BLAST[152]. Sequencing coverage was visualized using IGV[153].

2.5.2 Live cell imaging and interaction quantification

We characterized the interactions of Deep Red stained murine macrophages (cell line RAW 264.7; ATCC) and a GFP expressing *Candida albicans* strain (SC5314) interaction types *in vitro*, after 20 minutes and 4 hours after exposure. Cells were exposed to each other at an MOI of 1:1, in a 6 well tissue culture dish. Cells were co-incubated in C10 media at 37 degrees and 5% CO_2 using a microscope stage top incubator. Images were obtained under 40x magnification, in one field of view. 4 separate wells were analyzed for each condition. Interactions types were classified manually and counted using ImageJ plugin CellCounter.

2.5.3 Macrophage and *Candida albicans* infection assay

Primary, bone derived macrophages (BMDMs) were derived from bone marrow cells collected from the femur and tibia of C57BL/6, female mice. All mouse work was performed in accordance with the Institutional Animal Care and Use Committees (IACUC) and with relevant guidelines at the Broad Institute and Massachusetts Institute of Technology, with protocol 0615-058-1. Primary bone marrow cells were grown in "C10" media as previously described[154] and supplemented with macrophage colony stimulating factor (M-CSF) (ThermoFisher Scientific) at final concentration of 10 ng/ml, to promote differentiation into macrophages. C10 media was not supplemented with uridine; this led to reduced CAI4-F2 filamentation (compared to wild type SC5314) and allowed more *C. albicans* cells to be sorted. Cultures were then stained with F4/80 (Biolegend) to ensure that ~95% of the culture had differentiated into macrophages. For the infection RNA-sequencing

experiment, BMDMs were seeded in 6 well plates (Falcon). Two days prior to the start of the infection experiment, yeast strains were revived on rich media plates. One day prior to the infection experiment, yeast were grown in 3 ml overnight cultures in rich media at 30°C. On the day of the infection experiment, macrophages were stained with CellMask Deep Red plasma membrane stain (diluted 1:1000) (ThermoFisher Scientific). Macrophages and stain were incubated at 37°C for 10 minutes, then macrophages were washed twice in 1X PBS. 2 hours prior to infection, yeast cells were acclimated to macrophage media (RPMI 1640 no phenol red, plus glutamine, ThermoFisher Scientific) at 37°C prior to the infection. Yeast cells were then counted and seeded in a ratio of 1 *C. albicans* cell to 2 macrophage cells. Yeast and macrophages were then co-incubated at 37°C (5% CO_2). At the indicated time point, media was removed via aspiration, 1 ml of 1X TrypLE, no phenol red (ThermoFisher Scientific) was added to each well and incubated for 10 minutes. After vigorous manual pipetting, 2 wells for each time point were combined into one tube. Each time point was run in biological triplicate (except for 24 hours, which was collected in biological duplicate). Samples were then spun down at 37°C, 300g for 10 minutes and resuspended in 1 ml PBS + 2% FCS and placed on ice until FACS. RNAlater was used shortly after sorting, as this reagent leads to decreased GFP expression[155]. To control for gene expression changes that may be induced by FACS, only sorted samples were used in comparative analyses.

2.5.4 Fluorescence-activated cell sorting (FACS)

Samples were sorted on the BDSORP FACSAria running the BD FACSDIVA8.0 software into 1X PBS and then frozen at -80°C until RNA extraction. Cells were sorted into the following sub populations: (i) macrophages infected with live *C. albicans* (GFP+, mCherry+, Deep red+), (ii) macrophages infected with dead, phagocytosed *C. albicans* (GFP-, mCherry+, Deep red+), (iii)

macrophages exposed to *C. albicans* (GFP-, mCherry-, Deep red+) and (iv) *C. albicans* exposed to macrophages (GFP+, mCherry+, Deep red-). Single cells were sorted into 5 ul of RLT 1% β-Mercaptoethanol in a 96 well plate (Eppendorf) and frozen at -80°C .

2.5.5 RNA extraction, evaluation of RNA quality

RNA was extracted from population samples using the Qiagen RNeasy mini kit. All samples were resuspended in RLT (Qiagen) + 1% β-Mercaptoethanol (Sigma) and subjected to 3 minutes of bead beating with .5mm zirconia glass beads (BioSpec Products) in a bead mill. Single macrophages infected with *C. albicans* were directly with dead, phagocytosed by sorting cells into a 96 well plate containing 5 ul of RLT (Qiagen) + 1% β-Mercaptoethanol (Sigma).

2.5.6 cDNA synthesis and library generation

For population samples, the RT reaction was carried out with the following program, as described[156], with the addition of RNAse inhibitor (ThermoFisher) at 40U/ul. cDNA was generated from single cells based on the Smart-seq2 method as described previously[157], with the addition of RNase inhibitor was used at 40 U/ul (ThermoFisher) and 3.4 ul of 1 M trehalose was added to the RT reaction. All libraries were constructed using the Nextera XT DNA Sample Kit (Illumina) with custom indexed primers as described[157]. Infection subpopulation samples were sequenced on an Illumina NextSeq (37 x 38 cycles). *Candida* only samples were sequenced on an Illumina MiSeq (75 x 75 cycles). Single infected cells were sequenced on Illumina's NextSeq (75x75 cycles).

2.5.7 Read processing and transcript quantification of population-RNA-Seq

Basic quality assessment of Illumina reads and sample demultiplexing was done with Picard version 1.107 and Trimmomatic[158]. Samples profiling exclusively the mouse transcriptional response were aligned to the mouse transcriptome generated from the v. Dec. 2011 GRCm38/mm10 and a collection of mouse rRNA sequences from the UCSC genome website[159]. Samples profiling exclusively the yeast transcriptional response were aligned to transcripts from the *Candida albicans* reference genome SC5314 version A21-s02-m09-r10 downloaded from the *Candida* Genome Database (http://www.candidagenome.org).

Samples from the infection assay profiling in parallel host and fungal transcriptomes, were aligned to a "composite transcriptome" made by combining the mouse transcriptome described above and the *C. albicans* transcriptome described above. To evaluate read mappings, BWA aln (BWA version 0.7.10-r789, http://bio-bwa.sourceforge.net/)[160] was used to align reads, and the 'XA tag' was used for read enumeration and separation of host and pathogen sequenced reads. Multi-reads (reads that aligned to both host and pathogen transcripts) were discarded, representing only an average of 2.6% of the sequenced reads. Then, each host or pathogen sample file were aligned to its corresponding reference using Bowtie2[161] and RSEM (RNA-Seq by expectation maximization; v.1.2.21). Transcript abundance was estimated using transcripts per million (TPM). Since parallel sequencing of host and pathogen from single macrophages increased the complexity of transcripts measured compared to studies of only host cells alone, we detected lower number of transcripts of macrophages as compared with other studies using phagocytes and similar scRNA-seq methods[92,93,95].

2.5.8 Differential gene expression analysis of population-RNA-Seq

TMM-normalized 'transcripts per million transcripts' (TPM) for each transcript were calculated, and differentially expressed transcripts were identified using edgeR[162], all as implemented in the Trinity package version 2.1.1[163]. Genes were considered differentially expressed if they had a 4-fold change difference (> 4 FC) in TPM values and a false discovery rate below or equal to 0.001 (FDR < 0.001), unless specified otherwise.

2.5.9 Read processing and transcript quantification of single cell RNA-Seq

BAM files were converted to merged, demultiplexed FASTQ format using the Illumina Bcl2Fastq software package v2.17.1.14. For the RNA-Seq of sorted population samples, paired-end reads were mapped to the mouse transcriptome (GRCm38/mm10) or to the *Candida albicans* transcriptome strain SC5314 version A21-s02-m09-r10 using Bowtie2[161] and RSEM (RNA-Seq by expectation maximization; v.1.2.21)[164]. Transcript abundance was estimated using transcripts per million (TPM). For read mapping counts paired-reads were aligned to the 'composite reference' as described above.

For each single macrophage infected with *C. albicans*, we quantified the number of genes for which at least one read was mapped (TPM > 1). We filtered out low-quality macrophage or *C. albicans* cells from our data set based on a threshold for the number of genes detected (a minimum of 2,000 unique genes per cell for macrophages, and 600 unique genes per cell for *C. albicans*, and focused on those single infected macrophages that have good number of transcript detected in both host and pathogen (**Figure S9A**). For a given sample, we define the filtered gene set as the genes that have an expression level exceeding 10 TPM in at least 20% of the cells. After cell and gene

filtering procedures, the expression matrix included 3,254 transcripts for the macrophages and 915 transcripts for *C. albicans*. To estimate the number of *C. albicans* in each macrophage, we measured the correlation of GFP levels from FACs with the total number of transcripts detected in live, phagocytosed *C. albicans* cells (at least 1 TPM) but found only a modest correlation between these two metrics (R^2= 0.52).

To eliminate the non-biological associations of the samples based on plate-based processing and amplification, single cell expression matrices were log-transformed (log(TPM + 1)) for all downstream analyses, most of which were performed using the R software package Seurat (https://github.com/satijalab/seurat). In addition, we do not find substantial differences in the number of sequenced reads and detected genes between samples. We separately analyzed two comparisons of macrophages-*C. albicans* cells: *i*) single macrophages infected - live phagocytosed *C. albicans* cells at 2 and 4 hours (macrophages = 267; *C. albicans* cells = 215; macrophage-*C. albicans* = 156); and *ii*) macrophages infected and live or dead, phagocytosed *C. albicans* cells at 4 hours (macrophages = 142; *C. albicans* cells = 71). These numbers of macrophages and *C. albicans* cells are the total that met the described QC filters.

2.5.10 Detection of variation across single, infected cells

To examine if cell to cell variability existed across a wide range of population expression levels, we analyzed the variation and the intensity of non-unimodal distribution for each gene across single macrophages and *C. albicans* cells. Briefly, we determined the distribution of the average expression (μ), and the dispersion of expression (σ^2; normalized coefficient of variation), placing

each gene into bins, and then calculating a z-score for dispersion within each bin to control for the relationship between variability and average expression as implemented R package Seurat[134].

2.5.11 Detection of variable genes, cell clustering and trajectory

To classify the single-cell RNA-Seq from macrophages and *C. albicans*, the R package Seurat was used[134]. We first selected variable genes by fitting a generalized linear model to the relationship between the squared co-efficient of variation (CV) and the mean expression level in log/log space, and selecting genes that significantly deviated (p-value < 0.05) from the fitted curve, as implemented in Seurat, as previously described[93]. Then highly variable genes (CV > 1.25; p-value < 0.05) were used for principle component analysis (PCA), and statistically significant determined for each PC using JackStraw plot function. Significant PCs (p-value < 0.05) were used for two-dimension t-distributed stochastic neighbor embedding (tSNE) to define subgroups of cells we denominated host-pathogen co-states. We identified differentially expressed genes (corrected-p < 0.05) between co-states using a likelihood-ratio test (LRT) for single cell differential expression[135] as implemented in Seurat. To perform pseudo-time analysis we Monocle 2.8.0 method[136] using filtered normalized genes from Seurat. We reordered cells in pseudo-time along a trajectory using the top 50 differentially expressed genes, and identified genes with most significant changes as a function of progress along the trajectory using the branched expression analysis modeling (BEAM; q-value < 1e-4).

2.5.12 Detection of differential expression distributions and bimodality

To detect which genes have different expression distributions single infected macrophage and live C. albicans we compared the distributions of gene expression within single infected macrophage

and live C. albicans, across and between 2 and 4 hours, and identified genes showing evidence of differential distribution using a Bayesian modeling framework as implemented in scDD[137]. We used the permutation test of the Bayes Factor for independence of condition membership with clustering (n permutations = 1000), and test for test for a difference in the proportion of zeroes (testZeroes=TRUE). A gene was considered differentially distributed using Benjamini-Hochberg adjusted *p*-values test (p-value < 0.05).

2.5.13 Detection of differential splicing

To detect alternative splicing and determine variation in isoform usage between single macrophages during *C. albicans* infection we calculated exon splicing rates in individual macrophages using the data-dependent module of BRIE v0.2.0[138]. BRIE calls splicing at predefined cassette exons and quantifies splicing using exon reads in single cell data. We used the processed annotation file and sequence features for mouse (mm10 GRCm38.p5, ANNO_FILE = SE.gold, FACTOR_FILE = mouse_factors.SE.gold). Differential splicing was performed using the brie-diff module for all pairwise comparisons, and genes with differential splicing between single macrophages were defined as cell pairs > 2000 and Bayes factor > 200.

2.5.14 Functional biological enrichment analyses

For *C. albicans*, Gene ontology (GO) term analysis was performed in through the *Candida* Genome Database GO Term Finder and GO Slim Mapper (http://www.candidagenome.org[165]). GO terms were considered significantly enriched in a cluster or set of genes if we found a GO term corrected p-value lower than 0.05 using hypergeometric distribution with Bonferroni correction

For macrophages, Ingenuity Pathway analysis (IPA) was performed. We investigated biological relationships, canonical pathways and Upstream Regulator analyses as part of IPA software. This allowed us to assess the overlap between significantly DEGs and an extensively curated database of target genes for each of several hundred known regulatory proteins. Clusters or set of genes were considered significantly enriched if we found enriched a -log(*p-value*) greater than 1.3 (*i.e.* p-value ≤ 0.05) and z-score greater than 2 as recommended by the IPA software.

2.5.15 Data and biological materials availability statement

All sequence data for this project has been deposited in the SRA under Bioproject PRJNA437988. Raw and processed data for gene expression analysis was deposited in the GEO under GSE111731. Whole genome sequence data of CAI4-F2-Neut5L-*NAT1*-mCherry-GFP has been deposited at the NCBI SRA under SRX4924342. The CAI4-F2-Neut5L-NAT1-mCherry-GFP reporter strain is available upon request to RPR or from the ATCC BEI (accession pending).

2.6 Supplementary Notes and Figures

Supplementary Notes

Characterization of heterogeneous infection subpopulations in *ex vivo* macrophage and *Candida albicans* interactions

The number of RNA-Seq reads and transcripts detected for both host and fungal pathogen subpopulations was sufficient for differential expression analysis in most samples and to widely profile parallel transcriptional responses. We aligned reads to a composite reference of both mouse and *C. albicans* transcriptomes (**Methods**) and found that the fraction of mapped reads for host

and pathogen was correlated with the percent of sorted cells for each subpopulation (**Figures 2.1B**, **S2B**; **Table S1**). In subpopulations containing both host and pathogen, the fraction of reads averaged 87% macrophage and 13% *C. albicans* for macrophages infected with live fungal cells, and 95% macrophage and 5% *C. albicans* for macrophages containing dead fungal cells. For the subpopulation of macrophages infected with live fungal cells, between 1.4 and 34.0 million reads mapped to host transcriptome, while between 0.3 to 13.3 million reads mapped to *C. albicans* transcriptome. *Candida* reads in this subpopulation increased over the time course (0,1, 2 and 4 hours; **Figure 2.1B)**, reflecting ongoing engulfment and/or *C. albicans* division within the macrophage. In subpopulations of dead phagocytosed cells, read counts were lower than other subpopulations and counts increased over time, reflecting their smaller proportion of sorted cells that also increased over time (**Figure S2B**). In subpopulations of macrophages infected with live fungus, an average of 10,333 host and 4,567 *C. albicans* genes were detected (at least 1 fragment per replicate across all samples; **Methods**; **Figure S3A**; **Table S1**). Fewer transcripts were detected in subpopulations of macrophages infected with dead *C. albicans* (an average of 3,214 host transcripts and 983 fungal transcripts **Figure S3A**) and had modestly correlated biological replicates (*e.g.* Pearson's $r < 0.56$). Lower coverage was expected for this subpopulation, as only up to 3% of each sample collected at each time point was comprised of macrophages infected with dead *C. albicans* (**Figure S2B**). Therefore, we primarily focused the differential expression analysis on subpopulations of macrophages or *C. albicans* exposed, and macrophages infected with live *C. albicans*, which had high transcriptome coverage, and highly correlated biological replicates (*e.g.* Pearson's *r* 0.96 and 0.92 in macrophages and *Candida* at 4 hours, respectively; **Figure S3B**). These results indicate that we have established a robust system for measuring host and fungal pathogen transcriptional signal during phagocytosis in sorted infection subpopulations.

Subpopulations of phagocytosed *C. albicans* adapt to macrophages by switching metabolic pathways and regulating cell morphology

We next examined how *C. albicans* gene expression varied across unexposed, exposed and phagocytosed cells over time. Using *k-means* clustering, we identified sets of genes with similar expression patterns; the major patterns of expression across time were either induced (cluster 1, 2 and 3) or repressed (cluster 4 and 5) in the live, phagocytosed subpopulation relative to all other *C. albicans* infection fates (**Figures 2.2B**, **S4A**). A large number of genes were differentially expressed in phagocytosed *C. albicans* (732 genes relative to unexposed across all time points) compared with the number of DEGs found in exposed but un-engulfed *C. albicans* (82 genes relative to unexposed across all time points). Comparing the phagocytosed and un-engulfed *C. albicans* subpopulations at each time point, the major differential response was found at 1 hour, highlighting a rapid and specific transcriptional response upon macrophage phagocytosis. Many of these genes maintained high expression levels throughout the 4-hour infection time course (Table S2; **Figures 2.2B**, **E**).

Supplementary Figures

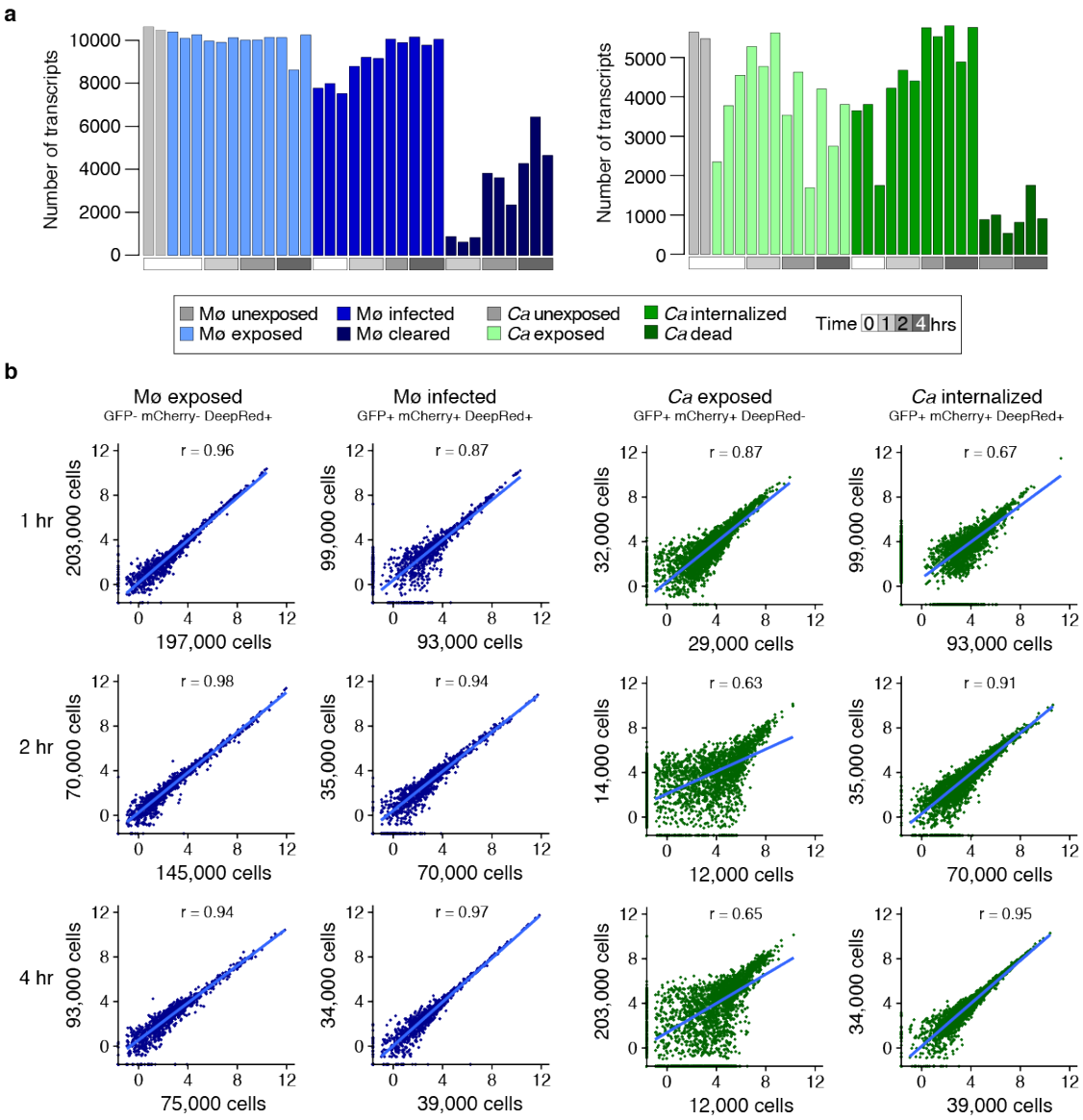

Figure S1. Parallel captured transcripts coverage and sample correlation in subpopulation of *Candida albicans*-macrophages infection outcome. (a) Number of transcripts simultaneously detected (TPM >1) in distinct subpopulations of macrophages (Mø; left; orange) and *C. albicans* (*Ca*; right; green). **(b)** Analysis of the correlation (Pearson correlation coefficient) of gene expression for sorted subpopulations replicates during macrophage-*C. albicans* interaction. The axes indicate number of sorted cell per replicate. Three subpopulations are depicted: macrophages infected-live *C. albicans* (GPF+, mCherry+, Deep Red+), macrophages exposed (GPF-, mCherry-, Deep Red+), and *C. albicans* exposed (GPF+, mCherry+, Deep Red-).

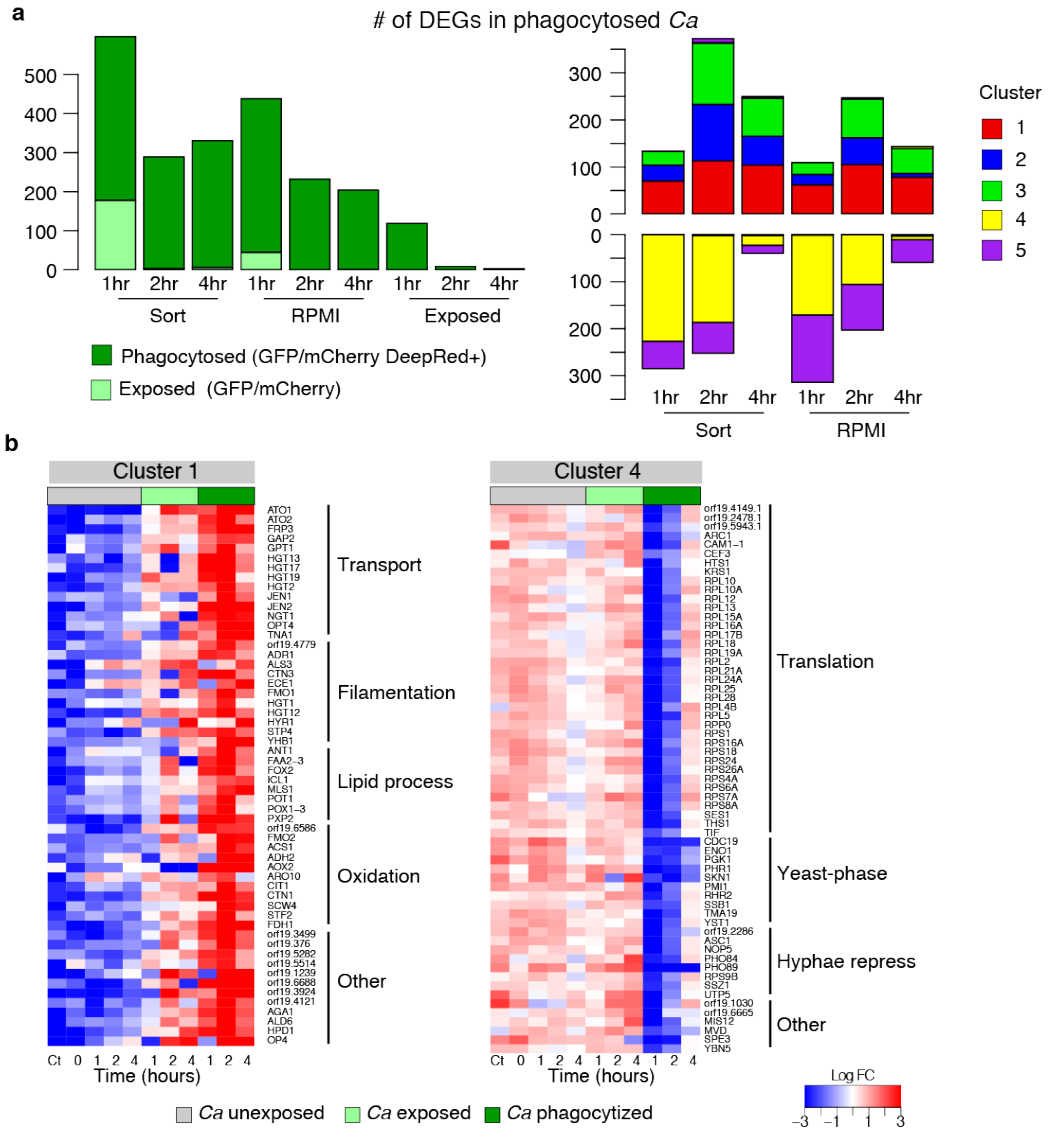

Figure S2. Differential gene expression in phagocytosed *Candida albicans* subpopulation during macrophages infection. (a) Number of genes differentially expressed (DEGs; FC >4, FDR < 0.001) in the live, phagocytosed subpopulation of *C. albicans* relative to exposed and unexposed (Sort, RPMI) subpopulations, and distribution of DEGs according the clusters of gene expression patterns. **(b)** Selection of most significantly DEGs from Cluster 1 (highly induced) and Cluster 4 (highly repressed) in *C. albicans* sorted populations (unexposed, exposed and phagocytosed) at 0, 1, 2 and 4 hours. Cluster 1 and 4 had the highest ratio of induction/repression upon phagocytosis. Synthesized functional biological categories were deducted from GO term enrichment and GO slim analyses (corrected-p < 0.05 hypergeometric distribution with Bonferroni correction).

Figure S3. Functional biological relationships of the transcriptional responses of macrophages to *Candida albicans* infection. (a) Canonical pathways, **(b)** Biological functions, and **(c)** Upstream regulators induced in macrophages with phagocytosed *C. albicans* (Infected) and macrophages with non-phagocytosed fungus (Exposed) in comparison to macrophages unexposed to *C. albicans*. Categories in bold are significantly enriched in both infected and exposed macrophages

Category		
cytokines	**Gene name**	**DE cluster**
C-C motif chemokine ligand 3 like 3	*Ccl3*	1
C-X-C motif chemokine ligand 3	*Cxcl2*	1
interleukin 1 receptor antagonist	*Il1rn*	1
secreted phosphoprotein 1	*Spp1*	2
interleukin 1 alpha	*Il1a*	3
oncostatin M	*Osm*	3
G-protein coupled receptors	**Gene name**	**DE cluster**
C-C motif chemokine receptor like 2	*Ccrl2*	1
complement C3a receptor 1	*C3ar1*	2
complement C5a receptor 1	*C5ar1*	2
C-X-C motif chemokine receptor 4	*Cxcr4*	3
calcitonin receptor like receptor	*Calcrl*	4
transcription regulators (cytoplasm)	**Gene name**	**DE cluster**
zinc finger and BTB domain containing 42	*Zbtb42*	3
transcription regulators (nucleus)	**Gene name**	**DE cluster**
CCAAT/enhancer binding protein beta	*Cebpb*	1
NFKB inhibitor delta	*Nfkbid*	1
NFKB inhibitor zeta	*Nfkbiz*	1
TATA-box binding protein associated factor 4b	*Taf4b*	1
inhibitor of DNA binding 1, HLH protein	*Id1*	2
JunB proto-oncogene, AP-1 transcription factor sul	*Junb*	2
AT-rich interaction domain 5A	*Arid5a*	3
BCL6 corepressor	*Bcor*	3
basic helix-loop-helix family member e40	*Bhlhe40*	3
BTG anti-proliferation factor 1	*Btg1*	3
cysteine and serine rich nuclear protein 1	*Csrnp1*	3
early growth response 1	*Egr1*	3
early growth response 2	*Egr2*	3
early growth response 3	*Egr3*	3
fem-1 homolog C	*Fem1c*	3
FosB proto-oncogene, AP-1 transcription factor su	*Fosb*	3
inhibitor of DNA binding 2, HLH protein	*Id2*	3
interferon regulatory factor 2 binding protein 2	*Irf2bp2*	3
Kruppel like factor 10	*Klf10*	3
MAF bZIP transcription factor F	*Maff*	3
regulator of calcineurin 1	*Rcan1*	3
SKI like proto-oncogene	*Skil*	3
SMAD family member 7	*Smad7*	3
CCAAT/enhancer binding protein delta	*Cebpd*	4
Fos proto-oncogene, AP-1 transcription factor subu	*Fos*	4
Kruppel like factor 3	*Klf3*	4
transmembrane receptors	**Gene name**	**DE cluster**
CD69 molecule	*Cd69*	1
intercellular adhesion molecule 1	*Icam1*	1
oxidized low density lipoprotein receptor 1	*Olr1*	1
toll like receptor 2	*Tlr2*	1
C-type lectin domain containing 7A	*Clec7a*	2
plasminogen activator, urokinase receptor	*Plaur*	2
CD5 molecule like	*Cd5l*	3
CD83 molecule	*Cd83*	3
interleukin 1 receptor type 1	*Il1r1*	3
mannose receptor C-type 1	*Mrc1*	3
stabilin 2	*Stab2*	3

Figure S4. Overall immune response genes to *Candida albicans* infection. Comparison of immune response genes found significantly differentially expressed in this study and previously reported in other studies of the host transcriptional response to *C. albicans* interaction.

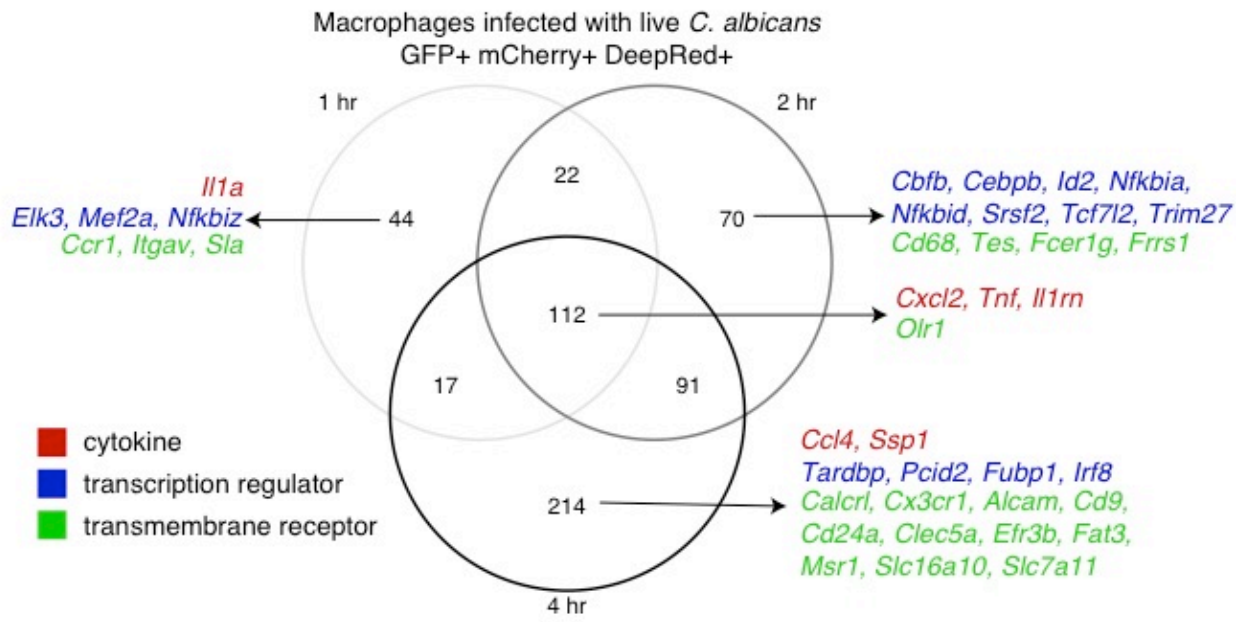

Figure S5. Overlap of the transcriptional response in infected macrophages. Different magnitude of induction for cytokines, transcriptional regulators, and transmembrane receptors differentially expressed in populations of macrophages infected with *Candida albicans* during 1, 2 and 4 hours post-infection.

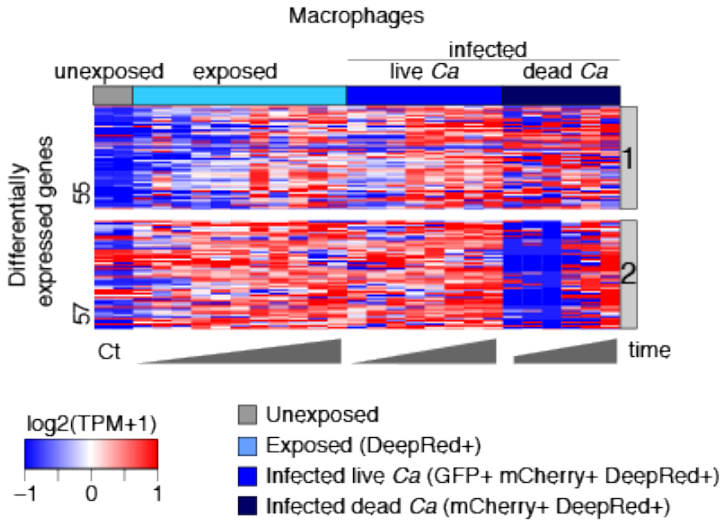

Figure S6. Highly induced genes in subpopulation of macrophages infected with dead *Candida albicans*. Heatmap depicts genes induced in subpopulation of macrophages infected with dead *Candida albicans* (clearance) relative to unexposed macrophages. Replicates for macrophages sorted subpopulations (unexposed, exposed and phagocytosed) at 0, 1, 2 and 4 hours post-infection are included. The gene expression is color coded from low expression (blue) to high expression (red) for each gene (row); two clusters of expression profile are shown (1 and 2, grey bars at right).

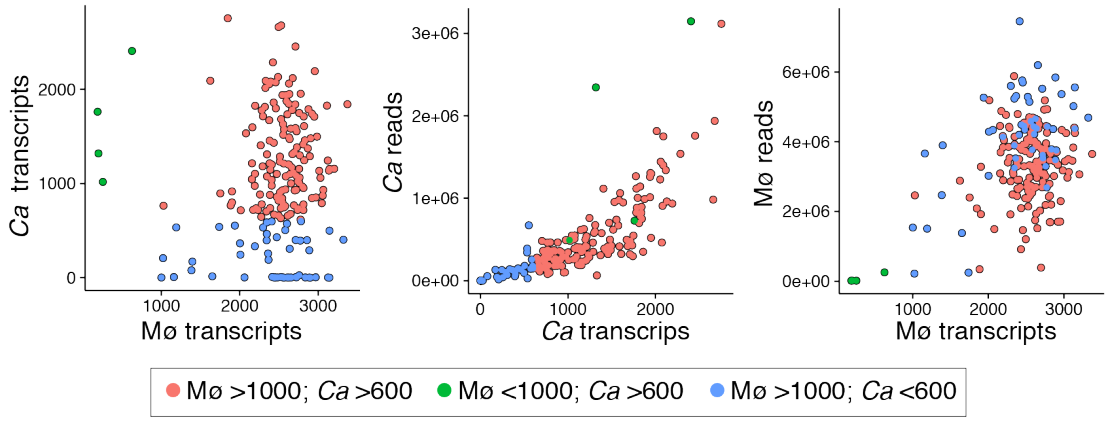

Figure S7. Transcripts detected at single infected cell level in macrophages and phagocytized *C. albicans*. Correlation plots of the number of detected transcripts (TPM > 1). Left: Comparison of the number of transcripts detected in macrophages and *C. albicans*. Middle and right: Comparison of transcripts detected (x-axis) and number of mapped reads (y-axis) for parallel infected macrophages with live *C. albicans* cells, respectively.

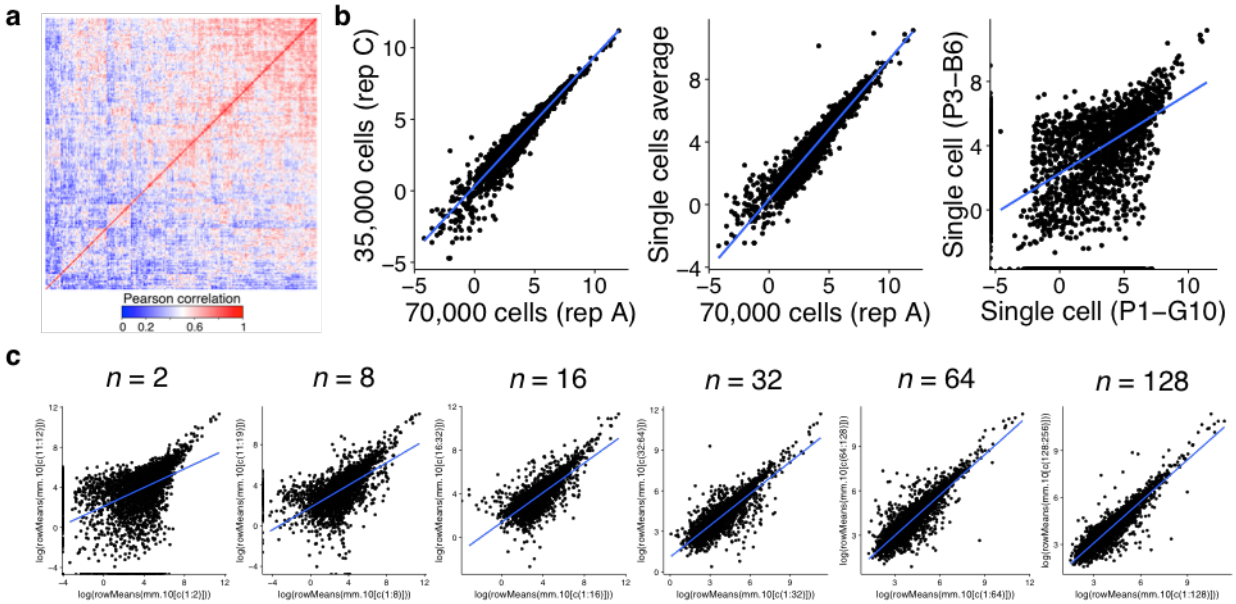

Figure S8. Gene expression correlation between individual macrophages with phagocytized *C. albicans*. (a) Heat-map displays Pearson's *r* values of the expression heterogeneity of the top 100 genes for 267 macrophages with phagocytized live *C. albicans* at 2 and 4 hrs. Red are highly correlated cells and blue are low correlated cells according the expression of the top 100 genes. **(b)** Plots of the gene expression correlation between two replicates of the sorted population of macrophages with phagocytized live *C. albicans* at 2 hrs. (left); between the single cells expression average and one replicate of the population (middle), and between two single cells (right). **(c)** Plots of the average gene expression correlation between increasing number of single-cells, *n* indicates the number of cells averaged.

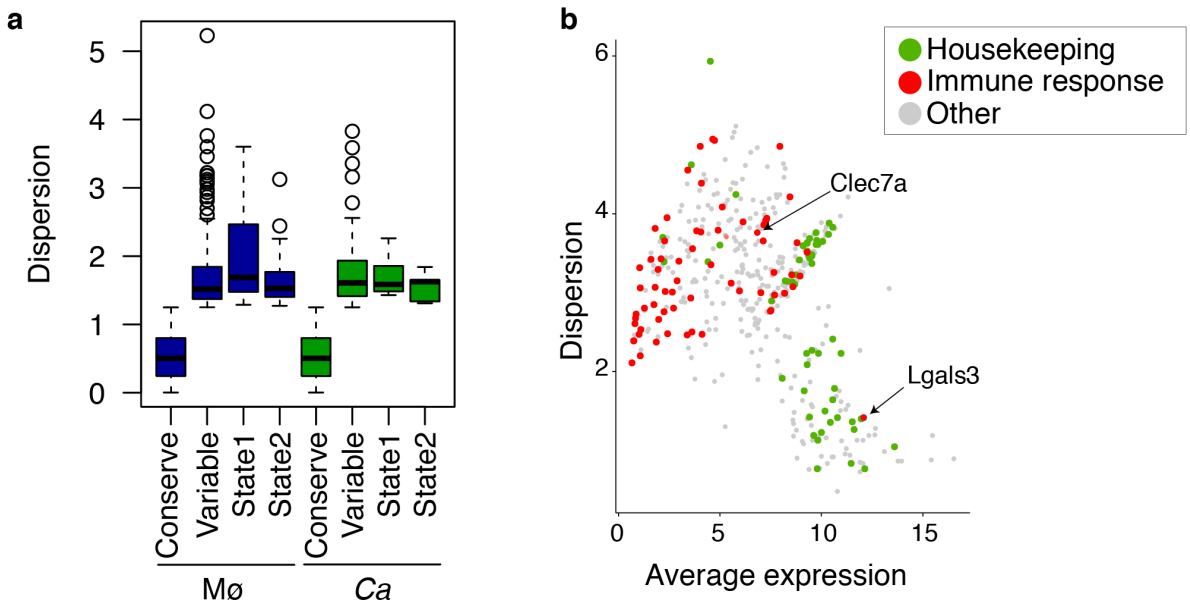

Figure S9. Expression variability at the single-cell level. (a) Boxplot of the dispersion of conserved and variable genes in macrophages infected and phagocytosed *C. albicans* for all genes, and for those in co-state1 and co-state2. **(b)** Relationship between average expression level in single macrophages with phagocytized *C. albicans* (*x axis*) and standard deviation (*y axis*) for ~3,000 genes. Red dots represent immune response genes, and green dots denote housekeeping genes. (**c**) Expression density distribution of five housekeeping genes across 267 macrophages with phagocytized live *C. albicans* cells at 2 and 4 hours. *y-axis*: cell density distribution; x-axis: single cell expression in log2(TPM +1) transformation.

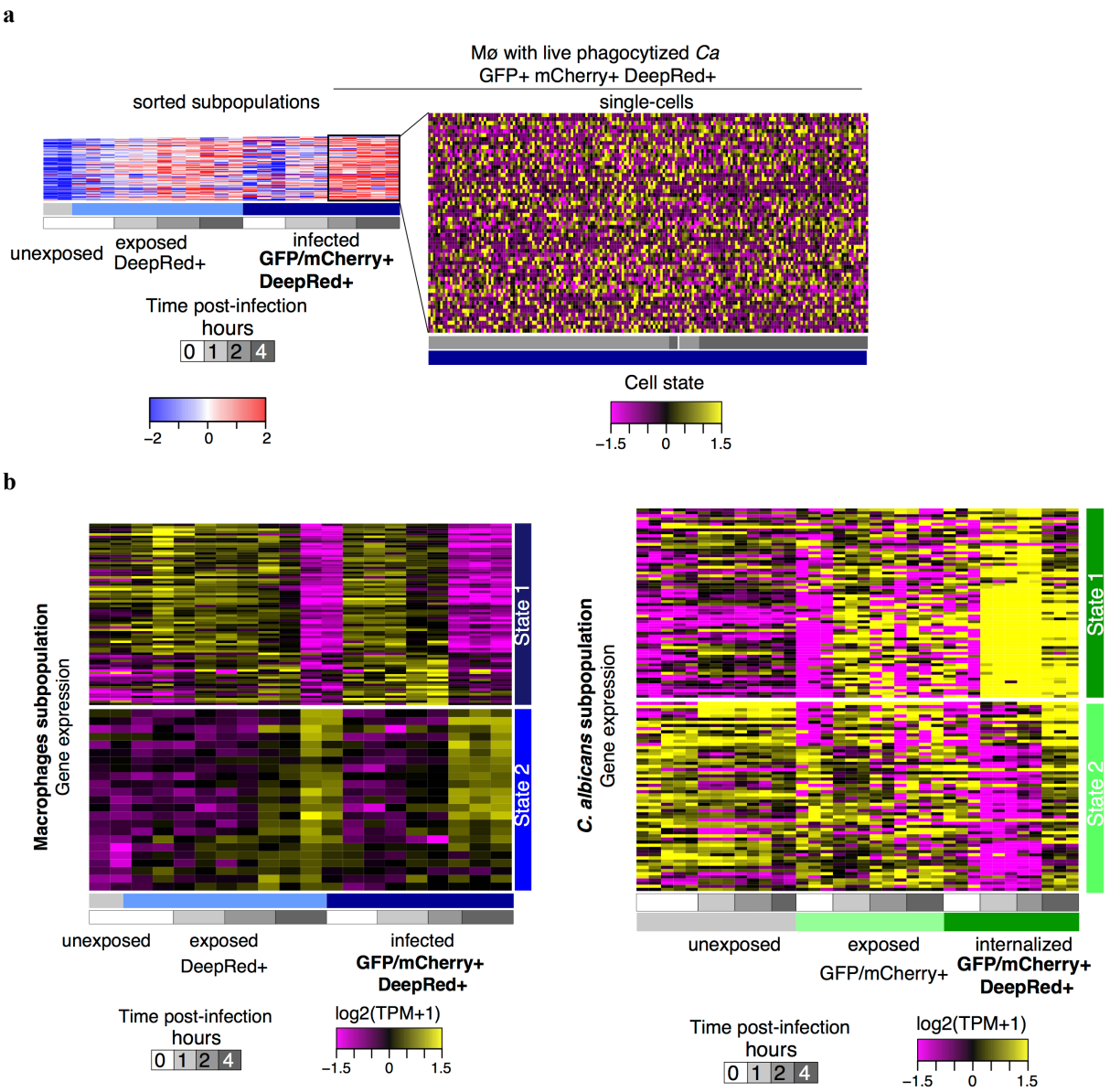

Figure S10. Expression variability at the single-cell level of highly expressed genes in the population of live *C. albicans*-infected macrophages at 2 and 4 hours. (a) Heatmap of DEGs in macrophages in sorted subpopulation samples (left) and expression of those genes in single infected macrophages (right). **(b)** Heatmap of the gene expression in the sorted subpopulations for the discriminative genes identified by single-cell analysis in co-state1 and co-state2 of single macrophages (left) infected with *C. albicans* (right).

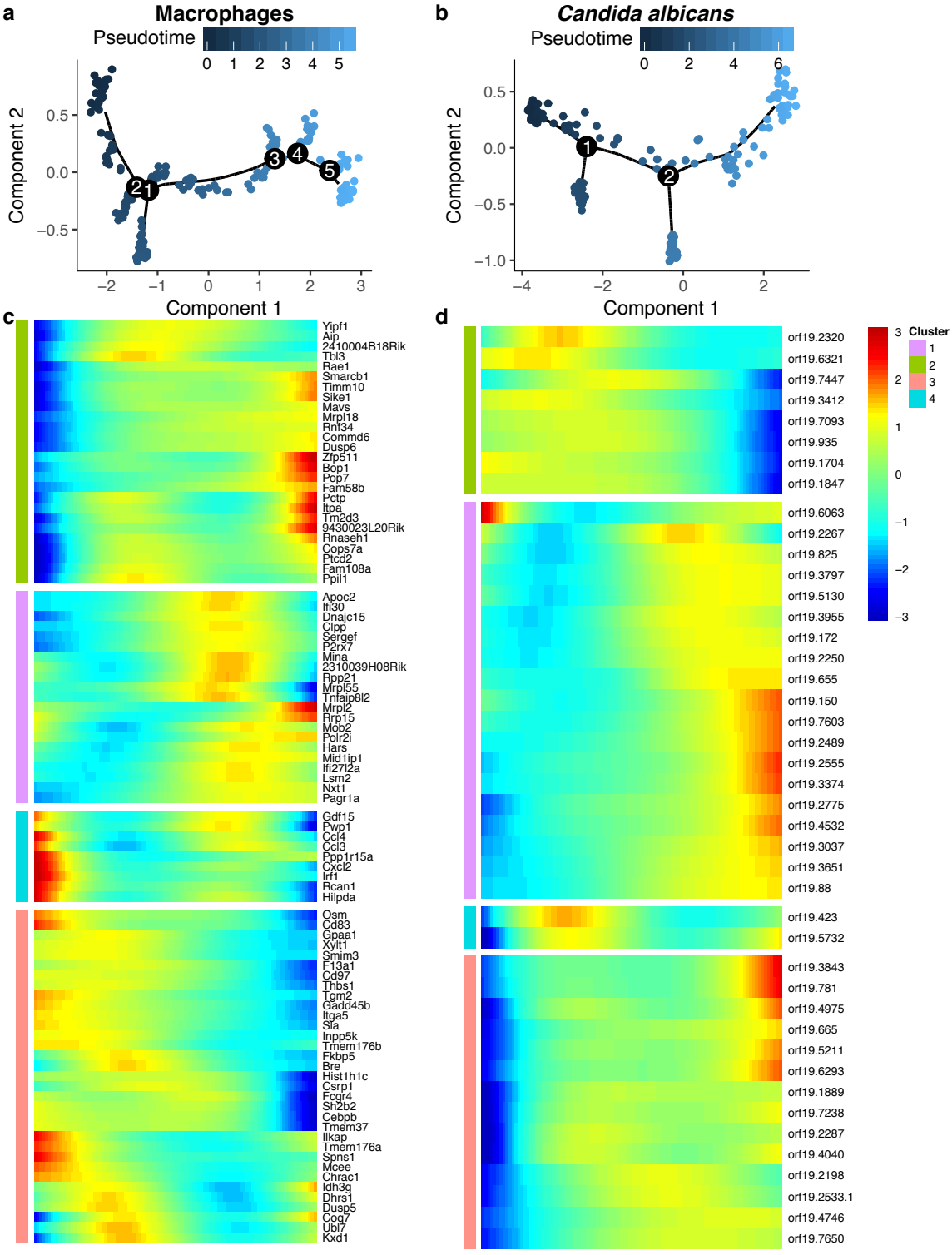

Figure S11. Pseudo-time analysis in single infected macrophages and *C. albicans*. tSNE plot for macrophages **(a)** and *C. albicans* **(b)** colored by the pseudo-time. Genes with most significant changes as a function of progress along the trajectory for **(c)** macrophages and **(d)** *C. albicans*.

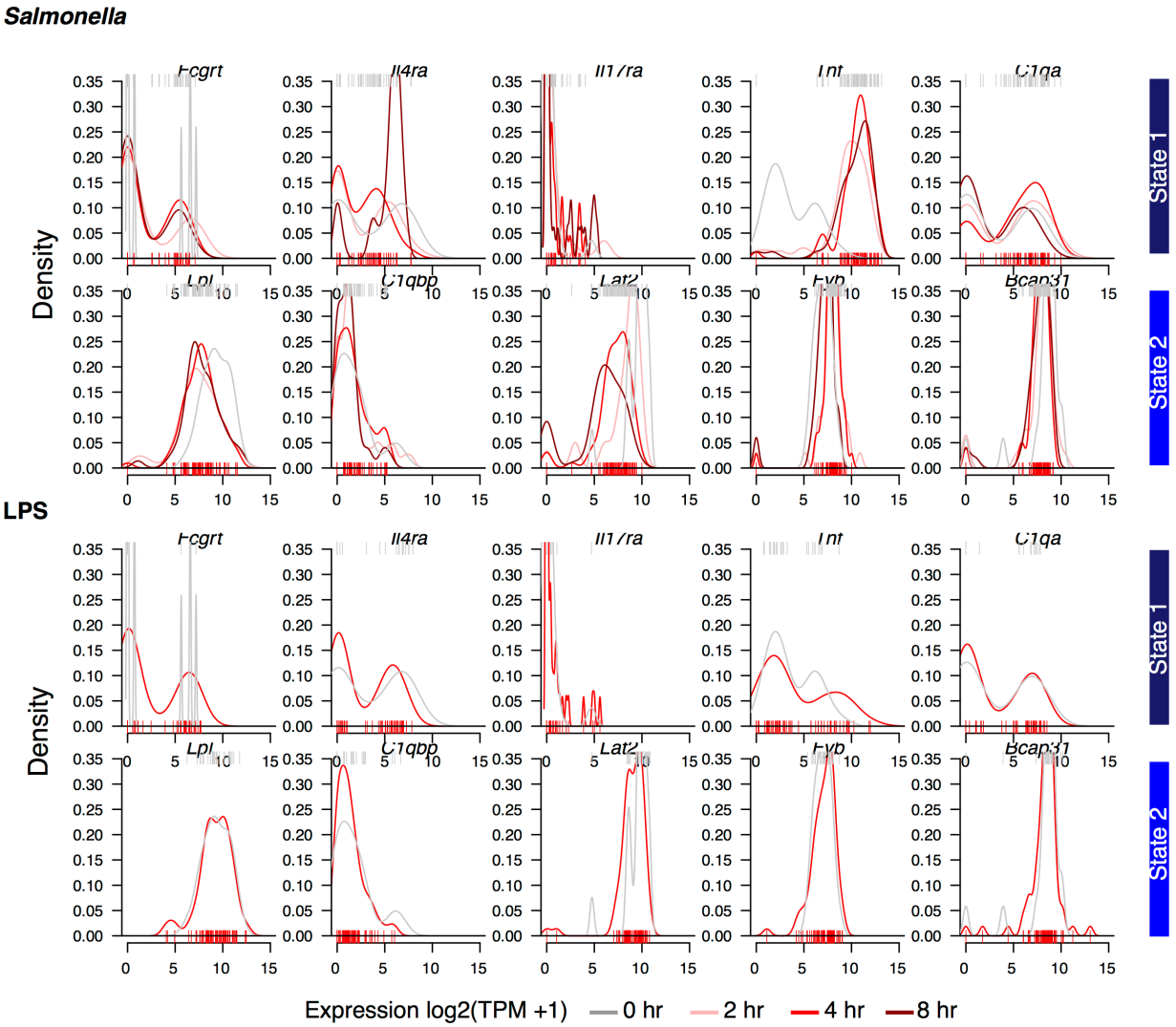

Figure S12. Comparison with single-cell RNA-seq macrophages response to other stimuli. Comparison of the expression density distributions for 5 top marker genes in co-state1 and co-state2 across macrophages-*C. albicans* single cells at 2 and 4 hours post infection (**Figure 5**) with the expression density of macrophages infected with *Salmonella* (top; $n = 94$) or macrophages exposed to LPS (bottom; $n = 94$) from Avraham, R., et al., 2015. Individual cells are plotted as bars underneath (red; 4 hours) or overhead (gray; 0 hours) for each distribution.

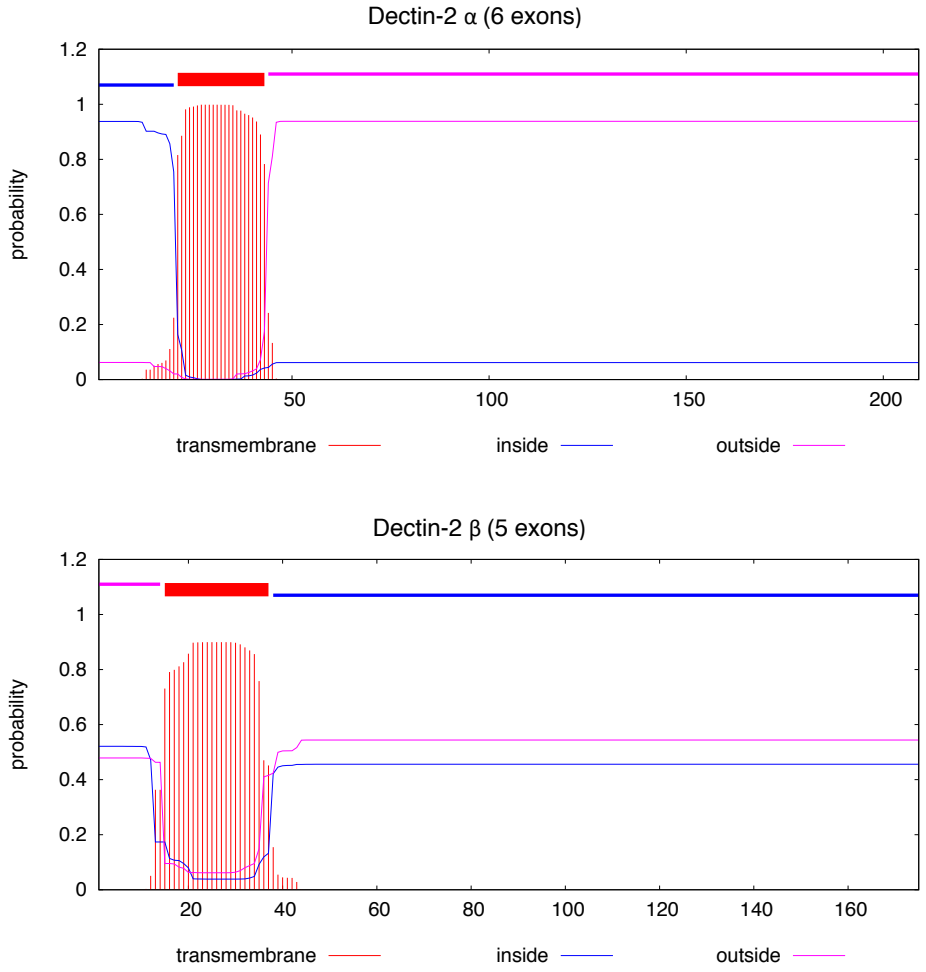

Figure S13. Prediction of transmembrane helices in Dectin-2. Prediction of transmembrane helices in the Dectin-2 α (Q9JKF4-1) and Dectin-2 β (Q9JKF4-2) amino acid sequences using TMHMM v2.0. and the hidden Markov model. The x -axis indicates the position in the amino acid sequence starting at the N-terminal end and the y -axis the probability of residing inside of, outside of, or in the membrane.

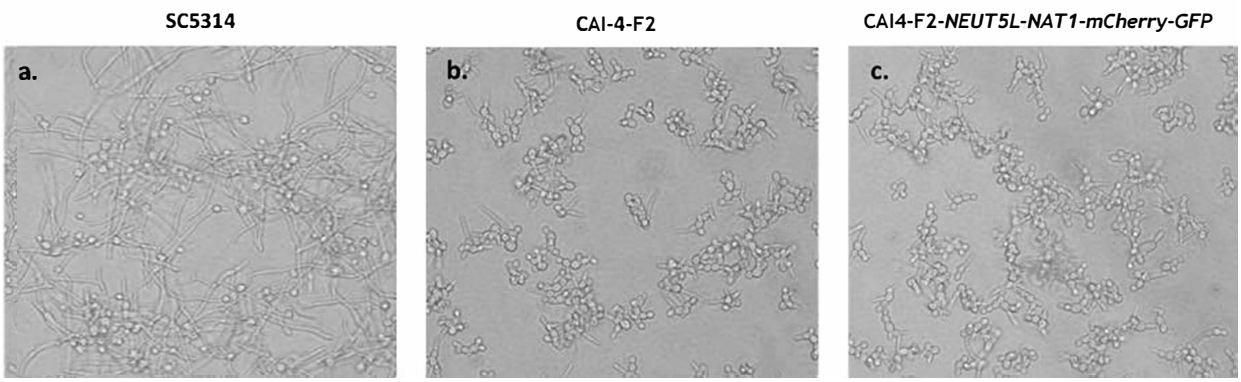

Figure S14. *Candida albicans* reporter strain displays reduced filamentation compared to SC5314. **(a)** SC5314, **(b)** CAI-F2 (reporter strain parent) and **(c)** CAI4-F2-*NEUT5L-NAT1-mCherry-GFP* strain (reporter strain) at 40x under identical growth conditions are phenotypically distinct. Fluorescent reporters were inserted into CAI-4-F2 **(b)** for our studies. Reduced filamentation made it possible to sort cells and monitor host interactions up to 4 hours at conditions that both mimic host environment *in vitro* and induce filamentation (RPMI medium, 10% fetal calf serum (not heat inactivated) at 37 °C and 5% CO_2 for four hours). **(c)** The resulting strain, CAI-4-F2-*NEUT5L-NAT1-mCherry-GFP*, is phenotypically similar to the parent and was verified as being resistant to Nourseothricin and fluorescent for both GFP and mCherry.

Chapter 3: Murine derived small intestine organoids and *Candida albicans* interactions

3.1 Abstract

While microbiome studies shed light on the ways in which bacterial species influence gut health, little is known about the ways in which the gut mycobiome alters the course of gastrointestinal health and disease. The most common fungus found in the adult intestine is *C. albicans*[22]. While *C. albicans* is not pathogenic in healthy adults, intestinal overgrowth of *C. albicans* is linked to Crohn's disease[166–170], which affects an estimated 1.6 million Americans[171–173]. The molecular mechanisms behind this overgrowth remain unknown. This project will examine complex transcriptional interactions between small intestine cells and fungi for the first time. This information could be critical for development of drugs or probiotics that effectively treat inflammatory bowel diseases, including Crohn's disease. Our goal is to develop a strategy to help define the transcriptional mechanisms by which *C. albicans* and small intestine epithelial cells interact with each other. In this work, we focus on *Candida albicans*, which is thought to be part of the resident microbiota of a healthy human intestine. We have developed a tool to simultaneously capture host and fungal pathogen gene expression at the population level, in murine derived small intestine organoids and *C. albicans*. We found that many of the same pathways upregulated in host epithelial cells exposed to *C. albicans* are also upregulated in macrophages exposed to or infected with *C. albicans.* However, epidermal growth factor binding was specifically upregulated in epithelial cells exposed to *C. albicans.* We found that *C. albicans* upregulates many of the same pathways when exposed to either epithelial cells and macrophages

but that *C. albicans* specifically downregulates genes involved in neurotransmitter catabolism when exposed to epithelial cells. Many uncharacterized genes were found to be specifically expressed by *C. albicans* when exposed to epithelial cells compared to macrophages, but expression of these genes was dependent on MOI. We also collected host expression data from single *C. albicans* infected epithelial cells. Applying this strategy to future work will facilitate our understanding of the interactions between *C. albicans* in and host gut, that could ultimately lead to therapeutic outcomes.

3.2 Introduction

Candida albicans is the sixth leading cause of hospital-acquired infections[2,3]. *Candida albicans* blood infections are associated with a ~30% mortality rate[8]. Blood infections can become disseminate *Candida albicans* throughout the body, eventually leading to infection of various organs such as the brain, liver, and kidneys. Immunocompromised individuals, such as those who have HIV/AIDs, have undergone chemotherapy treatments, or those on broad-spectrum antibiotics may succumb to these invasive *Candida albicans* infections. When patient treatment for *Candida albicans* blood infections is delayed for over 48 hours, associated mortality rates increases to 78%[10]. Many strains of *Candida albicans* have developed resistance to azoles[11,12], the major class of drugs used to treat both superficial and systemic infections.

Candida albicans is a dynamic organism that adapts to the host environment by changing morphology, transcriptional programs and the composition of its genome. *Candida* species are present in the guts of 70% of adults in the United States[22] and are generally commensal, non-

pathogenic residents. The most abundant fungal species in the gut is *C. albicans*[171]. *Candida albicans* over-growth is usually inhibited by the host immune system and other microflora. However, if a patient is on broad-spectrum antibiotics, for example, resident bacterial population dynamics can be altered and/or drastically reduced, resulting in the establishment of *Candida albicans* infections. Intestinal overgrowth of *C. albicans* is linked to Crohn's disease [166–170], and pathogenic forms of *C. albicans* can arise from benign strains that were once part of the commensal gut microbiota[172]. Pathogenic *Candida albicans* is phenotypically characterized as a strain with white-phase budding yeast cells, pseudo-hyphae and hyphae.[17,23]

Candida albicans can interact with the host epithelium by expressing adhesion proteins that allow cell-to-cell contact (adhesins) and promote biofilm formation[24]. This biofilm, a conglomerate of microorganisms, can also shield the *Candida albicans* pathogen associated molecular pattern (PAMP) molecules from host pathogen recognition receptors (PRRs), protecting the *Candida albicans* from opsonization. Additionally, when *Candida albicans* switch from the yeast to the hyphal form, the beta glucan molecules on the external cell wall are shielded from host PRRs, preventing host immune system detection[25]. The hyphal form of *Candida albicans* can result in organ damage, leading to dissemination of the fungus back into the bloodstream and eventually cause additional organ infections. While *Candida albicans* generally exists as a commensal organism in the small intestine of a majority of adults,[22] the genes important for maintaining commensalism and the host cues that may signal *Candida albicans* to switch to an invasive/pathogenic form are not well understood.

Candida albicans is able to not only survive and grow in commensal fashion inside of the host but is able to cause infections in diverse host niche environments like the oral cavity, vagina or gut. *Candida albicans* can also survive engulfment by phagocytic host macrophages, requiring rapid adjustment to changes in pH, oxygen levels, iron concentration, and other host molecules. Macrophages are phagocytic monocytes that are among the first immune cell responders to pathogens in the tissue and blood[28]. While many studies have characterized phagocyte interactions with *C. albicans*, less is known about how host cells interact with *C. albicans* in the intestines. However, it is documented that *C. albicans* hyphae formation can cause adhesion to epithelial cells and that fungal uptake can be a result of host cell endocytosis or active fungal penetration, leading to host cell damage and inflammation[174–177]. It was recently discovered that chronic (greater than 3 months) of sub-infection levels of *C. albicans* in the mouse gut can trigger a new *C. albicans* cell type, called GUT cells[17]; these cells were morphologically and transcriptionally different from initial wild type strain of *C. albicans* that was introduced into the mouse. Importantly, there is evidence that *Candida albicans* infections can originate from strains that were once commensal residents of the patients gut, rather than from exposure to a new, more pathogenic strain[108]. In order to better understand the signatures of *Candida albicans* exposed to epithelial cells and how these host cells respond in turn, we utilized small-intestine organoids, derived from the small intestines of mice, which better mimic the *in vivo* tissue structure where these dynamic interactions occur.

Mammalian, small intestine derived organoids in particular produce many of the major cell types found *in vivo,* despite the fact that the cell types in the small intestine vary greatly in terms of function and fate [113]. The small intestine contains villi structures to increase absorptive capabilities

and crypts, or invaginations at the end of villi which harbor stem cells. The small intestine contains a large reservoir of stem cells, allowing the organ to regenerate epithelial cells rapidly. It is estimated epithelial cells turn over every 3-5 days in mice and humans, for example. Lgr5 is used as marker for these stem cells, as it encodes a component of the WNT signaling pathway, a pathway indicated in cell renewal and stem cell control.[116] LGR5+ cells can be isolated from the small intestine or other gut tissue and used to initiate organoid culture. Specialized cell types with antimicrobial properties include mucous secreting Goblet cells, which secreting mucous, and cytokine and antimicrobial enzymes releasing Paneth cells. Specialized small intestine cell types have been shown to expand upon infection, and which cell types that expand are driven by pathogen type [178] (see **Introduction 1.6**) Because the organoids recapitulate some of the three-dimensional structure and cell types of the small intestine and are functional *in vivo* and dynamically respond to infection, they are more ideal for use in gut studies than traditional 2-D epithelial cell lines.

While *C. albicans* has been linked to gastrointestinal disease, it is much less studied that the "microbiome", which largely refers to the bacterial species that reside in the gut. We applied the dual host-fungal pathogen RNA-sequencing strategy (**Chapter 2**) to study the genetic interactions between *C. albicans* and small intestine epithelial cells, in two proof of concept experiments at specific MOIs; a predicted MOI of 1 *C. albicans* cell to 1 epithelial cell (intended to represent a more commensal *C. albicans* burden) and an MOI of 10 *C. albicans* cell to 1 epithelial cell (intended to represent a more pathogenic *C. albicans* burden).

3.3 Results

3.3.1 Murine small intestine organoid gene expression changes after exposure to *Candida albicans* bulk population level

The primary goal of this pilot experiment was to assess if *Candida albicans* could successfully penetrate a murine derived, small intestine organoid matrix and if we could detect both epithelial cell and *C. albicans* gene expression data simultaneously, using the same Smart-Seq2 strategy employed for macrophage and *C. albicans* population studies (**Chapter 2; Figure 2.1A**). The secondary goal of the pilot study was to perform differential gene expression analysis and tentatively characterize which genes are differentially expressed in both *C. albicans* and organoids during infection, relative to unexposed controls for both cell types (**Figure 3.2**). While using FACS to isolate specific infection populations is an ideal strategy to understand the heterogeneity within host-pathogen interactions, for this preliminary pilot study we did not sort cells prior to sequencing. Unlike *C. albicans* and host immune cells, gene expression networks of *C. albicans* and small intestine epithelial cells have not been characterized at the bulk, population level.

In order to assess if the *C. albicans* reporter strain expressing GFP and mCherry (see **Chapter 2**, **2.5 Methods**) could penetrate the epithelial cell organoid matrix, images were taken at 4 and 16 hours after organoids were exposed to the reporter strain, while both cell types were incubating in organoid media (**Figure 3.1**; **3.5 Methods**). At 4 hours, *Candida albicans* was beginning to line up at the boundaries of the organoid matrigel (**Figure 3.1A,B**). At 16 hours, *Candida albicans* was able to invade the matrigel (**Figure 3.1C,D)**

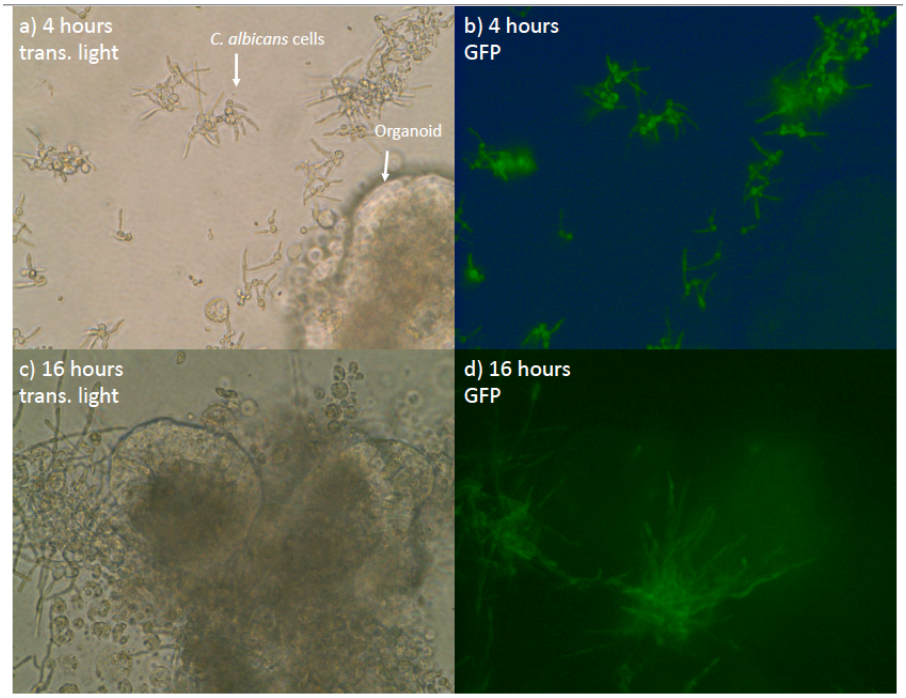

Figure 3.1: *Candida albicans* mCherry and GFP-expressing reporter strain interacts with 3-D murine derived, small intestine epithelial cell organoids in a gel matrix
Images taken at 40x after incubation at 37°C and 5% CO_2 with transmitted light or a 488 nm laser line to detect yeast expressing GFP at 4 hours (**a,b** respectively) and 16 hours (**c,d** respectively) at an estimated MOI of 10:1.

In order to assess if we could detect gene expression from both *C. albicans* and organoids at the unsorted population level and then perform differential gene expression analysis relative to unexposed controls for both cell types, we followed the workflow depicted in **Figure 3.2**.

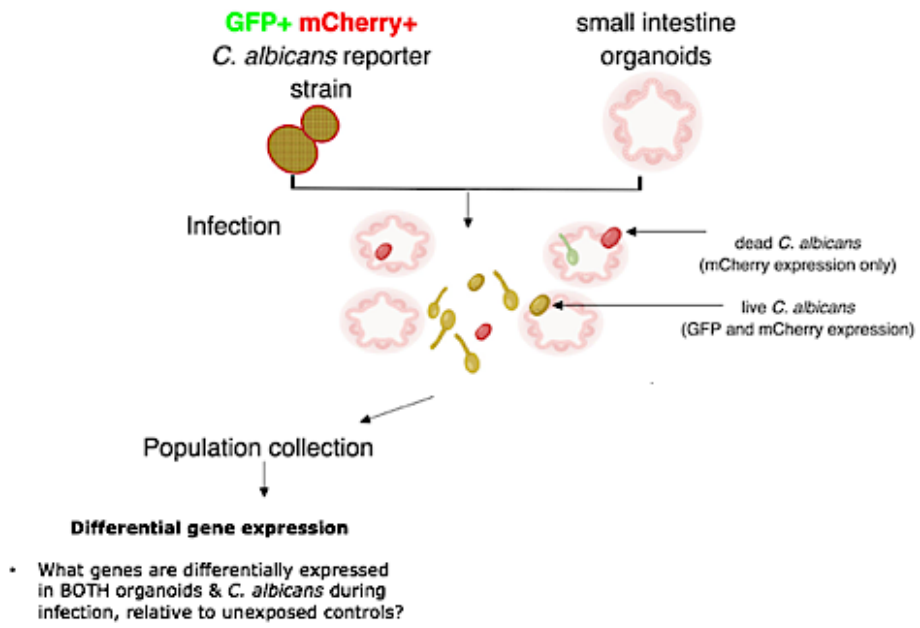

Figure 3.2: Murine organoid and *Candida albicans* population level RNA-seq pilot experiment design
A reporter strain of C. albicans constitutively expressing GFP and mCherry is exposed to small intestine organoid. For population level studies, samples are collected at specific time points (6, 12 and 24 hours for the pilot study) and then subjected to dual host-pathogen RNA-sequencing without sorting.

For this pilot experiment, we exposed organoids to *Candida albicans* in an estimated MOI of 10:1. Because a varying number of organoids can form in each well and organoid size can vary, it is difficult to estimate the number of host cells in each well. We estimated the number of epithelial cells in each well to be approximately 20,000 cells based on prior FACS data. Prior to sequencing, biological replicates of each time point were combined. We also tested sample lysis methods in this experiment. Samples containing *C. albicans* alone, epithelial samples alone and mixed samples were subjected to either: a) combination of chemical and mechanical lysis or b) chemical lysis only. However, mechanical lysis lead to a large increase in the number of transcripts detected

in *C. albicans* samples versus chemical lysis; samples that were mechanically lysed are labeled in bold font in **Figure 3.3**, and these samples were used for downstream analyses.

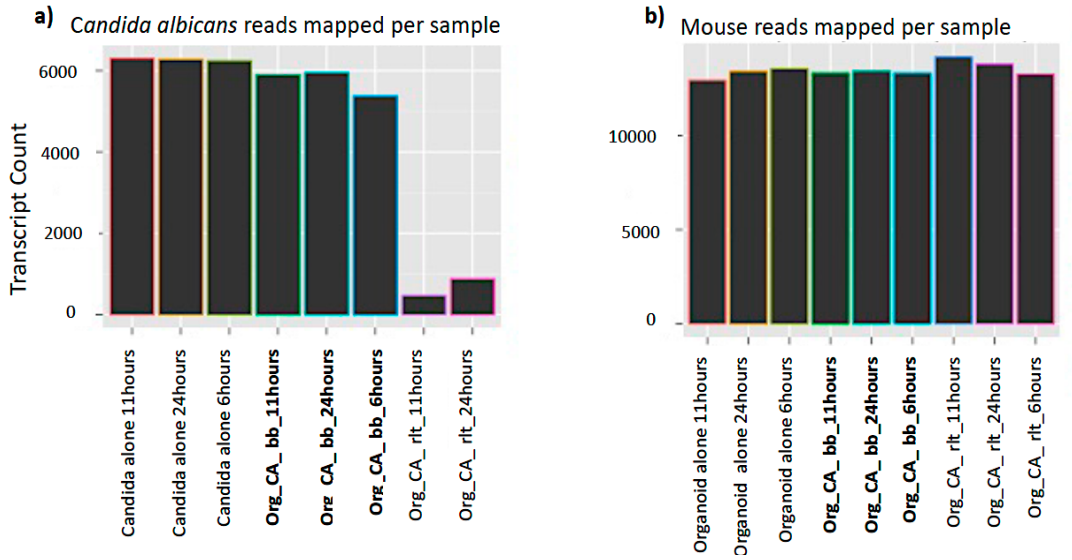

Figure 3.3: Both *Candida albicans* and mouse transcripts detected in each sample in the pilot study
C. albicans samples growing alone (labeled "Candida alone") samples were mechanically lysed via bead beading. Organoid epithelial cell samples growing alone were lysed chemically via RLT buffer (labeled "Organoid alone". Organoid and *C. albicans* mixed samples (labeled "Org_CA") were subjected to chemical lysis (labeled "rlt") or chemical and mechanical lysis, via bead beating (labeled "bb"). Samples were aligned to the *C. albicans* genome **(a)** and/or the mouse genome **(b)** and the number of transcripts detected (>1 reads RPKM) were plotted. Samples were collected at 6, 11 and 24 hours. The Y-axis indicates how many *C. albicans* and/or mouse transcripts were detected in each sample type.

3.3.2 *Candida albicans* gene expression changes after exposure to murine small intestine organoids at the bulk population level in a pilot experiment

Candida albicans upregulated and downregulated specific genes in response to host small intestine epithelial cells. First, we examined gene expression differences between sample conditions

collected over 6, 11 and 24 hours. Three biological replicates for each time point were pooled together prior to sequencing. The GO terms associated with the 20 most variably expressed genes among all *C. albicans* samples are "hyphal cell wall" (i.e. *DDR48, TDH3, PGA62, HWP1*) , and "Rho GTPase binding (i.e. *TEF1*, *TEF2*)" (**Figure 3.4**). These genes tended to be higher in *C. albicans* samples that were exposed to organoids. Interestingly, the *C. albicans* sample exposed to organoids at 6 hours clustered more closely to the *C. albicans* alone samples (6, 11 and 24 hours) than it did to the *C. albicans* samples exposed to organoids at 11 and 24 hours. Of note is uncharacterized gene orf19.2539, which is highly upregulated in *Candida albicans* samples exposed to organoids at 6 hours and upregulated in exposed samples at all time points relative to unexposed *C. albicans* controls.

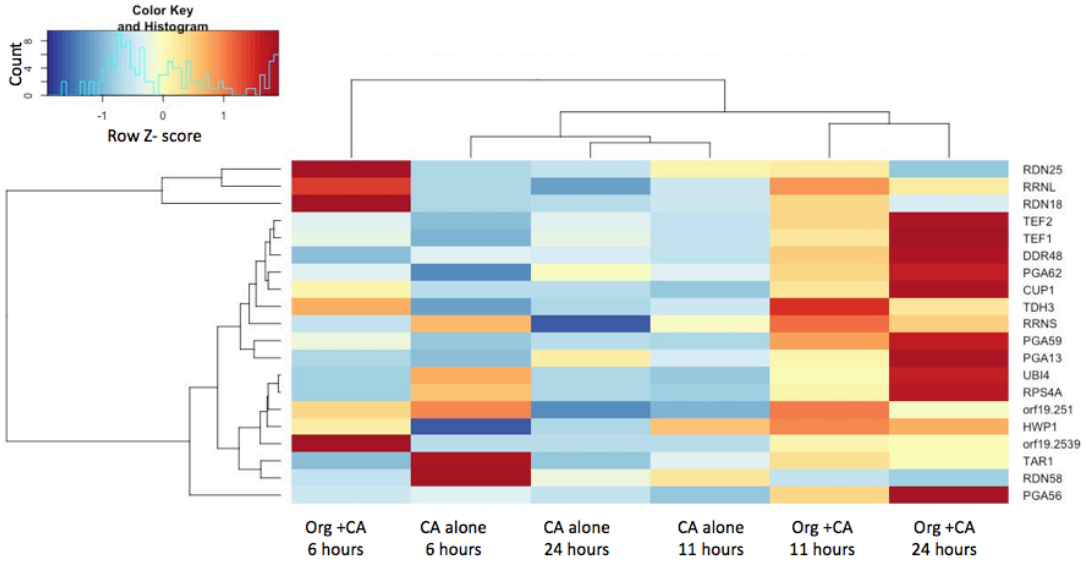

Figure 3.4: Top 20 variable genes across *Candida albicans* samples unexposed or exposed to organoids at an MOI of 10:1 and collected at 6, 11 and 24 hours in the pilot bulk RNA-sequencing experiment
Heatmap of sample row variance of the transformed expression values for the 20 most variable genes among samples. Each row is a gene and each column is a *C. albicans* sample. Sample time point and condition are labeled at the bottom of each column. "Org +CA" indicates sample consists of both *C. albicans* and organoid cells. The dendrogram on the left side of the heat map represents the hierarchical clustering of expression of the listed genes across samples. The dendrogram on the top of the heatmap represents the hierarchical clustering of the samples.

To perform differential expression analysis, exposed *C. albicans* samples at each time point (6,11, 24 hours) were pooled together and labeled as "infected", while unexposed controls were labeled "uninfected". Gene expression difference between populations of *C. albicans* exposed to epithelial cells versus unexposed *C. albicans* were assessed with the DESeq2 package[179] in R (**Methods section 3.5).** A volcano plot shows the log2 FC on the x-axis versus -log p-value on the y-axis (**Figure 3.5**). 32 genes were upregulated with the conservative thresholds of log2 FC of ≥ 2, adjusted p-value of ≤ 0.05 and FDR of < 0.05; gene expression levels meeting these criteria were considered upregulated in exposed samples relative to unexposed samples. A false discovery rate of < 0.05 was used to define a finding as significant. Here, we plot the p-value on a negative log scale so smaller p-values appear higher on the y-axis graph.

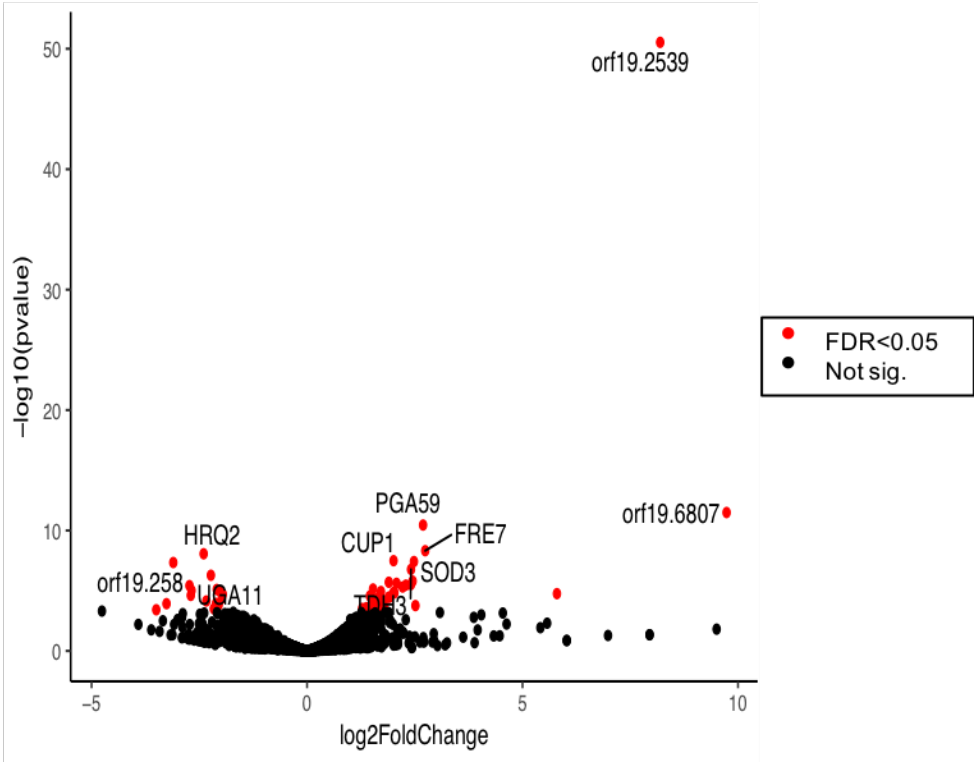

Figure 3.5: *Candida albicans* differential expression after exposure to murine small intestine organoids at an MOI of 10:1
C. albicans samples exposed to small intestine epithelial cells significantly upregulate and downregulate specific genes compared to unexposed controls.

In *C. albicans* exposed to organoids, 16 genes were upregulated with the thresholds of log2 FC of ≥ 2, adjusted p-value of ≤ 0.05 and FDR of < 0.05; gene expression levels meeting these criteria were considered upregulated in exposed samples relative to unexposed samples. These *C. albicans* genes included 6 uncharacterized genes (*orf19.2539, orf19.6807, orf19.3354, orf19.7077, orf19.4406, orf19.2662*) (**Table 3.1**). The GO terms associated with genes include "oxidoreductase activity", "filamentous growth", "pathogenesis", "response to stress", and "interspecies interaction between organisms".

When biological replicates for each time point were collapsed by condition and analyzed, *orf19.2539* counts were upregulated in exposed samples versus unexposed samples, which is what we expect based on transcript counts in each replicate (**Figure 3.6**) .

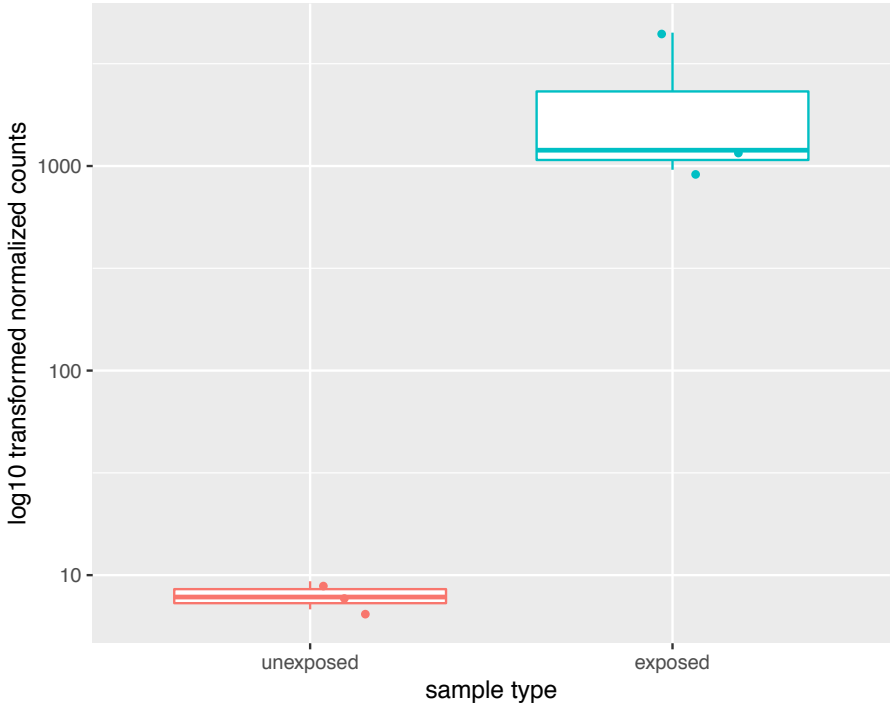

Figure 3.6: *Orf19.2539* transformed and counts when biological replicates are collapsed and analyzed by condition at an MOI of 10:1
A box plot depicting *C. albicans orf19.2539* expression in unexposed (pink) and exposed samples (blue) when *C.* albicans is exposed to epithelial cell at an MOI of 10:1.

Table 3.1 Six uncharacterized genes are significantly upregulated in populations of *Candida albicans* exposed to small intestine epithelial cells at an MOI of 10:1

Gene name	Normalized Mean Expression	Log2 fold change	p-value	FDR
orf19.2539	1110.18	8.19	2.92^{-51}	<0.05
orf19.6807	15.004	9.76	3.24^{-12}	<0.05
orf19.3354	99.026	2.24	1.59^{-06}	<0.05
orf19.7077	489.37	2.08	2.42^{-06}	<0.05
orf19.4406	237.41	2.03	1.53^{-05}	<0.05
orf19.2662	4.95	5.80	1.75^{-05}	<0.05

The GO terms associated with genes downregulated in *C. albicans* exposed small intestine epithelial cells include "neurotransmitter metabolic processes", "regulation of neurotransmitter levels" and "alpha-amino acid catabolic process". This was unique to organoid exposure and did not occur in the same *C. albicans* reporter strain when exposed to macrophages (**Chapter 2, section 2.3.2**).

3.3.3 Targeting *orf19.2539* for mutation via a *Candida albicans* optimized CRISPR-Cas9 system

In order to assess to further assess the function of these uncharacterized genes, we wanted to develop a rapid mutation and/or deletion approach. Because *C. albicans* is diploid, traditional chemical or electric transformation that relies on homologous recombination is more difficult than in other yeasts, such as haploid *S. cerevisiae*. New, *C. albicans* specific CRISPR methods have been developed recently[180,181]. As *Candida albicans* translates CUG to serine, Cas9 had to be recoded to work in this fungal clade[180]. Cas9 enhances the efficiency of homologous

recombination by introducing a cut. A synthetic repair template can then mutate or replace both copies of a gene in one transformation reaction. This system allows for mutation of gene families[180].

In order to investigate the function of uncharacterized genes found to be upregulated or down regulated in *C. albicans* exposed to organoids, we decided to test two CRISPR approaches, with the help of the Shapiro lab at the University of Guelph. The first approach we tested was to introduce a loss of function mutation into *orf19.2539*, by causing a frameshift mutation leading to an early stop codon, i.e. a loss of function (LOF) mutation. Because *C. albicans* can't maintain replicative plasmids, a guide RNA is cloned into a plasmid and then the plasmid is linearized. The resulting fragment contains the *CAS9* gene, the nourseothricin resistance gene, an *SNR52* promoter (which recruits Pol III) upstream of the guide RNA (gRNA) to gene of interest and homology to Neut5, a locus in the genome which has been shown to be devoid of open reading frames and amenable to large genetic insertions; inserting these large fragment at the site of the open reading frame could disrupt other genes. This gene block and a double stranded repair template with 90 base pairs of homology surrounding the gene cut site and containing mutation of interest are delivered into *C. albicans* cells via lithium acetate transformation. Although Cas9 will be constitutively expressed in the genome, once the genetic edit interest of interest is made, Cas9 will no longer cut the target site.

We used the CRISPR LOF approach to create mutations in both SC5314 (traditional, wild type *C. albicans*) as well as CAI4 (the genetic background of our fluorescent reporter strain used in the RNA-sequencing experiments). *Orf19.2539* was highly expressed in *C. albicans* exposed to

epithelial cells at 6 hours and upregulated in exposed samples at 12 and 24 hours in our pilot RNA-seq experiment at an estimated MOI of 10:1 (**Figure 3.5, Figure 3.6**). As such, we chose to target this gene via CRISPR. The first approach we tried was mutagenesis, as this strategy is more cost effective compared to a full knockout. The repair template delivered to *C. albicans* cells contained the first 90 bases of *orf19.2539* but with a 2 base pair deletion that would introduce a stop codon at amino acid 19 in the predicted protein product (**Methods**). **Figure 3.7** shows an alignment of the *orf19.2539* wild type target sequence and the mutated sequence containing the desired 2 base pair deletion, resulting in a premature stop codon at amino acid 19.

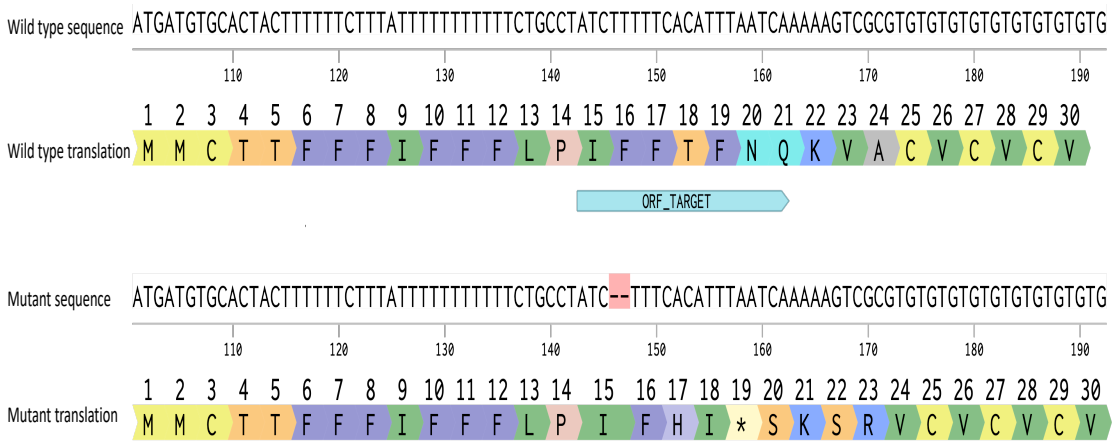

Figure 3.7: *Orf19.2539* wild type sequence and desired mutant sequence alignment
A portion of the wild type *orf19.2539* sequence and translation are visualized on the top two tracks. The open reading frame (orf) region targeted by CRISPR is indicated in blue. The desired mutated sequence with a 2 base pair deletion (indicated in red) is shown with the resulting amino acid translation below; the deletion leads to a premature stop codon at amino acid 19.

As part of the DNA fragment we transformed into cells contains a nourseothricin resistant gene, transformants were plated on rich media containing 200 ug/ml of nourseothricin . We observed varying colony sizes in both genetic backgrounds **(Figure 3.8)**

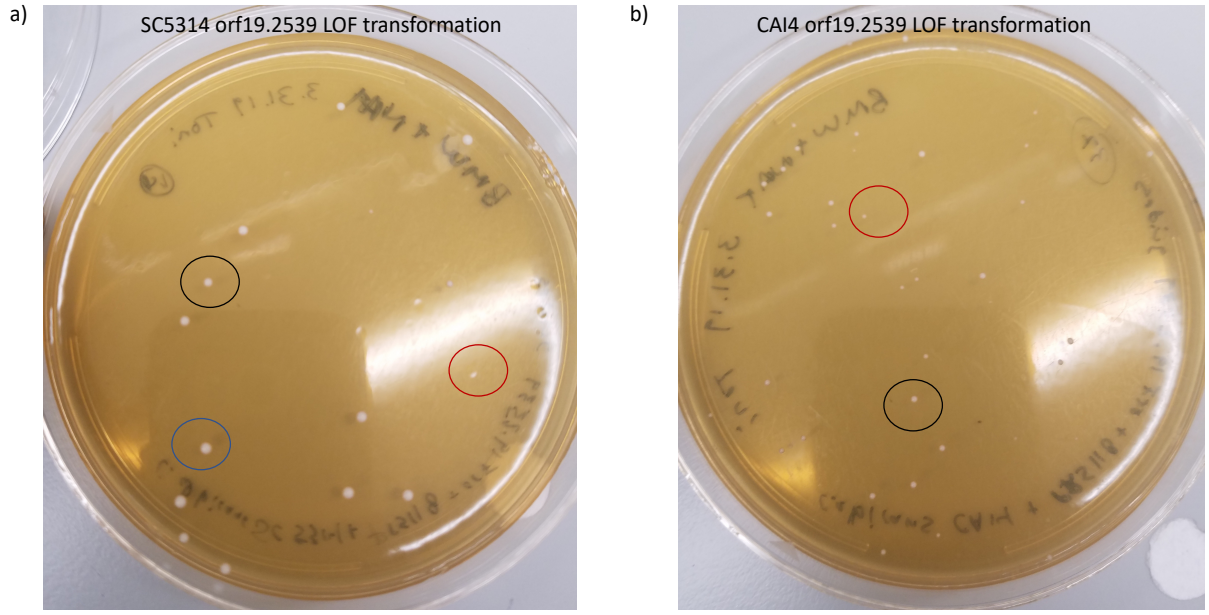

Figure 3.8: Colony size varies in two genetic backgrounds of *Candida albicans orf19.2539* LOF transformants
C. albicans was transformed to induce a loss of function mutation in *orf19.2539* via CRISPR. (a) After transformation in SC5314, three colony sizes were observed: larger than average (blue circle), average (black circle) and smaller than average (red circle). (b) After transformation in SC5314, two colony sizes were observed: average (black circle) and smaller than average (red circle).

Prior to Sanger sequencing, we amplified a region of the *CAS9* gene to assure it was integrated into the transformants genome **(Figure 3.9**). Only transformants that tested positive for Cas9 genome integration were screened via Sanger sequencing.

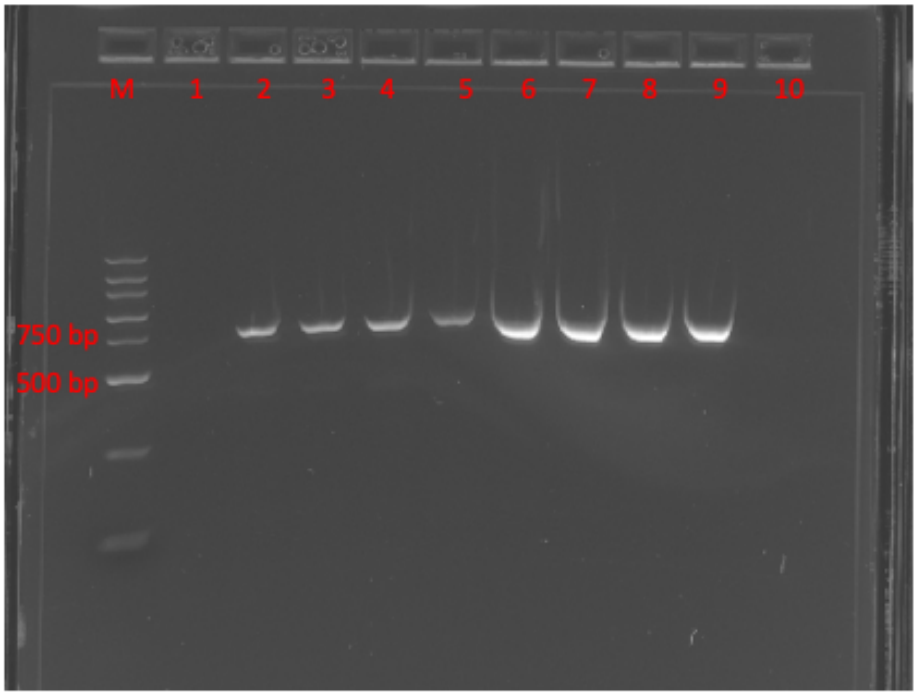

Figure 3.9: *CAS9* amplification in *orf19.2539* LOF transformants in strains SC5314 and CAI4
Amplification of a 760 base pair fragment of *CAS9* gene in colonies post transformation *were* run on a 1% gel. Lanes 2-5 are SC5314 transformants. Lanes 6-9 are CAI4 transformants. Lane M: Genscript 100 base pair ladder; 600 base pairs and 750 base pair ladder bands are labeled. Lane 1: colony #1, Lane 2: colony #2, Lane 3: colony #8, Lane 4: colony #10, Lane 5: colony #15, Lane 6: colony #6, Lane 7: colony #10, Lane 8: colony #14, Lane 9: colony #15, Lane 10: no DNA PCR control.

We amplified a 350 base pair fragment around the targeted *orf19.2539* cut site (~150 base pairs up and downstream and attempted to used Sanger sequencing to verify the sequence of this region in transformation clones. The *orf19.2539* transformation was repeated 4 times and 67 clones were screened with Sanger sequencing. However, this approach led to inconclusive results.

Next, we created whole genome libraries (Nextera XT, Illumina) of two transformed colonies, #15 and #12. Post transformation, colony #15 had a wild type colony size, while colony #12 had a small colony size. Samples were pooled together and sequenced at a length of 50 base pairs in one

direction. Samples were aligned to a 500 base pair reference containing 200 base pairs up and down stream of *orf19.2539* as well as the open reading frame sequence of *orf19.2539* containing the targeted 2 base pair deletion. Alignments were visualized in the Integrative Genomics Viewer (IGV)[182] (**Figure 3.10**). Sequencing indicated that there was a 2 base pair insertion (TT) at the targeted cut site (**Figure 3.10;** black box**)**

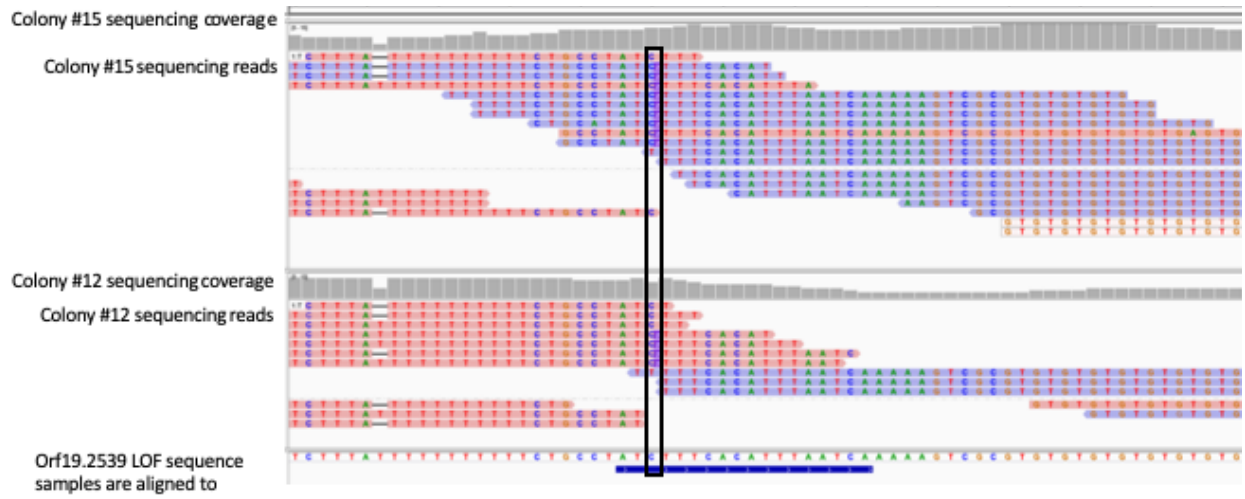

Figure 3.10: Sequencing reads from whole genome libraries of SC5314 *orf19.2539* LOF colony #15 and #12 aligned to the targeted mutated sequence.
SC5314 *orf19.2539* LOF transformants, colony #15 and colony #12, whole genome sequencing data was aligned to a 500 base pair, mutated reference sequence for this gene and aligned reads colored by strand (red for Watson strand and blue for Crick strand) were visualized by the Integrative Genomics Viewer. Sequencing coverage of the area is depicted by gray bars. The 50 base pair sequencing reads are displayed (red and blue rectangles) and each base call is shown. The bottom track shows the mutated reference sequence that each sample was aligned to. The blue bar indicates the location of the mutated open reading frame. The black box shows the target mutation site, and that most reads have a 2 base pair insertion (TT) at the mutated site.

These results clearly indicate that these transformants did not contain the desired two base pair deletion which would result in a loss of function mutation in *orf19.2539*.

We next created a SC5314 strain with a deletion of *orf19.2539*. This strategy was not used initially, as it requires synthesis of an approximately 1 kb fragment. The concept is similar to that of the first approach, but 2 guide RNAs are introduced that cut at the beginning and the end of the gene of interest and are flanked by homology arms to the upstream and downstream regions of *orf19.2539*. The gene block is integrated into the genome at *Neut5L*. After Cas9 induces double strand breaks, the homology arms flanking the guides are used as a repair template during homologous recombination. The result is a deletion of orf19.2539 at both alleles[183] (**Figure 3.11**).

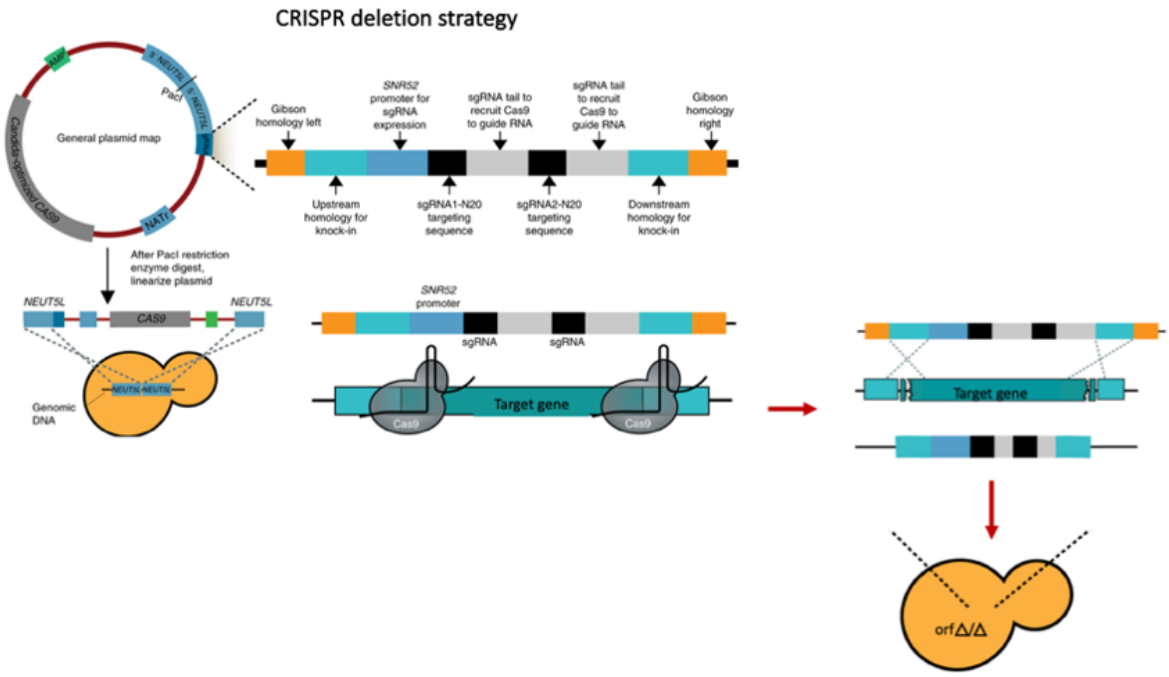

Figure 3.11: CRISPR gene deletion strategy used to create *orf19.2539* mutant in SC5314
Adapted from Shapiro et al., 2018. A gene block is synthesized and cloned into a general plasmid backbone via Gibson homology. The block, integrated into *Neut5L*, contains Cas9 and the nourseothricin resistance. The block also contains *SNR52* promoter to recruit Pol III. The two guide RNAs are flanked by homology to the gene region. The guide RNAs recruit Cas9 cut at 5' and 3' end of the target gene. After Cas9 induces double strand breaks, the target gene homology arms are used as a repair template during homologous recombination and both alleles of the gene (i.e. *orf19.*2539) are deleted.

A 50 base pair fragment of the 3' end of the gene (Crick strand) was not targeted for deletion, do to gene block synthesis design issues. We were able to successfully delete the majority *orf19.2539* and observed a small cell size in these mutants, compared to wild type (**Figure 3.12**). Four sequencing reads map to a repetitive, internal portion of *orf19.2539,* but these reads were mapped with low confidence and likely mapped in error; IGV displays reads mapped with low confidence in transparent colors.

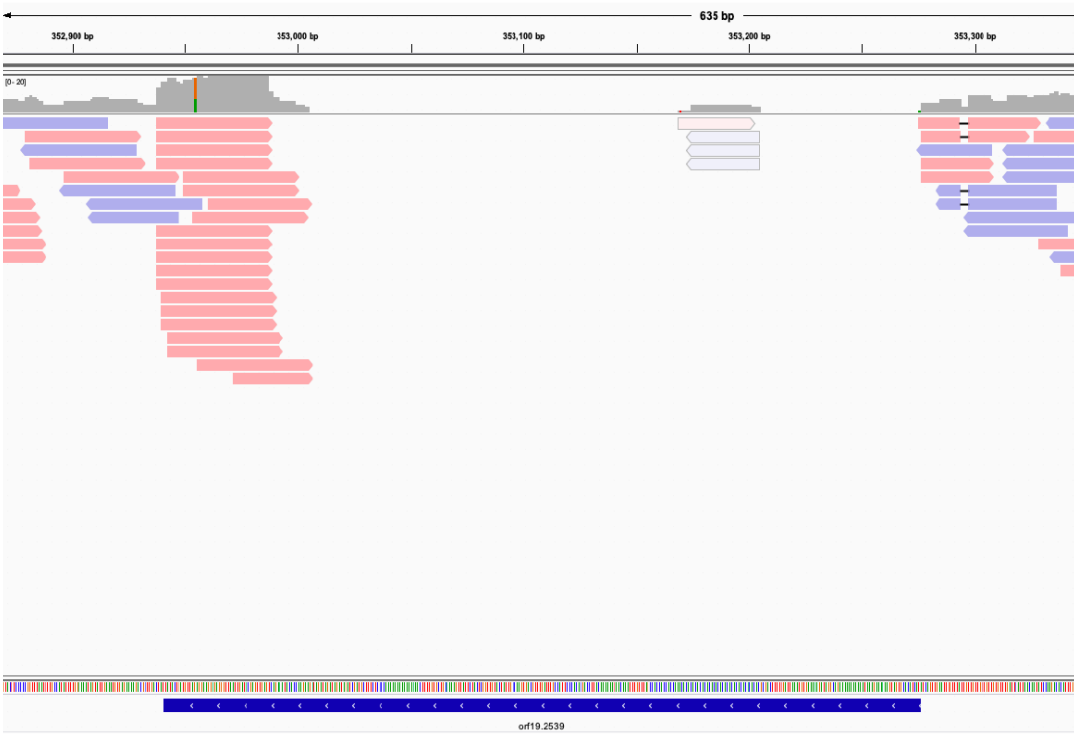

Figure 3.12: Sequencing reads from a whole genome library of SC5314 orf19.2539 knockout aligned to the *Candida albicans* genome
SC5314 *orf19.2539* deletion mutant whole genome sequencing data was aligned to the *Candida* albicans reference sequence. Sequencing reads colored by strand (red for Watson strand and blue for Crick strand) were visualized by the Integrative Genomics Viewer. Sequencing coverage of the area is depicted by gray bars. Genome location and the length of base pairs visualized in the window is indicated above the sequencing coverage track. The blue bar indicates the location of the wild type open reading frame. The black box shows the target mutation site, and that most reads have a 2 base pair insertion (TT) at the mutated site.

These results indicate that we successfully created an *orf19.2539* mutant in SC5314.

3.3.4 Gene expression difference between populations of *C. albicans* exposed to epithelial cells versus populations of macrophages

We previously examined how *C. albicans* gene expression varied across exposed and phagocytosed cells over time (**Chapter 2, Figure 2.5B**). At 1 and 2 hours upon phagocytosis (cluster 2) are involved in early fatty acid oxidation response and transmembrane transport (Opt and Hgt classes; Tables S2 and S3). These transporters differ from those in cluster 1 in that they show peak expression at 1 hour in phagocytosed cells and decrease in expression by 4 hours (**Figure 2.2B**). Clusters 1 and 2 contained 6 confirmed or putative transcription factors (*SUT1, STP4, TEA1, ADR1, ZCF38, TRY4*), all of which encode zinc finger containing proteins; zinc cluster transcription factors have been implicated in *C. albicans* virulence[128,129].

Notably, 49 out of 57 GO terms overlapped exclusively with GO terms associated with *Candida albicans* internalized by macrophages. All of these terms are associated with Cluster 4 (**Figure 2.5B**) , which were genes repressed during the first 2 hours of phagocytosis but upregulated 4 hours after phagocytosis. These overlapping GO terms are involved in nucleoside metabolic processes, gluconeogenesis and host adaptation. The subset of GO terms that were unique to *C. albicans* exposed to organoids are listed in **Table 3.2**. These results suggest that most of the pathways are upregulated when *C. albicans* is exposed to or phagocytosed by macrophages and when exposed to epithelial cells.

Table 3.2: GO terms uniquely assigned to upregulated genes in *Candida albicans* exposed to epithelial cells at an MOI of 10:1

***Candida albicans* GO term ID**	**GO term**
16052	carbohydrate catabolic process
6733	oxidoreduction coenzyme metabolic process
51186	cofactor metabolic process

3.3.5 *Candida albicans* gene expression changes after exposure to murine small intestine organoids at the bulk population level at an estimated MOI of 1:1

In an additional experiment, we tested the effects of organoid exposure to *C. albicans* at 8,12 and 24 hours. Three replicates were collected for each time point as well as time matched controls of *Candida albicans* growing alone in organoid media. For this test, we used an estimated MOI of 1:1. Tenfold less *Candida* were used in this experiment than were used in the pilot, to try to reduce organoid cell death and mimic the conditions of our experiments involving macrophages and *C. albicans*. For this experiment, the analysis was performed differently than in the pilot, because replicates were not pooled prior to sequencing **(Methods**). Similar to our pilot RNA-seq study, approximately 6,000 genes were detected in each *C. albicans* containing sample (**Figure 3.13**)

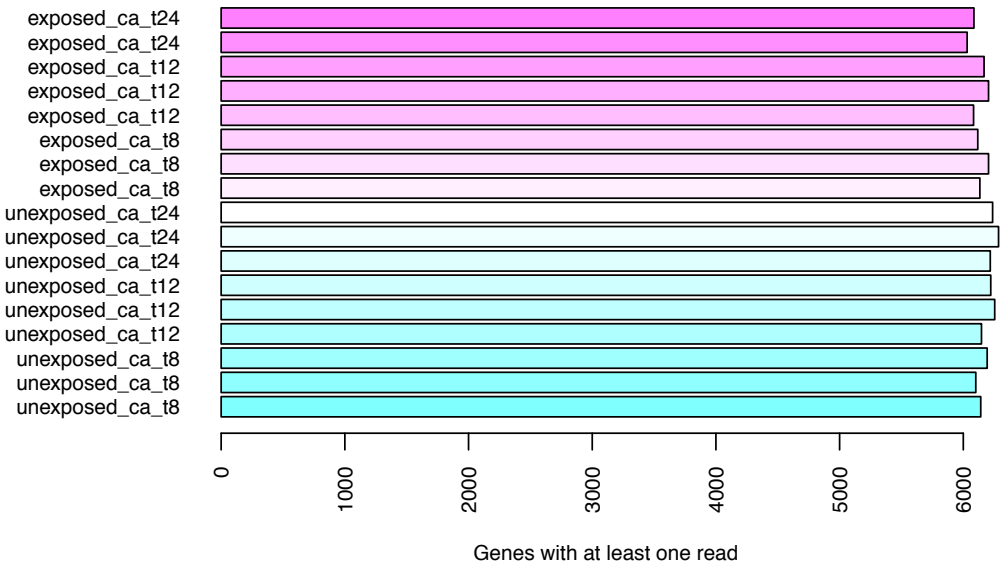

Figure 3.13: The number of genes with at least one transcript detected in each sample in *Candida albicans* and organoid RNA-sequencing experiment at MOI 1:1
Transcripts were aligned to the *C. albicans* and the number of transcripts detected (>1 reads FPKM) were plotted. The X-axis indicates how many *C. albicans* and/or mouse transcripts were detected (≥1 transcript detected) in each sample type. "Exposed" samples were exposed to small intestine epithelial cells at 8, 12 or 24 hours. Unexposed *C. albicans* samples were growing alone in organoid media and collected in parallel at 8, 12 or 24 hours.

PCA analysis shows us that at an MOI of 1:1, most of the variance in gene expression can be explained by time (**Figure 3.14**), rather than exposure status.

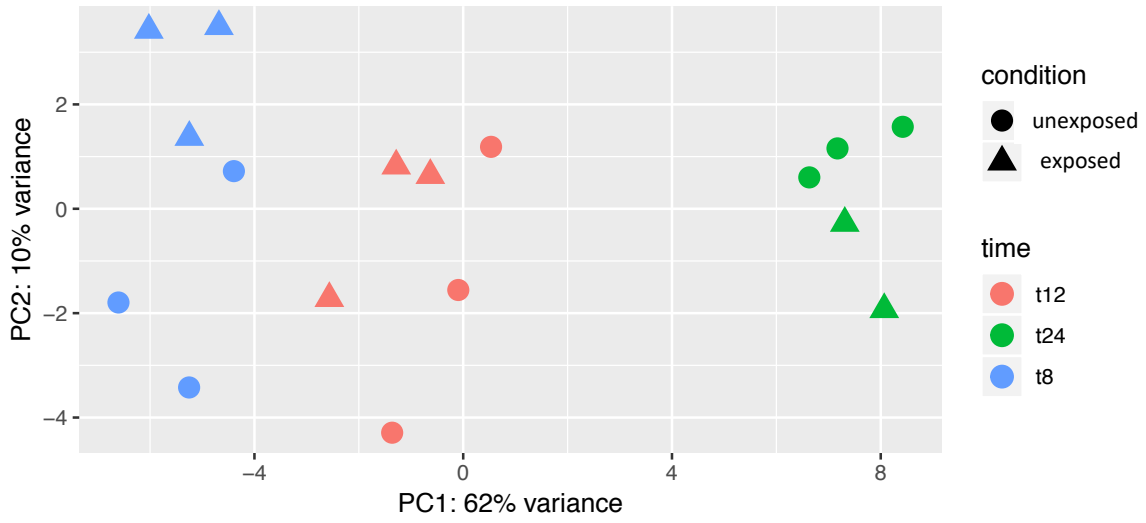

Figure 3.14: PCA of *Candida albicans* samples exposed to organoids at an MOI of 1:1
Principal component analysis (PCA) using the transcript abundance for *C. albicans* samples exposed to epithelial cells (circles) or unexposed (triangles at 8 hours (blue), 12 hours (pink) or 24 hours (green).

In order to generate p-values for any genes which were differentially expressed due to condition at any time, we used a Likelihood Ratio test (LRT) and a reduced model for condition; genes with small p-values here mean that expression was more likely affected by condition than time alone. If we examine the genes that have a p-value of <0.05 and LFC of at least 1, 32 genes are found to be upregulated in *C. albicans* exposed to organoids over all time points. We next used an adjusted p-value of <0.05 to test for significance; the adjusted p-value takes into account FDR and thus an adjusted p-value of less <0.05 indicates that less than 5% of significant findings will be false positives. Using this cut off, only 7 genes were found to be significantly upregulated at a LFC > 2, over all time points in *C. albicans* exposed to organoids versus unexposed (**Table 3.2**). at a predicted MOI of 1:1. GO terms associated with these genes (Methods) include: hydrolase activity, transporter activity, DNA metabolic processes, DNA binding, oxidoreductase activity, response to

stress, lipid and metabolic processes, as well as molecular function unknown and biological process unknown.

Table 3.3: Genes which are consistently, significantly upregulated from 8 to 24 hours in populations of *Candida albicans* exposed to small intestine epithelial organoid cells at an MOI of 1:1

Gene name	Mean Normalized Expression	Log2 fold change	Adjusted p-value
DAL7	7.455572	3.029383	0.000453
orf19.936	12.37869	3.035466	3.3E-09
orf19.4383	12.1955	4.038804	2.47E-10
orf19.1844	8.919061	4.656661	0.000419
snR41b	2.975429	2.119501	0.00681
LIP3	4.976392	4.561962	0.026114
orf19.3120	2.437112	9.299654	0.018541

Interestingly, we did observe increased expression of *orf19.2539* levels over time experiment in both unexposed and exposed *C. albicans* samples at MOI 1:1. However, *orf19.2539* expression levels were higher in exposed samples compared to unexposed samples (**Figure 3.15**) and the slope of the line fitted to exposed expression was significantly non-zero (p-value = 0.029). In our pilot experiment (MOI 10:1), *orf19.2539* was one of the most variably expressed genes among *C. albicans* exposed organoid samples (**Figure 3.4**).

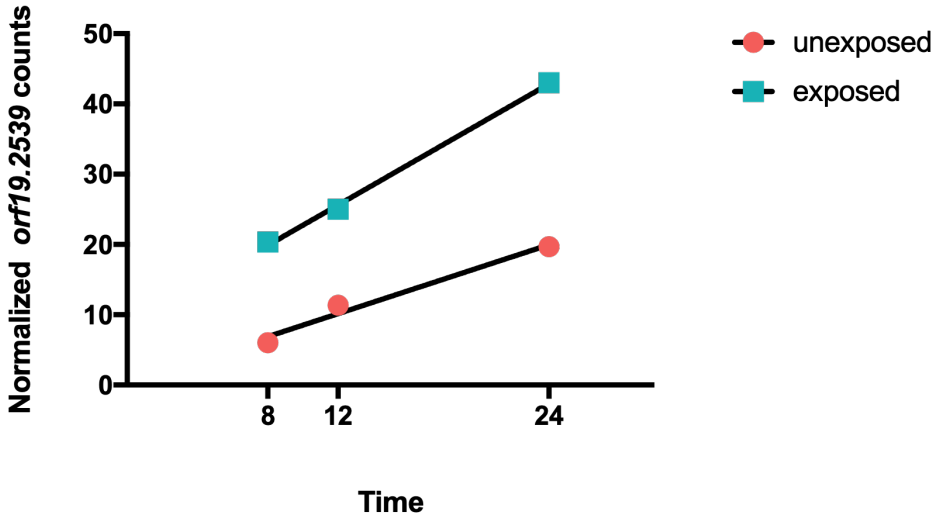

Figure 3.15: Normalized *orf19.2539* expression levels in *Candida albicans* sample type at an MOI of 1:1
C. albicans orf19.2539 expression increases in overtime and increases more in *C.* albicans exposed to epithelial cell at an MOI of 1:1. Sample means are plotted (n=2 for exposed samples at 24 hours; n=3 for all other time points). R square values for unexposed and exposed fitted lines are 0.9762 and 0.9979, respectively. The Y-intercept for unexposed and exposed trend lines are 0.3462 and 8.385, respectively.

We also saw expression variation among individual samples in this experiment (MOI 1:1), especially at 24 hours (**Figure 3.16**).

Figure 3.16: Normalized orf19.2539 expression levels in individual *Candida albicans* samples at MOI of 1:1
C. albicans orf19.2539 expression increases in overtime and increases more in *C.* albicans exposed to epithelial cell at an MOI of 1:1. Normalized counts for each sample are plotted (n=2 for exposed samples at 24 hours; n=3 for all other time points).

Despite this gene expression variation for *orf19.2539* among *C. albicans* samples exposed to epithelial cells for 24 hours, the sample Pearson correlation is high (0.9750648) between the 2 biological replicates (**Figure 3.17**)

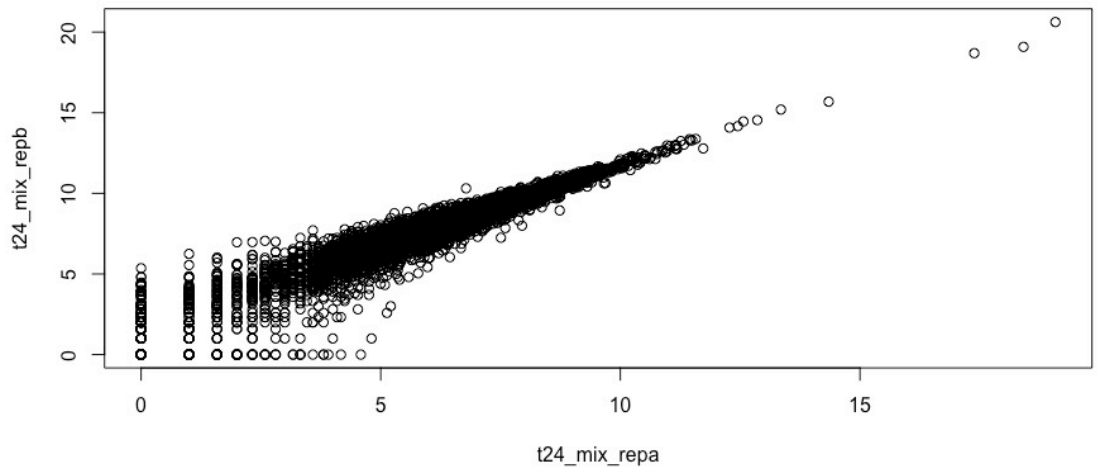

Figure 3.17: Correlation of transcript counts between exposed *Candida albicans* replicates at 24 hours
Log2 normalized counts for each gene detected in *C. albicans* samples exposed to epithelial cells at 24 hours. Transcript counts for biological replicate A are plotted on the x-axis. Transcript counts for biological replicate B counts are plotted on the Y axis.

Additionally, plotting the most variably expressed genes leads to sample time points clustering together, regardless of exposure status (**Figure 3.18**).

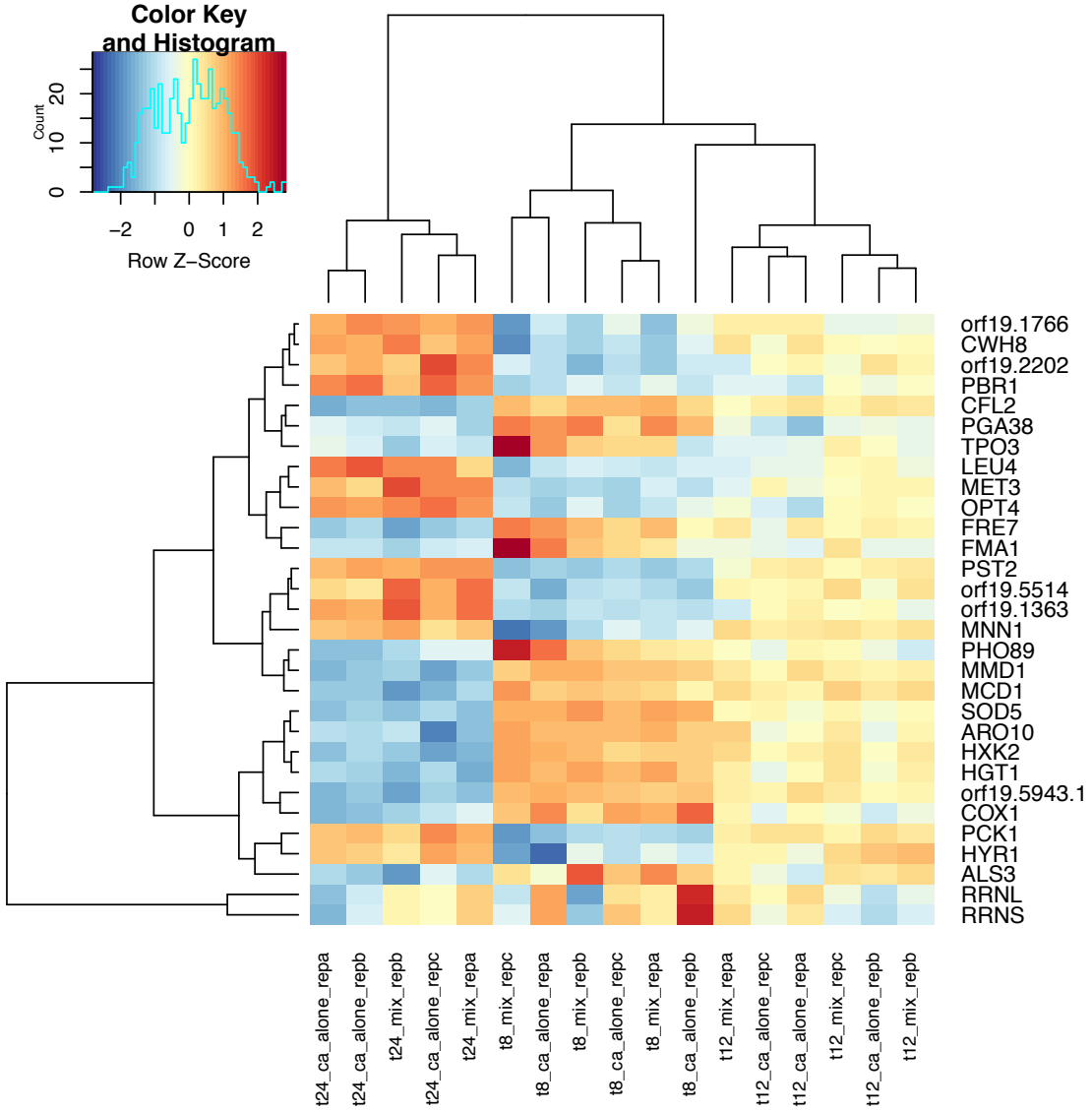

Figure 3.18: The most variable expressed genes across *Candida albicans* samples at an MOI of 1:1
Heatmap of sample row variance of the transformed expression values for the 20 most variable genes among samples. Each row is a gene and each column is a *C. albicans* sample. Sample time point and condition are labeled at the bottom of each column. The dendrogram on the left side of the heat map represents the hierarchical clustering of expression of the listed genes across samples. The dendrogram on the top of the heatmap represents the hierarchical clustering of the samples.

This indicates that at a 10x reduced MOI (relative to the pilot experiment), *C. albicans* may upregulate a similar set of genes when growing in organoid media alone or during organoid exposure (**Figure 3.18)**

To compare these data with the experiment that had an MOI of 10:1 in which biological replicates pooled prior to sequencing, we merged bam files for each replicate at each time point and re-ran the analysis in the same manner as was done for the pilot data (**Figure 3.19)**

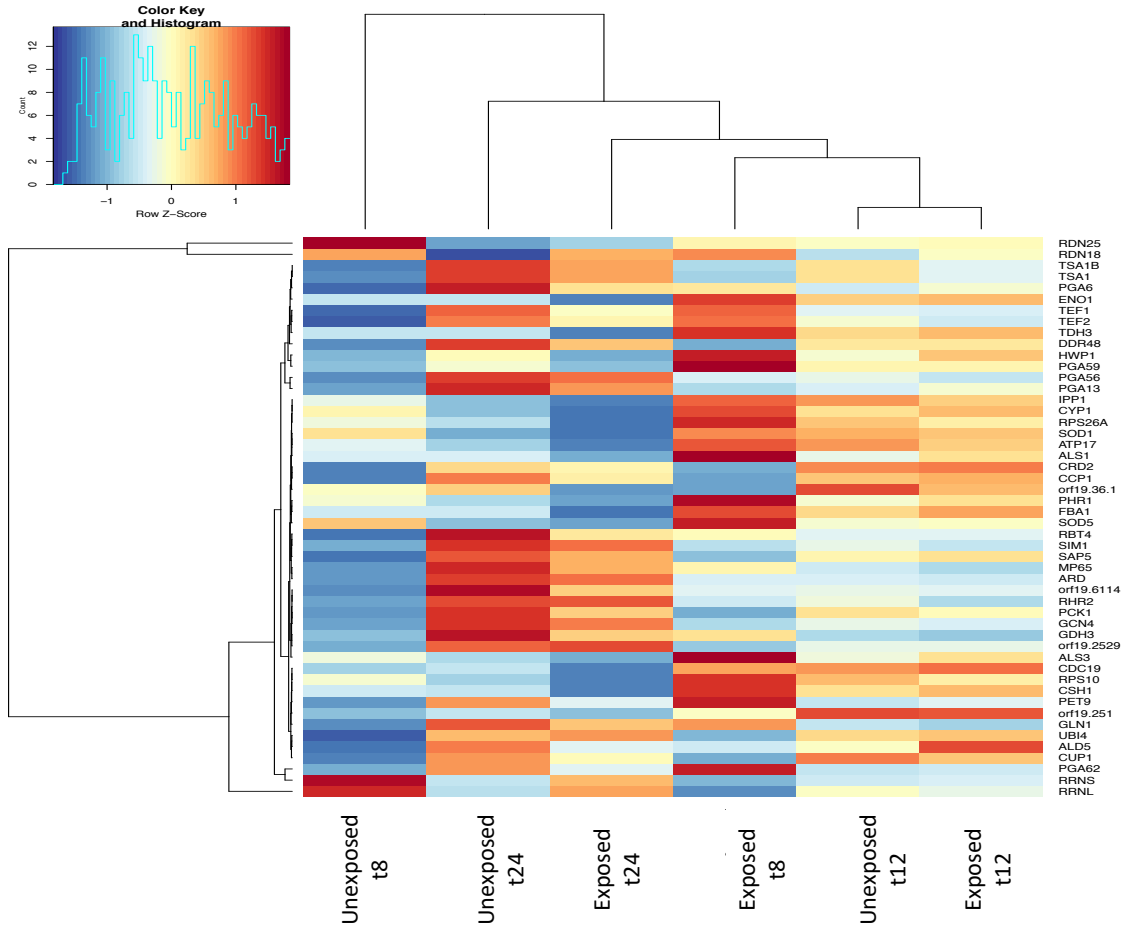

Figure 3.19: The top 50 most variable genes across *Candida albicans* samples when replicates are collapsed at an MOI of 1:1
Heatmap of sample row variance of the transformed expression values for the5 most variable genes among samples in which biological replicates are collapsed. Each row is a gene and each column is a *C. albicans* sample. Sample time point and condition are labeled at the bottom of each column. The dendrogram on the left side of the heat map represents the hierarchical clustering of expression of the listed genes across samples. The dendrogram on the top of the heatmap represents the hierarchical clustering of the samples.

Despite collapsing replicates, samples were still largely separated by time rather than exposure status (**Figure 3.19**)

Further supporting the idea that gene expression in samples collected at the same time points were more similar than that between samples in the same condition at a *C. albicans* to epithelial MOI

of 1:1, a volcano plot shows that there are very few statistically significantly differentially expressed genes between exposed and unexposed samples (**Figure 3.20**).

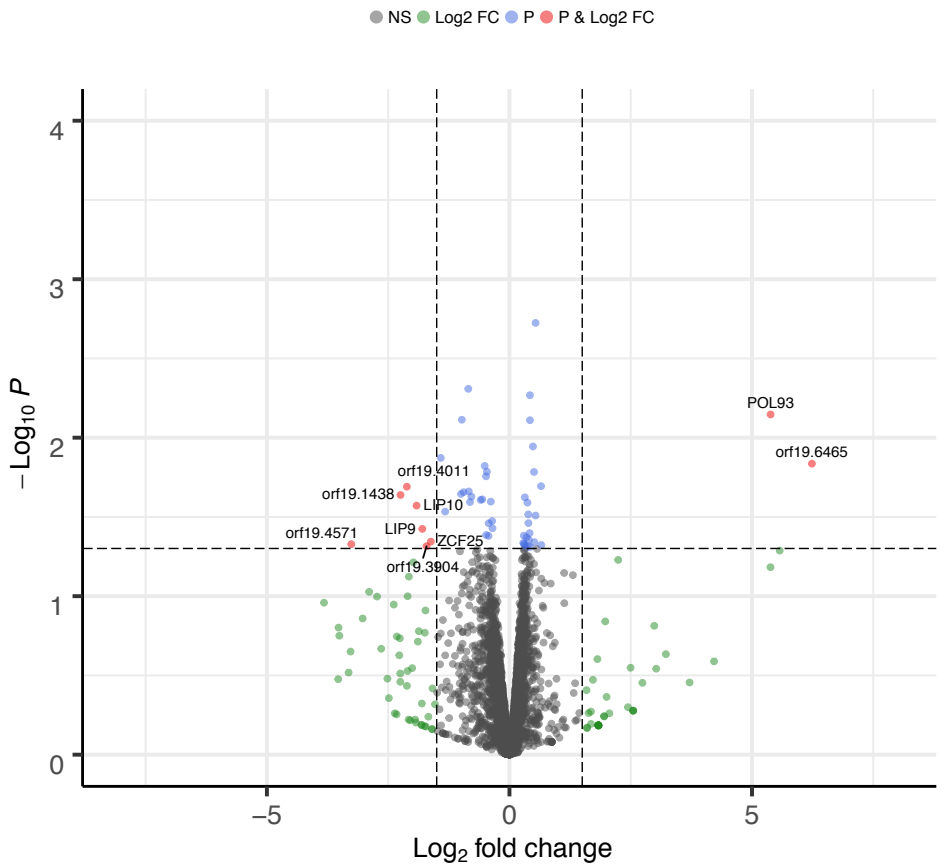

Figure 3.20: *Candida albicans* differential expression after exposure to murine small intestine epithelial cells at an MOI of 1:1
Biological replicates were collapsed and samples were grouped based only on condition. The x-axis is log2 fold change and the y-axis is -log10 p-values. Gray coloring indicates the genes was not statically different between exposed and unexposed samples. Green coloring indicates genes had a log2 fold change greater than 1. Blue coloring indicate a gene has a p-value < 0.05 and a pink dot means the gene has a log2FC greater than 1 and p-value < 0.05.

These results suggest that MOI has a significant effect on *C. albicans* expression.

3.3.6 Murine epithelial cell gene expression changes after exposure to *Candida albicans* at the bulk population level in a pilot experiment, at an estimated MOI 10:1

Gene expression difference between populations of epithelial cells exposed to *C. albicans* versus organoids growing alone were then assessed with the DESeq2 package[179] in R (**Methods section 3.5)**. Here, murine gene expression signatures were examined. The GO terms associated with the most 20 variable expressed genes across epithelial cell samples include macromolecule modification and ubiquitin dependent protein catabolic process; which are both metabolic processes that the organoids seem to me shutting down when exposed to *C. albicans*. We observed several alpha defenses (*Defa4,20,22,23,24*) to be among the top 20 most variably expressed genes in in the samples (**Figure 3.21**). These genes are generally more highly expressed in *C. albicans* exposed small intestine epithelial cells relative to unexposed controls (**Figure 3.22**).

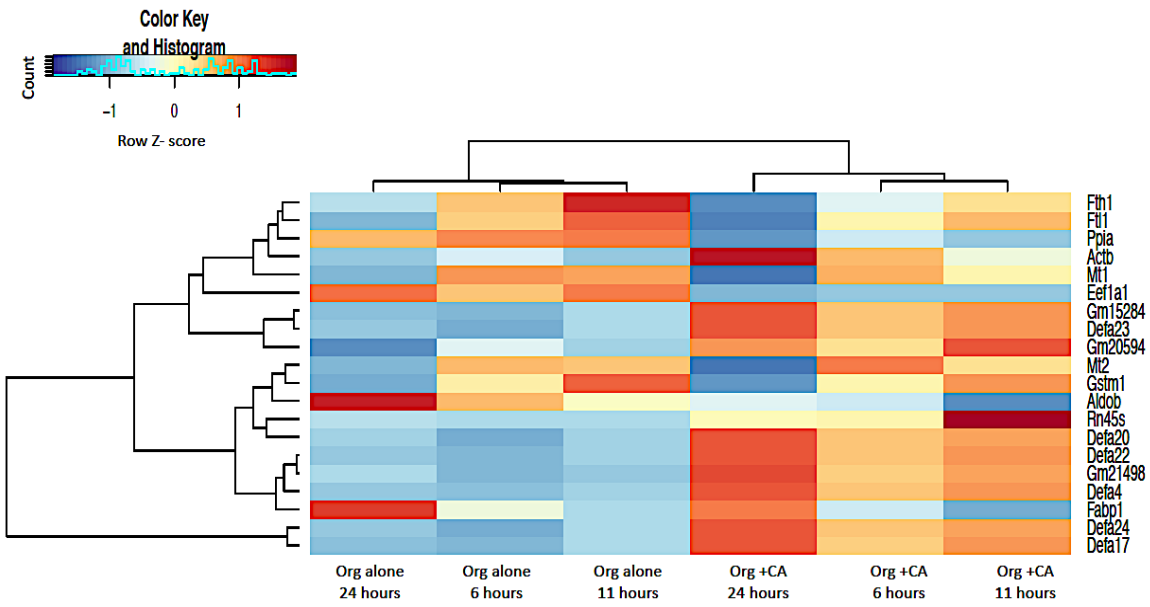

Figure 3.21: Top 20 variable genes across small-intestine epithelial cell organoid samples unexposed or exposed to *Candida albicans* at 6, 11 and 24 hours pilot, bulk RNA-sequencing experiment
Heatmap of sample row variance of the transformed expression values for the 20 most variable genes among samples. Each row is a gene and each column is an organoid derived small intestine epithelial cell sample. Sample time point and condition are labeled at the bottom of each column. "Org +CA" indicates sample consists of both *C. albicans* and organoid cells. The dendrogram on the left side of the heat map represents the hierarchical clustering of expression of the listed genes across samples. The dendrogram on the top of the heatmap represents the hierarchical clustering of the samples.

Several genes we found to be upregulated in murine small intestine epithelial cells exposed to *C. albicans*, including adrenomedullin (*Adm)* (**Figure 3.22**). The protein encoded by *Adm* is involved hormone secretion and antimicrobial activity, although it has previously been reported to only display antimicrobial effects on gram positive and negative bacteria and not *C. albicans*[184]. It was previously demonstrated that human oral epithelial cells do not regulate the adrenomedullin gene (*Am*) in response to *C. albicans* exposure[185]. The upregulation of the mouse ortholog (*Adm*) observed in this work may be specific to host niche, and in this case, intestinal epithelial cells.

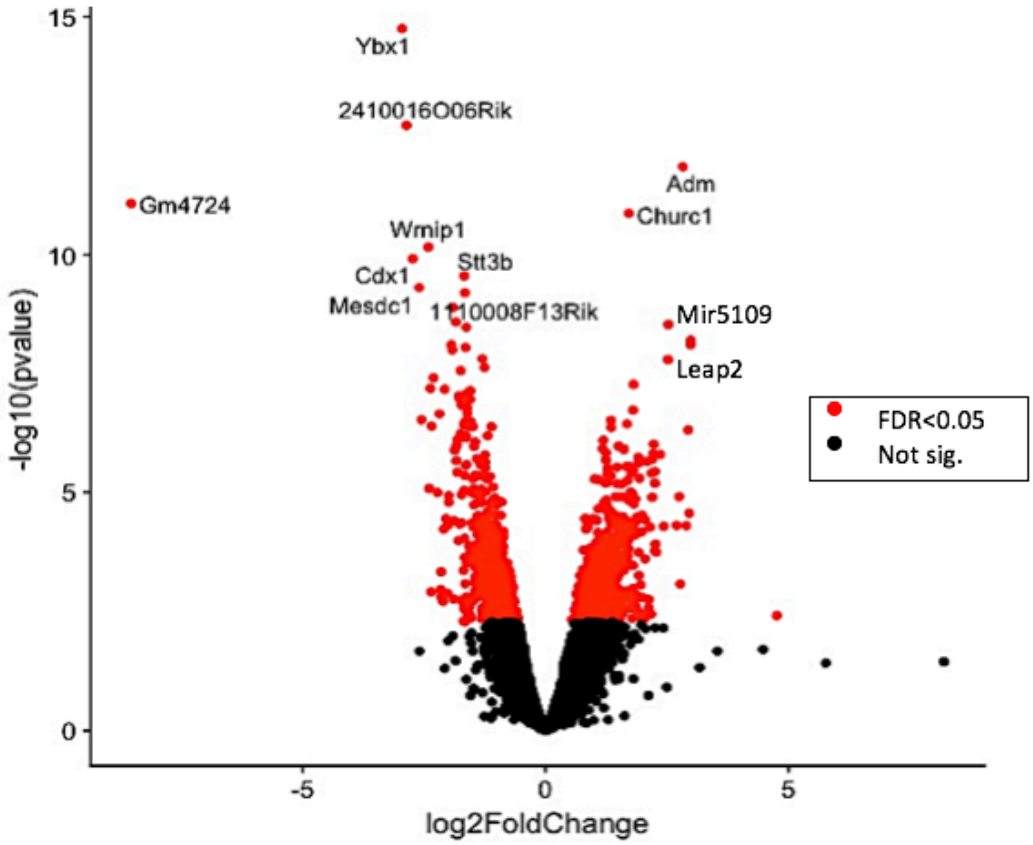

Figure 3.22: Organoid epithelial cell differential expression after exposure to *Candida albicans*
Small intestine epithelial cells samples exposed to *C. albicans* significantly upregulate and downregulate specific genes compared to unexposed controls. Genes are colored in red if the false discovery rate was less than 5%.

Similarly, *Leap2*, which encodes is an antimicrobial peptide that has activity against gram-positive bacteria and yeasts, was found to be upregulated in organoids exposed to *C. albicans*. *Leap2* has also shown to be significantly downregulated in IBD patient colitis tissue versus normal colons[186]. *Churc1* was also found to be upregulated in small intestine epithelial cells exposed to *C. albicans*, (log2 FC >1.7). *Churc1* is a target of the *WNT* and *VEGF* signaling pathways[187].

We hypothesized that Goblet, Paneth and Tuft cell gene expression would be higher in populations of organoids exposed to *C. albicans.* Alpha defensins (encoded by *Defa* genes) are class of

antimicrobial molecules secreted by Paneth cells. We did observe some canonical Goblet and Paneth cell markers were upregulated in murine organoids exposed to candida relative to those that were unexposed (i.e. *Tff3 and Defa21*, respectively), but these did not meet the significance criteria (log2 FC $\geq$ 2). There are no GO terms associated with downregulated genes (23 genes with log2 FC < -2, adjusted p-value $\leq$0.05 and FDR <0.05).

3.3.7 Gene expression difference between populations of epithelial cells versus populations of macrophages exposed to *Candida albicans*, at an estimated MOI 10:1

GO terms for genes found to be upregulated in small intestine epithelial cells overlap with many of those found to be upregulated in macrophages exposed to *Candida albicans* (**Chapter 2**), such as those related to the regulation of apoptosis and response to stress. However, we also saw unique GO terms, like those that relate to epithelial cell development and neurogenesis. Among GO terms associated with the most variable expressed genes among all epithelial cell samples that were more expressed in exposed samples versus controls, is the term "epidermal growth factor binding", previously shown to activate proliferation and inhibit apoptosis of intestinal stem/progenitor cells[188].

3.3.8 Gene expression in single, *Candida albicans* infected murine organoid epithelial cells at, an estimated MOI 10:1

We next generated pilot RNA-sequencing data with single, *C. albicans* infected epithelial cells using the high throughput Chromium Single Cell Gene Expression assay (10x Genomics). While we can currently process and multiplex up to 384 cells at a time using the Smart-Seq2 strategy, the high throughput droplet based strategy (Chromium Single Cell Gene Expression assay, 10x

Genomics) can provide data on up to ten thousand single cells from one sample. Based on the transcription bimodality we observed in single *C. albicans* infected macrophages (**Chapter 2, Figure 2.11**), we hypothesized that we would also observe transcriptional bimodality in infected small intestine organoid cells. Although these cells are not canonical epithelial cells, many small intestine epithelial cell types are involved in mucosal immunity. We also hypothesized that that Goblet, Paneth and Tuft cell gene expression would be higher in populations of organoids exposed to *C. albicans.* The sample collection strategy for the single cell pilot study is outlined below in **Figure 3.23**.

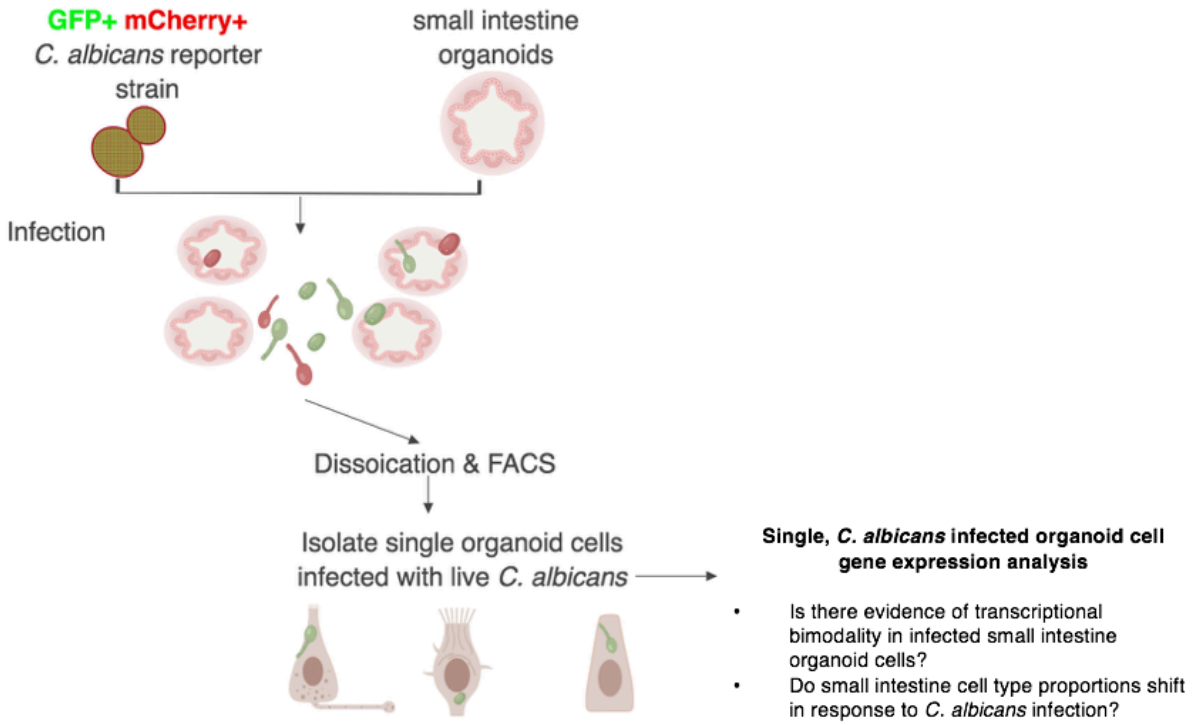

Figure 3.23: Single, *Candida albicans* infected epithelial cell RNA-seq pilot experiment design
A reporter strain of C. albicans constitutively expressing GFP and mCherry is exposed to small intestine organoid. For population level studies, samples are collected at specific time points (16 hours for the pilot study), dissociated and sorted to obtain single small intestine epithelial cells infected with *C. albicans.*

16 hours post infection, a total of 1,258 single epithelial cells were recovered However, no *Candida albicans* gene expression was recovered. 998 uninfected epithelial cells were recovered with a 3,127 median genes detected per cell. In comparison, 288 *C. albicans* infected epithelial cells were recovered with a median of 288 genes detected per cell. With these data, we produced t-distributed stochastic neighbor embedding (t-SNE) plot using the Seurat package in R (**Methods section 3.5.8**). A t-SNE plot is commonly used to visual high-dimensional data, like gene expression from thousands of single cells, in a 2-dimensional space. Just like principal component analysis plots, t-SNE plots do not quantitatively measure the discrete distance between two points and t-SNE axes

units are not meant to be quantitative; the axes units define the 2-D space into which the higher dimensional data is projected, while preserving relative proportional distances between points as much as possible. As such, t-SNE plots are a visualization technique which will group similar points together. Using our data, each point in the t-SNE plot is a cell (**Figure 3.24)**, and cells with similar gene expression patterns that cluster together. After cells are assigned to discrete clusters based on gene expression similarity, we identified the genes that were differentially expressed in each cluster relative to all others cells, to identify the small intestine epithelial cells types in each cluster. Using markers recently identified for small intestine specific epithelial cell types[117] in survey of 50,000 single cells, we identified canonical markers for 2 types of enterocytes, M cells, paneth cells, Tuft cells and one cluster that expressed markers for both Goblet/Paneth cells. Clusters 0-2 did not predominately express any canonical small intestine epithelial cell markers.

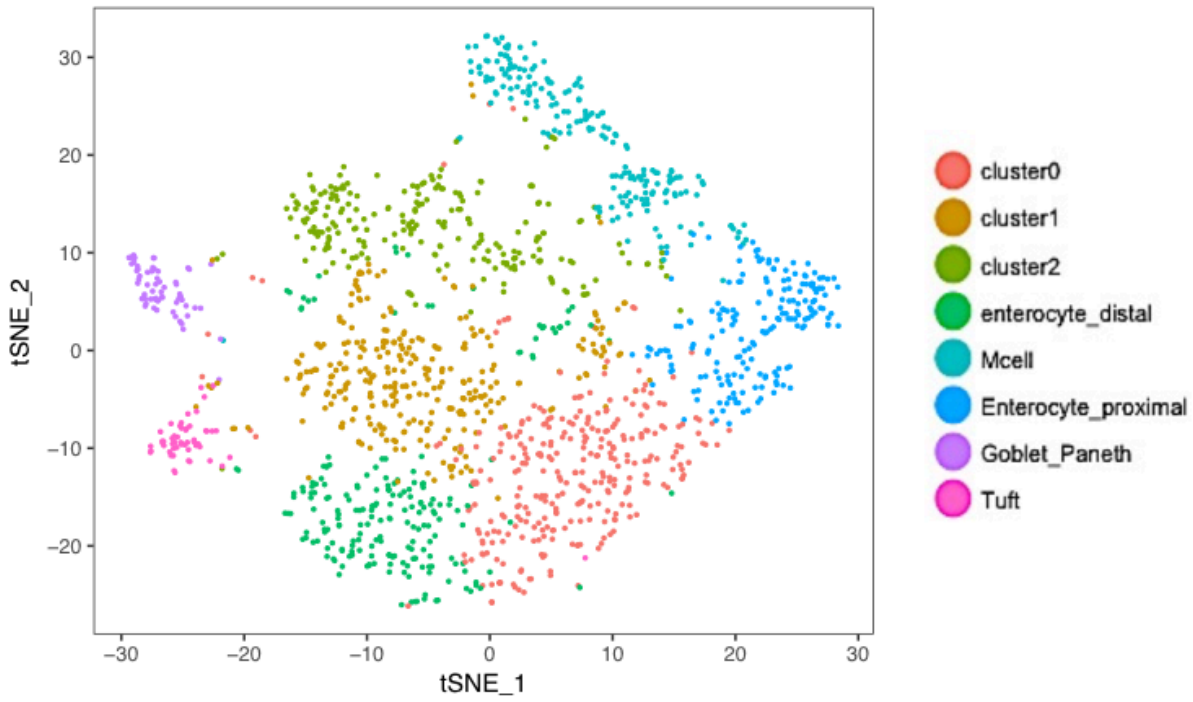

Figure 3.24: Murine small intestine organoids recapitulate many of the epithelial cell subtypes found *in vivo*
A t-SNE algorithm clusters cells together based on gene expression. Genes that were differentially expressed between clusters were used to determine small intestine cell type and rename cell clusters based on these expression markers. Clusters which did not strongly express canonical markers are numbered (clusters 0-2).

Next, we observed that infected cells appeared most frequently together in cluster 2, a cluster which did not specifically upregulate canonical small intestine cell type expression markers (**Figure 3.25**). The GO terms associated with genes that are upregulated in the infected cells in cluster 2 (relative to all other clusters) include fibroblast growth factor binding, chemotaxis and *FGF* signaling.

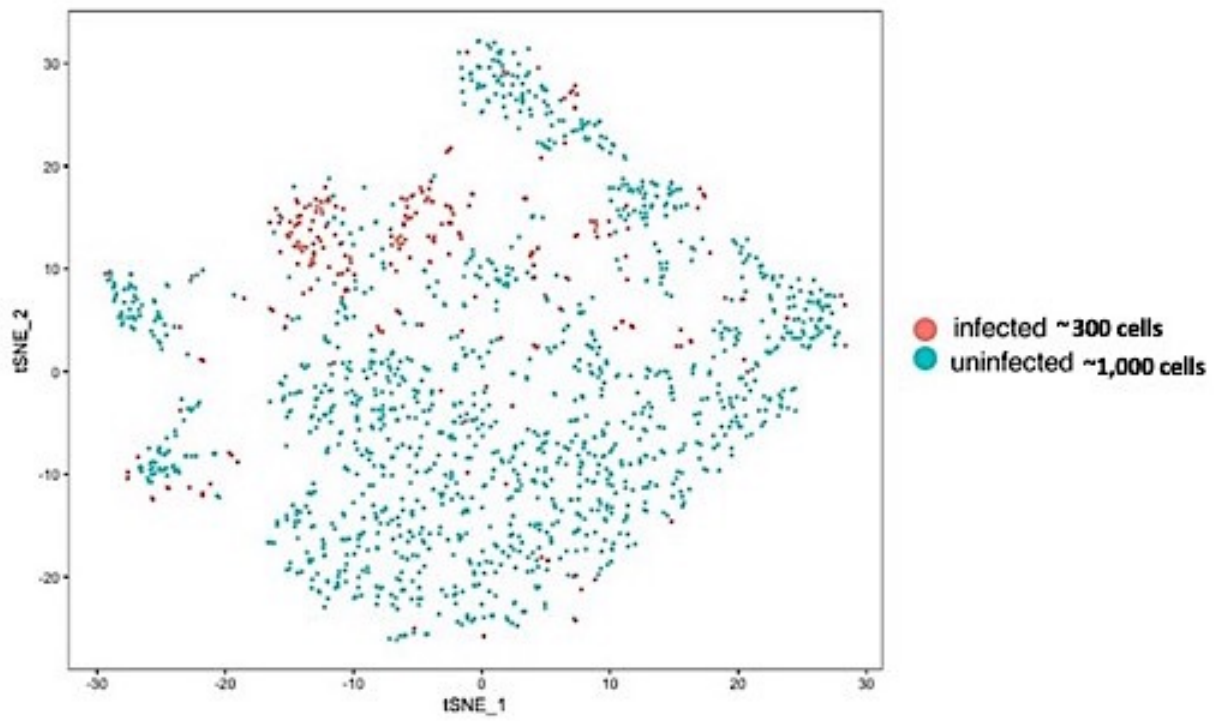

Figure 3.25: *Candida albicans* infected murine organoid epithelial cells cluster together
A t-SNE algorithm clusters cells together based on gene expression. Each cell is colored in accordance with sample classifier, either infected (pink) or uninfected (blue).

We next examined differentially expressed genes between *C. albicans* infected and uninfected epithelial cells (**Figure 3.26**). The most upregulated gene among *C. albicans* infected epithelial cells relative to uninfected epithelial cells was *AY036118*. Interestingly, *AY036118* a non-coding RNA known to be most highly expressed in the adult small intestine and has also been associated with telomere binding

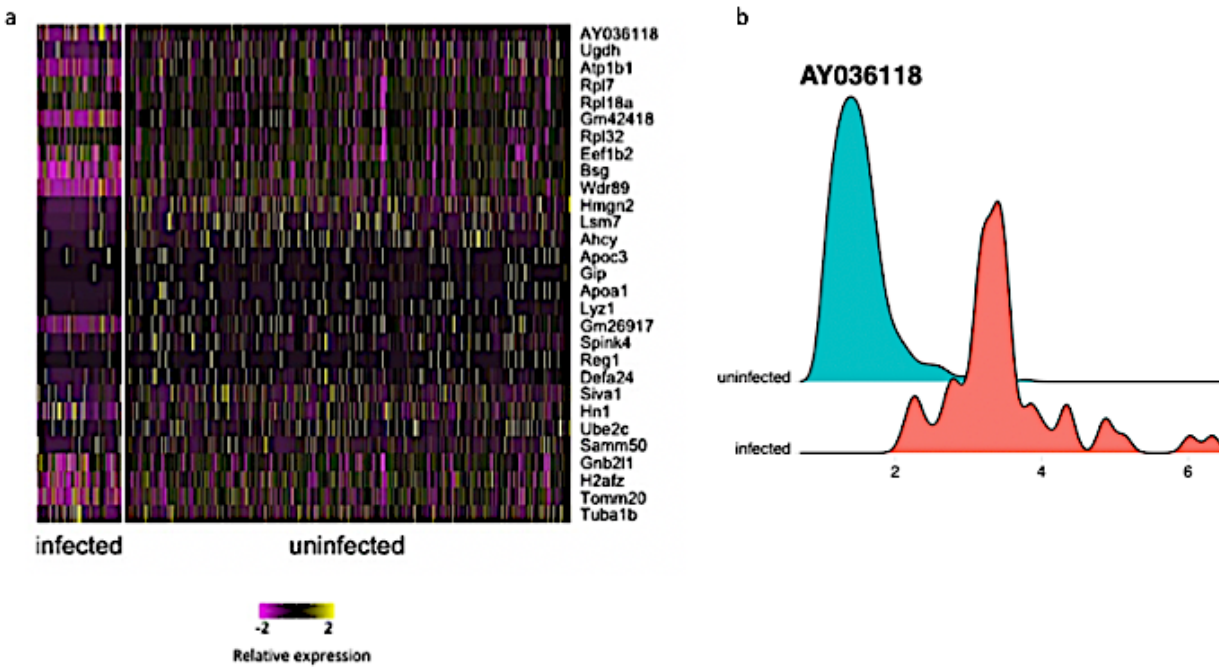

Figure 3.26: Gene expression profiles of single *Candida albicans* infected organoid epithelial cells
(a) A heatmap shows the relative expression of approximately 300 infected single epithelial versus approximately 1,000 single uninfected epithelial cells; each column is a cell and each row is relative normalized gene expression. **(b)** Density plots show probability density distribution of cells expressing AY0368118 (with a TPMI $\geq$ 1) for uninfected cells (blue distribution; n= 892) and infected cells (pink distribution; n=53 cells); the y-axis is the probability density and the x-axis is log normalized expression (TPM+1).

Single, *C. albicans* infected epithelial cells exhibit bi-modal or multi-modal expression patters for subsets of genes (**Figure 3.27**). Interesting examples include including hematological and neurological expressed 1 protein (*Hn1*) and *Churc1*, a transcriptional activator that mediates *FGF* signaling during neural development. *FGF* signaling is involved in small intestine development and in intestinal injury repair.

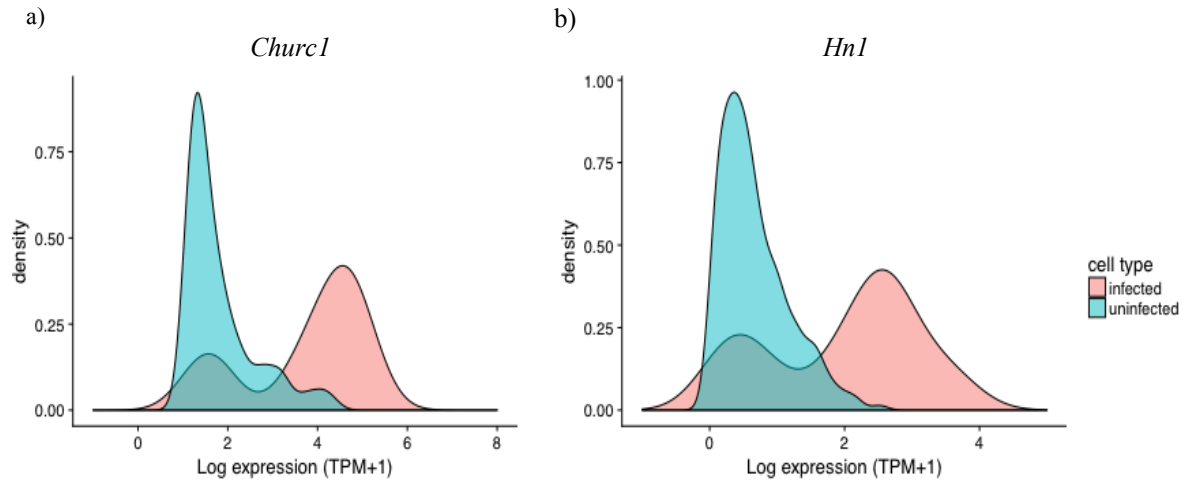

Figure 3.27: Gene expression bimodality in single *Candida albicans* infected organoid cells
Density plots on the show probability density distribution of cells expressing **(a)** *Churc1* and **(b)** *Hn1* (with a TPMI ≥ 1) for uninfected cells (blue distribution) and infected cells (pink distribution); the y-axis is density and the x-axis is log normalized expression (TPM+1). *Churc1* infected n= 23 cells, uninfected n = 192 cells; Hn1 infected n=69 cells, uninfected n = 717 cells.

Despite the fact that approximately 3 times more uninfected cells were recovered than *C. albicans* infected epithelial cells, we believe we have captured potentially biological relevant expression information. Similar to **Figure 3.27**, **Figure 3.28** is a violin plot, showing the probability density of expression of *Hn1* among single, *C. albicans* infected epithelial cells (pink distribution) and single, uninfected cells (blue distribution). However, this plot also illustrates each cell location in the distribution. Despite having fewer infected epithelial cells overall, *Hn1* appears to be significantly differentially expressed in a subset of infected cells, relative to uninfected cells (**Figure 3.28**).

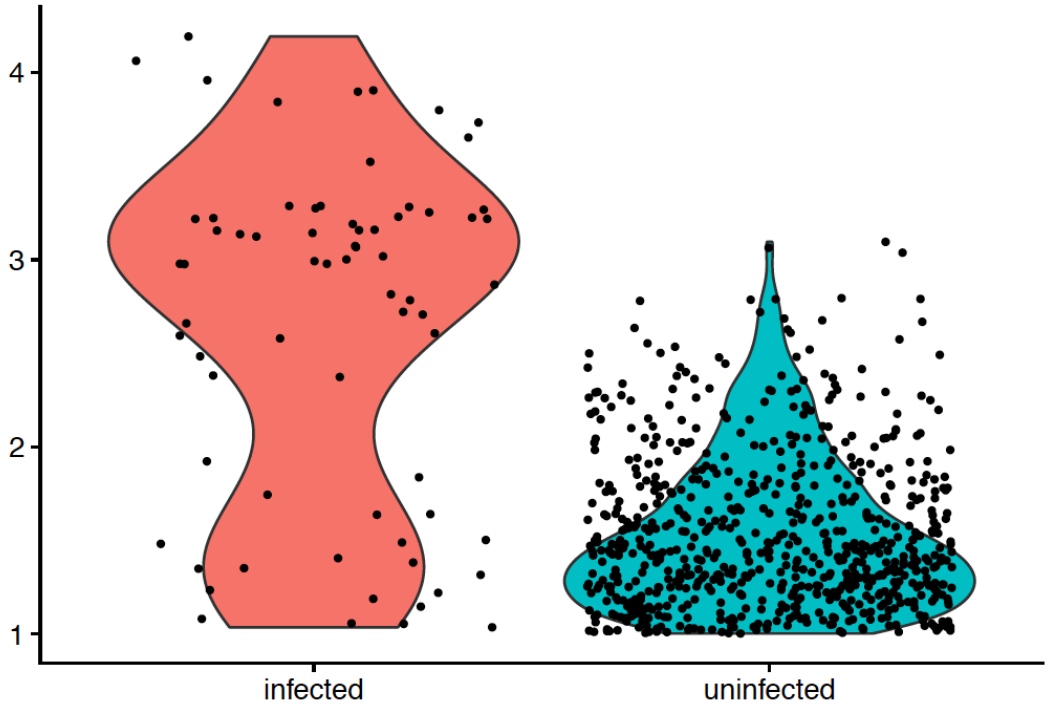

Figure 3.28: Evidence of *Hn1* expression bimodality among single, *Candida albicans* infected epithelial cells
Violin plot showing the probability distribution of cell expression (log TPM+1) *of Hn1* is plotted for single infected epithelial cells (dots in pink distribution) and uninfected epithelial cells (dots in blue distribution) .

These results suggest that there is expression bimodality or multi-modality among single epithelial cells infected with *C. albicans*.

3.4 Discussion

Although we saw rapid transcriptional changes in *C. albicans* exposed to macrophages over 4 hours (**Chapter 2**), we first chose to profile *C. albicans* exposed to organoids at time points greater than 4 hours because it took at least 6 hours for *C. albicans* to begin penetrating the organoid matrigel. When pooling time points and analyzing *Candida albicans* exposed to organoids an MOI 10 *C. albicans* cells to 1 epithelial cell (MOI 10:1), *Candida albicans* appears to upregulated many genes in the same pathways critical to *C. albicans* after phagocytosis at 4 hours, including hyphal cell wall formation, positive regulation by symbiont of host defense response and glycolytic processes. At an MOI of 10:1, 9 unique GO term IDs were associated with *C. albicans* exposed to organoids. Although the IDs were unique, 6 of these terms were very similar to GO terms found upregulated in phagocytosed *C. albicans*, the majority of which included carbohydrate purine nucleoside and ribonucleoside biosynthetic processes. The 3 most unique GO terms associated with genes upregulated in *C. albicans* exposed to organoids include "carbohydrate catabolic process", "oxidoreduction coenzyme metabolic process" and "cofactor metabolic process". Interestingly, carbohydrate metabolism was upregulated in phagocytosed *C. albicans* while carbohydrate catabolism was upregulated among *C. albicans* exposed to organoids. Because the organoid matrigel is rich with extracellular matrix proteins and epithelial cells have been shown to express carbohydrates on their surface[189], these changes in metabolism likely reflect the unique host cell environment that *C. albicans* are exposed to when penetrating the organoid matrigel and organoid epithelial cells.

At an estimated MOI of 10:1, upregulated genes in *C. albicans* include 6 uncharacterized genes (*orf19.2539, orf19.6807, orf19.3354, orf19.7077, orf19.4406, orf19.2662*). These results are,

intriguing, as *C. albicans* is less well studied in the context of the gut than in other host niches. As this work has identified several uncharacterized genes in *C. albicans* that may be important for host epithelial cell interaction, further development of the CRISPR loss of function mutation would allow for rapid screening and confirmation of LOF. Because the *C. albicans* genome is AT rich and some genomic regions are not amenable to Sanger sequencing, an approach like PCR with Derived Cleaved Amplified Polymorphic Sequences (dCAPS)[190] could be a more cost effective solution. With this approach, one or two base pairs are inserted into the genomic region of interest after PCR with an error prone polymerase (like Taq) and a restriction enzyme is created in the mutant genotype but not in wild type. Presence or absence of digestion can then be analyzed with an agarose gel. Software is available from Neff et al. to guide the selection of primers (http://helix.wustl.edu/dcaps/). We attempted to verify *orf19.2539* LOF mutants with this method, but results were inconclusive (**Methods 3.5.10**).; more development is needed to make this technique reliable.

The GO terms associated with genes downregulated in *C. albicans* exposed small intestine epithelial cells include "neurotransmitter metabolic processes", "regulation of neurotransmitter levels" and "alpha-amino acid catabolic process". This was unique to organoid exposure and did not occur in the same *C. albicans* reporter strain when exposed to macrophages. These gene were also not found to be downregulated in *C. albicans* exposed to macrophages[16], or GUT cells growing outside of the host in rich liquid media[17]. We hypothesize that reducing the amount of neurotransmitter break down could help maintain the niche homeostasis, *C. albicans* may better be able to avoid immune detection by the host. Gut neurotransmitters have been linked to host intestinal inflammation and immune system function[191,192]. To date, there are few studies

investigating *C. albicans* and host neurotransmitters interplay, but Gamma-aminobutyric acid (GABA), a neurotransmitter found in the GI tract and throughout the body[193], has been shown to increase hyphae in *C. albicans*[194]

Because the time points examined via RNA-sequencing in this Chapter (6, 8,12,24 hours) are later than those collected during the macrophage experiment (0,1,2,4 hours), it is hard to make a direct comparison between these two datasets. Additional time points were collected during small intestine epithelial cell and *C. albicans* exposure at an MOI of 1:1 (0 and 4 hours). However, RNA yields were considerably lower relative to recovered in the pilot experiment (MOI 10:1). For this reason, we chose not to test the 0 and 4 hour time points in this initial sequencing run but will be processed in the future.

The genes that are important for *C. albicans* survival may depend on infection MOI. We tested an estimated MOI of 1:1 and 10:1. In each case, different sets of genes were most highly upregulated at specific time points, although many of the pathways upregulated in both infections were similar to each other, and also to those pathways upregulated in *C. albicans* inside of macrophages. These results highlight that infection MOI is important to consider in the context of experimental results and interpretation, and should be considered when interpreting the importance of specific genes or pathways in the context of *C. albicans* survival in the host.

On the host side, we found Alpha defensins (*Defa4,20,22,23,24*), to be among the top 20 most variably expressed genes across organoid samples and upregulated in epithelial cells exposed to *C. albicans*. Defensins have previously been found to be expressed in mice and expression of these

defensins can be modulated by bacteria[195]. Interestingly, *Defa5* and *Defa6* have been shown to be upregulated in patient ulcerative colitis colon tissue and it's hypothesized that this is due to an increase of number of paneth cells in this region[196]. However, these alpha defensin genes have not previously reported to be upregulated in mice and we did not observe *Defa5* and *Defa6* upregulation in this work.

The *WNT* and *VEGF* signaling pathways have previously been implicated in stem cell and vascular development in the small intestine[197]. *Churc1* is a transcriptional activator that also plays a role in *FGF* signaling during neural development[198]. The intestine is often referred to as the second brain because it's home to an estimated 100-600 million neurons in the human intestine, and allows our stomach to communicate with our brain[199,200]; This nervous system is known as the enteric nervous system.

Intriguingly, we found evidence of expression bimodality among single, *C. albicans* infected epithelial cells, despite recovering 10 fold fewer infected cells than non-infected controls. This leads us to believe expression bimodality is a phenomenon that occurs after exposure to fungal pathogens in both immune cells, like macrophages (Chapter 2), as well as epithelial cells. Analysis of a larger data set of single, *C. albicans* epithelial cells would allow us to determine if specific epithelial cell subsets exhibit bimodal expression patterns or if this expression mode is ubiquitous among all epithelial cell types.

Using a dual-RNA-sequencing approach to study host and fungal gene expression would be beneficial when studying the role of *C. albicans* in the context of ulcerative colitis and Crohn's

disease, although fungal load could change fungal gene expression and should be considered in study design and analysis.

In the near the future, applying single cell RNA-sequencing approaches could also benefit IBD research. While the high throughput, droplet based Chromium technology (10x Genomics) currently does not allow for the detection of fungal gene expression, host single cell expression data would still allow us to understand if fungal exposure changes the relative abundance of specific cell types. Using this approach to study gene expression in IBD afflicted tissue (compared to healthy tissue) could help determine if host cells are overproducing certain molecules, like inflammatory cytokines. In order to achieve this goal, an estimated 20 thousand cells would need to analyzed[178]. Data on this order will allow for more well-defined cell clusters that can be more easily identified by canonical markers. In the pilot experiment presented here, we generated data on10 fold less *C. albicans* infected epithelial cells relative to uninfected cells. As we sorted epithelial cells based on size, we believe that we unwittingly sorted large filamentous *C. albicans* that were clumped together. Additional analysis of "single *C. albicans* infected epithelial cells" from this experiment which were processed via Smart-seq2, revealed that 3 out of 5 wells assessed had transcripts mapping back only to *C. albicans*. In future work, using a fluorescent antibody (i.e. Cy7 Fluor) to bind mammalian host cell surface markers (i.e. Apc) and/or epithelial cells specific markers (i.e. Epcam), prior to sorting could increase the number of single infected epithelial cells recovered. However, this approach will require development, as organoid derived epithelial cells are sensitive to physical manipulation after being released from the matrigel and have a higher cell death rate relative to cell line derived cells, in our hands.

The benefit of studying host and fungal gene expression simultaneously is that we can see what gene expression changes are important during cross-species interaction. In the future, we may be able to use this information to determine if an antifungal treatment, immunotherapy or a combination of both would most effectively help the patient. For understanding how fungus relates to bowel disease, organoids developed from IBD patients (and healthy tissue controls) would be the optimal host cell type to study these interactions. Fungi, including *C. albicans* and other *Candida* species, have been linked to Alzheimer's disease[201–204]; the approach developed here could be used to study host-fungal interactions in other patient samples and/ or other types of tissue derived organoids as well.

3.5 Methods

3.5.1 *Candida albicans* reporter strain construction

The reporter construct used in this study was prepared by integrating the GFP and mCherry fluorescent tags driven by the bi-directional *ADH1* promoter and a nourseothricin resistance (NAT^R) cassette at the Neut5L locus of *Candida albicans* strain CAI4-F2[149] (as described in section **2.5.1**)

3.5.2 *Candida albicans* and organoid infection assay

Murine organoids were derived from the crypts of Lgr5-EGFP-IRES-CreERT2 (Lgr5–GFP) mice (Jackson Laboratory), as previously described, in 24 well culture dishes[117]. All mouse work was performed in accordance with the Institutional Animal Care and Use Committees (IACUC) and

with relevant guidelines at the Broad Institute and Massachusetts Institute of Technology, with protocol IACUC 0055-05-15 and 0612-058-18 . Organoids were allowed to differentiate for 12 days. One day prior to the infection experiment, yeast were grown in 3 ml overnight cultures in rich media at 30°C. 2 hours prior to infection, yeast cells were acclimated to organoid media (without cytokines; DMEM/F-12, no phenol red, ThermoFisher Scientific). 2.5 million or 250,000 *Candida* cells were added to each organoid well, depending on the MOI of the experiment . Yeast and organoids were then co-incubated at 37°C (5% CO_2). Prior to sample collection, media was removed, 1 ml of 1X TrypLE (no phenol red; ThermoFisher Scientific) was added to each well and the plate was incubated at 37°C (5% CO_2) for 10 minutes. Samples were collected from each well with vigorous manual pipetting and scraping. 3 wells for each time point were combined into one tube for the bulk-RNA-seq pilot experiment that was carried out an MOI 10:1. For the bulk RNA-seq experiment carried out an MOI of 1:1, each biological replicate was collected and processed separately. For population samples that were not sorted, samples were spun down at 1 minute at 300g, media was aspirated, 600 ul of RLT + 1% BME was added and tubes flash frozen before storage at -80°C. For samples that were sorted, after media was aspirated, samples were resuspended in 1 ml PBS + 2% FCS and placed on ice until FACS. Samples were sorted into a new tube for Chromium Single Cell Gene Expression assay processing or into each well of a 96 well plate for Smart-Seq2 processing.

3.5.3 Fluorescence-activated cell sorting (FACS)

For the Chromium Single Cell Gene Expression assay (10x Genomics), samples were sorted on the Astrios flow cytometer (Beckman Coulter). 4% BSA–PBS. Cells were sorted into the following sub populations: (i) live epithelial cells infected with live *C. albicans* (GFP+, mCherry+, DAPI-,

cell size > 20 μm), (ii) uninfected epithelial cells (GFP-, mCherry-, DAPI-,cell size > 20 μm). For Smart-Seq2 single cell processing, singe epithelial cells infected with live *Candida albicans* and single uninfected epithelial cells were sorted into 5 ul of RLT 1% β-Mercaptoethanol in a 96 well plate (Eppendorf) and frozen at -80°C . One well of each 96 well plate contained a negative control (no cell, 5 ul of RLT 1% β-Mercaptoethanol only) and one positive control, containing 50 cells (infected or uninfected epithelial cells).

3.5.4 RNA extraction

RNA was extracted from population samples using the Qiagen RNeasy mini kit. All samples were resuspended in RLT (Qiagen) + 1% β-Mercaptoethanol (Sigma) and subjected to 3 minutes of bead beating with .5mm zirconia glass beads (BioSpec Products) in a bead mill. Single macrophages infected with *C. albicans* were directly with dead, phagocytosed by sorting cells into a 96 well plate containing 5 ul of RLT (Qiagen) + 1% β-Mercaptoethanol (Sigma).

3.5.5 cDNA synthesis and library generation with Smart-Seq2

For population samples, the RT reaction was carried out with the following program, as described[156], with the addition of RNAse inhibitor (ThermoFisher) at 40U/ul. cDNA was generated from single cells based on the Smart-seq2 method as described previously[157], with the addition of RNase inhibitor was used at 40 U/ul (ThermoFisher) and 3.4 ul of 1 M trehalose was added to the RT reaction. All libraries were constructed using the Nextera XT DNA Sample Kit (Illumina) with custom indexed primers as described[157]. For the bulk RNA-sequencing experiment carried out at an MOI of 10:1, sample populations, single infected cells and uninfected single cell controls processed via Smart-seq2 were sequenced on Illumina's NextSeq (75x75 cycles). For the bulk

RNA-sequencing experiment carried out at an MOI of 1:1, sample populations were sequenced on Illumina's NextSeq (38 x38 cycles).

3.5.6 cDNA synthesis and library generation with Chromium Single Cell Gene Expression assay

Single cells were processed with the Chromium Single Cell Gene Expression assay using the GemCode Gel Bead, Chip and Library Kits (10X Genomics). In brief, ~6,000 single *C. albicans* infected cells or single uninfected epithelial cells were added to each channel with an average recovery rate of 1,500 cells. The cells were then partitioned into Gel Beads in Emulsion in the GemCode instrument, where cell lysis and barcoded reverse transcription of RNA occurred, followed by amplification, shearing and 5′ adaptor and sample index attachment. Single infected cells and uninfected single cell controls were sequenced on Illumina's NextSeq (75x75 cycles).

3.5.7 Population level RNA-sequencing analysis

Population level RNA-sequencing samples were aligned to reference genomes *Candida albicans* reference genome SC5314 version A21-s02-m09-r10 and to *Mus musculus* GRCm38/mm10. Samples were aligned using the Bowtie2[205]/Tophat2[206] and transcript counts were obtained with RSEM in the Trinity suite[163] v2.20. Differential gene expression analysis was performed with the DESeq2 package in R[179]. GO terms for *Candida albicans* were determined using the GO Gene Ontology Tool Finder and the GO Term Slim Mapper on the *Candida* genome database (http://www.candidagenome.org/GOContents.shtml). GO terms for small intestine epithelial cells derived from *Mus musculus* were found using the MGI Gene Ontology Tool (http://www.informatics.jax.org/gotools/).

3.5.8 Single-cell RNA-sequencing analysis for pilot study

Chromium Single Cell Gene Expression assay data clustering was performed with tSNE as implemented in the R Seurat package version 2.2.1[156]. Briefly, the t-SNE algorithm models the probability distribution of neighboring points around each point to find a mapping that minimizes the differences between the original normal distribution the high dimensional data is and the t-distribution that is modeled in the 2-D space[207].

Expression bimodality was assessed as described previously (**Chapter 2**,**Methods 2.5.9**) Briefly, we determined the distribution of the average expression (μ), and the dispersion of expression (σ^2; normalized coefficient of variation), placing each gene into bins, and then calculating a z-score for dispersion within each bin to control for the relationship between variability and average expression as implemented R package Seurat[134].

3.5.9 CRISPR in *Candida albicans*

We cloned a double stranded guide RNA (IDT) to *orf19.2539* that contained SapI cloning sites (lower case bases) :

cgaTGATTAAATGTGAAAAAGAT
aacATCTTTTTCACATTTAATCA

We cloned the guides into bacterial plasmid pRS118 (unpublished, courtesy of the Shapiro lab), downstream of *SNR52*, a Pol III promoter. The plasmid also contains a *Candida* optimized Cas9 and a nourseothricin resistance gene[180]. Upstream of these components lies homology to the Neut5

locus in *C. albicans*, a "neutral", intergenic region that can be targeted for genomic insertions without effects on growth rate, cell size or other phenotypes[150]. Because we previously targeted this region for insertion of GFP and mCherry in our reporter strain and incorporated a nourseothricin resistance gene (**Chapter 2; Methods 2.5.1**), we couldn't delete this region in the same strain used for RNA-sequencing. As such, we attempted to mutate and fully delete *orf19.2539* in SC5314 (the traditional wild type *C. albicans* strain) and CAI4 (the parent of the report strain; **Chapter 2**; **Methods 2.5.1**). After cloning in the *orf19.2539* guide, the plasmid was linearized with PacI (NEB). The resulting fragment and a 90 base pair, single stranded repair template, containing homology to the guide RNA cut site and the desired 2 base pair deletion was transformed into *C. albicans* using a lithium acetate protocol[208]. The repair template sequences is as follows (IDT):

ATGATGTGCACTACTTTTTTCTTTATTTTTTTTTTTCTGCCTATC**TTT**CACATT**TAA**TC

AAAAAGTCGCGTGTGTGTGTGTGTGTGTGTG

The gene block synthesized (IDT) for a deletion of orf19.2539 in *C. albicans* SC5314 is as follows:

GAGATCCAGAAAACTGAATTGTGCTTGAATACCACTTGTTTAGCCAGTATTATTTTTT
ATCATACTTTTCAAAGAATAATCATTGGTGGGCTTGGCTCGAACGGTTCTTTTATCCA
AATCTTGCGGCACTATATTGGAATAAATATGGTTATTGGTAATTTTTAATTGATTTAG
TAGGTGCAGAACCATGATTGATTCAAATGTGTCTTGGATATTGTTAATGAATCAAAC
ACAACAACAACAATAAACAAATTCAACTGATATGTAAAAGTCATTGGTCAATAACT
ATACTCGAGTATTGCCTCATCAAAGAAACAATCAAATATTATAGATACTCACTCCAT
CACGTGATAATTTCACTGGTATGGAAAAGTGGAAAATTTTATAAAAAAAAATTTGAT
GCCTTTGGCATAGCTGAAACTTCGGCCCAATAGGATTGGAGAATATGTTTTCGCAGC
GTTCTTACAATTAAATTGTGGTGGAAGTTCGAGACTTGCGTAAACTATTTTTAATTTG
TGTGTGTGTGTGTGTGTTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTA
GTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGGTGCGCTGATGTTAGAGG
AAGGAGAGTTTTAGAGCTATGCTGAAAAGCATAGCAAGTTAAAATAAGGCAGTGAT
TTTTAATCCAGTCCGTACACAACTTGAAAAAGTGCGCACCGATTCGGTGCTTTTTTTT
TAATAGTCAAGCGTCTTTTCTCACATTTCAATACTATTTTCCAAAATATGGCAAAATT
AAACCGGGGGAAGTACAAAAAAAAAGGCTGCAATTGAGATATCTGTAATCATCTTA
TTGATCTGAATGATATTTTTGCCCCCCCCCCCCCCTAAATGGGCTTTGGTATATTGAA
ACAAACATTGTGCGGTAATCTAATGCATGGCGAACGTGGCGAGAAAGGAAGGGAAG
AAAGCGAAAGGAGCGGGCGCTAG

3.5.10 Screening colonies for 2 base pair deletions in *Candida albicans* via Sanger sequencing and directed PCR

Transformant DNA isolation, amplification and Sequencing

We first tried to verify mutation via Sanger sequencing. To do this, transformant DNA isolation for colony PCR was carried out as follows: A sterile pipette tip was used to transfer a single yeast colon to a 1.5 ml Eppendorf tube containing 30 ul of 0.2% SDS. Tubes were vortexed at full speed for 15 seconds Tubes were placed in a 90 degree heat block for 4 minutes. Tubes were then spun at max speed for 1 minute. Supernatant was removed to a new tube. 1 ul of supernatant was used as input for PCR.

Transformant DNA was also purified with the YeaStar genomic DNA kit (Zymo Research) prior to PCR for some colonies. All untransformed control DNA was also isolated via this method. While this method improved yield and overall DNA quality compared to colony PC DNA isolation, it did not improve the Sanger sequencing results.

We amplified a 426 base pair amplicon around the orf19.2539 cut site (using both types of DNA input described above) and the following PCR reaction: 10 ul of Phusion HF PCR master mix (NEB), 1 ul of DNA, 1 ul of 10 uM mixed primers, 8 ul of DNase/RNase free H20. A standard After a 5 minute initial denaturation at 98 degrees, a standard Phusion cycling program was run for 30 cycles, with an annealing temperature of 55 degrees . The primer sequences used are as follows:

Primer upstream orf19.2539:AATTGATTTAGTAGGTGC

Primer downstream orf91.2539: ATGTTAGAGGAAGGAGAG

Colonies were also screened for Cas9 integration with the following primers:

Cas9 forward primer: GTACTTAAACGCGGCTGGAG

Cas9 reverse primer: TGTGTGGGATAAGGGGAGAG

Four separate PCR reactions for each transformant were then pooled together and purified with a 3x ratio of Ampure XP beads (Agencourt) to pooled PCR sample volume. The resulting fragments purified fragments were imaged on a gel and then sent for Sanger sequencing. The following Sanger primers tested are listed below, but failed to give consistent, quality base calls, especially around the intended CRISPR cut site. This was true for genomic DNA isolated form a column (YeaStar Genomic DNA kit; Zymo Research) or via colony PCR. We also observed fluctuations in quality when re-sequencing samples and using the same primers.

Sanger primer 1: AATTGATTTAGTAGGTGCA

Sanger primer 2: AACCGAGTAGAATAACTT

Sanger primer 3: AGAACCATGATTGATTCAAATG

Sanger primer 4: GTTGTACTATTGCCAAGTGACG

Samples were sequenced at Eton Biosciences. Some samples were sequenced with a proprietary sequencing protocol but this did not improve the sequencing trace quality versus standard sequencing reactions. Sanger sequencing data was viewed, aligned and translated with the Benchling software (https://benchling.com). Consensus sequence alignments were performed with Snapgene (GSL biotech; https://www.snapgene.com/)

Whole genome libraries were created from transformant DNA (Nextera XT library kit; Illumina) and sequenced on the Miseq (Illumina) using 50 base pair, single end reads. Reads were aligned to either the *Candida albicans* reference genome SC5314 version A21-s02-m09-r10 downloaded from the *Candida* Genome Database (http://www.candidagenome.org) or the desired mutated *orf19.2539* sequences. Alignments were completed with BWA version 0.7.10-r789, http://bio-bwa.sourceforge.net/)[160] and visualized in the Integrative Genomics Viewer (IGV)[153].

Transformation status check using a directed PCR strategy

Transformant DNA was isolated as described above. Samples were amplified using a standard Taq reaction, for 30 cycles and with an annealing temperature of 55 or 58 degrees.

Two forward primers matching either the wild type or mutated target sequence were designed and both were used with the same reverse primer, approximately 900 base pairs downstream of the cut site:

Wild type forward primer: CTTTTTCACATTTAATCAAAAAGTCG

Mutant forward primer: CTTTCACATTTAATCAAAAAGTCG

Reverse primer: ACCAAGAAAAGACGATTCAA

Transformation status check using a dCAP strategy

A PCR was performed with Taq polymerase, and forward primer designed to introduce two base pairs into the cut site region. With this strategy, if the 2 base pair deletion is present, the insertion

will result in creation of a BccI restriction enzyme site, creating a 29 base pair fragment and 272 base pair fragment. In a wild type genotype, no cut site is present. And one, 301 base pair band should be visible.

Primers used (2 base pair insertions in lower case letters) (IDT):

dCAP Forward primer: ATTTTTTTTTTTCTGCCTATCcaT

dCAP Reverse primer: TGTTAGAGGAAGGAGAGGGTT

A small portion of some reactions were run on a gel to confirm implication after a standard Taq polymerase reaction. Next, the entire PCR reaction was used as input for a 25 ul BccI (NEB) digest. We attempted both 1 hour and 12 hour digests. These reactions were then loaded on a 4% ethidium bromide containing agarose gel. No bands were visible for any conditions, including undigested controls.

Chapter 4: Conclusions and Future directions

With this work, we developed a general strategy for simultaneously profiling both host and fungal cell gene expression. We found that populations of phagocytic macrophages exposed to the commensal and fungal pathogen *C. albicans* rapidly upregulate and downregulate similar sets of genes (between 1 and 4 hours). Conversely, we found that live *C. albicans* inside of macrophages upregulate specific genes compared to *C. albicans* cells which were exposed to macrophages but not engulfed. Future work should include dual host-pathogen profiling of fungal hyphal mutants and/or macrophage fungal receptor mutants, as both categories of genes were found to be important to populations of interacting fungal and host cells, respectively. For example, macrophage receptor Dectin-2 preferentially binds the hyphal form of *C. albicans*[209,210], but we know that both yeast and hyphal cell forms exist among populations of fungal cells interacting with host cells; understanding how host cells respond to a clonal population of non-filamentous *C. albicans* may help us better understand expression fates.

The strategy can also be used to profile interactions at the single cell level. We observed expression bimodality among single macrophages infected with *C. albicans.* Because we uniquely barcoded all transcripts from a single infected macrophage and the *C. albicans* inside of it, we are able to match gene expression in single host cells with that of the fungal cells it had phagocytosed. With this technique, we can begin to understand how individual host cells respond within an infection population and how these gene expression differences can influence *C. albicans* expression, or vice versa. We also observed evidence of *C. albicans* expression bimodality within single macrophages. In the future, experiments with host or fungal cell mutants may be able to parse if host cell expression programs influence *C. albicans* or if fungal cell expression changes host cell

expression. With this experimental setup, we also could not determine the number of *C. albicans* cells within each macrophage. Experiments with a lower MOI could ensure that the majority of macrophages contain only one fungal cell. Additionally, sorting infected cells on a FACS on an instrument with a camera (i.e. the Amnis Imaging Flow Cytometer), which allows the user to visualize and select specific cells for collection and/or allow the user to determine the number of fungal cells inside of each cell post collection, could permit the correlation of gene expression in host to phagocytosed fungal cell number.

Using single cell data, we were able to observe differential splicing in infected macrophages. We detected differential splicing between macrophages in 144 genes, including immune response genes. Interestingly, Dectin-2 had differential exon retention between macrophage infection stages; the Dectin-2 α isoform (6 exons) predominantly in stage 1, and Dectin-2 β isoform (5 exons) predominately in stage 2. The truncated isoform Dectin-2 β lacks part of the intracellular domain and most of the transmembrane domain of the receptor[139]. We found that Dectin-2 β has a lower posterior probability of transmembrane helix and the lack of this transmembrane region may lead to the secretion of a protein that can antagonize full-length Dectin-2[139]. As activation of Dectin-2 receptors on macrophages and dendritic cells by *C. albicans* leads to Th17 T cell differentiation [62,63], splicing variation among single macrophages might indicate that single cells have different potentials to respond to fungal infection within a population. In future work, it would be interesting to examine how differential splicing is affected by factors like hyphal mutants (as explained above) and antifungal drugs. In the future, this type of information could also be used when developing novel immunotherapies to fight fungal infections.

We believe that single cell or near single cell levels of gene expression are also important to examine in fungal cells. We observe phenotypic diversity among clonal fungal cells interacting with macrophages, as well as evidence of transcript bimodality in live, phagocytosed *C. albicans.* Surprisingly, we saw bimodal expression of filamentation genes, shown to be critical for *C. albicans* survival in host cells. For these reasons, we this information could be potentially used when considering new antifungal targets. Testing these antifungals at the single cell level would help researchers select a target that has the intended affect in that majority of the cell population. And with the dual host-pathogen gene expression profiling strategy, researchers could simultaneous measure how individual immune cells within a population respond to antifungal treatment to assess off target effects or host cell toxicity.

We also applied the dual host-fungal RNA-sequencing approach to *C. albicans* exposed to epithelial cells derived from murine, small intestine organoids. *C. albicans* is a commensal resident of most adult human intestines but the transcriptional programs that differ between *C. albicans* interacting with immune cells versus gut cells is not well defined. In this work, we present two RNA-sequencing experiments that captured both host and fungal cell gene expression at the population level. For these experiments, we did not sort infection populations in an effort to save resources and prove that this strategy could be applied to another host cell type (other than immune cells). In the future, applying the same FACS strategy of isolating populations of epithelial cells that have endocytosed live or dead *C. albicans*, as well as free epithelial cells and *C. albicans* cells that were exposed to each other but did not interact would be useful for understanding infection population dynamics and comparing them to what we now know about other fungal and host immune cell interactions. We found two interesting things with this preliminary work. Firstly,

some of the same pathways upregulated in *C. albicans* exposed to epithelial cells (MOI 10:1) are found to be upregulated in *C. albicans* exposed to macrophages (MOI 1:1). Secondly, at a predicted MOI 1:1, there seems to be little difference in gene expression between *C. albicans* exposed to epithelial cells rather than *C. albicans* growing alone in organoid media and samples tend to cluster by time point rather than exposure status. We observed much less filamentation and organoid penetration over time at this lower MOI at early infection time points (i.e. less than 24 hours), compared to the MOI of 10:1. As such, MOI should be an important consideration for designing experiments and it would be interesting to catalog the effects that MOI has on host and pathogen interactions in specific host niches.

Secondly, in both epithelial and *C. albicans* RNA-sequencing experiments, we identified many uncharacterized genes to be differentially upregulated in exposed *C. albicans.* It would be beneficial to further refine the LOF CRISPR mutation strategy, to have a quick and relatively cheap way to screen for phenotypes within LOF mutants. However, while costlier, inserting large pieces of DNA to delete a gene may be preferable to inducing LOF through one or two base pair mutations, as the AT rich genome of *C. albicans* makes confirming small insertions, deletions or mutations difficult. In future work, it will be interesting to characterize the *orf19.2539* mutant strain in the context of filamentation phenotype, invasion capabilities and host epithelial cell physical interactions. The magnitude of the cell size difference between wild type and the *orf19.2539* deletion mutant could be measured by a machine like a Coulter Counter (Beckman Coulter). Also, examining if and how host epithelial cell gene expression responds to an *orf19. 2539* deletion mutant or an *orf19.2539* overexpression mutant would be interesting. Because the levels of *orf19.2539* gene expression varied depending on MOI, further investigation into how the

orf19.2539 mutants physically and genetically respond to host cells at different MOIs will be important. A cost effective experimental framework allowing for rapid mutation or deletion of these uncharacterized orfs that seem to be important for gut epithelial cell interactions could help us better understand the role that fungus plays in the host microbiome and could have implications for IBD related disease, like Crohn's disease.

We used FACS to isolate single, *C. albicans* infected epithelial cells. We tested two different single cell sequencing approaches on these cells. The first was using the high-throughput droplet based strategy (10x genomics). Using this strategy we were not able to detect fungal signal. As traditional lysis buffers do not effectively lyse *C. albicans*, this was not surprising. Future work may include working with single cell companies who develop these technologies to test ways to effectively lyse fungal cells. Additionally, we recovered 10 fold fewer single *C. albicans* infected epithelial cells than single uninfected cells. Because organoid cells need to be released from the matrigel and tend to die quickly during physical manipulation, we chose to forgo extra processing for this initial test. However, future work should include a selection for epithelial cells during the FACS process. For our test, we relied on epithelial cell size and the fact that our *C. albicans* reporter strain has fluorescent markers to differentiate between uninfected and infected epithelial cells. However, we also processed single infected epithelial cells from this same sorting event with Smart-Seq-2 strategy (**Chapter 2).** With this strategy, we were able to detect both host and fungal gene expression but we found that approximately half of the wells we tested contained only *C. albicans* transcripts, leading us to believe that large filamentous *Candida* cells or clumps of *Candida* were mistakenly identified as single, *C. albicans* epithelial cells during sorting. Using an antibody (i.e. anti-Epcam with Cy7) based method to isolate single epithelial cells containing *C. albicans* will

be critical to increase the recovery of single, *C. albicans* infected epithelial cells for further analyses.

In the future, it will be important to examine the correlation of single cell gene expression levels to protein levels. Does bimodal gene expression within a population lead to bimodal levels of protein production between cells? It will be important to understand this dynamic, especially if we are to use this information for the development of fungal drugs or therapeutics. Currently, there are limited commercial options for protein detection at the single cell level, but the Milo Single cell western machine (Protein Simple) is an example of one such technology. As single cell transcriptomics continues to evolve, there will likely be more cost effective options for screening protein levels in single cells in the near future. On the host side, if we observe that some immune response genes are NOT being expressed in a population of *C. albicans* infected macrophages as uniformly and highly as we would expect, we may be able to target those genes/proteins with a new therapeutic. Several new single cell gene expression data based startups and non-profits are emerging, including some in Massachusetts, with the goal of using both single cell expression and genomic information to develop new drugs for cancer and other diseases. Critically, all of the epithelial cell and *C. albicans* expression analysis performed in this work (**Chapter 3**) should also be reviewed by a statistician and/ or bioinformation, as I am neither. It is important that the underlying assumptions I made in these analysis are reviewed for accuracy before these findings are used as the basis for future research.

Increasing our understanding of the basic biology of host and fungal pathogen interactions, including how these changes depend on MOI and host niche, will be the key to developing desperately needed novel antifungals to fight deadly and rapidly evolving, drug resistant fungi.

Bibliography

1. Bongomin, F., Gago, S., Oladele, R. O. & Denning, D. W. Global and Multi-National Prevalence of Fungal Diseases-Estimate Precision. *J. fungi (Basel, Switzerland)* **3**, (2017).
2. Sardi, J., Pitangui, N., Gullo, F., Almeida, A. & Giannini, M. A Mini Review of Candida Species in Hospital Infection: Epidemiology,Virulence Factor and Drugs Resistance and Prophylaxis. *Trop. Med. Surg.* **1**, 1–7 (2013).
3. Pfaller, M. A., Jones, R. N., Messer, S. A., Edmond, M. B. & Wenzel, R. P. National surveillance of nosocomial blood stream infection due to species of Candida other than Candida albicans: Frequency of occurrence and antifungal susceptibility in the SCOPE program. *Diagn. Microbiol. Infect. Dis.* **30**, 121–129 (1998).
4. Wisplinghoff, H. *et al.* Nosocomial Bloodstream Infections in US Hospitals: Analysis of 24,179 Cases from a Prospective Nationwide Surveillance Study. *Clin. Infect. Dis.* **39**, 309–317 (2004).
5. Zaoutis, T. E. *et al.* The Epidemiology and Attributable Outcomes of Candidemia in Adults and Children Hospitalized in the United States: A Propensity Analysis. *Clin. Infect. Dis.* **41**, 1232–1239 (2005).
6. Fridkin, S. K. Candidemia is Costly--Plain and Simple. *Clin. Infect. Dis.* **41**, 1240–1241 (2005).
7. Gudlaugsson, O. *et al.* Attributable Mortality of Nosocomial Candidemia, Revisited. *Clin. Infect. Dis.* **37**, 1172–1177 (2003).
8. Cleveland, A. A. *et al.* Changes in incidence and antifungal drug resistance in candidemia: Results from population-based laboratory surveillance in Atlanta and Baltimore, 2008-

2011. *Clin. Infect. Dis.* **55**, 1352–1361 (2012).

9. Pfaller, M. A. & Diekema, D. J. Epidemiology of invasive candidiasis: a persistent public health problem. *Clin. Microbiol. Rev.* **20**, 133–63 (2007).

10. Blot, S. I., Vandewoude, K. H., Hoste, E. A. & Colardyn, F. A. Effects of nosocomial candidemia on outcomes of critically Ill patients. *Am. J. Med.* **113**, 480–485 (2002).

11. Perea, S. *et al.* Prevalence of Molecular Mechanisms of Resistance to Azole Antifungal Agents in Candida albicans Strains Displaying High-Level Fluconazole Resistance Isolated from Human Immunodeficiency Virus-Infected Patients Prevalence of Molecular Mechanisms of Resista. *Antimicrob. Agents Chemother.* **45**, 2676–2684 (2001).

12. White, T. C. Increased mRNA Levels of ERG16 , CDR , and MDR1 Correlate with Increases in Azole Resistance in Candida albicans Isolates from a Patient Infected with Human Immunodeficiency Virus. **41**, 1482–1487 (1997).

13. Centers for Disease Control and Prevention. Antifungal Resistance | Fungal Diseases | CDC. Available at: https://www.cdc.gov/fungal/antifungal-resistance.html. (Accessed: 19th February 2019)

14. Ford, C. B. *et al.* The evolution of drug resistance in clinical isolates of candida albicans. *Elife* **2015**, 1–27 (2015).

15. Gow, N. A. R., Brown, A. J. P. & Odds, F. C. Fungal morphogenesis and host invasion Gow, Brown and Odds 367. 366–371

16. Lorenz, M. C., Bender, J. A. & Fink, G. R. Transcriptional response of *Candida albicans* upon internalization by macrophages. *Eukaryot. Cell* **3**, 1076–1087 (2004).

17. Pande, K., Chen, C. & Noble, S. M. Passage through the mammalian gut triggers a phenotypic switch that promotes *Candida albicans commensalism*. *Nat. Genet.* **45**, 1088–

1091 (2013).

18. Richardson, J. P. *et al.* Candidalysin drives epithelial signaling, neutrophil recruitment, and immunopathology at the vaginal mucosa. *Infect. Immun.* IAI.00645-17 (2017). doi:10.1128/IAI.00645-17

19. Fradin, C. *et al.* Stage-specific gene expression of Candida albicans in human blood. *Mol. Microbiol.* **47**, 1523–1543 (2003).

20. Hube, B. From commensal to pathogen: stage-and tissue-specific gene expression of Candida albicans. *Curr. Opin. Microbiol.* **7**, 336–341 (2004).

21. Selmecki, A. M., Dulmage, K., Cowen, L. E., Anderson, J. B. & Berman, J. Acquisition of Aneuploidy Provides Increased Fitness during the Evolution of Antifungal Drug Resistance. *PLoS Genet.* **5**, e1000705 (2009).

22. Schulze, J. & Sonnenborn, U. Yeasts in the Gut: From Commensals to Infectious Agents. *Dtsch.Ärztebl.* **106**, 837–842 (2009).

23. Lockhart, S. R., Wu, W., Radke, J. B., Zhao, R. & Soll, D. R. Increased Virulence and Competitive Advantage of a/α Over a/a or α/α Offspring Conserves the Mating System of Candida albicans. *Genetics* **169**, (2005).

24. Calderone, R. & Fonzi, W. Virulence factors of Candida albicans. *Trends Microbiol.* **9**, 327–35 (2001).

25. Erwig, L. P. & Gow, N. a. R. Interactions of fungal pathogens with phagocytes. *Nat. Rev. Microbiol.* **14**, 163–176 (2016).

26. Selmecki, A., Forche, A. & Berman, J. Aneuploidy and Isochromosome Formation in Drug-Resistant Candida albicans. *Science (80-.).* **313**, (2006).

27. Sanglard, D., Ischer, F., Parkinson, T., Falconer, D. & Bille, J. Candida albicans mutations

in the ergosterol biosynthetic pathway and resistance to several antifungal agents. *Antimicrob. Agents Chemother.* **47**, 2404–12 (2003).

28. Calderone, R. A. & Clancy, C. J. *Candida and candidiasis*. (American Society for Microbiology Press, 2011).

29. Bensasson, D. *et al.* Diverse lineages of Candida albicans live on old oaks. *Genetics* **211**, 277–288 (2018).

30. Odds, F. C. *Candida and Candidosis*. (Bailliere Tindall, 1988).

31. Sudbery, P., Gow, N. & Berman, J. The distinct morphogenic states of Candida albicans. *Trends Microbiol.* **12**, 317–324 (2004).

32. *Candida Albicans: Cellular and Molecular Biology*. (Springer International Publishing, 2017). doi:10.1007/978-3-319-50409-4

33. Barelle, C. J. *et al.* Asynchronous Cell Cycle and Asymmetric Vacuolar Inheritance in True Hyphae of. *Society* **2**, 398–410 (2003).

34. Gow, N. A. R. & Gooday, G. W. A model for the germ tube formation and mycelial growth form of *Candida albicans*. *Med. Mycol.* **22**, 137–144 (1984).

35. Yokoyama, K. & Takeo, K. Differences of asymmetrical division between the pseudomycelial and yeast forms of Candida albicans and their effect on multiplication. *Arch. Microbiol.* **134**, 251–253 (1983).

36. Gow, N. A. R., Van De Veerdonk, F. L., Brown, A. J. P. & Netea, M. G. Candida albicans morphogenesis and host defence: Discriminating invasion from colonization. *Nat. Rev. Microbiol.* **10**, 112–122 (2012).

37. Miller, M. G. & Johnson, A. D. White-opaque switching in Candida albicans is controlled by mating-type locus homeodomain proteins and allows efficient mating. *Cell* **110**, 293–

302 (2002).

38. Sasse, C., Hasenberg, M., Weyler, M., Gunzer, M. & Morschhäuser, J. White-opaque switching of Candida albicans allows immune evasion in an environment-dependent fashion. *Eukaryot. Cell* **12**, 50–58 (2013).

39. Lohse, M. B. & Johnson, A. D. Differential phagocytosis of white versus opaque Candida albicans by Drosophila and mouse phagocytes. *PLoS One* **3**, (2008).

40. Lohse, M. B. & Johnson, A. D. White-opaque switching in Candida albicans. **12**, 650–654 (2010).

41. Rikkerink, E. H., Magee, B. B. & Magee, P. T. Opaque-white phenotype transition: a programmed morphological transition in Candida albicans. *J. Bacteriol.* **170**, 895–899 (1988).

42. Lohse, M. B. & Johnson, A. D. Temporal anatomy of an epigenetic switch in cell programming: The white-opaque transition of C. albicans. *Mol. Microbiol.* **78**, 331–343 (2010).

43. Hernday, A. D. *et al.* Structure of the transcriptional network controlling white-opaque switching in Candida albicans. *Mol. Microbiol.* **90**, 22–35 (2013).

44. Zordan, R. E., Galgoczy, D. J. & Johnson, A. D. Epigenetic properties of white-opaque switching in Candida albicans are based on a self-sustaining transcriptional feedback loop. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 12807–12 (2006).

45. Zordan, R. E., Miller, M. G., Galgoczy, D. J., Tuch, B. B. & Johnson, A. D. Interlocking Transcriptional Feedback Loops Control White-Opaque Switching in Candida albicans. *PLoS Biol.* **5**, e256 (2007).

46. Huang, G., Srikantha, T., Sahni, N., Yi, S. & Soll, D. R. CO2 Regulates White-to-Opaque

Switching in Candida albicans. *Curr. Biol.* **19**, 330–334 (2009).

47. Ramírez-Zavala, B., Reuß, O., Park, Y.-N., Ohlsen, K. & Morschhäuser, J. Environmental Induction of White–Opaque Switching in Candida albicans. *PLoS Pathog.* **4**, e1000089 (2008).

48. Lee, J. K., Bhakta, S., Rosen, S. D. & Hemmerich, S. *Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-sulfotransferase That Is Highly Restricted to Intestinal Tissue*. (1999).

49. Sicard, J.-F. *et al.* N-Acetyl-glucosamine influences the biofilm formation of Escherichia coli. *Gut Pathog.* **10**, 26 (2018).

50. Bertin, Y. *et al.* Carbohydrate utilization by enterohaemorrhagic Escherichia coli O157:H7 in bovine intestinal content. *Environ. Microbiol.* **15**, 610–22 (2013).

51. Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502**, 96–9 (2013).

52. Brown, G. D. Innate antifungal immunity: the key role of phagocytes. *Annu. Rev. Immunol.* **29**, 1–21 (2011).

53. Huffnagle, G. B. & Noverr, M. C. The emerging world of the fungal microbiome. *Trends Microbiol.* **21**, 334–341 (2013).

54. Frohner, I. E., 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.* **71**, 240–252 (2009).

55. Brown, G. D. & Gordon, S. Immune recognition. A new receptor for beta-glucans. *Nature* **413**, 36–37 (2001).

56. Ferwerda, B. *et al.* Human Dectin-1 Deficiency and Mucocutaneous Fungal Infections. *N.*

Engl. J. Med. **361**, 1760–1767 (2009).

57. Plantinga, T. S. *et al.* Early Stop Polymorphism in Human DECTIN-1 Is Associated with Increased *Candida* Colonization in Hematopoietic Stem Cell Transplant Recipients. *Clin. Infect. Dis.* **49**, 724–732 (2009).

58. Rosentul, D. C. *et al.* Genetic Variation in the Dectin-1/CARD9 Recognition Pathway and Susceptibility to Candidemia. *J. Infect. Dis.* **204**, 1138–1145 (2011).

59. Sato, K. *et al.* Dectin-2 is a pattern recognition receptor for fungi that couples with the Fc receptor γ chain to induce innate immune responses. *J. Biol. Chem.* **281**, 38854–38866 (2006).

60. Mcgreal, E. P. *et al.* The carbohydrate-recognition domain of Dectin-2 is a C-type lectin with specificity for high mannose. *Glycobiology* **16**, 422–430 (2006).

61. Ifrim, D. C. *et al.* The Role of Dectin-2 for Host Defense Against Disseminated Candidiasis. *J. Interf. Cytokine Res.* **36**, 267–276 (2016).

62. Saijo, S. *et al.* Dectin-2 recognition of α-mannans and induction of Th17 cell differentiation is essential for host defense against candida albicans. *Immunity* **32**, 681–691 (2010).

63. Robinson, M. J. *et al.* Dectin-2 is a Syk-coupled pattern recognition receptor crucial for Th17 responses to fungal infection. *J. Exp. Med.* **206**, 2037–2051 (2009).

64. Seider, K., Heyken, A., Lüttich, A., Miramón, P. & Hube, B. Interaction of pathogenic yeasts with phagocytes: survival, persistence and escape. *Curr. Opin. Microbiol.* **13**, 392–400 (2010).

65. de Bernardis, F. *et al.* The secretion of aspartyl proteinase, a virulence enzyme, by isolates of Candida albicans from the oral cavity of HIV-infected subjects. *Eur. J. Epidemiol.* **8**, 362–367 (1992).

66. Wu, T., Samaranayake, L. P., Cao, B. Y. & Wang, J. In-vitro proteinase production by oral Candida albicans isolates from individuals with and without HIV infection and its attenuation by antimycotic agents. *J. Med. Microbiol.* **44**, 311–316 (1996).

67. Agatensi, L. *et al.* Vaginopathic and proteolytic Candida species in outpatients attending a gynaecology clinic. *J. Clin. Pathol.* **44**, 826–30 (1991).

68. Cassone, A., De Bernardis, F., Mondello, F., Ceddia, T. & Agatensi, L. Evidence for a correlation between proteinase secretion and vulvovaginal candidosis. *J. Infect. Dis.* **156**, 777–83 (1987).

69. De Bernardis, F. *et al.* Evidence for a role for secreted aspartate proteinase of candida albicans in vulvovaginal candidiasis. *J. Infect. Dis.* **161**, 1276–1283 (1990).

70. Léchenne, B., Foundling, S., Beggah, S., Reichard, U. & Monod, M. Intra- and intermolecular events direct the propeptide-mediated maturation of the Candida albicans secreted aspartic proteinase Sap1p. *Microbiology* **146**, 2765–2773 (2000).

71. Abu-Elteen, K., Elkarmi, A. & Hamad, M. Characterization of phenotype-based pathogenic determinants of various Candida albicans strains in Jordan. *Japanese J. Infect.* (2001).

72. De Bernardis, F. *et al.* Elevated aspartic proteinase secretion and experimental pathogenicity of Candida albicans isolates from oral cavities of subjects infected with human immunodeficiency virus. *Infect. Immun.* **64**, 466–71 (1996).

73. de Bernardis, F. *et al.* High aspartyl proteinase production and vaginitis in human immunodeficiency virus-infected women. *J. Clin. Microbiol.* **37**, 1376–80 (1999).

74. Naglik, J. R., Challacombe, S. J. & Hube, B. Candida albicans secreted aspartyl proteinases in virulence and pathogenesis. *Microbiol. Mol. Biol. Rev.* **67**, 400–28, table of contents (2003).

75. Hube, B., Naglik, J., Hube, B. & Naglik, J. Candida albicans proteinases: resolving the mystery of a gene family. *Microbiology* **147**, 1997–2005 (2001).

76. Dunkel, N. & Morschhäuser, J. Loss of heterozygosity at an unlinked genomic locus is responsible for the phenotype of a Candida albicans sap4Δ sap5Δ sap6Δ mutant. *Eukaryot. Cell* **10**, 54–62 (2011).

77. Schaller, M. *et al.* Secreted aspartic proteinase (Sap) activity contributes to tissue damage in a model of human oral candidosis. *Mol. Microbiol.* **34**, 169–180 (1999).

78. Watts, H. ., Cheah, F. S. ., Hube, B., Sanglard, D. & Gow, N. A. . Altered adherence in strains of *Candida albicans* harbouring null mutations in secreted aspartic proteinase genes. *FEMS Microbiol. Lett.* **159**, 129–135 (1998).

79. Schaller, M., Schackert, C., Korting, H. C., Januschke, E. & Hube, B. Invasion of Candida albicans correlates with expression of secreted aspartic proteinases during experimental infection of human epidermis. *J. Invest. Dermatol.* **114**, 712–717 (2000).

80. Ibrahim, A. S., Filler, S. G., Sanglard, D., Edwards, J. E. & Hube, B. Secreted aspartyl proteinases and interactions of Candida albicans with human endothelial cells. *Infect. Immun.* **66**, 3003–5 (1998).

81. Kretschmar, M. *et al.* Germ tubes and proteinase activity contribute to virulence of Candida albicans in murine peritonitis. *Infect. Immun.* **67**, 6637–42 (1999).

82. Hube, B. *et al.* Disruption of each of the secreted aspartyl proteinase genes SAP1, SAP2, and SAP3 of Candida albicans attenuates virulence. *Infect. Immun.* **65**, 3529–38 (1997).

83. Sanglard, D., Hube, B., Monod, M., Odds, F. C. & Gow, N. A. R. A triple deletion of the secreted aspartyl proteinase genes SAP4, SAP5, and SAP6 of Candida albicans causes attenuated virulence. *Infect. Immun.* **65**, 3539–3546 (1997).

84. Borg-von Zepelin, M., Beggah, S., Boggian, K., Sanglard, D. & Monod, M. The expression of the secreted aspartyl proteinases Sap4 to Sap6 from Candida albicans in murine macrophages. *Mol. Microbiol.* **28**, 543–554 (1998).

85. Smeekens, S. P. *et al.* Functional genomics identifies type i interferon pathway as central for host defense against *Candida albicans*. *Nat. Commun.* **4**, (2013).

86. Quintin, J. *et al.* Candida albicans infection affords protection against reinfection via functional reprogramming of monocytes. *Cell Host Microbe* **12**, 223–232 (2012).

87. Hebecker, B. *et al.* Dual-species transcriptional profiling during systemic candidiasis reveals organ-specific host-pathogen interactions. *Sci Rep* **6**, 36055 (2016).

88. Liu, Y. *et al.* New signaling pathways govern the host response to *C. albicans* infection in various niches. *Genome Res.* **125**, 679–689 (2015).

89. Niemiec, M. J. *et al.* Dual transcriptome of the immediate neutrophil and *Candida albicans* interplay. *BMC Genomics* **18**, 696 (2017).

90. Tierney, L. *et al.* An interspecies regulatory network inferred from simultaneous RNA-seq of *Candida albicans* invading innate immune cells. *Front. Microbiol.* **3**, 1–14 (2012).

91. Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–40 (2013).

92. Avraham, R. *et al.* Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. *Cell* **162**, 1309–1321 (2015).

93. Saliba, A.-E. E. *et al.* Single-cell RNA-seq ties macrophage polarization to growth rate of intracellular Salmonella. *Nat. Microbiol.* **2**, 1–8 (2016).

94. Avital, G. *et al.* scDual-Seq: mapping the gene regulatory program of *Salmonella* infection by host and pathogen single-cell RNA-sequencing. *Genome Biol.* **18**, 200 (2017).

95. Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363–9 (2014).

96. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. *Cell Rep.* **2**, 666–673 (2012).

97. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444**, 643–646 (2006).

98. Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nat. Genet.* **38**, 636–643 (2006).

99. Goetz, J. J. & Trimarchi, J. M. Transcriptome sequencing of single cells with Smart-Seq. *Nat. Biotechnol.* **30**, 763–765 (2012).

100. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–81 (2014).

101. Petersen, M. & Wengel, J. LNA: a versatile tool for therapeutics and genomics. *Trends Biotecnol.* **21**, 74–81 (2003).

102. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, (2017).

103. Nadal-Ribelles, M. *et al.* Sensitive high-throughput single-cell RNA-seq reveals within-clonal transcript correlations in yeast populations. doi:10.1038/s41564-018-0346-9

104. Baumstark-Khan, C. *et al. Green Fluorescent Protein (GFP) as a Marker for Cell Viability After UV Irradiation*. *Journal of Fluorescence* **9**, (1999).

105. Tapani, E., Taavitsainen, M., Lindros, K., Vehmas, T. & Lehtonen, E. Toxicity of Ethanol in Low Concentrations. *Acta radiol.* **37**, 923–926 (1996).

106. Neville, B. A., d'Enfert, C. & Bougnoux, M. E. Candida albicans commensalism in the

gastrointestinal tract. *FEMS Yeast Res.* **15**, 1–13 (2015).

107. Cohen, R., Roth, F. J., Delgado, E., Ahearn, D. G. & Kalser, M. H. Fungal Flora of the Normal Human Small and Large Intestine. *N. Engl. J. Med.* **280**, 638–641 (1969).

108. Odds, F. C. *et al.* Candida albicans strain maintenance, replacement, and microvariation demonstrated by multilocus sequence typing. *J. Clin. Microbiol.* **44**, 3647–3658 (2006).

109. Samonis, G. *et al.* Prospective study of the impact of broad-spectrum antibiotics on the yeast flora of the human gut. *Eur. J. Clin. Microbiol. Infect. Dis.* **13**, 665–667 (1994).

110. Kennedy, M. J. & Volz, P. A. Ecology of Candida albicans gut colonization: inhibition of Candida adhesion, colonization, and dissemination from the gastrointestinal tract by bacterial antagonism. *Infect. Immun.* **49**, 654–63 (1985).

111. Chen, C., Pande, K., French, S. D., Tuch, B. B. & Noble, S. M. An Iron Homeostasis Regulatory Circuit with Reciprocal Roles in Candida albicans Commensalism and Pathogenesis. *Cell Host Microbe* **10**, 118–135 (2011).

112. Lorenz, M. C., Bender, J. a & Fink, G. R. Transcriptional Response of Candida albicans upon Internalization by Macrophages. *Eukaryot. Cell* **3**, 1076–1087 (2004).

113. Barker, N., Van Oudenaarden, A. & Clevers, H. Identifying the stem cell of the intestinal crypt: Strategies and pitfalls. *Cell Stem Cell* **11**, 452–460 (2012).

114. Jung, C., Hugot, J.-P. & Barreau, F. Peyer's Patches: The Immune Sensors of the Intestine. *Int. J. Inflam.* **2010**, 823710 (2010).

115. A.S., D., U., A., D.M., A. & A., R.-H. Meta-analysis of the turnover of intestinal epithelia in preclinical animal species and humans. *Drug Metab. Dispos.* **42**, 2016–2022 (2014).

116. Nusse, R. Wnt signaling and stem cell control. *Cell Res.* **18**, 523–527 (2008).

117. Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**,

(2017).

118. Wallach, T. E. & Bayrer, J. R. Intestinal Organoids: New Frontiers in the Study of Intestinal Disease and Physiology. *J. Pediatr. Gastroenterol. Nutr.* **64**, 180–185 (2017).

119. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262–265 (2009).

120. Yui, S. *et al.* Functional engraftment of colon epithelium expanded in vitro from a single adult Lgr5+ stem cell. *Nat. Med.* **18**, 618–623 (2012).

121. Lo, H.-J. *et al.* Nonfilamentous C. albicans Mutants Are Avirulent. *Cell* **90**, 939–949 (1997).

122. Andrutis, K. A., Riggle, P. J., Kumamoto, C. A. & Tzipori, S. Intestinal lesions associated with disseminated candidiasis in an experimental animal model. *J. Clin. Microbiol.* **38**, 2317–23 (2000).

123. Johannes, F.-J. *et al.* In vitro reconstructed human epithelia reveal contributions of Candida albicans EFG1 and CPH1 to adhesion and invasion. *Microbiology* **148**, 497–506 (2002).

124. Bareiss, P. M. *et al.* Organotypical tissue cultures from adult murine colon as an in vitro model of intestinal mucosa. *Histochem. Cell Biol.* **129**, 795–804 (2008).

125. Prieto, D. *et al.* A Biphasic Innate Immune MAPK Response Discriminates between the Yeast and Hyphal Forms of Candida albicans in Epithelial Cells. *Cell Host Microbe* **8**, 225–235 (2010).

126. Moyes, D. L., Richardson, J. P. & Naglik, J. R. Candida albicans- epithelial interactions and pathogenicity mechanisms: Scratching the surface. *Virulence* **5594**, 00–00 (2015).

127. Martin, R. *et al.* A core filamentation response network in *Candida albicans* is restricted to eight genes. *PLoS One* **8**, e58613 (2013).

128. Issi, L. *et al.* Zinc cluster transcription factors alter virulence in Candida albicans. *Genetics*

205, 559–576 (2017).

129. Du, H. *et al.* Roles of Candida albicans Gat2, a GATA-Type Zinc Finger Transcription Factor, in Biofilm Formation, Filamentous Growth and Virulence. *PLoS One* **7**, e29707 (2012).

130. Bruno, V. M. *et al.* Transcriptomic Analysis of Vulvovaginal Candidiasis Identifies a Role for the NLRP3 Inflammasome. *MBio* **6**, e00182-15 (2015).

131. Lionakis, M. S. *et al.* CX3CR1-dependent renal macrophage survival promotes Candida control and host survival. *J. Clin. Invest.* **123**, 5035–51 (2013).

132. Leonardi, I. *et al.* CX3CR1+mononuclear phagocytes control immunity to intestinal fungi. *Science* **359**, 232–236 (2018).

133. Van Der Maaten, L. J. P. & Hinton, G. E. Visualizing high-dimensional data using t-sne. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

134. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).

135. McDavid, A. *et al.* Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. *Bioinformatics* **29**, 461–467 (2013).

136. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

137. Korthauer, K. D. *et al.* A statistical approach for identifying differential distributions in single-cell RNA-seq experiments. *Genome Biol.* **17**, 222 (2016).

138. Huang, Y. & Sanguinetti, G. BRIE: transcriptome-wide splicing quantification in single cells. *Genome Biol.* **18**, 123 (2017).

139. Gavino, A. C. P., Chung, J.-S., Sato, K., Ariizumi, K. & Cruz, P. D. Identification and

expression profiling of a human C-type lectin, structurally homologous to mouse dectin-2. *Exp. Dermatol.* **14**, 281–288 (2005).

140. Reales-Calderón, J. A., Aguilera-Montilla, N., Corbí, Á. L., Molero, G. & Gil, C. Proteomic characterization of human proinflammatory M1 and anti-inflammatory M2 macrophages and their response to *Candida albicans*. *Proteomics* **14**, 1503–1518 (2014).

141. Jouault, T. *et al.* Host responses to a versatile commensal: PAMPs and PRRs interplay leading to tolerance or infection by *Candida albicans*. *Cell. Microbiol.* **11**, 1007–15 (2009).

142. Westman, J., Moran, G., Mogavero, S., Hube, B. & Grinstein, S. Candida albicans Hyphal Expansion Causes Phagosomal Membrane Damage and Luminal Alkalinization. *MBio* **9**, e01226-18 (2018).

143. Bain, J. M. *et al. Candida albicans* Hypha Formation and Mannan Masking of β-Glucan Inhibit Macrophage Phagosome Maturation. *MBio* **5**, e01874-14 (2014).

144. Wellington, M., Koselny, K., Sutterwala, F. S. & Krysan, D. J. *Candida albicans* triggers NLRP3-mediated pyroptosis in macrophages. *Eukaryot. Cell* **13**, 329–340 (2014).

145. Kasper, L. *et al.* The fungal peptide toxin Candidalysin activates the NLRP3 inflammasome and causes cytolysis in mononuclear phagocytes. *Nat. Commun.* **9**, 4260 (2018).

146. de Jong, I. G., Haccou, P. & Kuipers, O. P. Bet hedging or not? A guide to proper classification of microbial survival strategies. *BioEssays* **33**, 215–223 (2011).

147. Lewis, K. Persister Cells. *Annu. Rev. Microbiol.* **64**, 357–372 (2010).

148. Kerscher, B., Willment, J. A. & Brown, G. D. The Dectin-2 family of C-type lectin-like receptors: An update. *Int. Immunol.* **25**, 271–277 (2013).

149. Selmecki, A., Bergmann, S. & Berman, J. Comparative genome hybridization reveals widespread aneuploidy in Candida albicans laboratory strains. *Mol. Microbiol.* **55**, 1553–

1565 (2005).

150. Gerami-Nejad, M., Zacchi, L. F., McClellan, M., Matter, K. & Berman, J. Shuttle vectors for facile gap repair cloning and integration into a neutral locus in *Candida albicans*. *Microbiol. (United Kingdom)* **159**, 565–579 (2013).

151. Safonova, Y., Bankevich, A. & Pevzner, P. A. dipSPAdes: Assembler for Highly Polymorphic Diploid Genomes. in *Journal of Computational Biology* **22**, 265–279 (Springer, Cham, 2014).

152. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–402 (1997).

153. Thorvaldsdottir, H. *et al.* Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–92 (2013).

154. Stubbs, M. *et al.* MLL-AF9 and FLT3 cooperation in acute myelogenous leukemia: development of a model for rapid therapeutic assessment. *Leukemia* **22**, 66–77 (2008).

155. Zaitoun, I., Erickson, C. S., Schell, K. & Epstein, M. L. Use of RNAlater in fluorescence-activated cell sorting (FACS) reduces the fluorescence from GFP but not from DsRed. *BMC Res. Notes* **3**, 328 (2010).

156. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. **33**, (2015).

157. Villani, A.-C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science (80-.).* **356**, (2017).

158. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

159. Karolchik, D. *et al.* The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32**,

493D – 496 (2004).

160. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–95 (2010).

161. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

162. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–40 (2010).

163. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–512 (2013).

164. Li, B. B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

165. Inglis, D. O. *et al.* 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**, D667–D674 (2012).

166. Kumamoto, C. Inflammation and gastrointestinal Candida colonization. *Curr Opin Microbiol* **14**, 386–391 (2012).

167. Trojanowska, D. *et al.* The role of Candida in inflammatory bowel disease. Estimation of transmission of C. albicans fungi in gastrointestinal tract based on genetic affinity between strains. *Med. Sci. Monit.* **16**, CR451-7 (2010).

168. Standaert-Vitse, A. *et al.* Candida albicans Colonization and ASCA in Familial Crohn's Disease. *Am. J. Gastroenterol.* **104**, 1745–1753 (2009).

169. Lewis, J. D. D. *et al.* Inflammation, Antibiotics, and Diet as Environmental Stressors of the

Gut Microbiome in Pediatric Crohn's Disease. *Cell Host Microbe* **18**, 489–500 (2015).

170. Li, Q. *et al.* Dysbiosis of gut fungal microbiota is associated with mucosal inflammation in Crohn's disease. *J. Clin. Gastroenterol.* **48**, 513–23 (2014).

171. Ananthakrishnan, A. N. *et al.* Differential Effect of Genetic Burden on Disease Phenotypes in Crohn's Disease and Ulcerative Colitis: Analysis of a North American Cohort. *Am. J. Gastroenterol.* **109**, 395–400 (2014).

172. Loftus, E. V. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology* **126**, 1504–1517 (2004).

173. Molodecky, N. A. *et al.* Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* **142**, 46-54.e42 (2012).

174. Naglik, J. R., Moyes, D. L., Wächtler, B. & Hube, B. Candida albicans interactions with epithelial cells and mucosal immunity. *Microbes Infect.* **13**, 963–76 (2011).

175. Wächtler, B., Wilson, D., Haedicke, K., Dalle, F. & Hube, B. From attachment to damage: Defined genes of Candida albicans mediate adhesion, invasion and damage during interaction with oral epithelial cells. *PLoS One* **6**, (2011).

176. Zhu, W. & Filler, S. G. Interactions of Candida albicans with epithelial cells. *Cell. Microbiol.* **12**, 273–282 (2010).

177. Moreno-Ruiz, E. *et al.* Candida albicans internalization by host cells is mediated by a clathrin-dependent mechanism. *Cell. Microbiol.* **11**, 1179–1189 (2009).

178. Adam L. Haber, Moshe Biton, Noga Rogel, Rebecca H. Herbst, Karthik Shekhar, Christopher Smillie, Grace Burgin, Toni M. Delorey, Michael R. Howitt, Yarden Katz, Semir Beyaz, Danielle Dionne, Mei Zhang, Raktima Raychowdhury, Wendy Garrett, Hai Ning, A. R. A comprehensive single-cell atlas of the small intestinal epithelium during

homeostasis and pathogenic infection. *Nature* **In Review**, (2016).

179. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

180. Vyas, V. K., Barrasa, M. I. & Fink, G. R. A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families. (2015).

181. Halder, V., Porter, C. B. M., Chavez, A. & Shapiro, R. S. Design, execution, and analysis of CRISPR–Cas9-based deletions and genetic interaction networks in the fungal pathogen Candida albicans. *Nat. Protoc.* 1 (2019). doi:10.1038/s41596-018-0122-6

182. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–6 (2011).

183. Halder, V., Porter, C. B. M., Chavez, A. & Shapiro, R. S. Design, execution, and analysis of CRISPR–Cas9-based deletions and genetic interaction networks in the fungal pathogen Candida albicans. *Nat. Protoc. 2019* 1 (2019). doi:10.1038/s41596-018-0122-6

184. Allaker, R. P., Zihni, C. & Kapas, S. An investigation into the antimicrobial effects of adrenomedullin on members of the skin, oral, respiratory tract and gut microflora. *FEMS Immunol. Med. Microbiol.* **23**, 289–293 (1999).

185. Kapas, S. *et al.* Adrenomedullin expression in pathogen-challenged oral epithelial cells. *Peptides* **22**, 1485–1489 (2001).

186. Arijs, I. *et al.* Mucosal Gene Expression of Antimicrobial Peptides in Inflammatory Bowel Disease Before and After First Infliximab Treatment. *PLoS One* **4**, e7984 (2009).

187. Klein, D. *et al.* Wnt2 acts as a cell type-specific, autocrine growth factor in rat hepatic sinusoidal endothelial cells cross-stimulating the VEGF pathway. *Hepatology* **47**, 1018–1031 (2008).

188. Suzuki, A., Sekiya, S., Gunshima, E., Fujii, S. & Taniguchi, H. EGF signaling activates

proliferation and blocks apoptosis of mouse and human intestinal stem/progenitor cells in long-term monolayer cell culture. *Lab. Investig.* **90**, 1425–1436 (2010).

189. Zhang, D. *et al.* Human intestinal organoids express histo-blood group antigens, bind norovirus VLPs, and support limited norovirus replication. *Sci. Rep.* **7**, 12621 (2017).

190. Neff, M. M., Neff, J. D., Chory, J. & Pepper, A. E. dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. *Plant J.* **14**, 387–392 (1998).

191. Carabotti, M., Scirocco, A., Maselli, M. A. & Severi, C. The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. *Ann. Gastroenterol.* **28**, 203–209 (2015).

192. Clarke, G. *et al.* The microbiome-gut-brain axis during early life regulates the hippocampal serotonergic system in a sex-dependent manner. *Mol. Psychiatry* **18**, 666–673 (2013).

193. Hyland, N. P. & Cryan, J. F. A Gut Feeling about GABA: Focus on GABA(B) Receptors. *Front. Pharmacol.* **1**, 124 (2010).

194. Reyes-García, M. G., Hernández-Hernández, F. & García-Tamayo, F. Gamma-aminobutyric acid (GABA) increases in vitro germ-tube formation and phospholipase B1 mRNA expression in Candida albicans. *Mycoscience* **53**, 36–39 (2012).

195. Menendez, A. *et al.* Bacterial Stimulation of the TLR-MyD88 Pathway Modulates the Homeostatic Expression of Ileal Paneth Cell α-Defensins. *J. Innate Immun.* **5**, 39–49 (2013).

196. Noble, C. L. *et al.* Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis. *Gut* **57**, 1398–405 (2008).

197. Schlieve, C. R. *et al.* Vascular Endothelial Growth Factor (VEGF) Bioavailability Regulates Angiogenesis and Intestinal Stem and Progenitor Cell Proliferation during Postnatal Small

Intestinal Development. *PLoS One* **11**, e0151396 (2016).

198. Sheng, G., dos Reis, M. & Stern, C. D. Churchill, a zinc finger transcriptional activator, regulates the transition between gastrulation and neurulation. *Cell* **115**, 603–13 (2003).

199. Goyal, R. K. & Hirano, I. The Enteric Nervous System. *N. Engl. J. Med.* **334**, 1106–1115 (1996).

200. Furness, J. B., Callaghan, B. P., Rivera, L. R. & Cho, H.-J. The Enteric Nervous System and Gastrointestinal Innervation: Integrated Local and Central Control. in 39–71 (Springer, New York, NY, 2014). doi:10.1007/978-1-4939-0897-4_3

201. Pisa, D., Alonso, R., Rábano, A., Rodal, I. & Carrasco, L. Different Brain Regions are Infected with Fungi in Alzheimer's Disease. *Sci. Rep.* **5**, 15015 (2015).

202. Alonso, R. *et al.* Fungal Infection in Patients with Alzheimer's Disease. *J. Alzheimer's Dis.* **41**, 301–311 (2014).

203. Sotgiu, S., Musumeci, S., Marconi, S., Gini, B. & Bonetti, B. Different content of chitin-like polysaccharides in multiple sclerosis and Alzheimer's disease brains. *J. Neuroimmunol.* **197**, 70–73 (2008).

204. Castellani, R. *et al.* Chitin-like Polysaccharides in Alzheimers Disease Brains. *Curr. Alzheimer Res.* **2**, 419–423 (2005).

205. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

206. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).

207. Hoare, J. How t-SNE works and Dimensionality Reduction | Displayr. Available at: https://www.displayr.com/using-t-sne-to-visualize-data-before-prediction/. (Accessed:

13th March 2019)

208. Walther, A. & Wendland, J. An improved transformation protocol for the human fungal pathogen Candida albicans. *Curr. Genet.* **42**, 339–343 (2003).

209. Taylor, P. R. *et al.* Dectin-2 is predominantly myeloid restricted and exhibits unique activation-dependent expression on maturing inflammatory monocytes elicitedin vivo. *Eur. J. Immunol.* **35**, 2163–2174 (2005).

210. Drummond, R. A., Saijo, S., Iwakura, Y. & Brown, G. D. The role of Syk/CARD9 coupled C-type lectins in antifungal immunity. *Eur. J. Immunol.* **41**, 276–281 (2011).

Appendices

Appendix A Characterization of heterogenous subpopulations ex vivo during macrophage and *Candida albicans* interactions

Macrophage and *Candida albicans* interactions are heterogenous, even among macrophages from the same genetic background and clonal fungal cells. We characterized macrophage and *Candida albicans* interaction types *in vitro*, after 20 minutes and 4 hours after exposure. Images were obtained under 40x magnification, in one field of view. Four separate wells were analyzed for each condition (**Figure A.0.1**).

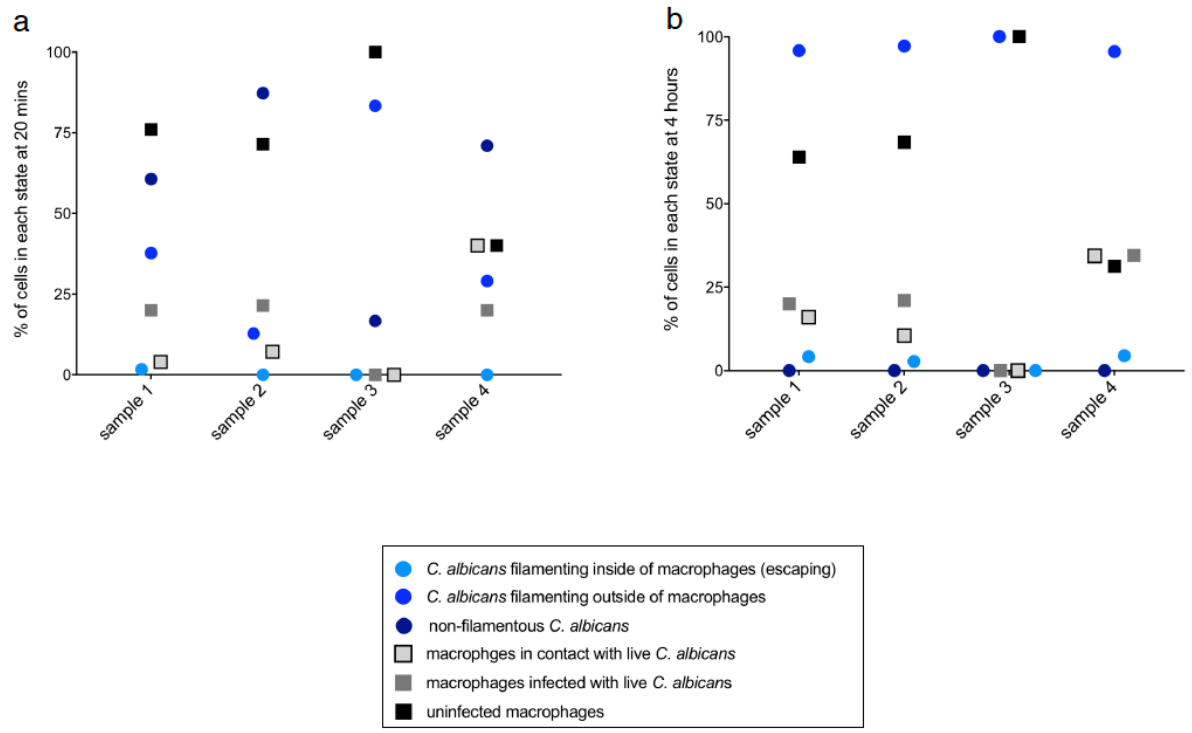

Figure A.0.1: The percentage of *Candida albicans* and macrophage interaction types after 20 minutes and 4 hours of interaction
Murine macrophages (RAW 264.7) and *Candida albicans* interaction types *in vitro*, after 20 minutes (**a**) and 4 hours after exposure (**b**). Images were obtained under 40x magnification, in one field of view. One field of view in 4 separate wells were manually scored.

The percentage of the cells in one field of view in each well were diverse at 20 minutes and 4 hours; when these interactions were then averaged over the four wells, the standard error of the mean (SEM) was large for most interaction types (**Figure A.0.2**).

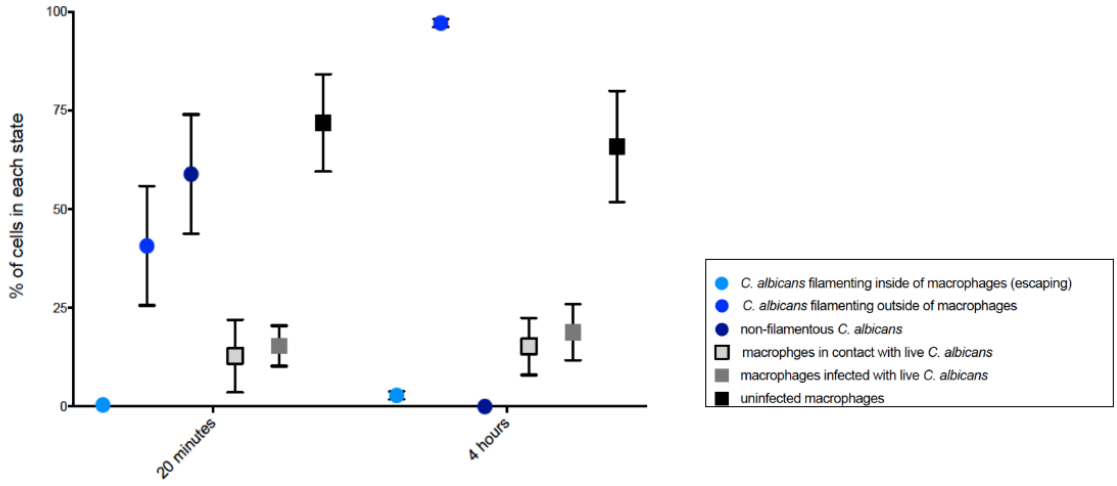

Figure A.0.2: The average of the percentage of *Candida albicans* and macrophage interaction types after 20 minutes and 4 hours of interaction
The average and SEM of the percentage of macrophage and *Candida albicans* interaction types shown in **Figure 2.1** are reported.

These heterogenous cell interaction fates were also observed in microwells, which contained between 2 to 10 cells per well (**Figure A.0.3**). In **Figure A.3A**, a GFP expressing *C. albicans* cells is phagocytosed by macrophages (stained with Deep Red) and is effectively killed (loss of GFP signal) and in **Figure A.0.3B**, *C. albicans* is phagocytosed by macrophage but are still able to form hyphae and can penetrate the macrophage cell wall.

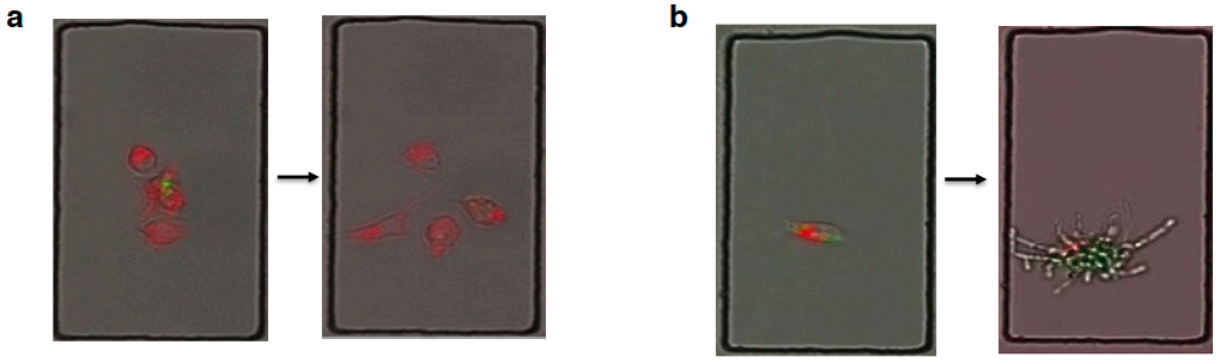

Figure A.0.3: Diverse infection fates during *Candida albicans* and macrophage interactions
Image taken by Dr. Christopher Ford, Broad Institute. Stained cells from a mouse macrophage cell line (red) and a clonal, GFP-expressing *Candida albicans* strain (green) were imaged in microwells over 9 hours. **(a)** In some cases, macrophages are able to engulf and effectively kill *C. albicans* (loss of green signal) while in other cases, **(b)** macrophages are unable to control *C. albicans* growth. Images were obtained under 40x magnification, in a chamber that was incubated on the microscope at 37°C and 5% CO_2.

These results indicate that macrophage and *C. albicans* interactions are highly variable within an infection population.

Appendix B *Candida albicans* sap mutant survival after exposure to mammalian macrophages at early time points

B.1 *C. albicans* Sap mutant and mammalian macrophage survival Results

Upregulation of the *SAP5* gene was observed in unexposed, exposed and phagocytosed *C. albicans* over the length of the RNA-seq time course, with the largest upregulation in phagocytosed C. *albicans* **(Figure B.0.4)**.

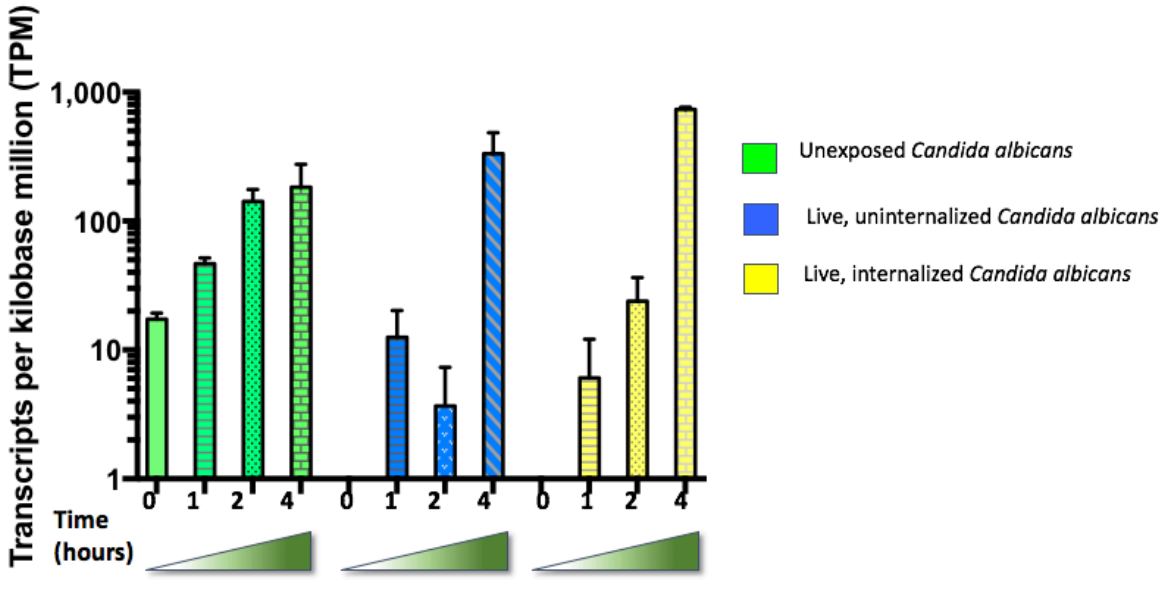

Figure B.0.4: *SAP5* is most highly expressed in populations of live, phagocytosed *Candida albicans*
Normalized expression (TPM) of *SAP5* among unexposed, macrophage exposed and phagocytosed live, *C. albicans* at 0,1,2,4 hours. Sample mean (n=3) and SEM are reported.

We studied the functional consequences of individual and triple *SAP* gene disruption mutants (denoted as sap mutants) on *C. albicans* survival after exposure to two murine macrophage cell lines (C57BL/6 and RAW 264.7) and in human unpolarized (M0), polarized M1 and M2 macrophages (**Figure B.0.5**; **Figure B.0.6)**. Because Sap proteins are proteases that can break down host cells, and an increase in *SAP5* expression in live *C. albicans* phagocytosed by macrophages was observed, it was hypothesized a sap5 mutant would show decreased survival after exposure to mouse and human macrophages at 2, 4, and 8 hours; 8 hours is twice as long as our last RNA-sequencing time point. There was no statistical difference in wild type (SC5314) or sap5 or sap4 mutant *C. albicans* survival after exposure to mouse macrophage (MØ) cell lines (**B.0.5**; p-value=0.1556; two-way ANOVA, Dunnett's multiple comparisons test) or when exposed to human unpolarized M0, or M1 or M2 polarized macrophages for 4 hours. There was a statistically significant increase in survival of sap5 mutants after exposure to human M2 macrophages compared to wild type *C. albicans* (SC5314) at 4 hours.

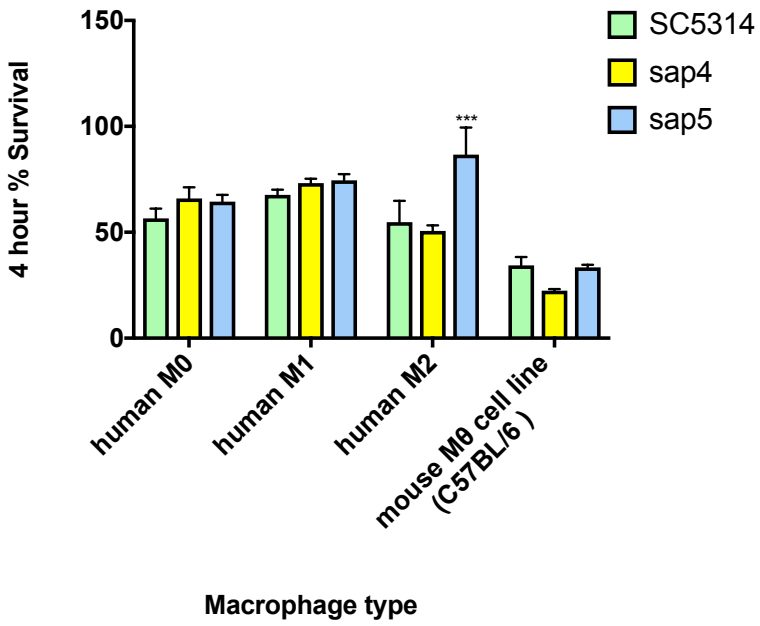

Figure B.0.5: *Candida albicans* survival after 4 hours of exposure to human, unpolarized (M0), M1 or M2 polarized macrophages and mouse macrophage cell line (RAW264.7).
Percent survival (% Survival) is calculated relative to the number of *C. albicans* only control colony numbers for each strain at 4 hours. MOI 15:1. The mean of 3 biological replicates and SEM reported. Significance was calculated by comparing the mean percent survival for sap4 and sap5 mutants to wild type (SC5314) mean survival after exposure to each macrophage cell type; two-way ANOVA, Dunnett's multiple comparison test; ***= $p < 0.0005$.

We also tested *C. albicans* survival after exposure to human macrophages at 8 hours (**Figure B.0.6**). These mutants were found to survive *more* when exposed to human macrophages than when growing in isolation, in some cases; any bar that appears above the hashed gray line in **Figure B.0.6** indicates that *Candida* survived or replicated more after macrophages exposure than when growing alone in macrophage media. Wild type *C. albicans* (SC5314) also grew more after human M1 exposure than when growing alone.

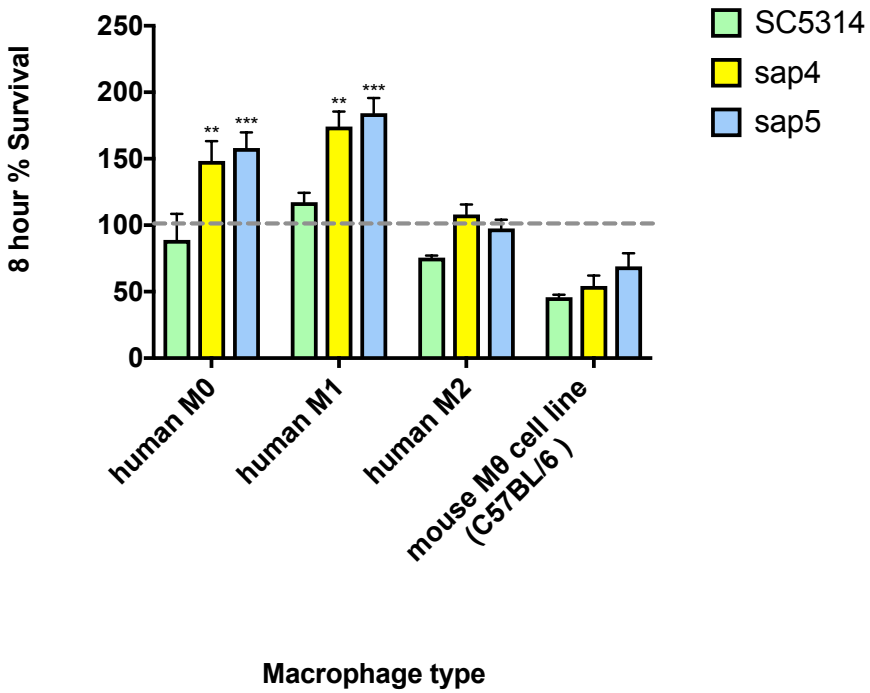

Figure B.0.6: *Candida albicans* survival after 8 hours of exposure to human, unpolarized (M0), M1 or M2 polarized macrophages.
Percent survival (% Survival) is calculated relative to the number of *C. albicans* only control colony numbers for each strain at 8 hours. MOI 15:1. The mean of 3 biological replicates and SEM are reported. The grey hashed line indicates 100% survival; survival greater than 100% indicates that the *Candida* strain survived more in the presence of macrophages than when growing alone. Significance calculated by comparing the % survival mean for sap4 and sap5 mutants to wild type (SC5314) mean survival in each macrophage cell type; two-way ANOVA, Dunnett's multiple comparison test; ***=p <0.0005, **=p <0.005, *=p <0.05.

It was therefore hypothesized that long sample processing time in this survival screen of sap mutants and human and mouse macrophages (96 samples; approximately 1.5 hours of processing time at each time point (**Figure B.0.5; Figure B.0.6**) may have led to the equal or increased survival of sap mutants when exposed to macrophages compared to wild type *C. albicans*. To determine if this was the case, an experiment with the same line of mouse macrophages (C57BL/6) was exposed to just one mutant, sap4,at the same MOI (15:1) for 8 hours (**Figure B.0.4**). We chose to test sap4 first, as this mutant tended to survive less than sap5 mutants in the human macrophage screen, but still more than wild type (**Figure B.0.5; Figure B.0.6**). We found no statistically

significant difference between sap4 mutant survival and wild type *C. albicans* survival (**Figure B.0.7**). These findings are in line with our results in the human macrophage screen (**Figure B.0.5; Figure B.0.6**), wherein sap4 mutants survive more than wild type *C. albicans* exposed to mouse macrophages but the difference between the two means was not significantly different.

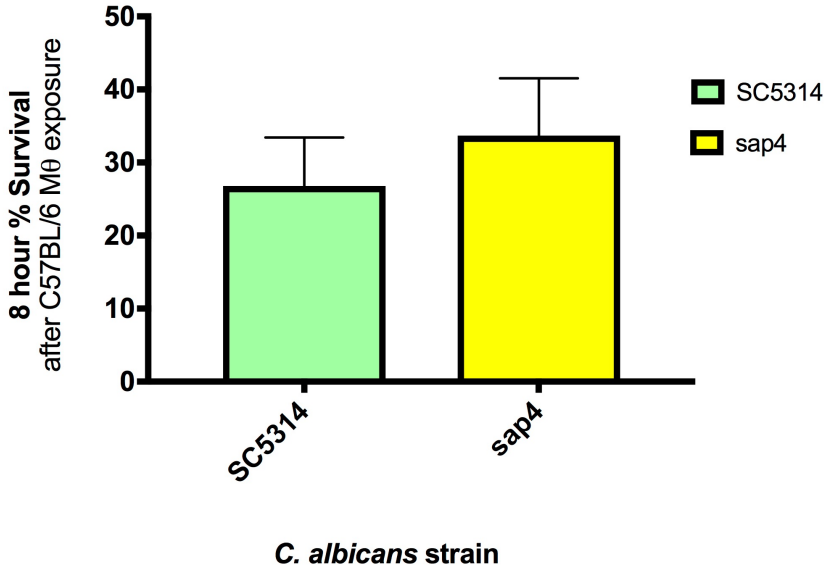

Figure B.0.7: *Candida albicans* sap4 mutant and wild type survival after 8 hours of exposure to macrophage cell line (C57BL/6).
Percent survival (% Survival) is calculated relative to the number of *C. albicans* only control colony numbers for each strain at 8 hours. MOI 15:1. The mean of 3 biological replicates and SEM are reported. No statistically significant difference between the means of wild type (SC5314; green) and sap4 mutant (yellow) survival after 8 hours of exposure; unpaired t test with Welch's correction, p=0.5730.

It was then hypothesized that the mouse macrophage cell line utilized in these studies may itself have accounted for the differences in sap mutant survival, as previous studies have reported that a sap4-6 *C. albicans* triple mutant is eliminated 53% percent more than wild type by the mouse macrophage line C57BL/6 (**Table1.2**). There was no statistically significant difference between the survival of the sap4-6 mutant and wild type *C. albicans* when exposed to the macrophage mouse cell line RAW264.7, even at an increased macrophage to *C. albicans* MOI of 1:1 (**Figure B.0.8**), the same MOI used in the *C. albicans* and macrophage RNA-sequencing experiment in which

SAP5 was found to be drastically upregulated in live *Candida* inside of macrophages at 1, 2 and 4 hours (**Figure B.0.4**).

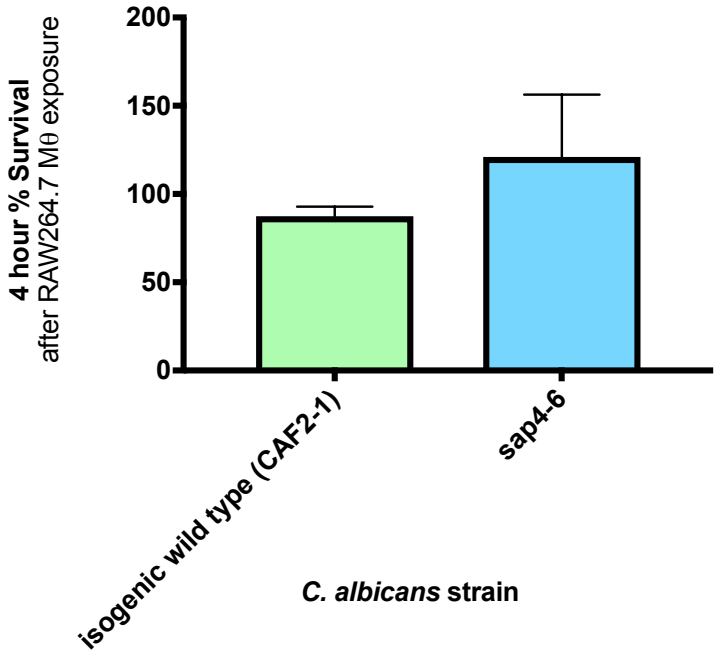

Figure B.0.8: *Candida albicans* sap4-6 mutant and wild type survival after 4hours of exposure to macrophage cell line (RAW264.7).
Percent survival (% Survival) is calculated relative to the number of *C. albicans* only control colony numbers for each strain at 4 hours. MOI 1:1. The mean of 2 biological replicates and SEM are reported. No statistically significant difference between wild type (CAF2-1; green) and sap4 mutant (blue) survival after 4 hours of exposure; unpaired t test with Welch's correction, p=0.5730

We next tested the ability of the single sap5 mutant to survive after exposure to mouse macrophages after 4 hours at an MOI of 1:1. Because *SAP5* was highly upregulated in live *Candida* inside of macrophages over 4 hours (**Figure B.0.4**), we hypothesized that a sap5 mutant would survive less after macrophage exposure compared to wild type. Instead of comparing the sap5 mutant survival to that of the conventional wild type strain, SC5314, we compared sap5 mutant survival to the wild type strain from which the mutant was derived (CAF2-1) (**Figure B.0.9**). In

accordance with our previous results, we observed no statistical difference between the mean survival rate of sap5 mutants after exposure to macrophages compared to wild type *C. albicans*.

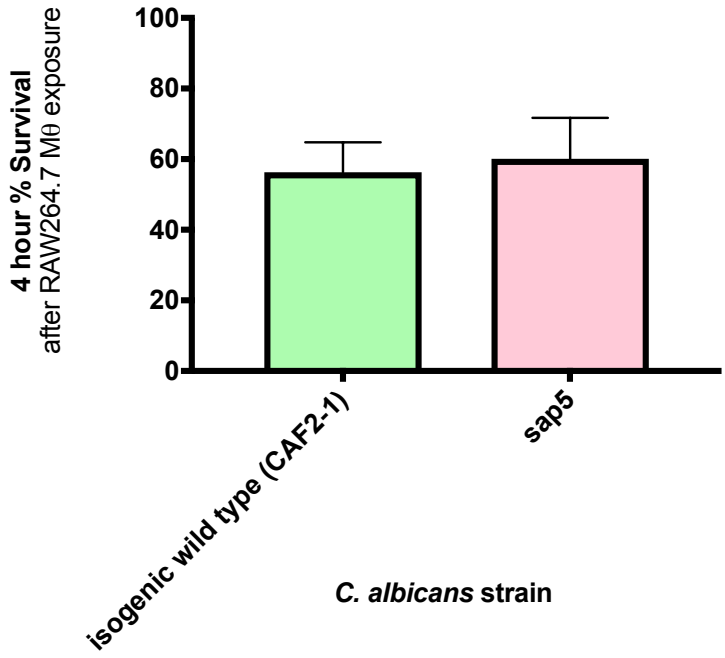

Figure B.0.9: *Candida albicans* sap5 mutant and wild type survival after 4 hours of exposure to a macrophage cell line (RAW264.7)
Percent survival (% Survival) is calculated relative to the number of *C. albicans* only control colony numbers for each strain at 4 hours. MOI 1:1. The mean of 9 biological replicates and SEM are reported. No statistically significant difference between mean wild type (green) and sap5 mutant (pink) survival after 4 hours of exposure; unpaired t test with Welch's correction, p=0.7926

We also analyzed the percentage of mouse macrophage (RAW264.7) survival after 2 hours and 24 hours of exposure to sap4-6 mutant and wild type *C. albicans* (SC5314) (**Figure B.0.10A,B**). Because secreted aspartyl proteases are able to break down host proteins, it was hypothesized that macrophages exposed to sap mutants would survive *more* than those exposed to wild type *C. albicans*. Relatively early and late time points were analyzed in this assay, as we hypothesized that a lack of Sap protease activity likely wouldn't have a significant effect on the survival of

macrophages at 2 hours. In contrast, we hypothesized that we would observe a large increase in macrophage survival after 24 hours of exposure to sap mutants, relative to the viability of those macrophages exposed to wild type *C. albicans*. At 2 hours, 39% of macrophages survived after exposure wild type *C. albicans*, compared to 75% of macrophages surviving after sap4-6 mutant exposure, and 65% of macrophages surviving after sap5 mutant exposure (**Figure B.0.10A**). However, the increase in macrophage survival when exposed to sap4 mutants versus wild type *C. albicans* was not statistically significant (p-value=0.1070; one-way ANOVA, Dunnett's multiple comparisons test). 87% of macrophages were alive in the macrophage only control, with the observed death being attributed to loss during the plating process (**Figure B.0.10A)**.

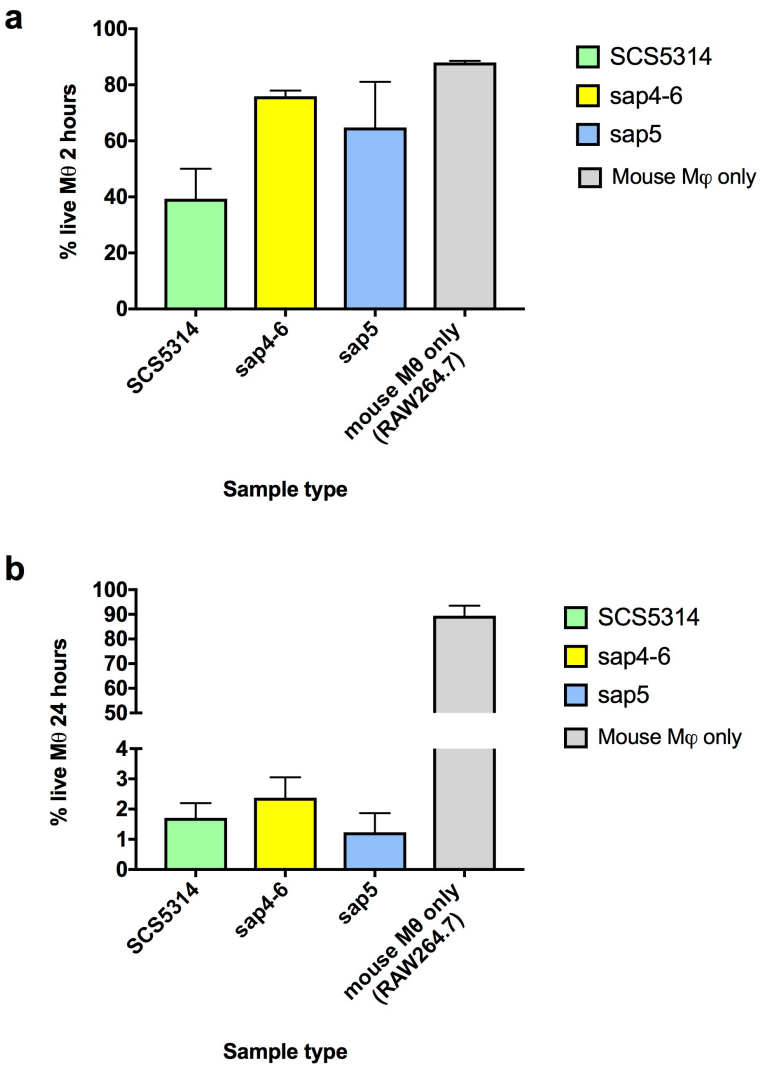

Figure B.0.10: Macrophage survival after exposure to wild type *Candida albicans* (SC5314*)* and *Candida albicans* sap mutants for 2 hours and 24 hours.
%Survival relative to the percentage of live macrophages in macrophage only controls at each time point. MOI 1:1. Mean of 3 biological replicates and SEM reported. No statistical difference between **(a)** macrophage survival after exposure to *C. albicans* sap4-6 mutants or sap5 mutants compared to wild type (SC5314) at 2 hours (p=0.1070 and p=0.2668, respectively) or **(b)** 24 hours (p=0.6711 and p=0.8077, respectively); one-way ANOVA, Dunnett's multiple comparison test.

After 24 hours, 90% of macrophages in the macrophage only control were viable, while only between 1.2-2.4% of *C. albicans* exposed macrophages were viable. There was no statistically significant difference in viability in those macrophages exposed to the sap4-6 mutant or the sap5

mutant compared to wild type *C. albicans* at 24 hours (p-value=0.6711 and 0.877, respectively; one-way ANOVA, Dunnett's multiple comparisons test) (**Figure B.0.10B**).

B.2 *C. albicans* Sap mutant and mammalian macrophage survival Discussion

When examining genes that were highly upregulated among live, *C. albicans* inside of macrophages, we observed high expression of secreted aspartyl protease 5 (*SAP5*). It is well documented in the literature that the 12 member *SAP* gene family plays a critical role in virulence and host cell break down, but that the role that these proteases play is host niche specific (**1.3.1 Secreted Aspartyl Proteinases (*SAP*) genes in *Candida albicans***). Because *SAP5* was specifically highly upregulated in phagocytosed, live *C. albicans*, we hypothesized that a *SAP5* mutant would not survive as well as wild type *C. albicans* when exposed to macrophages (despite overall *SAP* gene redundancy). While there is some evidence that this is true at an MOI of 15:1 (Rao lab, unpublished data), we did not observe a statistically significant decrease in *SAP5* mutant survival or triple *SAP4-6* mutant survival when these mutants were exposed to macrophage at an MOI of 1:1 used in the RNA-seq experiment (**Chapter 2**). We also observed in an increase in *SAP* mutant survival after exposure to human macrophages, relative to both wild type *C. albicans* and to *SAP* mutants growing alone in macrophage media. This finding could indicate that *SAP* mutants respond very differently to human versus mosue macrophages; further experimentation is required to verify this surprisng result. If these results are replicated, it is possible that lack of Sap secretion allows for *C. albicans* evasion from human macrophages in a specific way, relative to mouse macrophages.

B.3 Methods for *Candida albicans* survival assays with macrophages

Human THP-1 monocytes (ATCC) were seeded into tissue culture plates (one ml at 500,000 cells/ml) and 100 ng/ml of phorbol 12-myristate 13-acetate (PMA) was added to each well to induce macrophage differentiation. Cells were then incubated for 16-24 hours and then media was aspirated and replaced with non-PMA containing RPMI-1640 (ThermoFisher Scientific). M0 macrophages received no additional treatment. To polarize monocytes toward M1 macrophages, 20 ng/ml of IFN-gamma was added to each well and cells were allowed to incubated for 48 hours before a media change (RPMI only). To polarize monocytes toward M2 macrophages, 20 ng/ml of IL-4 was added to each well and cells were allowed to incubated for 48 hours before a media change (RPMI only). Two T75 flasks of human (unpolarized (M0), M1 or M2 polarized THP-1 monocytes) or murine macrophages (RAW 264.7 or C57BL/6 cell lines) were grown to 80–90% confluence in DMEM (Caisson labs) or RPMI 1640 + 10% FBS. Macrophage cells were allowed to adhere to the plate surface for at least 5 hours at 37°C, 5% CO_2. *C. albicans* overnight cultures were counted, resuspended in macrophage media and added into each well, MOI 15:1 or 1:1. *Candida albicans* only controls were also seeded into each plate. Each sample condition was run in biological duplicate or triplicate. Tissue culture plates were then incubated at 37°C, 5% CO_2 until the experimental time point. For sample collection, cells were scraped from each well, and resuspended in 10 ml 0.02% Triton-X 100 (v/v) to osmotically lyse macrophages. The solution was then spun down at 4,000 rpm for 5 minutes, and the pellet resuspended in 1 ml sterile water. Cells were diluted 10^{-2} and spread on rich media plates using sterile glass beads or a sterile cell spreader. After incubating the plates overnight at 30°C, yeast colonies were counted on each plate. The percent survival of *Candida albicans* was determined by dividing the number of colonies

generated from the *Candida albicans* and macrophage wells by the average number of colonies generated from the *Candida albicans* only controls.

B.4 Methods for macrophage survival assay with *Candida albicans*

Two T75 flasks of murine macrophages (RAW 264.7 cell line) were grown to 80–90% confluence in RPMI 1640 + 10% FBS. Macrophage cells were allowed to adhere to the plate surface for at least 5 hours at 37°C, 5% CO_2. *C. albicans* overnight cultures were counted, resuspended in macrophage media and added into each well containing macrophages at a 1:1 MOI or 15:1. Macrophage only controls were also seeded into each plate. Each sample condition was run in biological triplicate. Tissue culture plates were then incubated at 37°C, 5% CO_2 until the experimental time point. Media from each well was collected into a tube and 1.0mL TrypLE was added to each well and plates were incubated at 37°C incubator for 10 minutes. Macrophages were confirmed to be no longer adherent to the plate surface by visualizing cells under a microscope. Cells in each well were pipetted vigorously and collected into the sample tube. Wells were then washed with 500 ul warm (37°C) PBS and this was added to each tube. Tubes were spun down at 1000 rcf for 2 minutes at room temperature. Supernatant was removed with an aspirator and the pellet was resuspended in 100 ul of PBS +2% FBS. 50 nanograms of C11b antibody (anti-mouse/human CD11b Antibody Biolegend; cell surface marker for macrophages) was added to each sample and the tube was mixed gently by flicking. Samples were then incubated on ice and protected from light for 20 minutes. Tubes were then spun for 5 mins at 2200 rpm at 4°C to remove unconjugated AB. Supernatant was removed and cells were resuspended in 190 ul of PBS + 2% FBS. Add an additional 100 nanograms of C11 antibody to each sample on ice, protected from light. Transfer to a cold room and rotate the tubes in the dark for 14 minutes. Tubes were again

spun for 5 mins at 2200 rpm at 4°C and then resuspend in 500 ul of PBS +2% FBS. Cells were strained into a glass FACS tube (Corning; 35 μm filter). 10 ul of Propidium Iodide (ThermoFisher; stains dead cells) was added to each sample and samples were incubated on ice and protected from light for 5 minutes. Samples were then analyzed on the BD ACCURI C6 flow cytometer (BD Biosciences).