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

Abstract

The control of fungal infections depends on interactions with innate immune cells including macrophages, which are among the first host cell types to respond to pathogens such as *Candida albicans*. This fungus is a member of the healthy human microbiome, although also a devastating pathogen in immunocompromised individuals. Consistent with recent findings from studies of other pathogens, we observed that within a population of interacting macrophages and *C. albicans*, there are distinct host-pathogen subpopulations reflecting cell specific trajectories and infection outcomes. Little is known about the molecular mechanisms that control these different fates. To address this, we developed an experimental system to isolate the major host-fungal pathogen subpopulations observed during *ex vivo* infection using fluorescent markers. We separated subpopulations of macrophages infected with live *C. albicans* from uninfected cells and assessed the variability of gene expression in both host and fungal pathogen for each subpopulation across time using RNA-Seq. In infected cells, we observed a coordinated, time-dependent shift in gene expression for both host and fungus. The early response in macrophages was established upon exposure to *C. albicans* prior to engulfment and involved up-regulation of pathways and regulatory genes required for cell migration, pathogen recognition, activation of engulfment, and phagocytosis; this pro-inflammatory response declined during later time points in parallel with expression changes in *C. albicans*. After phagocytosis, the initial response of *C. albicans* was to up-regulate genes related to survival in the nutrient-limited and stressful environment within macrophages; at later time points, gene expression shifted to initiate hyphal growth and escape. To further probe the heterogeneity seen observed in host-pathogen interactions, we performed RNA-Seq of single macrophages infected with *C. albicans*. We observed that some genes show higher levels of heterogeneity in both host and fungal pathogen cells that we could not detect in subpopulation samples; we observed that the time shift in expression is asynchronous and that expression changes in both the host and pathogen are tightly coupled. This work highlights how

analysis of subpopulations and single host-pathogen pairs can resolve population heterogeneity and trace distinct trajectories during host interactions with fungal pathogens.

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 infection by fungi[1]. *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[2]. In addition, *C. albicans* can withstand harsh host environments, including in response to macrophage engulfment by regulating metabolic pathways and cell morphology[3,4]. 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 survive within and kill macrophages upon escape, some cells lyse in the phagosome and other cells may evade direct host cell interaction or phagocytosis[5].

Previous studies of *C. albicans* and immune cell interactions in bulk, non-sorted populations have identified key pathways by characterization of either the fungal or host transcriptional response during these interactions[3,6,7]. More recently, dual transcriptional profiling of host-fungal pathogen interactions have examined populations of cells, which measures the average the transcriptional signal of millions of cells which may include diverse infection fates[8–11]. In addition, even in a clonal population of phagocytes, many immune cells may not engulf any fungal cells, while others can phagocytose up to ten fungal cells[12]. The maturation of single-cell RNA sequencing technologies has demonstrated that substantial variation in gene expression between cells can be detected within stimulated or infected immune cell populations[13–15]. For example, single-cell RNA-seq revealed how variability in infection leads to expression bimodality of the interferon-response among single macrophages that were exposed to LPS or *Salmonella*[14]. A recent study measured both host and pathogen gene expression in parallel, in single host cells infected with the bacteria *S. typhimurium*; however, as sufficient coverage of pathogen transcriptomes was difficult to recover, pathogen transcriptional data was pooled prior to analysis[16]. 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 parallel, in sorted subpopulations and in single, infected macrophages. We have focused on four distinct infection outcomes: (i) macrophages infected with live *C. albicans*, (ii) macrophages infected with dead *C. albicans*, (iii) macrophages exposed to *C. albicans* that remained uninfected and (iv) *C. albicans* exposed to macrophages that remained un-engulfed. As mixed populations of cells show phenotypic heterogeneity over the course of *in vitro* infection experiments, we hypothesized that we would observe transcriptional variation in genes important for these interactions among single, macrophages infected with *C. albicans*. Additionally, we hypothesized that transcriptional variability among single host cells may be correlated with gene expression differences in phagocytosed *C. albicans.* Here, we isolated single macrophages infected with *C. albicans* and adapted methods to measure gene expression of the host and pathogen cells in parallel. Sorting infection subpopulations prior to RNA-sequencing resolved the signal from different infection fates within a population, especially in phagocytosed *C. albicans*. By comparing heterogeneity in the transcriptional profiles of *C. albicans* and murine macrophages at both the subpopulation and single infected cell levels, we characterized how gene expression varies in close coordination between the host and pathogen across distinct infection fates. We found cell-to-cell variability when we analyzed single infected macrophages, and that key genes involved in host immune response and in fungal morphology and adaptation show expression bimodality that can be resolved as tightly coupled time-dependent transcriptional responses between the host and fungal pathogen.

Results

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

To capture heterogeneous infection subpopulations and examine parallel shifts in host and pathogen interactions, we developed a system for fluorescent sorting of both live and dead *Candida albicans* with macrophages during phagocytosis (**Methods**). We constructed a *C. albicans* strain that constitutively expresses Green Fluorescent Protein (GFP) and mCherry; when *C. albicans* cells lyse in the acidic macrophage phagosome, GFP loses fluorescence upon change in pH[15] whereas mCherry remains stable for up to 4 hours in this environment as visualized by microscopy (**Figure S1**). We confirmed that the reporter construct was present at *NEUT5* locus via whole genome sequencing of the SC5314-*NEUT5L-NAT1-mCherry-GFP* strain (**Methods**). This *C. albicans* reporter strain was co-incubated with

103 macrophages and sorted using fluorescent activated cell sorting (FACS) at time intervals (0,1, 2 and 4
104 hours) to isolate different host-fungal pathogen subpopulations and single-cell pairs (**Figure 1A**). These
105 time points were selected to capture the rapid transcriptional changes of *C. albicans* in response to
106 macrophages[3]. Primary bone derived macrophages were stained with CellMask Deep Red plasma
107 membrane stain prior to sorting. To examine gene expression profiles, both host and fungal RNA were
108 extracted and adapted for Illumina sequencing using Smart-Seq2 (**Methods**). Both microscopy and
109 FACS analysis revealed that four distinct infection subpopulations can be isolated (**Figure S2A**): (i)
110 macrophages infected with live *C. albicans* (GFP+, mCherry+, Deep red+), (ii) macrophages infected
111 with dead *C. albicans* (GFP-, mCherry+, Deep red+), (iii) macrophages exposed to *C. albicans* (GFP-,
112 mCherry-, Deep red+) and (iv) *C. albicans* exposed to macrophages (GFP+, mCherry+, Deep red-;
113 **Figure 1A**).
114
115 The number of RNA-Seq reads and transcripts detected for both host and fungal pathogen
116 subpopulations was sufficient for differential expression analysis in most samples and to widely profile
117 parallel transcriptional responses. We aligned reads to a composite reference of both mouse and *C.*
118 *albicans* transcriptomes (**Methods**) and found that the fraction of mapped reads for host and pathogen
119 was correlated with the percent of sorted cells for each subpopulation (**Figures 1B**, **S2B**; **Table S1**). In
120 subpopulations containing both host and pathogen, the fraction of reads averaged 87% macrophage
121 and 13% *C. albicans* for macrophages infected with live fungal cells, and 95% macrophage and 5% *C.*
122 *albicans* for macrophages containing dead fungal cells. For the subpopulation of macrophages infected
123 with live fungal cells, between 1.4 and 34.0 million reads mapped to host transcriptome, while between
124 0.3 to 13.3 million reads mapped to *C. albicans* transcriptome. *Candida* reads in this subpopulation
125 increased over time, reflecting ongoing phagocytic activity during the time course (0,1, 2 and 4 hours;
126 **Figure 1B**). In subpopulations of dead phagocytosed cells, read counts were lower than other
127 subpopulations and counts increased over time, reflecting their smaller proportion of sorted cells that
128 also increased over time (**Figure S2B**). In subpopulations of macrophages infected with live fungus, an
129 average of 10,333 host and 4,567 *C. albicans* genes were detected (at least 1 fragment per replicate
130 across all samples; **Methods**; **Figure S3A**; **Table S1**). Fewer transcripts were detected in
131 subpopulations of macrophages infected with dead *C. albicans* (an average of 3,214 host transcripts
132 and 983 fungal transcripts **Figure S3A**) and had modestly correlated biological replicates (*e.g.*
133 Pearson's $r < 0.56$). Lower coverage was expected for this subpopulation, as only up to 3% of each
134 sample collected at each time point was comprised of macrophages infected with dead *C. albicans*
135 (**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.

Next, we examined the major expression profiles in both the host and fungal pathogen. We identified 1,218 and 2,226 differentially expressed genes (DEGs; fold change (FC) > 4; false discovery rate (FDR) < 0.001) among all pairwise comparisons in *C. albicans* and macrophages, respectively (**Methods**; **Figure 2**; **Data set 1** and **2**). To determine major patterns of infection-fate specific or interaction-time specific in *C. albicans* and macrophages, 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 of both macrophages and *C. albicans* could be more finely characterized using sorted subpopulations and how this response varied over time. For both host and pathogen, PC1 separates the samples by subpopulation whereas PC2 separates the samples by time (**Figures 2A**, **2C**). Examining sorted infection subpopulations revealed a close relationship between the infected and exposed macrophage subpopulations at the same time point where as distinct *C. albicans* subpopulations appeared more separate as described below. This highlights how cell sorting can resolve the signal from different infection fates within a population.

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 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 2B**, **2E**).

Genes highly induced in phagocytosed *C. albicans* are involved in adaptation to the macrophage environment. This highlights major changes in metabolic pathways; these genes (cluster 1*;* **Figure 2B**, **2E**) 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 *C. albicans* exposed to macrophages reported that similar changes in metabolism and nutrient uptake allow *Candida* to utilize the limited spectrum of nutrients available in the phagosome[3]. With RNA-Seq data and sorted infection fates, we observed upregulation of additional genes related to these functions and confirmed which expression changes were specific to the phagocytosed *C. albicans* subpopulation (**Table S2**). Genes involved in glyoxylate metabolism, the beta-oxidation cycle and transmembrane transport were significantly induced in phagocytosed *C. albicans* relative to exposed cells. By contrast, multiple classes of transporters were highly up-regulated 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 2B**; **Tables S2** and **S3**). Cluster 1 includes genes involved in pathogenesis and genes associated with the formation of hyphae, including core filamentous response genes (*ALS3*, *ECE1*, *HTG2*, *ORF19.2457*; **Table S2**)[17]. Overall, gene expression levels in cluster 1, including filamentation genes, increased over the time course with the highest expression at 4 hours, regardless of whether *C. albicans* cells were phagocytosed or remained un-engulfed, suggesting that these genes were induced by the presence of macrophages. While media containing serum can also induce *C. albicans* filamentation, we found that those genes were more highly induced upon phagocytosis (**Table S2**; **Figure S4B**). We also identified genes induced in live, phagocytosed *C. albicans* and repressed in un-engulfed cells relative to the unexposed subpopulation (cluster 2; **Figure 2B**). Cluster 2 includes several secreted aspartyl proteases and additional genes involved in transmembrane transport (Opt and Hgt classes; **Tables S2** and **S3**). These transporters differ with those found in cluster 1, as the scale of their induction after phagocytosis was more modest. Other sets of genes up-regulated during *C. albicans* phagocytosis in cluster 3 had lower induction relative to cluster 1 or cluster 2. This set included genes related to oxidation-reduction processes, including dehydrogenases, mitochondrial respiratory response, and transcription factors (enriched GO terms, corrected-$P < 0.05$, hypergeometric distribution with

Bonferroni correction; **Figure 2B**; **Table S3**). Cluster 1, 2, and 3 encompassed the major transcriptional modules up-regulated in *C. albicans* upon phagocytosis.

Other sets of genes were specifically down-regulated in live, phagocytosed *C. albicans* at the earliest time point, whereas expression levels of these genes did not change in non-phagocytosed *C. albicans* (both unexposed and exposed) over the time course (cluster 4 and 5; **Figures 2B**, **2E**). In cluster 4, which displays the strongest signature of repression, genes down-regulated in phagocytosed *C. albicans* recovered their expression levels at 4 hours. This cluster includes 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 S4B**). Repression of the translation machinery was previously noted as a response of *C. albicans* to macrophage interaction[3]. Here, we demonstrated that down-regulation of ribosomal proteins, chaperones and translation-regulator transcription factors is specific to phagocytosed *C. albicans* and that expression of these genes recovered at later time points (**Figures 2E**, **S4B**). Cluster 4 also encompassed highly repressed yeast-phase specific genes, including those involved in ergosterol biosynthesis, cell growth and cell wall synthesis (**Figures 2B**, **S4B**; **Table S2**). Cluster 5, which was not as strongly repressed as cluster 4, includes a set of genes down-regulated in phagocytosed *C. albicans* that did not recover expression over the length of the time course (**Table S2**). These genes are largely involved in nucleoside metabolic processes and host adaptation (enriched GO terms, corrected-$P < 0.05$, hypergeometric distribution with Bonferroni correction; **Figure 2B**; **Table S3**). A subset of these down-regulated genes were involved in morphological and cell surface remodeling, including three essential negative regulators of filamentation (*SSN6*, *MFG1*, and *TUP1*)[18–21]. This subset also includes additional repressors of filamentation[22] and transcription factors that regulate the white-opaque phenotypic switch (*WOR1* and *WOR2*)[23,24] and commensalism in the mammalian gut (*WOR1*)[25] (**Figures 2B**, **S4**). These results highlight that a large part of the observed transcriptional variation is likely due to the down-regulation of these transcriptional regulators.

Together, the expression patterns in the phagocytosed *C. albicans* subpopulation suggest a rapid shift in gluconeogenic metabolism, amino acids uptake, reorganization of the cell wall and initiation of a hyphal program. In particular, the transcriptional repression program triggered by phagocytosis provides new insight into the time-specific adaptation of *C. albicans* phagocytosed by macrophages. These gene expression patterns reflect changes in *C. albicans* metabolism and growth from within

macrophages. While some of these results recapitulated those from non-sorted, bulk transcriptional studies of *C. albicans* and phagocytic cells[3,10,26,27], analysis of sorted subpopulations differentiated gene expression changes that are specific to *C. albicans* phagocytosis from those that may reflect only exposure to immune cells.

Subpopulations of macrophages showed major pathogen recognition and pro-inflammatory response to *C. 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 2,226 DEGs (FC > 4; FDR < 0.001; **Data set 2**), which grouped into four clusters with similar expression patterns. PC analysis showed that exposed and infected macrophages clustered closely together, separately from unexposed macrophages along the PC1 and PC2, indicating that exposed and infected macrophages had similar transcriptional responses, which primarily varied across time in the first and second PCs (cluster 1 and 2; **Figure 2C**). This suggests that the early (1 to 4 hours) host transcriptional response to *C. albicans* is set during exposure and maintained during phagocytosis. For both exposed and infected macrophages, a major difference was found between 2 and 4 hours along PC2, which highlights genes that were differentially induced or repressed at 4 hours in these subpopulations (clusters 3 and 4; **Figure 2C**,**D**).

In all four clusters, infected and exposed macrophages have similar patterns of differential expression relative to unexposed macrophages (**Figure 2D**). Overall, these clusters were significantly enriched in activation, migration, phagocytosis, 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; **Figures 2D**, **S5**). Activation of some of these pathways is consistent with previous analysis of phagocyte transcriptional responses to *C. albicans* infection (**Figure S6**)[8–10,28] however, sorting distinct infection subpopulations during early time points of infection established that host cells up or down regulate subsets of genes upon *C. albicans* exposure and these expression patterns are largely maintained after *C. albicans* phagocytosis (**Figures 2D**, **S5**).

In exposed and infected macrophages, many of the genes induced at 1 hour remained relatively stable at 2 and 4 hours (cluster 1; **Figure 2D**). These genes are related to defense mechanisms such as pro-

inflammatory cytokine production and fungal recognition via transmembrane receptors. Up-regulated genes related to pro-inflammatory cytokines included tumor necrosis factor (*Tnf*), interleukin 1 receptor antagonist (*Il1rn*), and chemokines (*Ccl3*, *Cxcl14*, *Cxcl2*; **Figures 2D**, **S7**; **Table S4**). Other genes in cluster 1 include the transmembrane receptors intercellular adhesion molecule 1 (*Icam1*), interleukin 11 receptor subunit alpha (*Il11ra1*), macrophage scavenger receptor 1 (*Msr1*), oxidized low density lipoprotein receptor 1 (*Olr1*), toll like receptor 2 (*Tlr2*), and the interferon regulatory factor 1 (*Ifr1*; **Figures 2D**, **2F**, **S7**; **Table S4**). The chemokine *Cxcl2* and the intracellular receptor *Icam1* have also been previously shown to be induced during *C. albicans* interactions with other host cells, including neutrophils *in vitro*[10], murine kidney[8], a murine vaginal model[28], and mouse models of hematogenously disseminated candidiasis and of vulvovaginal candidiasis in humans[29], highlighting the role of these genes in host defense against *C. albicans* infection of different tissues (**Figure S6**).

A second set of genes was up-regulated in both exposed and infected macrophages, with peak expression at 2 hours in exposed macrophages (cluster 2; **Figure 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 *Clec7a* (also known as *Dectin-1*), lectin-like receptors such as galectin 1 (*Lgals1*) and galectin 3 (*Lgals3*; **Figure 2D**; **Tables S4** and **S5**), and other transmembrane receptors (*Cd36*, *Cd74*, *Fcer1g*, *Ifnar2*, *Igsf6*, *Itgb2*, *Gm14548*, *Trem2*; **Table S4** and **S5**). In addition, several complement proteins were found more highly induced in exposed macrophages, including complement factor properdin (*Cfp*), extracellular complement proteins (*C1qa*, *C1qb*, and *C1qc*), and complement receptor proteins (*C3ar1* and *C5ar1*). These results highlight the importance of *C. albicans* recognition and activation of engulfment in the exposed macrophage subpopulation. Since these genes maintained high expression in infected macrophages, they may also play an important role during phagocytosis or allow for uptake of additional *C. albicans* cells.

We also examined subpopulations of macrophages infected with dead *C. albicans;* this data was analyzed separately, as the total number of cells sorted and therefore the transcriptome coverage was low (*i.e.* 31% relative to macrophages infected with live *C. albicans*; see above). While we did not have sufficient data to examine *C. albicans* transcripts (*i.e.* 21% transcriptome coverage relative to exposed and live-phagocytosed subpopulations; see above), our analysis of the host transcripts revealed a small set of highly induced genes. This set includes pro-inflammatory cytokines such as *Ccl3*, *Cxcl2*, *Cxcl14*, *Il1rn*, and *Tnf*, and transcription regulators such as *Cebpb*, *Irf8*, and *Nfkbia* (**Figure S8**). These genes were also induced in macrophages infected with live *C. albicans* (clusters 1 and 2; **Figure 2D**),

299 indicating that maintaining expression of these genes may be important for pathogen clearance after
300 phagocytosis.
301
302 Another major shift in gene expression occurred at 4 hours, with sets of genes highly repressed or
303 highly induced at this later time point (cluster 3 and 4, respectively; **Figures 2D**, **2F**) in both the
304 exposed and infected macrophage subpopulations. Highly repressed genes at 4 hours (cluster 3) are
305 enriched in cytokines (*Il1a*, *Irf4*) and transmembrane receptors, including intracellular toll-like receptors
306 (*Tlr5* and *Tlr9*), and interleukin receptors (*Il1r1*, *Irf4*, *Il17ra*). This cluster was also enriched in categories
307 associated with proliferation and immune cell differentiation (*P*-value < 0.05 right-tailed Fisher's Exact
308 Test; **Table S5**). By contrast, macrophage genes highly induced at 4 hours (cluster 4) were enriched in
309 categories related with phagocytosis and programmed cell death mechanisms, including the chemokine
310 *Cxcl10*, and transcriptional regulators that play a role in inflammation and programmed cell death
311 (*Cebpd, Fos, Irf2bp1, Irf8, Irf9, Nfkbie, Card9, Bcl2*; (*P*-value < 0.05 right-tailed Fisher's Exact Test;
312 **Table S5**). This suggests that during phagocytosis of *C. albicans* there was a strong shift toward a
313 weaker pro-inflammatory transcriptional response by 4 hours. Our approach not only identified genes
314 specifically induced in infected macrophages, but also highlighted expression transitions related to time
315 of *C. albicans* exposure. These patterns were conserved among subpopulations of exposed and
316 infected macrophages and corresponded with shifts in pathogen gene expression (**Figures 2A**, **2B**; see
317 below).
318
319 **Detection of host-pathogen transcriptional responses from single macrophages infected with**
320 ***C. albicans***
321
322 Even in sorted populations, individual cells may not have uniform expression patterns, as they can
323 follow different trajectories over time. To address this, we next examined the level of single cell
324 transcription variability during infection. We collected sorted, single macrophages infected with live or
325 dead *C. albicans*, at 2 and 4 hours, and adapted the RNA of both the host and pathogen for Illumina
326 sequencing using Smart-seq2 (**Figure 1A**; **Methods**). With this approach, each infected macrophage
327 and the corresponding phagocytosed *C. albicans* received the same sample barcode, allowing us to
328 pair transcriptional information for host and pathogen at the single, infected cell level. While we can
329 successfully isolate single infected macrophages via FACS, we cannot control for the number of *C.*
330 *albicans* cells inside of each macrophage with this approach, since macrophages phagocytose variable
331 numbers of *C. albicans* cells[12] (**Methods**). 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, over 99% of which passed our sequence quality-control filters (**Figure S9A**; **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 3A**). Although parallel sequencing of host and pathogen decreases the sensitivity to detect both transcriptomes, the number of transcripts recovered (> 1 Transcripts Per Million, TPM) for macrophages and *C. albicans* was sufficient for cell clustering and differential expression (see below). 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 (3,904 on average), and 162 (72.3%) had at least 600 *C. albicans* transcripts detected (1,435 on average; **Figures 3A**, **S9A**). 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 subpopulations of macrophages infected with live *C. albicans*, this represents a single-infected-cell transcriptome coverage sensitivity of 38% and 31% for host and *C. albicans* transcripts respectively, indicating that we obtained sufficient 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 3B**) was reduced when we aggregated the expression of 32 single-cells (**Figure S10**). These results are consistent with previous single cell studies of immune cells[13,14,30], and indicate that we can accurately detect gene expression in single infected macrophages and phagocytosed *C. albicans*.

Dynamic host-pathogen co-states defined by analysis of single macrophages infected with live *C. albicans*

To finely map the basis of heterogeneous responses during infection, we clustered cells by differential expressed genes in single macrophages infected with live, phagocytosed *C. albicans.* Since we measured both host and pathogen gene expression changes in single infected cells, we identified host-pathogen co-states in groups of single infected macrophages and phagocytosed *C. albicans* pairs that showed similar gene expression profiles. 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)[31] as implemented in Seurat[32] (**Methods**). We identified differentially expressed genes (corrected-$P < 0.05$) using a likelihood-ratio test (LRT) for

single-cell differential expression[33] (**Methods**). The transcriptional response among single infected macrophages exhibited two time-dependent states (state 1^M and state 2^M) associated with expression shifts from 2 to 4 hours (**Figure 4**, **top**; **Figure S11A**). While cells were largely separated into these two states by time, a few macrophages were assigned to the alternate cell state by unsupervised clustering and appeared to be either early or delayed in the initiation of the transcriptional shift in the immune response. Immune cells that display asynchrony in their transcriptional response have previously been reported in LPS-stimulated dendritic cells, where a few "precocious" cells expressed high levels of immune response genes early, leading to cytokine secretion and activation of other cells in the population[30]. However, precocious host cell expression can also be related to pathogen state[14]. In single infected macrophages, we found that 88 and 70 differentially expressed genes (likelihood-ratio test (LRT), $P < 0.001$) that showed higher expression in the single, infected macrophages assigned to state 1^M, and assigned to state 2^M, respectively (**Table S6**). Genes in state 1^M are related with pro-inflammatory response, and their expression significantly decreases in state 2^M; at 4 hours (**Figure 4**, **top**). 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 4**, **top**). Many genes variably expressed in these single, infected macrophages were also found to be up-regulated in subpopulations of infected and exposed macrophages (*e.g. Tnf*, *Orl1*, **Figures S7**, **S12**); however, significant differences in expression of these genes between the 2 and 4 hours subpopulations samples were not detected, as subpopulation RNA-Seq average the gene expression of thousands of cells. These results highlight that cell-to-cell variability within each time point can only be observed when we analyzed single infected macrophages.

As each single, infected macrophage received a unique sample barcode and host and fungal transcription were measured simultaneously, we matched expression from each macrophage with that of the live, phagocytosed fungus (**Figure 4**, **bottom**). Notably, independent analysis of the parallel fungal transcriptional response identified two pathogen stages that were also primarily distinguished by time. In live phagocytosed *C. albicans*, we found a total of 168 differentially expressed genes (likelihood-ratio test (LRT), $P < 0.001$), 80 and 86 genes showed higher expression in *C. albicans* phagocytosed by macrophages assigned to state 1^M, and state 2^M, respectively (**Table S7**). In *C. albicans*, genes in state 1^C were enriched in organic acid metabolism ($P = 6.63e\text{-}11$; enriched GO term, corrected-$P < 0.05$, hypergeometric distribution with Bonferroni correction; **Table S8**), including a strong upregulation of transporters (*HGT13*) and glyoxylate cycle genes, specifically those from beta-oxidation

metabolism (*ECI1*, *FOX3*, *FOX2*, *PXP2*; **Figure 4**, **bottom**). Most macrophages infected by *C. albicans* in state 1^M induced a strong pro-inflammatory response (co-state 1; **Figure 4**). At 4 hours, expression of these genes was reduced and we observed an up-regulation of genes enriched in carbon metabolism (P = 2.43e-05; enriched GO term, corrected-P < 0.05, hypergeometric distribution with Bonferroni correction; **Table S8**), including genes related with a shift to glycolysis and gluconeogenesis (*PGK1*), fatty acid biosynthesis (*FAS1*, *ACC1*), and genes associated with filamentation, such as *ECE1*, *HWP1*, *OLE1*, *RBT1*; whereas the majority of infected macrophages had down-regulated expression of pro-inflammatory cytokines (co-state 2; **Figure 4**). In summary, these 2 host-pathogen co-states largely correspond to time of infection and highlight an asynchronous shift from a strong to a weak pro-inflammatory gene expression profile in the host that in turn correlated with the induction of genes related to filamentation and metabolic adaptation to host in the fungal pathogen.

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

We next examined expression patterns in infected macrophages, which may lead to or result from distinct expression programs in phagocytosed *C. albicans*. Previous work demonstrated that single dendritic cells stimulated with LPS displayed expression bimodality in a subset of immune response genes[13,30]. To further examine heterogeneity in gene expression, we first characterized unimodal and bimodal expression profiles, and then compared these distributions in host-pathogen single cells across and within 2 and 4 hours, using a normal mixture model and Bayesian modeling framework as implemented in scDD[34] (**Methods**). Overall, an average of 84.5% of the genes that met the filtering criteria in infected macrophages (**Methods**) displayed unimodal distributions (**Figure S11**; **Table S9**). Highly expressed (top 5%) unimodal genes with similar expression profiles at 2 and 4 hours encompassed genes involved in the immune response to *C. albicans* infection, including genes enriched in pathogen recognition, opsonization, and activation of engulfment (cluster 2; **Figure 2D**), such as complement proteins (*C1qb*, *C1qc*) and galectin receptors that recognize beta-mannans (*Lgals1*, *Lgals3*; **Figure S11**). In phagocytosed *C. albicans*, most genes (an average of 76% of the genes detected; **Methods**) also had unimodal expression patterns (**Figure S11**; **Table S10**). We also found a set of unimodal genes with similar expression profiles at both time points, which 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 results

demonstrated uniformly gene expression patterns among all single host and phagocytosed fungal cells for core genes involved in host immune response to fungus and pathogen virulence, respectively.

A subset of the genes highly expressed in co-states of single infected macrophages and phagocytosed *C. albicans* showed evidence of transcriptional heterogeneity, forming bimodal expression distributions. As expression bimodality in single infected cells can signify distinct immune cell levels[13,30], we examined whether subgroups of single macrophages infected by *C. albicans* could be defined by shared bimodality of genes involved in the immune response. We characterized genes that showed patterns of bimodal distribution among or within 2 and 4 hours in infected macrophages (an average of 15% of the genes; exceeded Bimodality Index (BI) threshold; Dirichlet Process Mixture of normals model; **Table S9**, **Figure S11**). We found that genes involved in pathogen intracellular recognition and pro-inflammatory response (*e.g. Olr1*, *Il4ra*, *Il21r*, *Il1rn*, and *Tnf*) exhibited expression heterogeneity and bimodality within and among single infected macrophages at 2 and 4 hours (**Figure S13**, **top**; **Table S9**). In addition, we further identified 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[34]. We found a subset of immune genes that showed differential distribution ($P < 0.05$, Benjamini-Hochberg adjusted Fisher's combined test). For example, the *Olr1* lectin-like receptor, the tumor necrosis factor receptor superfamily member 12a (*Tnfrsf12a*), and the transmembrane marker *Cd83*, had bimodal expression distribution at 2 hours but not at 4 hours; the interleukin 4 receptor, alpha (*Il4ra*) and the transmembrane marker *Cd300a* had bimodal distribution at 4 hours but not at 2 hours; and the interleukin 21 receptor (*Il21r*) had differential distribution and is bimodal at both 2 and 4 hours (**Table S9**). Prior work demonstrated that macrophages infected with *Salmonella* and stimulated with LPS also exhibited heterogeneity and bimodal expression patterns in subsets of genes, including some immune response genes[14]. Similar to macrophages infected with *Salmonella* and stimulated with LPS, *Tnf* and *Il4ra* exhibited bimodal expression patterns in single macrophages infected with *C. albicans*[14] (**Figure S13**, **top**). Additionally, we found unique subsets of genes displaying differential distributions in single macrophages infected with *C. albicans,* but not in response to *Salmonella* or LPS, including *Il17ra* and other lectin-like receptors (**Figure S14**). These results highlight heterogeneity in gene expression patterns among single macrophages infected with live *C. albicans* that might indicate distinct cell levels and trajectories in response to fungal infection and suggest that some expression heterogeneity in macrophages is pathogen specific, likely a result of variably expressed pathogen-specific receptors.

We next hypothesized that *C. albicans* may also demonstrate expression heterogeneity and bimodality that is correlated 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 (**Table S10**; **Figure S11**). We also further examined if genes in phagocytosed *C. albicans* exhibited differential distributions (*e.g.* shifts in mean(s) expression, modality, and proportions of cells) across and within 2 and 4 hours using scDD[34]. We found a subset of virulence-associated genes that showed differential distribution ($P < 0.05$, Benjamini-Hochberg adjusted Fisher's combined test). Genes showing bimodal expression only at 2 hours were enriched in regulation of defense response (enriched GO term, corrected-$P < 0.05$, hypergeometric distribution with Bonferroni correction; **Table S10**), including genes associated with *C. albicans* response to phagocytosis (*CAT1*, *IMH3*, *ADH1*, *SSB1*, *PCK1*). Meanwhile, genes with a bimodal distribution detected only at 4 hours were enriched in pentose/xylose transport (*HGT10*, *HGT12*). Other subsets of genes that showed differential distributions 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 S13**, **bottom**). These results suggest that expression heterogeneity and bimodality of key genes involved in host immune response and fungal morphology and adaptation are tightly coupled during host- fungal pathogen responses and might result in different levels of immune response and virulence.

Dual RNA-Seq of subpopulations and single infected cells revealed tightly coupled macrophage and *Candida albicans* transcriptional responses

By parallel sequencing of host and pathogen transcriptomes in sorted infection subpopulations and in single infected macrophages, we can directly contrast gene expression changes in the host and fungal pathogen to characterize these interactions. While sorted subpopulations allow for detection of patterns across a group of cells, comparing the host and pathogen transcripts from single infected macrophages allows more precise detection of how host and pathogen gene expression co-vary, and the consequences of transcriptional heterogeneity in the infection outcome. Gene expression changes in subpopulations of exposed and infected macrophages highlighted immune gene functions that were induced early during these interactions and stable during the time course of infection (between 1 and 4 hours); this included genes important for pathogen recognition (*e.g. Tlr2*, *Clec7a, Lgal3*) and pro-

inflammatory cytokine production and phagocytosis (*e.g. Tnf*, *Cxcl10*, *Irf1*). Other sets of genes shifted in gene expression in subpopulations of macrophages exposed to or infected with live *C. albicans* from 2 to 4 hours, including up-regulation of genes required for the activation of the inflammosome (*e.g. Nod2*, *Card9*), and the down-regulation of pro-inflammatory cytokines (*e.g. Il1a*, *Il1r1*, *Il17ra*) and intracellular receptors (*e.g. Tlr5*, *Tlr9*; **Figures 2C**, **2D**). Live phagocytosed *C. albicans* within these macrophages showed a rapid shift in gluconeogenic metabolism, amino acid uptake, cell wall remodeling, and initiation of filamentation (**Figure 2A**, **B**). Up-regulation of gene clusters increased over the time course with the highest expression at 4 hours, and down-regulation of gene clusters recovered their expression levels at 4 hours, when macrophages up-regulated inflammosome-associated genes and down-regulated of pro-inflammatory cytokines and intracellular receptors (**Figure 2**). Transcriptional variability in these subpopulations could be further resolved into distinct transcriptional states and cell trajectories at the single-cell level that govern infection fate decisions. We observed that the gene expression in single macrophages infected with live *C. albicans* at 2 and 4 hours was tightly coupled with that in phagocytosed *C. albicans* and could be described as two, time dependent host-pathogen co-states. At 2 hours, most single infected macrophages induced genes related to a pro-inflammatory response and the phagocytosed *C. albicans* upregulated transporters and glyoxylate cycle genes, specifically those involved in beta-oxidation metabolism (co-state 1; **Figure 4**). However, a subset of the cells at two hours was more similar to the major co-state found at 4 hours, where expression of these transporters and metabolic *C. albicans* genes were reduced and we observed an up-regulation of genes related with filamentation, such as *ECE1*, *ALS3*, *OLE1*, *RBT1*; in parallel, macrophages down-regulated pro-inflammatory genes (co-state 2; **Figure 4**)**.** Notably, we observed expression heterogeneity of some infection-induced genes that was only detected at the level of single host-pathogen cells, which could have affect the outcome of the immune response to *C. albicans*, and virulence levels in response to phagocytosis. Measuring dual species gene expression in sorted infection subpopulations and at the single infected cell levels reveals the genes involved in and tightly coupled patterns of heterogeneity and transcriptional co-states among the host and fungal pathogen. This approach can further enhance our understanding of distinct infection fate decision and the correlated gene regulation that governs host cells and fungal pathogens.

Discussion

Host and fungal cell interactions are heterogeneous, even among clonal cell populations. Resolving this heterogeneity requires subdividing these populations by infection fate or stage, to measure gene

529 expression changes in specific stages of these interactions. Additionally, measuring host and fungal
530 pathogen gene expression levels using dual RNA-sequencing can provide insight as to how both
531 species respond to each other in specific infection states. Here, we developed a generalizable strategy
532 to isolate distinct host and fungal pathogen infection fates over time and comprehensively measured
533 subpopulation and single infected cell gene expression changes in both host and pathogen. Our
534 approach of using sorted cell populations facilitates more precise definition of genes important for host
535 cell survival, pathogen clearance and fungal virulence.
536
537 This approach builds on prior RNA-Seq studies of interactions between microbial pathogens and
538 immune cells. For fungi, previous RNA-Seq studies have largely measured gene expression profiles of
539 either the host or the fungal pathogen[3,6,7], and have measured transcription profiles across infection
540 outcomes[8–11]. Recent similar approaches have measured gene expression in populations of single
541 phagocytes infected by bacterial pathogens using a similar GFP reporter and FACS isolation in
542 combination with single-cell RNA-Seq[14–16]. However, these studies mainly focused on the
543 transcriptional response of the host, as bacterial transcriptomes can be difficult to measure due to their
544 relatively low number of transcripts[14,16]. By contrast, our method is capable of recovering sufficient host
545 and *C. albicans* RNA for differential expression measurement; further studies will be needed to examine
546 if the same or modified approaches can be applied to other microbial pathogens.
547
548 Our approach demonstrated that distinct infection fates within heterogeneous host and fungal pathogen
549 interactions can be disambiguated at both the subpopulation and single infected cell level. By
550 examining single infected macrophages, we showed that host and pathogen transcriptional co-states
551 are tightly coupled during an infection time course, providing a high-resolution view of host-fungal
552 interactions. This also revealed that expression heterogeneity in key genes in both infected
553 macrophages and in phagocytosed *C. albicans* may contribute to infection outcomes. We identified two,
554 time dependent co-states of host-fungal pathogen interaction. The initial state is characterized by
555 induction of a pro-inflammatory host profile after 2 hours of interaction with *C. albicans* that then
556 decreased by 4 hours. A previous study indicated that a balance between pro-inflammatory and anti-
557 inflammatory responses is necessary for *C. albicans* to establish infection[35]. A decrease in the pro-
558 inflammatory response after *C. albicans* exposure was also found in human macrophages, where pro-
559 inflammatory macrophages that interact with *C. albicans* for 8 hours or longer skew toward an anti-
560 inflammatory proteomic profile[36]. In addition, recent single-cell RNA-Seq analysis of macrophages
561 containing growing *Salmonella* showed that macrophages shifted to an anti-inflammatory state by 20

562 hours; here, the authors hypothesized that fast-growing intracellular *Salmonella* overcame host
563 defenses by reprogramming macrophage gene expression[15]. We found that single macrophages
564 infected with *C. albicans* assigned to co-state 2 decreased expression of pro-inflammatory cytokine
565 genes and upregulated genes involved in the activation of inflammasomes; this response was coupled
566 with activation of filamentation and cell-wall remodeling in phagocytosed *C. albicans*. Previous work has
567 shown that *C. albicans* can escape from macrophage phagosomes by lytic or non-lytic mechanisms.
568 For example, *C. albicans* can escape by rupturing the macrophage membrane during intra-phagocytic
569 hyphal growth[37], or by the activation of one of the macrophage programmed cell death pathways,
570 including formation of inflammasomes and pyroptosis[38]. Our work suggests that induction of
571 filamentation and cell-wall remodeling programs in *C. albicans* are coupled to a down-regulation of the
572 pro-inflammatory state of the host cells.
573
574 Remarkably, both single infected macrophages and the corresponding phagocytosed *C. albicans* cells
575 exhibit expression bimodality for a subset of genes. The expression bimodality observed for the host as
576 well as the pathogen is consistent with the evolutionary concept of bet hedging[39]. Both cell types may
577 rely on stochastic diversification of phenotypes to minimize their risks and improve their survival rate in
578 the event of an encounter with the other cell type. For instance, clonal populations of *C. albicans* that
579 find themselves in the unpredictable, changing environment of a host phagocyte could increase the
580 chance of survival by exhibiting various phenotypes. Fungal cells with maladapted phenotypes are
581 eliminated so as to increase fitness of a particular genotype that is fit for that environment. A similar
582 scenario may provide an advantage for phagocytes that have to hedge their bets when they encounter
583 pleomorphic *C. albicans*. These strategies have been described within microbial populations, where a
584 small fraction of “persister” cells might be transiently capable of surviving exposure to lethal doses of
585 antimicrobial drugs as a bet-hedging strategy[40]. While gene expression patterns are correlated in host
586 and fungal pathogen, it is unclear whether expression heterogeneity among individual infected
587 macrophages results from or results in expression bimodality among phagocytosed *C. albicans*.
588
589 Our approach represents a novel method to query host-fungal pathogen interaction. The
590 characterization of genes involved in heterogeneous responses is important to consider in the selection
591 of novel antifungal drug targets, as designing therapeutics to target products of genes expressed
592 uniformly among fungal cells in a population may be more effective than targeting the products of
593 genes expressed by only a subset of cells. Additionally, designing combination immunotherapies to
594 affect genes discriminative of distinct infection subpopulations, such as those genes upregulated in

595 macrophages infected with dead *C. albicans*, could help to increase the proportion of host cells that
596 clear the fungal pathogen. Parallel host-fungal pathogen expression profiling could also allow
597 researchers to not only measure the effects of new drug treatments on the pathogen, but also collect
598 information on how these treatments affect host cells. As single-cell RNA-Seq microfluidic platforms
599 continue to develop and become more cost effective to profile thousands of single cells, it will become
600 tractable to interrogate larger cohorts of host cells infected with fungus. This approach will allow
601 investigation of fungal phenotypic heterogeneity as a driver of different host responses and provide a
602 systems view of these interactions.
603
604 **Methods**
605
606 ***Candida albicans* reporter strain construction**
607 The reporter construct used in this study was prepared by integrating the GFP and mCherry fluorescent
608 tags driven by the bi-directional *ADH1* promoter and a nourseothricin resistance (NAT^R) cassette at the
609 Neut5L locus of *Candida albicans* strain SC5314 (**Figure S15**). Briefly, the pUC57 vector containing
610 mCherry driven by *ADH1* promoter (Bio Basic) was digested and this portion of the plasmid was ligated
611 into a pDUP3 vector[41] containing GFP, also driven by *ADH1* promoter, a NAT^R marker, and homology
612 to the Neut5L locus. The resulting plasmid was linearized and introduced via homologous
613 recombination into a neutral genomic locus, Neut5L using chemical transformation protocol with lithium
614 acetate. Transformation was confirmed via colony PCR and whole-genome sequencing. A whole
615 genome library was created from strain SC5314-Neut5L-*NAT1*-mCherry-GFP using Nextera-XT
616 (Illumina) and sequenced on Illumina's Miseq (150x150 paired end sequencing). Sequencing reads
617 were *de novo* assembled using dipSPAdes[42]. Scaffolds were aligned back to the plasmid used to
618 transform wild type SC5314 using BLAST[43]. Sequencing coverage was visualized using IGV[44].
619
620 **Macrophage *Candida albicans* infection assay**
621 Primary, bone derived macrophages (BMDMs) were derived from bone marrow cells collected from the
622 femur and tibia of C57BL/6, female mice. All mouse work was performed in accordance with the
623 Institutional Animal Care and Use Committees (IACUC) and with relevant guidelines at the Broad
624 Institute and Massachusetts Institute of Technology, with protocol 0615-058-1. Primary bone marrow
625 cells were grown in "C10" media as previously described[45] and supplemented with macrophage colony
626 stimulating factor (M-CSF) (ThermoFisher Scientific) at final concentration of 10 ng/ml, to promote
627 differentiation into macrophages. Cultures were then stained with F4/80 (Biolegend) to ensure that

~95% of the culture had differentiated into macrophages. To visualize the interactions between BMDMs and our reporter strain of *C. albicans*, we captured images over time using microwells (**Figure S1**). BMDMs were incubated in 35 micron by 70 micron by 10 micron microwells (SU-8 spin-coated) for 4 hours, until adherent and elongated. GFP expressing *C. albicans* (SC5314) was added to the microwells at a multiplicity of infection of 1:1. Microwells were imaged every 3 minutes on an Olympus IX-83 microscope, with on-stage incubation at 37 °C in RPMI supplemented with FCS for 6 hours. For the infection 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 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% C02). 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. 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. Unexposed controls for both species were collected as described above, not sorted and resuspended in 600 µl of buffer RLT (Qiagen) + 1% β-Mercaptoethanol (Sigma).

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 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 *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.

659 **RNA extraction, evaluation of RNA quality**

660 RNA was extracted from population samples using the Qiagen RNeasy mini kit. All samples were
661 subjected to 3 minutes of bead beating with .5mm zirconia glass beads (BioSpec Products) in a bead
662 mill. Single macrophages infected with *C. albicans* were directly lysed by sorting cells into a 96 well
663 plate containing 5 ul of RLT (Qiagen) + 1% β-Mercaptoethanol (Sigma).
664
665 **cDNA synthesis and library generation**
666 For population samples, the RT reaction was carried out with the following program, as described[46],
667 with the addition of RNAse inhibitor (ThermoFisher) at 40U/ul. cDNA was generated from single cells
668 based on the Smart-seq2 method as described previously[47], with the addition of RNase inhibitor was
669 used at 40 U/ul (ThermoFisher) and 3.4 ul of 1 M trehalose was added to the RT reaction. All libraries
670 were constructed using the Nextera XT DNA Sample Kit (Illumina) with custom indexed primers as
671 described[47]. Infection subpopulation samples were sequenced on an Illumina Nextseq (37 x 38 cycles).
672 *Candida* only samples were sequenced on an Illumina Miseq (75 x 75 cycles). Single infected cells
673 were sequenced on Illumina's Nextseq (75x75 cycles).
674
675 **Read processing and transcript quantification of population-RNA-Seq**
676 Basic quality assessment of Illumina reads and sample demultiplexing was done with Picard version
677 1.107 and Trimmomatic[48]. Samples profiling exclusively the mouse transcriptional response were
678 aligned to the mouse transcriptome generated from the v. Dec. 2011 GRCm38/mm10 and a collection
679 of mouse rRNA sequences from the UCSC genome website[49]. Samples profiling exclusively the yeast
680 transcriptional response were aligned to the *Candida albicans* transcriptome strain SC5314 version
681 A21-s02-m09-r10 downloaded from *Candida* Genome Database (http://www.candidagenome.org).
682
683 Samples from the infection assay profiling in parallel host and fungal transcriptomes, were aligned to a
684 "composite transcriptome" made by combining the mouse transcriptome described above and the *C.*
685 *albicans* transcriptome described above. To evaluate read mappings, BWA aln (BWA version 0.7.10-
686 r789, http://bio-bwa.sourceforge.net/)[50] was used to align reads, and the 'XA tag' was used for read
687 enumeration and separation of host and pathogen sequenced reads. Multi-reads (reads that aligned to
688 both host and pathogen transcripts) were discarded, representing only an average of 2.6% of the
689 sequenced reads. Then, each host or pathogen sample file were aligned to its corresponding reference
690 using Bowtie2[51] and RSEM (RNA-Seq by expectation maximization; v.1.2.21). Transcript abundance
691 was estimated using transcripts per million (TPM). Since parallel sequencing of host and pathogen from

692 single macrophages increased the complexity of transcripts measured compared to studies of only host
693 cells alone, we detected lower number of transcripts of macrophages as compared with other studies
694 using phagocytes and similar scRNA-seq methods[14,15,30].
695
696 **Differential gene expression analysis of population-RNA-Seq**
697 TMM-normalized 'transcripts per million transcripts' (TPM) for each transcript were calculated, and
698 differentially expressed transcripts were identified using edgeR, all as implemented in the Trinity
699 package version 2.1.1[52]. Genes were considered differentially expressed if they had a 4-fold change
700 difference (> 4 FC) in TPM values and a false discovery rate below or equal to 0.001 (FDR < 0.001),
701 unless specified otherwise.
702
703 **Read processing and transcript quantification of single-cell RNA-Seq**
704 BAM files were converted to merged, demultiplexed FASTQ format using the Illumina Bcl2Fastq
705 software package v2.17.1.14. For the RNA-Seq of sorted population samples, paired-end reads were
706 mapped to the mouse transcriptome (GRCm38/mm10) or to the *Candida albicans* transcriptome strain
707 SC5314 version A21-s02-m09-r10 using Bowtie2[51] and RSEM (RNA-Seq by expectation maximization;
708 v.1.2.21)[53]. Transcript abundance was estimated using transcripts per million (TPM). For read mapping
709 counts paired-reads were aligned to the 'composite reference' as described above.
710
711 For each single macrophage infected with *C. albicans*, we quantified the number of genes for which at
712 least one read was mapped (TPM > 1). We filtered out low-quality macrophage or *C. albicans* cells from
713 our data set based on a threshold for the number of genes detected (a minimum of 2,000 unique genes
714 per cell for macrophages, and 600 unique genes per cell for *C. albicans*, and focused on those single
715 infected macrophages that have good number of transcript detected in both host and pathogen (**Figure**
716 **S9A**). For a given sample, we define the filtered gene set as the genes that have an expression level
717 exceeding 10 TPM in at least 20% of the cells. After cell and gene filtering procedures, the expression
718 matrix included 3,254 transcripts for the macrophages and 915 transcripts for *C. albicans*. To estimate
719 the number of *C. albicans* in each macrophage, we measured the correlation of GFP levels from FACs
720 with the total number of transcripts detected in live, phagocytosed *C. albicans* cells (at least 1 TPM) but
721 found only a modest correlation between these two metrics (R^2= 0.52).
722
723 To eliminate the non-biological associations of the samples based on plate based processing and
724 amplification, single-cell expression matrices were log-transformed (log(TPM + 1)) for all downstream

725 analyses, most of which were performed using the R software package Seurat
726 (https://github.com/satijalab/seurat). In addition, we do not find substantial differences in the number of
727 sequenced reads and detected genes between samples. We separately analyzed two comparisons of
728 macrophages-*C. albicans* cells: *i*) single macrophages infected - live phagocytosed *C. albicans* cells at
729 2 and 4 hours (macrophages = 267; *C. albicans* cells = 215; macrophage-*C. albicans* = 156); and *ii*)
730 macrophages infected and live or dead phagocytosed *C. albicans* cells at 4 hours (macrophages = 142;
731 *C. albicans* cells = 71). These numbers of macrophages and *C. albicans* cells are the total that met the
732 described QC filters.
733
734 **Detection of variation across single, infected cells**
735 To examine if cell to cell variability existed across a wide range of population expression levels, we
736 analyzed the variation and the intensity of non-unimodal distribution for each gene across single
737 macrophages and *C. albicans* cells. Briefly, we determined the distribution of the average expression
738 (μ), and the dispersion of expression (σ^2; normalized coefficient of variation), placing each gene into
739 bins, and then calculating a z-score for dispersion within each bin to control for the relationship between
740 variability and average expression as implemented R package Seurat[32].
741
742 **Detection of variable genes and cell clustering**
743 To classify the single cell RNA-Seq from macrophages and *C. albicans*, the R package Seurat was
744 used[32]. We first selected variable genes by fitting a generalized linear model to the relationship between
745 the squared co-efficient of variation (CV) and the mean expression level in log/log space, and selecting
746 genes that significantly deviated (p-value < 0.05) from the fitted curve, as implemented in Seurat, as
747 previously described[15]. Then highly variable genes (CV > 1.25; p-value < 0.05) were used for principle
748 component analysis (PCA), and statistically significant determined for each PC using JackStraw.
749 Significant PCs (p-value < 0.05) were used for two-dimension t-distributed stochastic neighbor
750 embedding (tSNE) to define subgroups of cells we denominated host-pathogen co-states. We identified
751 differentially expressed genes (corrected-p < 0.05) between co-states using a likelihood-ratio test (LRT)
752 for single-cell differential expression[33] as implemented in Seurat.
753
754 **Detection of differential expression distributions and bimodality**
755 To detect which genes have different expression distributions single infected macrophage and live *C.*
756 *albicans* we compared the distributions of gene expression within single infected macrophage and live
757 *C. albicans*, across and between 2 and 4 hours, and identified genes showing evidence of differential

758 distribution using a Bayesian modeling framework as implemented in scDD[34]. We used the permutation
759 test of the Bayes Factor for independence of condition membership with clustering (*n* permutations =
760 1000), and test for test for a difference in the proportion of zeroes (testZeroes=TRUE). A gene was
761 considered differentially distributed using Benjamini-Hochberg adjusted *p*-values test (*p*-value < 0.05).
762
763 **Functional biological enrichment analyses**
764 For *C. albicans*, Gene ontology (GO) term analysis was performed in through the *Candida* Genome
765 Database GO Term Finder and GO Slim Mapper (http://www.candidagenome.org[54]). GO terms were
766 considered significantly enriched in a cluster or set of genes if we found a GO term corrected *p-value*
767 lower than 0.05 using hypergeometric distribution with Bonferroni correction
768
769 For macrophages, Ingenuity Pathway analysis (IPA) was performed. We investigated biological
770 relationships, canonical pathways and Upstream Regulator analyses as part of IPA software. This
771 allowed us to assess the overlap between significantly DEGs and an extensively curated database of
772 target genes for each of several hundred known regulatory proteins. Clusters or set of genes were
773 considered significantly enriched if we found enriched a -log(*p-value*) greater than 1.3 (*i.e. p*-value ≤
774 0.05) and z-score greater than 2 as recommended by the IPA software.
775
776 **Data availability statement**
777 All sequence data for this project has been deposited in the SRA under Bioproject PRJNA437988.
778 Raw and processed data for gene expression analysis was deposited in the GEO under GSE111731.
779
780
781 **Acknowledgments**
782 We thank Aviv Regev and members of her lab for providing support for all the experimental work in this
783 paper. We also thank Raktina Raychowdhury in Nir Hacohen's lab for help with preparation of the
784 primary BMDM cells and Anh Hoang, Mehment Toner, and Daniel Irima for providing microwells. This
785 project has been funded in whole or in part with Federal funds from the National Institute of Allergy and
786 Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under
787 award N°: U19AI110818. CBF was supported by a Helen Hay Whitney postdoctoral fellowship, TD was
788 supported by NIAID and WPI.
789
790 **Author Contributions**

CBF, DAT, RPR, and CAC designed the study. TD, CBF and BYL carried out experiments. JFM analyzed the data and prepared figures and tables. JFM, TD, RPR and CAC wrote the initial draft of the manuscript, which was revised with input from all authors. All authors read and approved final manuscript.

Additional information

Competing interests: The authors declare no competing interests.

Figures

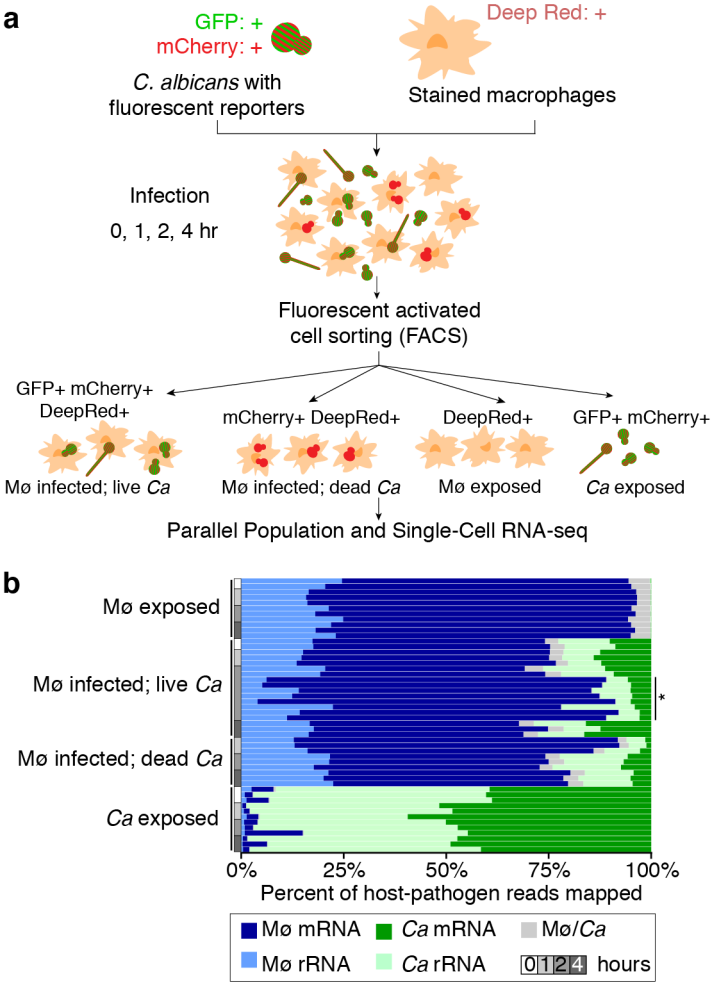


Figure 1. Heterogeneous populations of macrophage-*Candida* encounters are captured by cell sorting sequencing and expression analysis. (a) Schematic representation of the experimental model, using primary, bone marrow-derived macrophages (BMDMs) incubated with *Candida albicans* reporter strain SC5314-mCherry-GFP, sampling at time intervals, and sorting to separate subpopulations of interacting cells. **(b)** 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* (SC5314; 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.

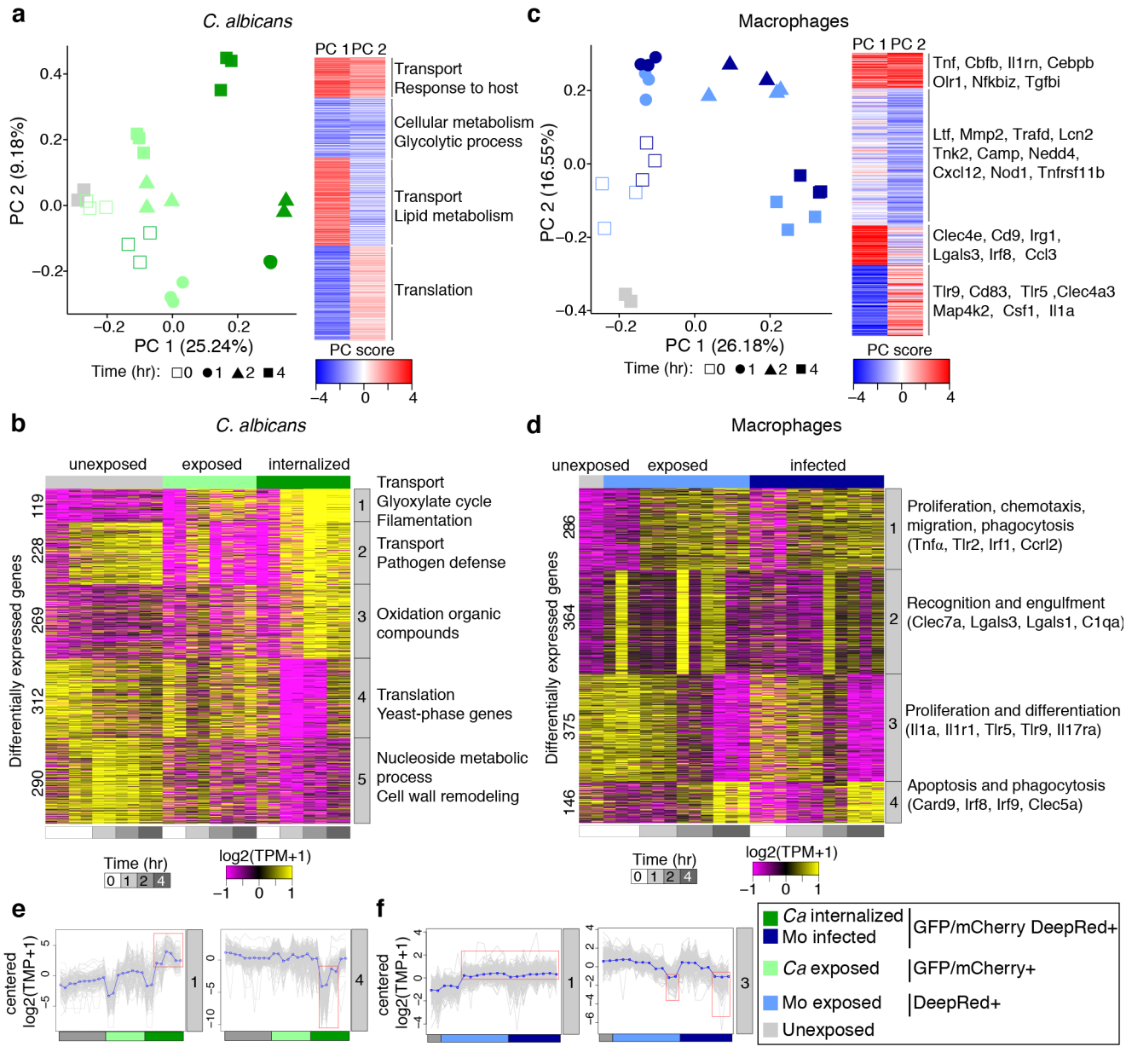
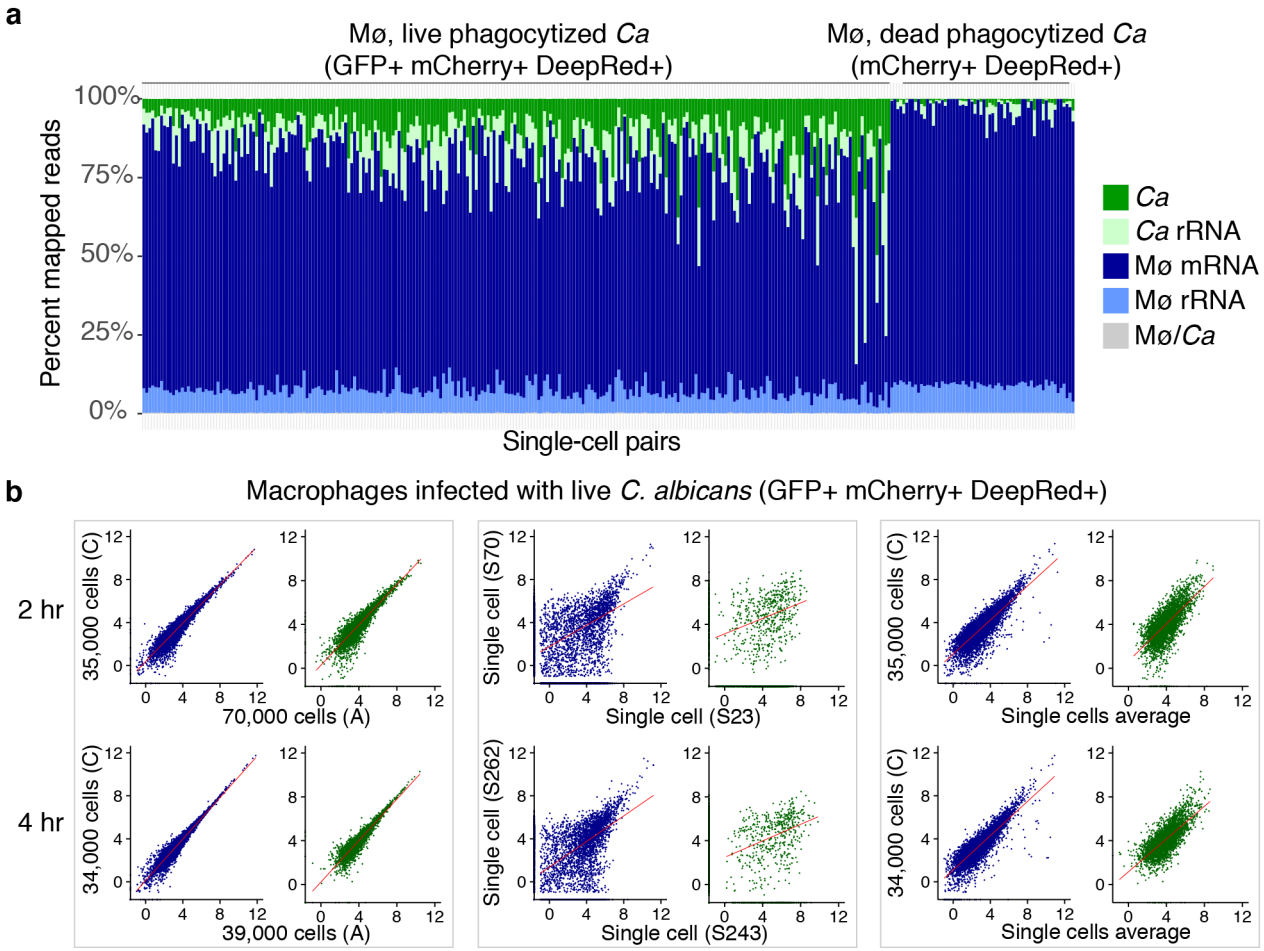


Figure 2. Subpopulations of infection outcome reveal dynamic parallel host-pathogen orchestration of transcriptional response. Principal component analysis (PCA) using the transcript abundance for the pathogen (**a**; *C. albicans*) and the host (**c**; macrophages) using significantly differentially 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 (unexposed, 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

825 relationships using Gene Ontology (GO) terms (corrected $P < 0.05$). **(d)** Transcriptional response in
826 subpopulations of macrophages. Heatmap depicts significantly differentially expressed genes (DEGs;
827 FC >4, FDR < 0.001) for replicates of each macrophage-sorted populations (unexposed, exposed and
828 infected) at 0, 1, 2 and 4 hours post-infection, clustered by *k-means* (similar expression patterns). Each
829 cluster includes synthesized functional biological relationships using IPA terms (-log(p-value) > 1.3; z-
830 score > 2). Expression patterns for **(f)** clusters 1 and 4 in *C. albicans* and **(g)** cluster 1 and 3 in
831 macrophages. Each gene is plotted (gray) in addition to the mean expression profile for that cluster
832 (blue).
833



834

835 **Figure 3. Processing and evaluation of parallel host-pathogen single-cell RNA-seq. (a)** Plot of the
836 percent of mapped reads to the composite reference transcriptome (mouse messenger and ribosomal
837 RNA + *Candida albicans* messenger and ribosomal RNA collections) of 314 single macrophages with
838 live (mCherry + GFP+) or dead (mCherry + GFP-) phagocytosed *C. albicans* cells. **(b)** Plots of the gene
839 expression correlation between (left) two replicates of the sorted subpopulation of macrophages (blue)
840 with phagocytosed live *C. albicans* (green) at 2 and 4 hours post-infection; (middle) between two single
841 macrophages and two phagocytosed *C. albicans*, and (right) between the single cells expression
842 average and one replicate of the population.
843

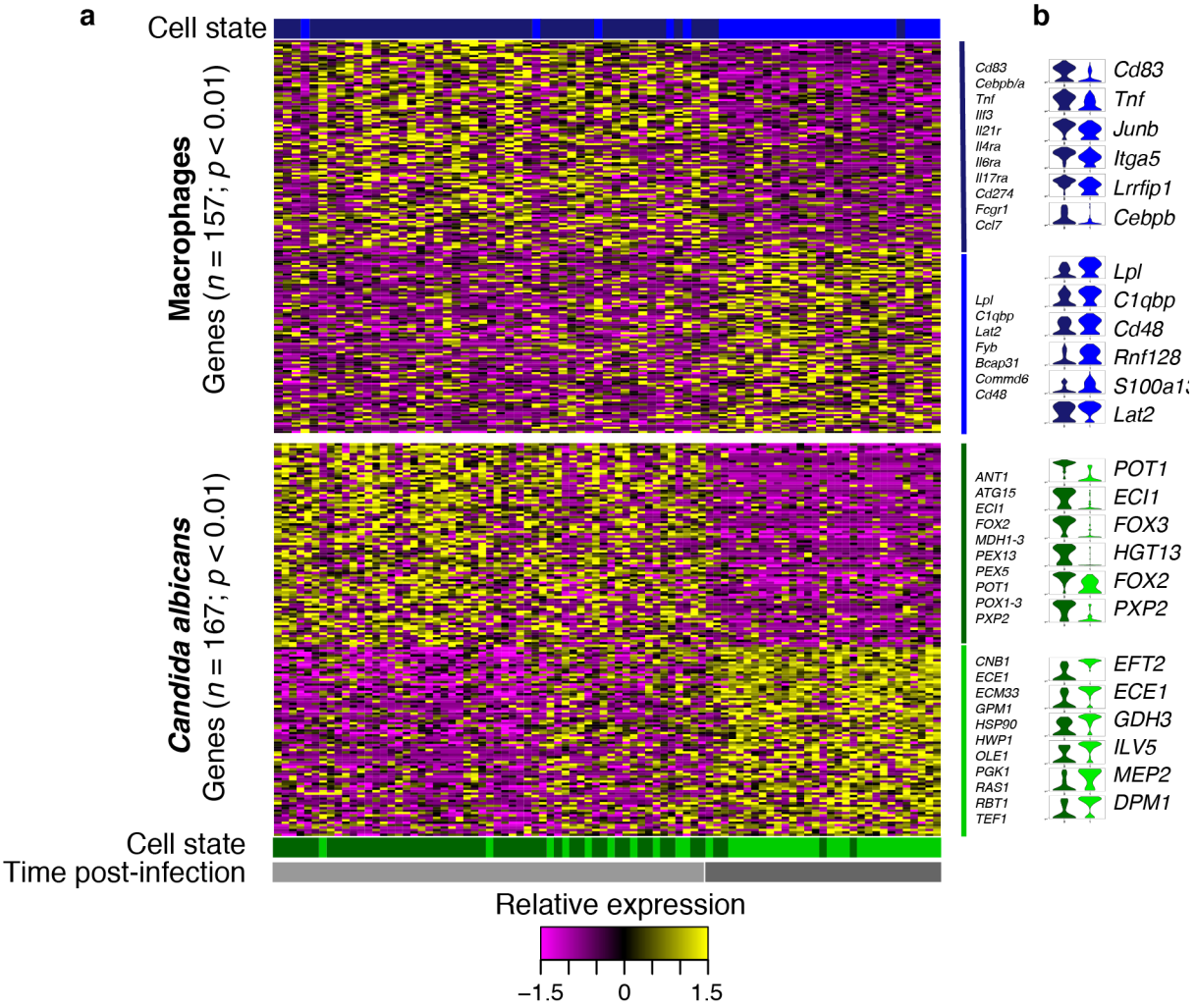


Figure 4. Parallel host-pathogen single-cell RNA-seq profiling reveals infection outcome shifts in pro-inflammatory response and pathogen immune evasion. **(a)** Heat maps report parallel scaled expression [log TPM (transcripts per million) +1 values] of differentially expressed genes for each co-state 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)** 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.

References

1. Brown, G. D. Innate antifungal immunity: the key role of phagocytes. *Annu. Rev. Immunol.* **29,** 1–21 (2011).
2. Huffnagle, G. B. & Noverr, M. C. The emerging world of the fungal microbiome. *Trends Microbiol.* **21,** 334–341 (2013).
3. Lorenz, M. C., Bender, J. A. & Fink, G. R. Transcriptional response of *Candida albicans* upon internalization by macrophages. *Eukaryot. Cell* **3,** 1076–1087 (2004).
4. 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).
5. 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).
6. Smeekens, S. P. *et al.* Functional genomics identifies type i interferon pathway as central for host defense against *Candida albicans*. *Nat. Commun.* **4,** (2013).
7. Quintin, J. *et al.* Candida albicans infection affords protection against reinfection via functional reprogramming of monocytes. *Cell Host Microbe* **12,** 223–232 (2012).
8. Hebecker, B. *et al.* Dual-species transcriptional profiling during systemic candidiasis reveals organ-specific host-pathogen interactions. *Sci Rep* **6,** 36055 (2016).
9. Liu, Y. *et al.* New signaling pathways govern the host response to *C. albicans* infection in various niches. *Genome Res.* **125,** 679–689 (2015).
10. Niemiec, M. J. *et al.* Dual transcriptome of the immediate neutrophil and *Candida albicans* interplay. *BMC Genomics* **18,** 696 (2017).
11. 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).
12. Erwig, L. P. & Gow, N. a. R. Interactions of fungal pathogens with phagocytes. *Nat. Rev. Microbiol.* **14,** 163–176 (2016).
13. Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498,** 236–40 (2013).
14. Avraham, R. *et al.* Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. *Cell* **162,** 1309–1321 (2015).
15. Saliba, A.-E. *et al.* Single-cell RNA-seq ties macrophage polarization to growth rate of intracellular *Salmonella*. *Nat. Microbiol.* **2,** 16206 (2016).
16. 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).

17. Martin, R. *et al.* A core filamentation response network in *Candida albicans* is restricted to eight genes. *PLoS One* **8,** e58613 (2013).
18. Murad, A. M. A. *et al.* Transcript profiling in *Candida albicans* reveals new cellular functions for the transcriptional repressors CaTup1, CaMig1 and CaNrg1. *Mol. Microbiol.* **42,** 981–993 (2001).
19. Braun, B. R. & Johnson, A. D. Control of filament formation in *Candida albicans* by the transcriptional repressor TUP1. *Science (80-.).* **277,** 105–109 (1997).
20. Hwang, C.-S., Oh, J.-H., Huh, W.-K., Yim, H.-S. & Kang, S.-O. Ssn6, an important factor of morphological conversion and virulence in *Candida albicans*. *Mol. Microbiol.* **47,** 1029–1043 (2003).
21. Ryan, O. *et al.* Global Gene Deletion Analysis Exploring Yeast Filamentous Growth. *Science (80-.).* **337,** 1353–1356 (2012).
22. O'Meara, T. R. *et al.* Global analysis of fungal morphology exposes mechanisms of host cell escape. *Nat. Commun.* **6,** 6741 (2015).
23. Huang, G. *et al.* Bistable expression of WOR1, a master regulator of white-opaque switching in *Candida albicans*. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 12813–8 (2006).
24. 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).
25. 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).
26. Fradin, C. *et al.* Stage-specific gene expression of Candida albicans in human blood. *Mol. Microbiol.* **47,** 1523–1543 (2003).
27. Fradin, C. *et al.* Granulocytes govern the transcriptional response, morphology and proliferation of *Candida albicans* in human blood. *Mol. Microbiol.* **56,** 397–415 (2005).
28. Bruno, V. M. *et al.* Transcriptomic Analysis of Vulvovaginal Candidiasis Identifies a Role for the NLRP3 Inflammasome. *MBio* **6,** e00182-15 (2015).
29. Liu, Y. *et al.* New signaling pathways govern the host response to *C. albicans* infection in various niches. *Genome Res* **25,** 679–89 (2015).
30. Shalek, A. K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510,** 363–9 (2014).
31. Van Der Maaten, L. J. P. & Hinton, G. E. Visualizing high-dimensional data using t-sne. *J. Mach. Learn. Res.* **9,** 2579–2605 (2008).

924 32. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using
925 Nanoliter Droplets. *Cell* **161,** 1202–1214 (2015).
926 33. McDavid, A. *et al.* Data exploration, quality control and testing in single-cell qPCR-based gene
927 expression experiments. *Bioinformatics* **29,** 461–467 (2013).
928 34. Korthauer, K. D. *et al.* A statistical approach for identifying differential distributions in single-cell
929 RNA-seq experiments. *Genome Biol.* **17,** 222 (2016).
930 35. Jouault, T. *et al.* Host responses to a versatile commensal: PAMPs and PRRs interplay leading
931 to tolerance or infection by *Candida albicans*. *Cell. Microbiol.* **11,** 1007–15 (2009).
932 36. Reales-Calderón, J. A., Aguilera-Montilla, N., Corbí, Á. L., Molero, G. & Gil, C. Proteomic
933 characterization of human proinflammatory M1 and anti-inflammatory M2 macrophages and their
934 response to *Candida albicans*. *Proteomics* **14,** 1503–1518 (2014).
935 37. Bain, J. M. *et al. Candida albicans* Hypha Formation and Mannan Masking of β-Glucan Inhibit
936 Macrophage Phagosome Maturation. *MBio* **5,** e01874-14 (2014).
937 38. Wellington, M., Koselny, K., Sutterwala, F. S. & Krysan, D. J. *Candida albicans* triggers NLRP3-
938 mediated pyroptosis in macrophages. *Eukaryot. Cell* **13,** 329–340 (2014).
939 39. de Jong, I. G., Haccou, P. & Kuipers, O. P. Bet hedging or not? A guide to proper classification of
940 microbial survival strategies. *BioEssays* **33,** 215–223 (2011).
941 40. Lewis, K. Persister Cells. *Annu. Rev. Microbiol.* **64,** 357–372 (2010).
942 41. Gerami-Nejad, M., Zacchi, L. F., McClellan, M., Matter, K. & Berman, J. Shuttle vectors for facile
943 gap repair cloning and integration into a neutral locus in *Candida albicans*. *Microbiol. (United*
944 *Kingdom)* **159,** 565–579 (2013).
945 42. Safonova, Y., Bankevich, A. & Pevzner, P. A. in 265–279 (Springer, Cham, 2014).
946 doi:10.1007/978-3-319-05269-4_21
947 43. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database
948 search programs. *Nucleic Acids Res.* **25,** 3389–402 (1997).
949 44. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-
950 performance genomics data visualization and exploration. *Brief. Bioinform.* **14,** 178–92 (2013).
951 45. Stubbs, M. *et al.* MLL-AF9 and FLT3 cooperation in acute myelogenous leukemia: development
952 of a model for rapid therapeutic assessment. *Leukemia* **22,** 66–77 (2008).
953 46. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-
954 cell gene expression data. **33,** (2015).
955 47. Villani, A.-C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells,
956 monocytes, and progenitors. *Science (80-.).* **356,** (2017).

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

49. Karolchik, D. *et al.* The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32,** 493D–496 (2004).

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

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

52. 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).

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

54. 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).