Comparative transcriptomics of THP-1 monocytes in response to different pathogens

Ernst Thuer[1,2] and Toni Gabaldón[1,2,3]

[1] *Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain*
[2] *Universitat Pompeu Fabra (UPF). 08003 Barcelona, Spain.*
[3] *ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain.*

Undifferentiated human monocytes encounter various pathogens while present in the bloodstream. They are considered a primary responder and regulator for human immune reactions. As such, experiments investigating responses to pathogens are often reliant on monocyte cell cultures. For reproducibility reasons, immortalized cell lines are used. One of the most important cell lines used to model pathogen interactions is THP-1, which has been used in a variety of high throughput transcriptomics experiments. Yet as a cancer derived cell line it may no longer maintain its orignial functionality in detecting and responding to pathogens. Using available large scale transcriptomics datasets, we compare the response of THP-1 to a variety of human pathogens; viruses, bacteria, protozoa and fungi. Our approach focuses on the behavior of THP1 in its response to the different pathogens. Our aim is to provide comparative insights into the cell lines, which may serve to potentially improve future experimental design.

1 Introduction

Undifferentiated human monocytes reside in the bloodstream for up to three days before differentiating and moving into tissue. During this time, they are a primary responder to any invasive pathogen entering the bloodstream. Due to their primary role in coordinating host responses, they have been suggested as targets for immune augmentations strategies e.g during fungal infections (Segal, 2007). Most experiments investigating bloodstream infections to pathogens rely on the use of immortalized cell lines to reliably and reproducibly model potential interactions between humans and pathogens e.g (Leland and Ginocchio, 2007). A potential downside of such cell lines, usually derived from cancer lines, is that they are established after human cells have already mutated to exhibit unnatural behaviour (Kaur and Dufour, 2012).

A common cell line used to study the behavior of monocytes is the leukemia derived cell line THP-1 (Tsuchiya et al., 1980). This cell line has been used to study a variety of pathogens using RNA sequencing based transcriptomics. Experiments using THP-1 interaction models involve interactions with viruses, Ebola and Marburgvirus (Martinez et al., 2013), Zika (Hanners et al., 2016) and bacteria such as *Coxiella burnetii* (Millar et al., 2015), *Mycobacteria spp.* (Reyes et al., 1999; Zakharova et al., 2010) as well as the protist *Leishmania mexicana* (Millar et al., 2015). In a recent study, Toth et al. (Tóth et al., 2017) Investigated the response of the pathogenic yeast *Candida parapsilosis* to THP-1. Additionally, available data includes exposure to ethanol, calcitriol and Tissue-type Plasminogen Activator (TPA) (Barendsen, Mueller, and Chen, 1990). To our knowledge, no direct investigation into the comparative response behavior of the cell line THP-1 has been carried out so far. In this study we hope to provide insights into the behavior of THP-1 if

exposed to different human pathogens, and evaluate its ability to develop specific responses. To address how THP1 cells respond to the different stimuli, we investigated transcriptional profiles via RNA sequencing for the human THP1 cell line after exposure to the above mentioned pathogens and chemicals. To reduce analytical bias from the individual experiments, our analysis began with raw RNA sequencing data available at the NCBI Sequence Read Archive for the individual projects. A common, standardized pipeline was applied to carry out the data processing. We focused the subsequent analysis on the inter-project response for the individual pathogens. In order to overcome the very different experimental setups, we relied on more global approaches for tanscriptomic analysis. An important focus was the ability of the cancer derived THP-1 cell lines general response to distinguish the individual pathogens. Specific responses have been investigated in the individual experiments, yet such a variety of pathogens is expected to trigger substantially different overall response pathways. We investigated the impact of the individual pathogens via dimensionality reduction based clustering, and comparative GO term enrichment. Both on large scale to compare the overall behavior of the cells, and on the two available time course analyses to investigate the more minute temporal dynamic of transcription shifting. The available time course analyses comprise the bacterium and intracellular pathogen *Mycobacterium abscessus* and the yeast *Candida parapsilosis*. The *mycobacteria* species consist of a range of bacteria best studied for causing tuberculosis. They are documented to be fast growing and potentially multidrug resistant. The *M. abscessus* complex is also resistant to disinfectants and, therefore, can cause postsurgical and postprocedural infections (Lee et al., 2015). Candida species cause common nosochomial infections (Casadevall and Pirofski, 1999). As yeast, their mechanisms of pathogenicity differs significantly from that of bacteria. The yeast potentially inducing a much weaker response, testing the limits of THP-1 transcriptome adaptation.

2 Material and Methods

2.1 RNA sequencing data processing

RNA sequencing data was obtained in its raw read format from the Sequence read archive (Leinonen, Sugawara, and Shumway, 2011), with the exception of the data for *C. parapsilosis*, which was provided by the authors of Toth et al. (Tóth et al., 2017), the full list of sequence runs can be found in Table 1. A shell script to initiate the full download can be found at the github repository https://github.com/Gabaldonlab/THP1data . After extraction using the sratoolkit. Trimming, for quality pre-processing, of the reads was performed via Trimmomatic v0.32 (Bolger, Lohse, and Usadel, 2014). We mapped the reads using the STAR (Dobin et al., 2013)

mapper against the hg38 human reference genome. Human genomic data hg38 v 81., both the reference and annotation files were obtained from UCSC (Ucsc and Browser, 2003). Reads were counted using the htseq package (Anders, Pyl, and Huber, 2015). For the analysis only annotated exons were considered. An overview of samples considered is presented in Table 1.

2.2 Data processing

Read count normalization was performed via transcript per million TPM normalization. R libraries were used to investigate the Principal Components underlying the data variability. The R built in prcomp module and the library FactoMineR (Le, Josse, and Husson, 2008) were used to compute the PCA and cluster estimation respectively. Tree based hierarchical clustering was carried out using the python scipy library. Gene enrichment was analyzed via python scripts available on the projects github https://github.com/Gabaldonlab/THP1data , generating a background model of variance. The enrichment compared to the full human background was carried out using the GOrilla tool (Eden et al., 2009)

2.3 Visualization

Visualization was performed via the R module ggbiplot, based on the ggplot2 library, as well as the FactoMineR and superheat plotting function for the respective R scripts. Visualization in python was produced via matplotlib.

2.4 Enrichment analysis

Expression enrichment for unregulated genes was computed for each gene on the variance over the non pathogen derived conditions, uninfected cells and separately against the chemicals ethanol and calcitriol. Outliers were tested against a normal distribution using student t-test. The method described by Benjamini & Hochberg (Benjamini and Hochberg, 1995) was used to correct for multiple testing and evaluate false discovery rate (FDR), in order to correct the resulting p-values. Adjusted p-values of < 0.05 were considered significant and analyzed by GOrilla against a total background.

3 Results and Discussion

As shown in Figure 1 the Principal Component Analysis showed a clear and distinct response to the individual conditions.

This can be considered an important sign that the THP1 cell line has retained its potential for detection of the individual pathogens in initiating individual responses. In a PCA, the abstract underlying effects are quantified and shown on components or axes, with relative strength per axis denoted in percent of variance explained. Individual principal components can show multidimensional response factors. Dimensions are

Comparative transcriptomics of THP-1 monocytes in response to different pathogens

Pathogen	Sample count	Time course	Quality control	Archive
Mycobacterium bovis	14	No	failed	ERR5604
Mycobacterium abscessus	19	Yes	pass	SRR23160
Zika	2	No	pass	SRR51901
Ebola	2	No	failed	SRR16602
Marburgvirus	2	No	failed	SRR16367
Leishmania mexicana	4	No	pass	SRR1562
Coxiella burnetii	4	No	pass	SRR1562
Candida parapsilosis	16	Yes	pass	from author
Staphylococcus aureus	9	No	failed	ERR50285
Non Pathogen	5	No	pass	SRR16365

Table 1: *Table of RNA seq runs processed in the analysis. The columns indicate, in this order: number of samples, pass of the mentioned quality check and accession number.*
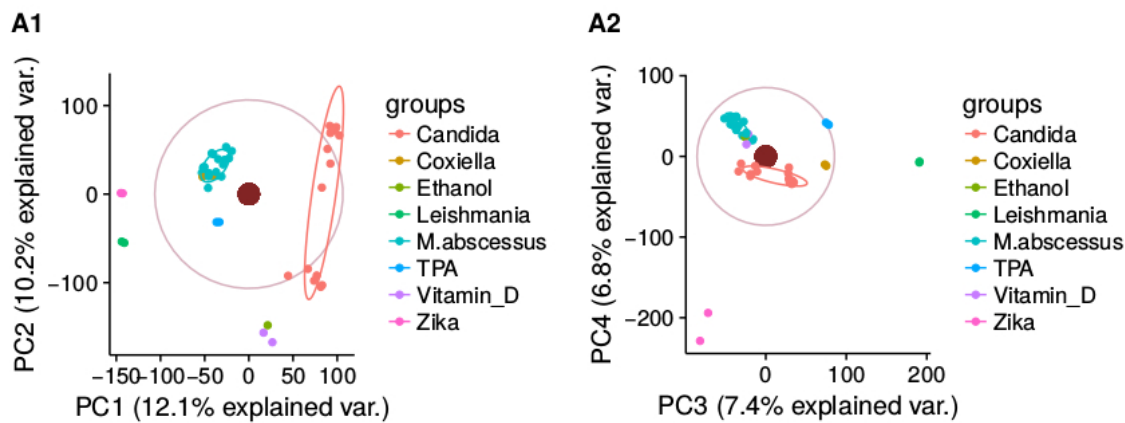

Figure 1: *Principal component analysis for all genes, normalized to TPM. Due to the relative low variation per component the first four dimensions are displayed in two 2dimensional plots. Components are displayed in two plots PC1 and 2 in plot A1 and PC3 and 4 in A2. In A1, a cluster separation between the different pathogens and stressors can be observed on the first component. Two clusters comprised of the timeline experiments on* M. abscessus *and* C.parapsilosis *are visible along the first axis. A2 seperates viruses more clearly from the protist, as well as the intracellular bacteria* Coxiella bruneii *and* M. abscessus.

visualized in Figure 1 Due to the complexity of analysis, the first four dimensions were considered to explain sufficient variance, collectively accounting for 36.5% of the observed variance. Similar responses were observed between the intracellular bacterial pathogens *Coxiella burnetii* and *Mycobacterium abscessus*, derived from experiments performed by independent groups [Table 1], suggesting that the first components are not influenced by the sequencing but directly by the THP-1 response, the profile of the two strains diverges in the third component showing more nuanced differences in response of THP-1 between the two pathogens.

Overall four distinct response clusters can be observed. With separation of clusters occurring for virus to yeast in the first component, and a distinction of bacteria over the second and third. The intracellular mycobacterium *M. abscessus* and the yeast *C. parapsilosis* are the most robust groups due to a larger sample size of 19 and 16 runs respectively, and replicates over

a time course of infection assay. Factorial analysis to investigate time point responses was therefore limited to those two species. To gain a more detailed view on the minute behavior we investigated the response to the two larger time course analysis. Data was produced for rough and smooth morphologies during exposure times of 1, 4 and 24 hours. Although the analysis shows a clear separation (see Figure 2), the primary component derives from the effect between the replicates. The response to *C. parapsilosis* was less homogeneous, most likely due to the lack of replicates and the overall lower pathogenicity of *Candida* as compared to *Mycobacteria*.

In the next step, we quantified overall transcriptional responses against background noise models visualized in Figure 3. Two Noise models were designed. In the first, (background) we used the average counts per gene in uninfected samples, to evaluate the expression compared to an uninfected baseline. For the second model (stressors) we evaluated the average gene count for samples treated with TPA, ethanol and the Vita-
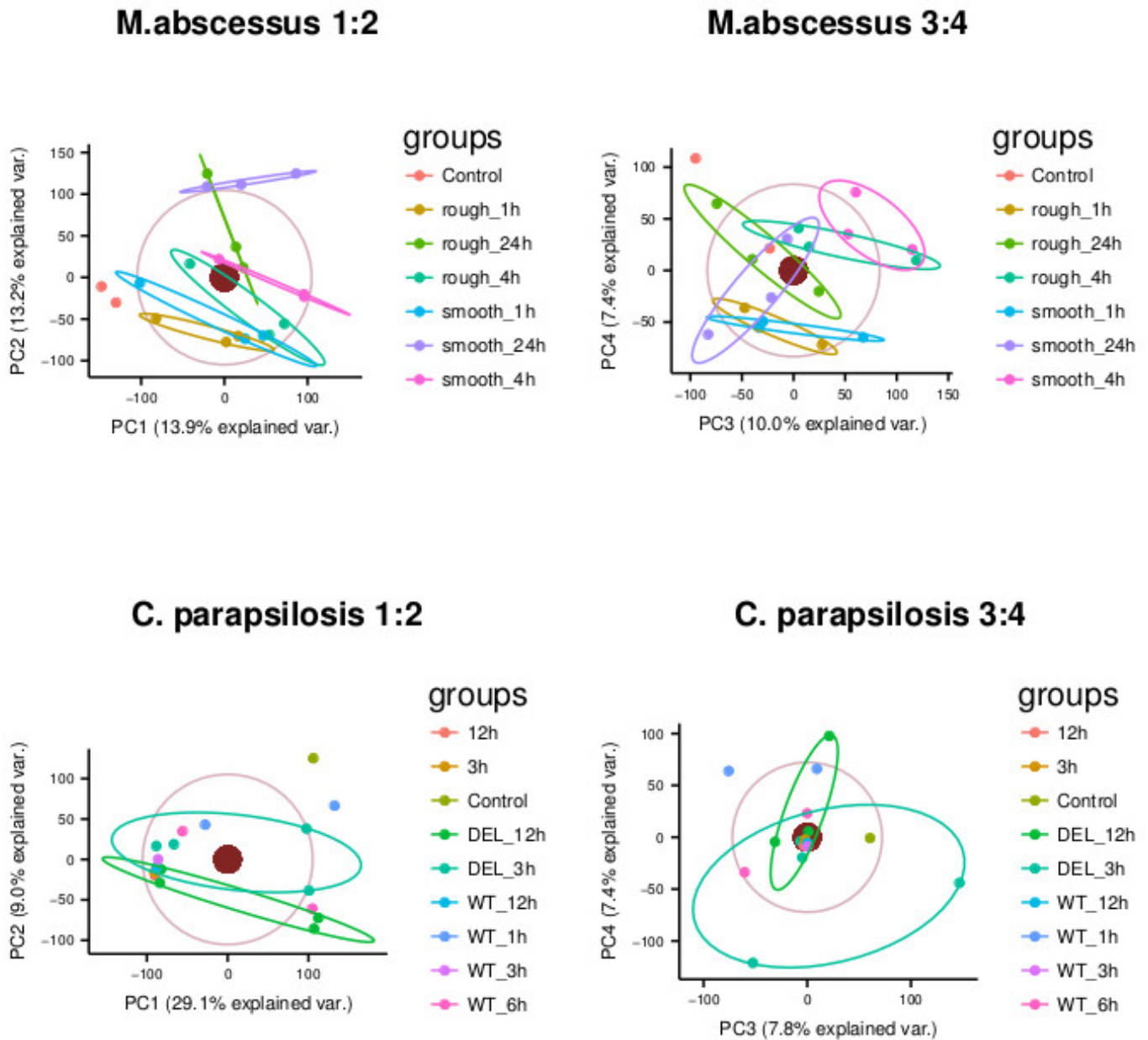
Figure 2: *PCA of the time course analysis for the infividual sets of* M. abscessus *(above) and* C. parapsilosis *(below). Both studies show variance between technical replicates to be responsible for their first component, suggesting a lack of regulation by the THP-1 cells, or rather a lack of designated response. While* M. abscessus *shows a distinguishable variability over the time course, clustering on* C. parapsilosis *does not separate the individual conditions.*

Mycobacterium abscessus		
RNA biosynthetic/metabolic process	defense response to virus	type I interferon signaling pathway
Zika virus		
forebrain development	triglyceride catabolic process	neutral lipid catabolic process
Leishmania mexicana		
transmembrane transport	ion transport	organic acid transmemb. transport
Coxiella brunetti		
	None	
Candida parapsilosis		
mitotic cell cycle process	leukocyte activation	neutrophil activation

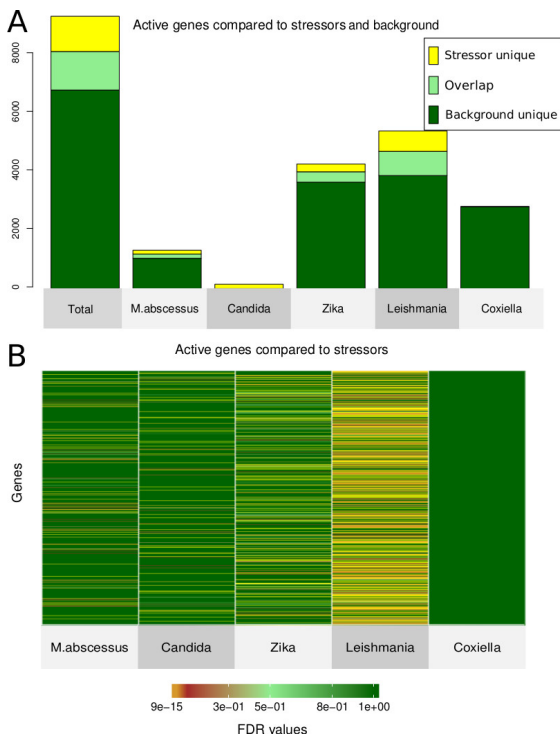**Table 2:** *GO enrichment of pathogen responses against a stressor background*

Figure 3: *(A) Barplot visualizing the quantification of response against the two background models. Unique responses in yellow and dark green are only observed against the specific background. Overlap in light green shows responses similar between stress and uninfected cells. (B) shows a heatmap of individual genes actively regulated in the 5 species against the stress response. Notably,* Coxiella burnetii *shows no significantly enriched genes against the stressor subset.*

min D metabolite calcitriol to evaluate basic stress responses. Individual genes were tested for upregulation only against the model genes, returning a value of significance per gene and pathogen. This methods ignores experimental design in order to generalize the various experiments. Figure 3 shows the overlap between the background and stressor comparison. In total, 9302 genes were activated in any pathogen response compared to both backgrounds, with 1318 genes overlapping between the noise models. 6780 and 1204 genes

were unique to the background and stressors model, respectively. This indicates an active response by THP-1, and its ability to distinguish uninfected surroundings to chemical stimuli.The cells show a clear distinction between the responses to the individual pathogens. Yet, the THP-1 response to *C. parapsilosis* shows no significant difference to the background model, but does present a distinction to the stressors model. *C. parapsilosis*, as a pathogenic yeast, is often commensal e.g (Gabaldon, Naranjo-Ortiz, and Marcet-Houben, 2016), and seemingly does not provoke a general strong transcriptomic response in the host. Interestingly, compared to the stressor background, only the *M. abscessus* cells show a classified defense response. According to GO terms, the response is antiviral [Table 2].

An antiviral response to *M. abscessus* is partially expected, due to the related *M. tuberculosis* ability to trigger Interferon responses (Prabhakar et al., 2003), and Interferon production in general T-cell reponses (Belardelli and Gresser, 2010). *L. mexicana* shows the strongest overall response. With 3861 more genes activated than in the background, and a response of 694 unique genes to the *L. mexicana* stress compared to the stressor background. GO enrichment using GOrilla visualized in Table 2 shows unique responses for the pathogens. Most notably, the most enriched GO term for Zika is GO:0030900, pertaining to anatomical structure and forebrain development. This is in line with Zikas clinical symptoms, such as microcephalus as described in this review (Paixo et al., 2016). Enrichment for *L. mexicana* showed active genes involved in iron transport, an observation in line with Huynh et aliis (Huynh, Sacks, and Andrews, 2006) discovery of iron transporters being essential for parasitic reproduction. Active Iron transportation could therefore be an expected host response.

4 Conclusion

Our analysis suggests that the THP-1 cell line is capable of distinguishing various cellular stresses, and provide individual responses to various chemicals and pathogens. It accurately portrays gene enrichment for e.g Zika clinical symptoms. The large scale transcrip-

tomic response is uniquely different for each analyzed experiment, and the cells show similar responses to the intracellular pathogens e.g *Coxiella burnetii* and *Mycobacterium abscessus*. Yet, the experimental resolution is less pronounced in more detailed experiments. Time course analysis using THP-1 showed a stronger variation between technical replicates than the actual experimental course. A recent study by Schurch et al (Schurch et al., 2016) estimates the number of true positives in RNA sequencing with 3 replicates to be between 20 and 40%. Using the first components as indicators of variance, we estimate that at least 12% of the total variance are attributed to technical variation. By using general probabiliy, we can estimate the true positives for triplicates to be between 13.6 and 27.26% using THP-1 cells (according to $(1 - (1 - \alpha)^n)$). There-fore, especially for the investigation of pathogens with an expected mild response, additional replicates are strongly recommended.

Bibliography

Anders, Simon, Paul Theodor Pyl, and Wolfgang Hu-ber (2015). "HTSeq-A Python framework to work with high-throughput sequencing data". In: *Bioinformatics* 31.2, pp. 166–169. ISSN: 14602059. DOI: 10.1093/bioinformatics/btu638.

Barendsen, N, M Mueller, and B. Chen (1990). "In-hibition of TPA-induced monocytic differentiation in THP-1 human monocytic leukemic cells by stau-rosporine, a potent protein kinase C inhibitor." In: *Leuk Res.* 5, pp. 467–74.

Belardelli, Filippo and Ion Gresser (2010). "The ne-glected role of type I interferon in the T-cell response: implications for its clinical use". In: *Immunology Today* 8, pp. 369–72.

Benjamini, Yoav and Yosef Hochberg (1995). "Con-trolling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing". In: *Journal of the Royal Statistical Society* 57.1, pp. 289–300.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel (2014). "Trimmomatic: A flexible trimmer for Il-lumina sequence data". In: *Bioinformatics* 30.15, pp. 2114–2120. ISSN: 14602059. DOI: 10.1093/bioinformatics/btu170.

Casadevall, A and L Pirofski (1999). "Host-pathogen interactions: redefi ning the basic concepts of viru-lence and pathogenicity". In: *Infection and immunity* 67.8, pp. 3703–3713. ISSN: 0019-9567.

Dobin, Alexander et al. (2013). "STAR: Ultrafast uni-versal RNA-seq aligner". In: *Bioinformatics* 29.1, pp. 15–21. ISSN: 13674803. DOI: 10.1093/bioinformatics/bts635.

Eden, Eran et al. (2009). "GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists". In: *BMC Bioinformatics* 10.1, p. 48. ISSN: 1471-2105. DOI: 10.1186/1471-2105-10-48. URL: http://www.biomedcentral.com/1471-2105/10/48.

Gabaldon, Toni, Miguel A. Naranjo-Ortiz, and Marina Marcet-Houben (2016). "Evolutionary genomics of yeast pathogens in the Saccharomycotina". In: *FEMS Yeast Research* 16.6, pp. 1–10. ISSN: 15671364. DOI: 10.1093/femsyr/fow064.

Hanners, Natasha W. et al. (2016). "Western Zika Virus in Human Fetal Neural Progenitors Persists Long Term with Partial Cytopathic and Limited Immuno-genic Effects". In: *Cell Reports* 15.11, pp. 2315–2322. ISSN: 22111247. DOI: 10.1016/j.celrep.2016.05.075. URL: http://dx.doi.org/10.1016/j.celrep.2016.05.075.

Huynh, Chau, David L. Sacks, and Norma W. An-drews (2006). "A <i>Leishmania amazonensis</i> ZIP family iron transporter is essential for para-site replication within macrophage phagolysosomes". In: *The Journal of Experimental Medicine* 203.10, pp. 2363–2375. ISSN: 0022-1007. DOI: 10.1084/jem.20060559. URL: http://www.jem.org/lookup/doi/10.1084/jem.20060559.

Kaur, Gurvinder and Jannette M Dufour (2012). "Cell lines: Valuable tools or useless artifacts." In: *Sper-matogenesis* 2.1, pp. 1–5. ISSN: 2156-5554. DOI: 10.4161/spmg.19885. URL: http://www.ncbi.nlm.nih.gov/pubmed/22553484$\backslash$nhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3341241.

Le, Sebastien, Julie Josse, and Francois Husson (2008). "FactoMineR: An R Package for Multivariate Anal-ysis". In: *Journal of Statistical Software* 1. DOI: 10.18637/jss.v025.i01.

Lee, Meng Rui et al. (2015). "Mycobacterium abscessus complex infections in humans". In: *Emerging Infec-tious Diseases* 21.9, pp. 1638–1646. ISSN: 10806059. DOI: 10.3201/eid2109.141634.

Leinonen, Rasko, Hideaki Sugawara, and Martin Shumway (2011). "The sequence read archive". In: *Nucleic Acids Research* 39.SUPPL. 1, pp. 2010–2012. ISSN: 03051048. DOI: 10.1093/nar/gkq1019.

Leland, Diane S. and Christine C. Ginocchio (2007). "Role of cell culture for virus detection in the age of technology". In: *Clinical Microbiology Reviews* 20.1, pp. 49–78. ISSN: 08938512. DOI: 10.1128/CMR.00002-06.

Martinez, O. et al. (2013). "Ebola Virus Exploits a Monocyte Differentiation Program To Promote Its Entry". In: *Journal of Virology* 87.7, pp. 3801–3814. ISSN: 0022-538X. DOI: 10.1128/JVI.02695-12. URL: http://jvi.asm.org/cgi/doi/10.1128/JVI.02695-12.

Millar, Jess A et al. (2015). "Coxiella burnetii and Leishmania mexicana residing within similar para-sitophorous vacuoles elicit disparate host responses." In: *Frontiers in microbiology* 6.August, p. 794. ISSN: 1664-302X. DOI: 10.3389/fmicb.2015.00794. URL: http://journal.frontiersin.org/article/10.3389/fmicb.2015.00794/abstract.

Paixo, Enny S. et al. (2016). "History, epidemiology, and clinical manifestations of Zika: A systematic review". In: *American Journal of Public Health* 106.4, pp. 606–612. ISSN: 15410048. DOI: 10.2105/AJPH.2016.303112.

Prabhakar, Savita et al. (2003). "Inhibition of Response to Alpha Interferon by Mycobacterium tuberculosis". In: *Infect Immun* 5, 2487–2497.

Reyes, Leticia et al. (1999). "Effects of Mycoplasma fermentans incognitus on differentiation of THP-1 cells". In: *Infection and Immunity* 67.7, pp. 3188–3192. ISSN: 00199567.

Schurch, Nicholas J. et al. (2016). "How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?" In: *Rna* 22.6, pp. 839–851. ISSN: 1355-8382. DOI: 10.1261/rna.053959.115. arXiv: 1505.02017. URL: http://rnajournal.cshlp.org/lookup/doi/10.1261/rna.053959.115.

Segal, Brahm H (2007). "Role of macrophages in host defense against aspergillosis and strategies for immune augmentation." In: *The oncologist* 12 Suppl 2.Supplement 2, pp. 7–13. ISSN: 1083-7159. DOI: 10.1634/theoncologist.12-S2-7. URL: http://theoncologist.alphamedpress.org/content/12/suppl{_}2/7.full.

Tóth, Adél et al. (2017). "Investigation of Candida parapsilosis virulence regulatory networks following host-pathogen interaction". In: *PLOS Pathogen, submitted*.

Tsuchiya, S et al. (1980). "Establishment and characterization of a human acute monocytic leukemia cell line (THP-1)". In: *Int J Cancer*, pp. 171–6.

Ucsc, The and Genome Browser (2003). "The UCSC Genome Browser". In: *Curr Protoc Bioinformatics*, pp. 1–23. DOI: 10.1002/0471250953.bi0104s28. The.

Zakharova, Ekaterina et al. (2010). "Mycoplasma suppression of THP-1 cell TLR responses is corrected with antibiotics". In: *PLoS ONE* 5.3, pp. 3–6. ISSN: 19326203. DOI: 10.1371/journal.pone.0009900.