Step I: White Paper Application

Application Guidelines

1. *The application should be submitted electronically per requirements via the web site of any of the NIAID Genomic Sequencing Centers for Infectious Diseases. Include all attachments, if any, to the application.*
2. *There are no submission deadlines; white papers can be submitted at anytime.*
3. *GSC personnel at any of the three Centers can assist / guide you in preparing the white paper.*
4. *Investigators can expect to receive a response within 4-6 weeks after submission.*
5. *Upon approval of the white paper, the NIAID Project Officer will assign the project to a NIAID GSC to develop a management plan in conjunction with the participating scientists.*

White Paper Application

Project Title: Human and *Toxoplasma gondii* Genetics and Cellular/Molecular Interactions

Authors: Rima McLeod, Fiona Henriquez, Ying Zhou, and Kamal El Bissati (please also see collaborators and their letters; As discussed with NIH DMID Program [F.Lee Hall, S.Rosenthal, and Deirdre Joy] this will be a companion for our NIH grant and application). Huan Ngo also contributed to some parts of the formative stages of this work. Discussed with Hernan Lorenzi and Rempert Pieper.

Primary Investigator Contact:

Name	Rima McLeod
Position	Professor
Institution	University of Chicago
Address	5841 S Maryland Ave
State	Illinois
ZIP Code	60637
Telephone	773-834-4130
Fax	773-834-3577
E-Mail	rmcleod@uchicago.edu

1.Executive Summary *(397 words.*"added value" = activities not in scope of work / deliverables of project)

The overarching goal of the proposed work is to define interactions of differing isolates and molecules of *T.gondii* with biologically relevant human host cell molecules, addressing both critical tachyzoite and encysted bradyzoite parasite life cycle stages. The purpose of this work is to identify key host and parasite molecules that interact in pathogenesis of and protection against human toxoplasmosis. Our methodology will involve infecting relevant human host cells with parasites of differing lineages, of differing life cycle stages and focusing on relevant host cell types from stem to differentiated cells. We will generate transcriptional profiles. We will also utilize laser capture of encysted organisms and tachyzoites within neuronal cells in human brain and eye tissue to characterize transcriptomes of parasites, host and contiguous cells in situ. This will characterize how genetically different parasites alter host and parasite proteins and genes. Functional genomics data will be acquired using high-throughput transcriptional profiling (mRNASeq and miRNASeq) and, with more limited scope, proteomics of host and pathogen preparations and studies of host-pathogen interactions using yeast two hybrid. Following completion of these experiments, molecular networks of host-pathogen interactions will be established. The data from these analyses will be made available to the research community through their deposition in GenBank and ToxoDB, in accordance with the NIAID guidelines. This study will identify pathways that influence disease outcomes from the infection. These data concerning affected pathways in the host cells herein thus can then be compared with relevant pathways in the host cell identified through separate genetic studies of families. These families have a child with congenital toxoplasmosis in the United States National Collaborative Chicago Based Congenital Toxoplasmosis Study (US NCCCTS) or twin cohorts. Of interest, in a pilot run for the first genes approached in this manner that were imprinted, we identified many novel variants, mutations, deletions, duplications when only a few would otherwise have been predicted. Our expectation concerning the important impact of this work is that transcriptomics, proteomics, and interactomics data generated will allow us to establish key signatures for relevant human cells and their interaction with *T. gondii* cells, molecules and life cycles stages. We will learn how critical human and parasite genes and cell types/strains influence these signatures. It will provide a beginning relevant database and reference source for all the *T. gondii* research community and a basis for continuing work.

2. Justification

Provide a succinct justification for the sequencing or genotyping study by describing the significance of the problem and providing other relevant background information.

This section is a key evaluation criterion.

1. *State the relevance to infectious disease for the organism(s) to be studied; for example the public health significance, model system etc.*

2. *Are there genome data for organisms in the same phylum / class / family / genus? What is the status of other sequencing / genotyping projects on the same organism including current and past projects of the NIAID GSC? Provide information on other characteristics (genome size, GC content, repetitive DNA, pre-existing arrays etc.) relevant to the proposed study. Have analyses been performed on the raw data already generated/published? If additional strains are proposed for a species, please provide a justification for additional strains?*

3. *If analyses have been conducted, briefly describe utility of the new sequencing or genotyping information with an explanation of how the proposed study to generate additional data will advance diagnostics, therapeutics, epidemiology, vaccines, or basic knowledge such as species diversity, evolution, virulence, etc. of the proposed organism to be studied.*

1.Relevance:

Toxoplasmosis is a devastating disease. It is the most common infection of the retina, causing loss of sight. This parasite is life threatening and damages the eyes and brain when it infects a fetus *in utero* before the immune system is mature, if untreated. Almost all such children have damage at birth or to their eyes by the time they are teenagers. Many are reported to have cognitive impairment and some develop seizures or motor abnormalities secondary to infection by early adolescence. Disease is also devastating and life threatening in immune compromised persons such as those with malignancies or their treatment, transplantation, autoimmune disease, or AIDS.

Toxoplasma gondii is the most common parasitic infection worldwide. Thirty percent of all persons throughout the world are infected. Consequences of carrying this parasite in brain across a lifetime for otherwise healthy adults is unknown. Although a number of chronic diseases and behavioral abnormalities have been ascribed to the infection, cause and effect is not proven. The active form of the parasite can be treated, although available medicines are limited by hypersensitivity and toxicity. No available medicines eliminate dormant, latent parasites present in chronic infection. Thus recurrences are possible from the dormant organism at any time. There is no vaccine for humans.

The transcriptomics, proteomics and interactomics data generated by this project will allow the identification of novel host and parasite genes and pathways that play a key role during infection. This will provide the bases for new lines of research for the development of new therapies, anti-parasitic medicines and vaccines.

2. Existing Genome Data:

Toxoplasma gondii:
There are data in the ToxoDB database and GenBank in which genomes, proteomes, cell cycle and stage-specific transcriptomes for parasites have been collected and made public.

As part of the*Toxoplasma gondii* MSC and GSC sequencing projects three canonical parasite strains have been sequenced (GT1, ME49 and VEG) and genome and transcriptome sequencing is underway for additional genetically divergent parasites. Parasites with differing genetics cause different phenotypes in tissue culture, and in mice.

Human Host: There is an ongoing series of studies and plans to accumulate additional host data in an interactome with signatures from different cells and parasite strains. This has not been accomplished yet.

The parasite takes over its host cells subverting their normal functions. Some candidate parasite molecules for determining resistance and susceptibility by interacting host proteins in murine macrophages or human fibroblasts have been identified. There is one study of a human brain tumor cell line and one with monocytes [2, 3].

Other analyses:

We have found recently that parasite and host genetics influence outcomes of infection profoundly in people in the US. In addition, using transmission disequilibrium testing, and then using the NIH disease association bioinformatics tools and our preliminary studies in which we have generated transcriptomes, we have found pathways of human genes (Figure 1) critical for susceptibility/resistance to human congenital toxoplasmosis.
Some of the genes are obvious candidates, i.e., genes from innate or adaptive immunity

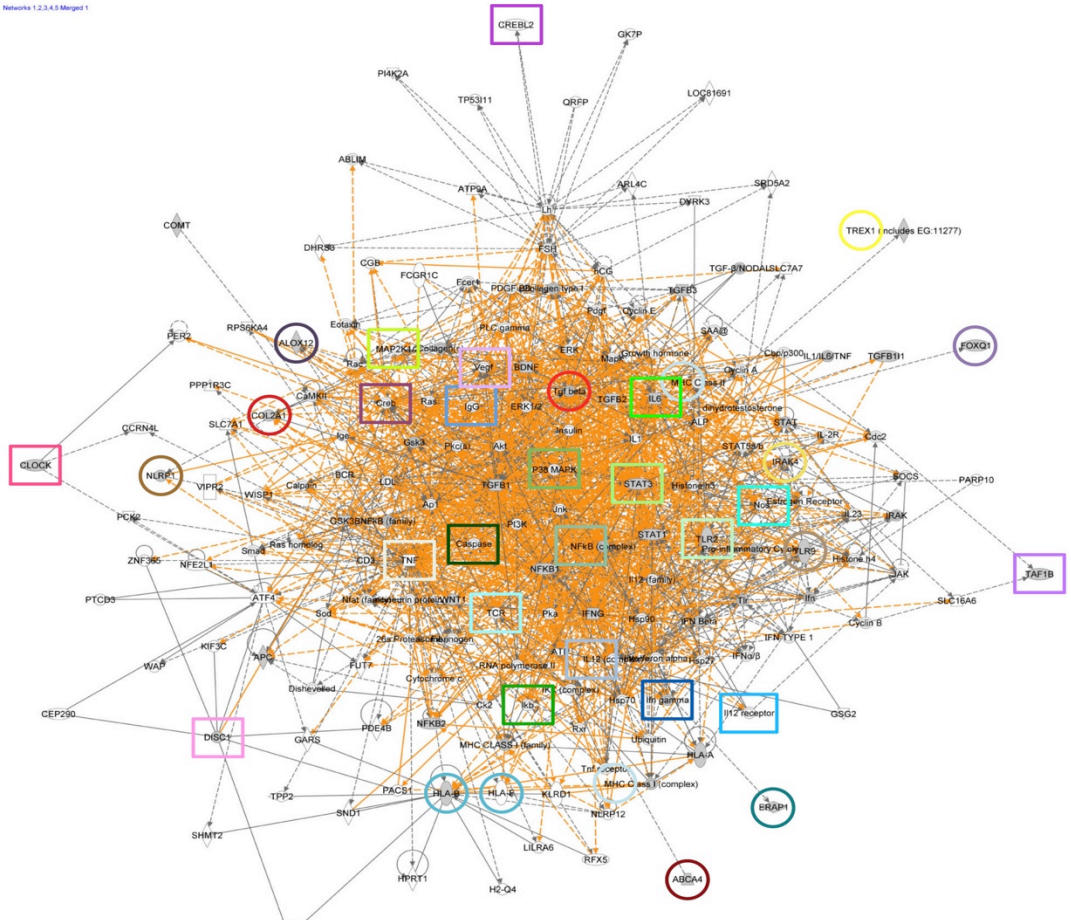


Figure 1. Pathways in resistance and susceptibility to human toxoplasmosis

pathways that have been defined in mice, or for other diseases, but some are completely unexpected. For example ***COL2A*** and ***ABC4R*** were discovered because they produce similar clinical phenotypes for other diseases (e.g, ***COL2A*** and ***ABC4R*** for hydrocephalous, or ***TREX*** for calcifications) or because experiments of others suggested

their relevance, e.g., ***GRA 15*** and ***NFκB*** [4]. Figure 2 shows these genes.

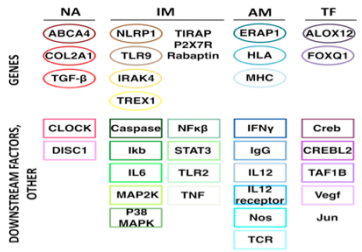


Figure 2. Genes identified by candidate gene approach in resistance to toxoplasmosis

3.Utility of new sequencing

3.1. Transcriptomics

Herein, we plan to generate miRNA-seq and mRNA-seq data of biologically relevant human cells and tissues infected with *T. gondii* to further identify and characterize pathways critical for resistance/susceptibility to toxoplasmosis. This will be accomplished using differences in parasite genetics and host cell phenotypes to understand pathogenesis. We will sequence transcriptomes of human neuronal stem cells, neuronal cells, and monocytic cells. These cells comprise the primary and most relevant target organs in humans associated with the most devastating consequences. They will be uninfected (controls) or infected with *T. gondii* strains from different haplotypes (I, II, III, IV, Brazilian [TgPBr] and Wild Amazonian [GUY[RUB]]) strains. These are listed in Table 1 with their phenotypes. We propose to define transcriptional signatures that will allow the identification of new candidate genes and pathways responsible for modulating host susceptibility or resistance to toxoplasmosis. We will also use parasites available now with key genes knocked out or overexpressed found to modulate the same pathways we identify through genotyping performed earlier or separately later (Table 1).

We will generate as well RNA-seq data of different transgenic human monomac cell lines carrying knockdowns for some of the genes shown in Figure 2.These are associated with susceptibility/resistance to human toxoplasmosis. They include ALOX12, NALP1, NFκβ, AP-1 and TGFβ. This is important because these genes are key central nodes in pathways. Thus they likely regulate susceptibility/resistance networks manipulated by secreted *T.gondii* proteins that determine parasite virulence and pathogenicity. Knockdown of ALOX12 and NALP1 show increased parasite replication and pathogenicity by type I parasites in monomacs cells.

In addition, to shed light into biology of *Toxoplasma* cysts and their interaction with the host we will do transcriptomics from individual bradyzoites cysts and corresponding contiguous human cells dissected from infected human tissues (eye, brain) using laser capture microscopy.

3.2. Proteomics and Yeast Two Hybrid

In contrast to transcriptional profiling, proteomics does not allow multiplexed global analyses of host and parasite proteins. Instead of generating a proteomic complement of the RNA-seq profiles of numerous strains/monocyte cell types, we propose proteomic efforts of limited scope which add functional insights into RNA-seq-based pathway analysis. One of these efforts will focus on global phosphorylation of signaling proteins, the other effort will focus on protein-protein interactions. Both studies are clearly important for molecular pathogenesis (see Figures 4P1 and 2); both studies add value to the host-pathogen network even if performed only on a small number of experimental conditions/genes. A phosphoproteomic study will compare the MonoMac 6 cell line infected with two different *T. gondii* strains (wild type and an insertional mutant that escapes the resistance phenotype in rat and human cells). If this study yields significant insights into pathogenesis, it will allow us to initiate large-scale efforts in proteomics with the *Toxoplasma* community. It is hypothesized that genetic variations among *T. gondii*

strains trigger altered phosphorylation-dependent signaling cascades, e.g. that pertaining to NFkB, thus influencing the outcome of the infection.
The protein-protein interaction analysis will be the first true interaction analysis of the *GRA10* gene, delineating involvement of this gene in escape from host defense in conjunction with transcriptome data. It should be feasible to recover at least 10 million *T.gondii* tachyzotes in each condition making it possible to study both parasite and host phosphoproteomes.
Key parasite molecules modify relevant host cells in ways essential for the parasite to take over human cells when the cells are infected by *T. gondii*. Some of these parasite genes/proteins thereby may become key targets for vaccines and medicines to interrupt the processes in which *T. gondii* takes over its human host cells during infection.

3. Rationale for Strain Selection

4. *Provide the rationale behind the selection of strains and the number of strains proposed in the study. The focus of the program is on potential agents of bioterrorism or organisms responsible for emerging or re-emerging infectious diseases. Non-select agents or non-pathogenic organisms will be considered when they can provide insight into these scientific areas.*

4. Rationale

Parasites

We are focusing on the Class B bioterrorism agent *Toxoplasma gondii* that also is emerging with isolates of hyper virulence and considerable variability among isolates in the US. With global warming and travel, the potential for these emerging hyper-virulent isolates to have greater medical importance outside the Amazon region has developed.
Eighteen parasite strains (8 that occur in nature [RH, GT1, ME49, Prugneand, VEG, TgPBr, RAY, GUY-2004-Ang] and 19 transgenic) were chosen for the work proposed herein because they represent the range of phenotypes and pathogenicity for this parasitic infection and include the hypervirulent strains from South America (TgPBr.GUY 2004-Ang). It is now clear that in the US additional strains, not present in Europe, have been detected recently in domestic and wild animals and also as causing human disease (McLeod et al, CID, In Press). Table 1 shows strains we will study herein, their pathogenicity and haplotype.

Table 1: T. gondii strains, type,pathogenicity and genes modified

Type	Strain Names	Pathogenicity/virulence
I	RH, GT1	Virulent(lethal) in mice as tachyzoites
II	Prugneaud, Me49	Less virulent(lethal),makes most brain cysts in mice
III	Veg or Cepre	Intermediate virulence and cyst production
IV (HgXII)	RAY	Pathogenicity in mice not well characterized
I/III	TgPBr	Unknown in mice, eye disease in humans in Brazil
Amazonian-wild	GUY (RUB)	Hypervirulent in humans

Parasite Genes Knocked Out	Parental parasites	Mechanism of knockout and control
GRA 1	RH	Knockout
GRA 10	RH	Tetracycline repressor,promoter tet O
GRA 15	Prugneaud	Domain swap
ROP 5	RH	KO, KO Complemented
ROP 16	RH	KO, KO Complemented
ROP 18	RH	KO, KO Complemented
ROP 38	Veg	Wildtype, overexpressed
RPS 13	RH	Tetracycline repressor, promoter tet O
ENR	RH	Tetracycline repressor, promoter tet O

A number of strains have already been sequenced (RH, GT1, ME49, Prugneaud [PRU], VEG, RAY). Some genes that are responsible for the different pathogenicity of these

parasites also have been identified recently. Parasites with these key genes knocked down and/or overexpressed or swapped between strains, where they differ, will be used (Table 1). This will be a paradigm signature set and facilitate many future studies in which a library of parasite crosses between strains may be used. These can include those already available or those that will be made because of their specific phenotypes in order to identify new genes associated with different degrees of pathogenicity.

Life cycle stages in tissues.
What happens within a cyst and what happens around a cyst is really not known for any cells, and certainly not for human cells. Single cell transcriptomes are now feasible.

The following shows the results of laser capture microscopy of single cysts or single cells. Cysts are definitively identified with dolichos staining and contiguous samples can be identified with Hematoxylin and Eosin staining. With either the indicated cyst can be extracted and with frozen samples single cell RT-PCR has been performed successfully in our laboratory (see Figure 3 and Table 2 below). The perimeter of the cyst can also be identified and therefore cells in the vicinity of the cyst can be analyzed as well. This will be the first time that bradyzoite *in situ* transcriptomes have ever been obtained.

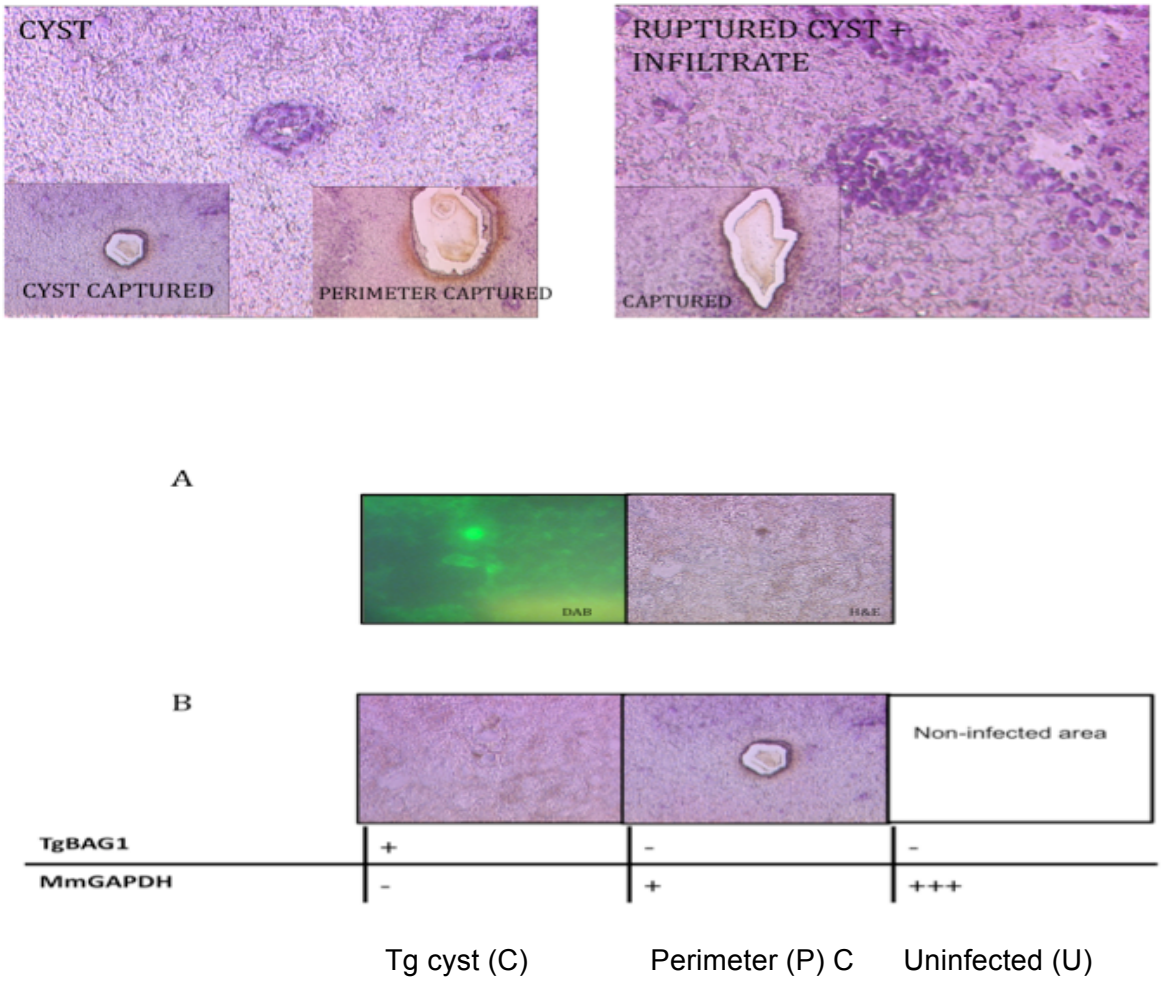


Figure 3: example of a dissected bradyzoite cyst and perimeter cells from a frozen mouse brain by laser capture microscopy.

PCR primers for:	*Tg* Cyst	PC Brain	U-Brain
TgBAG1	+	-	-
MmGAPDH	-	+	+++

Table 2: RT-PCR amplification results from bradyzoite cysts and perimeter cells isolated from frozen infected mouse brains by laser capture microscopy. *Tg Cyst,* dissected *T. gondii cyst*; PC Brain, dissected perimeter cells; U Brain = Uninfected mouse brain tissue; TgBAG1, Bradyzoite antigen 1 gene; MmGAPDH = constitutively expressed mouse GAPDH gene; +, RT-PCR positive; -, RT-PCR negative.

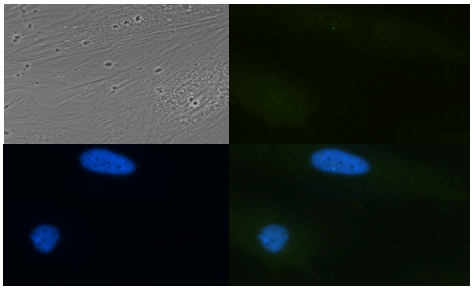
Human cells: Types of human cells are the other selection criteria we used based on those most relevant to human diseases caused by *T.gondii.* Thus, we have chosen 3 types: neuronal stem cells, neuronal differentiated cells and mononuclear phagocytic cells (Table 3). We have successfully cultured and infected these cells in the past and found differences between the stem cells and neuronal cells: (i) tachyzoites invade better and grow more rapidly in the stem cells than in the differentiated neuronal cells, and (ii) tyrosine hydroxylase (the rate limiting enzyme in the synthesis of dopamine by the parasite) is localized to the perimeter of the nucleus in infected differentiated neuronal cells but not in infected stem cells (see Figure 4 below). The parasites have dopamine which is expressed in parasites in both stem and differentiated human neurons. We now can capture isolated bradyzoites from within cysts with a laser and perform RT-PCR to determine genes transcribed in the cyst. Some images showing these findings are as follows:

Stem cells differentiating into neurons:

Figure 4. Stem cells differentiating to neurons, effect on T.gondii, stained for tyrosine hydroxylase

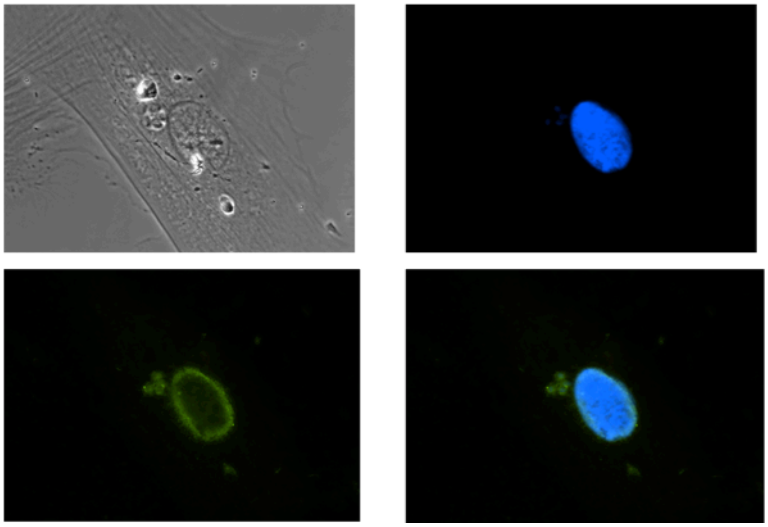
YFP expressing parasites

UNINFECTED STEM CELLS (CONTROLS)

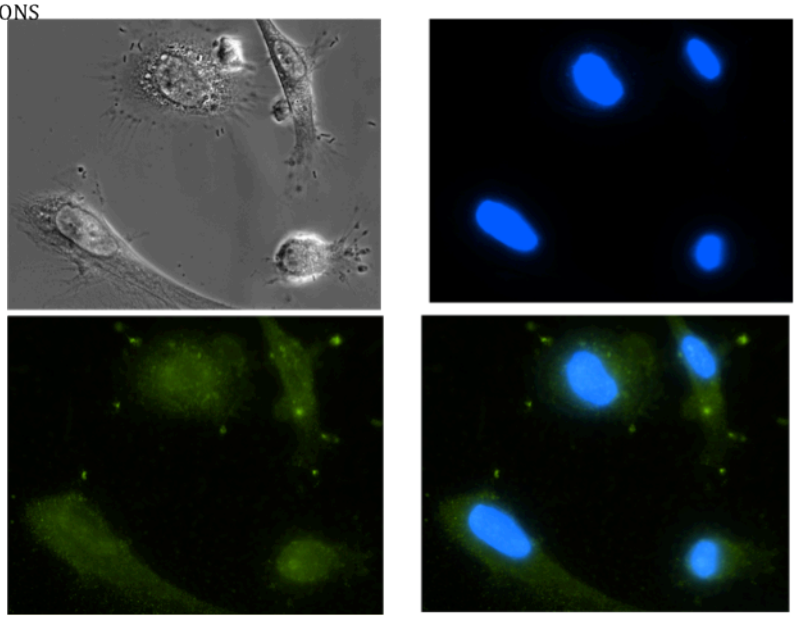


Staining for tyrosine hydroxylase (green) and nucleus (blue):

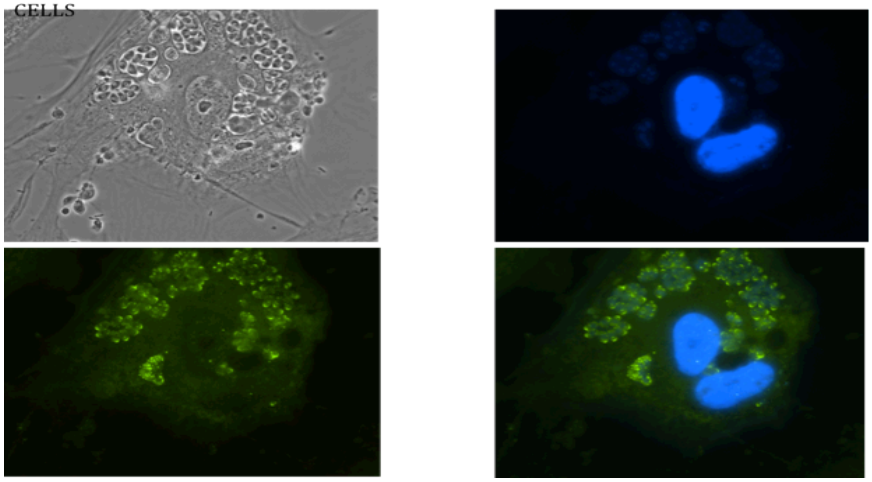
INFECTED NEURONS



UNINFECTED NEURONS



INFECTED STEM CELLS



Monomacs:

We have knockdowns of key susceptibility genes such as ALOX12 and NALP1 in monomacs cells. Thus, transcriptomes will also be performed plus and minus these genes (Table 3). The RNA has already been collected from these monomac cells infected with three types of parasites (I, II, III).

4a. Approach to Data Production: **Data Generation**

5. State the data and resources planned to be generated. (e.g draft genome sequences, finished sequence data, SNPs, DNA/protein arrays generation, clone generation etc.)

5. Data and Resources to be generated.

Parasite and host RNA-seq data generated from mRNA extracted from the following infections:

Table 3. RNA-seq datasets that will be generated by this white paper.

	Human cells (phase 1)			Human knockdown monomac cells (phase 2)				
Parasite types	NSC	NDC	Hmφ	Hmφ kd-ALOX12	Hmφ kd-NALP1	Hmφ kd-NFkBeta	Hmφ kd-AP1	Hmφ kd- TGFBeta
None	X	X	X	XC	X	X	X	X
I	X	X	X	XC	X		X	X
II	X	X	X			XC		
III	X	X	X					
IV	X	X	X					
BrazilianTgPBR	X	X	X					
Guyana TgGUY__RUB	X	X	X					
KO-GRA1	-	XC	XC					
KO-GRA10	-	XC	XC					
KO-GRA15	XC	-	XC					
KO-ROP5	-	-	XC					
KO-ROP16	-	XC	XC					
KO-ROP18	-	XC	-					
KO-RPS13	XC	XC	XC					
KO-ENR	XC	XC	XC					
KO-ROP38	-	XC	-					

Table 3: mRNA-seq/miRNA-seq datasets that will be generated by this white paper. NSC, neuronal stem cell; NDC, neuronal differentiated cell; Hmφ, monomac cells; kd, knockdown; KO, knockout, X/C assay conditions that will be carried out for mRNA/miRNA (X) transcriptional profiling of wild type stem cells, differentiated neurons and monomacs together with their respective controls when appropriate (C); Controls for GRA 1 and insertional mutants are wildtype parasites. Controls for GRA 10, RPS 13, ENR and LDH2 are conditional knockouts with a tet repressor in the presence (gene on) or absence (gene off) of anhydrotetracycline. Control for GRA 15 is domain swap parasite on a Prugneaud background. Controls for Rop5, 16, 18 are complemented knockout parasites. Control for ROP 38 overexpressing parasite is the wildtype parental parasite. Monomac controls are wildtype or have a lentivirus for the gene being knocked down, or an irrelevant construct. Irrelevant condition and wildtype are present once.

Infections/experiments will be performed in triplicates and **only two** of the replicates will be sequenced, resulting in a total of 136 mRNA-seq and 136 miRNA-seq datasets, including controls. The third separate replicate will be held in reserve if needed for repetitions if there is disparity between the two replicate experiments. It is anticipated that this might occur 6 times, totaling 142 mRNA-seq and 142 miRNA-seq datasets.

We will also generate in situ transcriptomics data from bradyzoites in cysts, tachyzoites, and host cells perimeter and uninfected, isolated using laser capture microscopy, initially comparing fixed and frozen and fresh tissue to assure that fixation does not alter transcription as described in other systems in the literature. These 4 additional laser capture experiments will be done in duplicates totaling 8 datasets.

In summary, the transcriptomics study proposed herein will generate 142 datasets (cases + controls) for mRNA-Seq and 142 for miRNA-seq plus 8 mRNA-seq datasets from laser-captured cells from infected tissue.

Complementing transcriptomics studies we will build phosphoproteomic profiles from human MonoMac-6 cells infected with wildtype or *GRA10* knockout parasites. *GRA10* knockout parasites can escape of cell death pheonotype in MonoMac 6 cells. Identification and quantitative determination of specific phosphoproteins, comparing phosphorilation profiles between MonoMac cells infected with wild type or *GRA10* mutant *Toxoplasma* strains, will allow us to interrogate hypotheses derived from transcriptome data that pertain to functional pathways involving phosphorylation-dependent signal cascades. These data further strengthen the evidence as to which pathways in the mammalian cell are perturbed by *Toxoplasma*, potentially providing novel targets for therapeutic intervention.

We will also generate Yeast-two-Hybrid (Y2H) interactomics data between the human proteome and the following ten *T. gondii* genes, known to be relevant during host-cell infection: *GRA10*, *GRA1*, *GRA15* (from Type I and Type II strains), *Rop16* (with/without mutation, see [5]), Adaptin 3 Beta, *Nalp1* (Brazilian and Guyana strains) and *Ap2*. For example, Y2H will allow the characterization of the interaction of *GRA10* with proteins encountered in the intracellular host environment. As in the case of phosphoproteomics, this will further enhance the network analysis derived from expansive transcriptomics datasets. The resulting data matrix will characterize the interactions of proteins and, with integration of data from known host pathways, will result in evidence as to where the *GRA10* virulence factor interferes with host cell signaling and trafficking. Interactomics data will be integrated with transcriptome data and deposited in ToxoDB and any other public repository requested by NIAID.

Bioinformatic analyses to identify diferentially expressed genes and pathways as well as changes in the phosphorilation pattern of infected MonoMac-6 cells will be performed at JCVI.

4b. Approach to Data Production: **Data Analysis**

6. Briefly describe the analysis (value-add) envisioned to be performed subsequently by the community and the potential to develop hypotheses driven proposals given the datasets and resources produced by this work.

6. Analysis and value added

Transcriptomic data sets will be extremely valuable as a resource for the community as they will present the phenotypes from infections with these genetically different parasites in biologically relevant human cells and life cycle stages. They can be compared with other datasets generated. They will become the reference source for these interactomes (e.g., changes in transcription elicted by the parasite or modification of specific parasite molecules or host molecules) for the community of toxoplasmosis researchers.

Complementing transcriptomics studies we will build phosphoproteomic profiles from human MonoMac-6 cells infected (or not) with 2 types of parasites (WT and *GRA10* KO) plus Y2H interactomics data for ten *T. gondii* genes that play relevant roles during infection. These analyses will add an extra dimension to the transcriptomics data proposed herein, providing evidence of how *T. gondii* infection modulates the cellular response of the host at the post-translational level (through changes in the phosphorilation state of proteins and/or by protein-protein interactions).

All the data generated by this white paper will certainly lead to new hypothesis and the development of many new lines of research in Toxoplasmosis.

Some specific more detailed information about analyses follows:

Regarding the protocols for miRNA and mRNA extraction. For total RNA isolation, including miRNA, we will use the following kit from Epicentre: MasterPure™ RNA Purification Kit. For further miRNA processing we will use a kit from Epicentre: ScriptMiner™ Small RNA-Seq Library Preparation Kit. This kit also generates the sequencing library so the final product is ready for sequencing. For mRNA, we will use an adaptation of a protocol for RNAseq from Illumina. The kit from Illumina is the following: TruSeq RNA Sample Preparation Kits v2.
Since RNA extractions from laser-captured tissues will probably give a low yield of RNA, sequencing library constructions will require of a previous amplification step following cDNA synthesis using SMARTer™ Ultra Low RNA Kit for Illumina® Sequencing. If this method does not work we could potentially measure differential expression of up to 800 host/parasite genes of interest per sample using NanoString technology (http://www.nanostring.com/). There is a possibility that JCVI acquires a NanoString nCounter gene expression system in the future. NanoString Technology allows measuring mRNA levels without any previous amplification and is similar in sensitivity to real-time PCR.

Analysis of RNA-seq datasets for the identification of differentially expressed host genes during *T. gondii* infection.
Identification of genes that are differentially expressed in the host cell during T. gondii infection will be carried out by comparing the expression profiles of control (non-infected) and treatment (infected) RNA-seq datasets using the Bioconductor package edgeR [6]. Briefly, after sequencing, short reads from each sample will be mapped to the reference human genome with CLC Genomics Workbench, and the number of reads mapped to each gene will be recorded. Afterwards, using edgeR, we will normalize every expression profile for compositional bias in sequenced libraries and for differences between libraries in sequencing depth. Once normalized, we will use the generalized linear model likelihood ratio test for computing the p-values for each gene of interest in order to accept/reject the null hypothesis that there is no significant expression differences among control and treatment samples for each tested gene. Calculated p-values will be adjusted for multiple testing using the Benjamini and Hochberg's approach [7]. The same approach that has been described by Boothroyd and Saeij et al [8] will be used for determining whether characteristic signatures are present in transcriptomes and in defining novel signatures.

Pathway analysis and functional annotation: Briefly, as described by Boothroyd and Saeij et al [8], gene set enrichment analysis (GSEA, www.broadinstitute.org/gsea/) will be used to find candidate transcription factors and canonical pathways that are activated, induced or repressed upon infection. This program makes use of defined gene sets that are generated experimentally, computationally, or by curation of literature. It then allows for comparison of ranked lists of genes to these reference sets and determines whether members of these reference sets are randomly distributed throughout the ranked lists (suggesting no overlap in the biology of these sets) or primarily found at the top or bottom of that list (suggesting enrichment and correlation with observed phenotypes). For the purposes of hypothesis-generation, gene sets enriched with a false discovery rate (FDR) <0.25 will be considered

significant. The following gene sets from the GSEA Molecular Signatures Database will be evaluated for enrichment: c2.cgp.v3.0 (gene sets derived from literature where cells were subjected to either chemical or genetic perturbations), c2.kegg (gene sets derived from KEGG canonical pathway lists), and c3.tft.v3.0 (gene sets predicted on the basis of a common cis-regulatory motif conserved in the human, mouse, rat, and dog genomes). Identification of functionally-related gene groups enriched in gene sets of interest will be performed using DAVID 6.7, available at http://david.abcc.ncifcrf.gov/.http://david.abcc.ncifcrf.gov .

Regarding details of Proteomics Studies.

Phosphoproteomics of *Toxoplasma gondii* in the human MonoMac-6 monocyte model of infection. Mass spectrometry (MS)-based proteomics is a successful technology to determine a functionally important post-translational modification, phosphorylation, of proteins on a global level. The objective is to identify and quantify MonoMac-6 cell and *T.gondii* phosphoproteins with shotgun proteomics, comparing the data in two stages of infection with wild type parasites and *GRA10* insertional mutant *Toxoplasma* strain with the remarkable phenotype that it can escape a profoundly important killing mechanism pathway. This will allow us to interrogate hypotheses derived from transcriptome data that pertain to functional pathways involving phosphorylation-dependent signal cascades.

The JCVI team has the technology resources and extensive experience in shotgun proteomic analyses to be used for the study. Phosphoproteomic experiments will be performed with a modified LC-MS/MS technology platform (displayed in Figure 4-P1) that has been successfully used in previous studies [9-11].

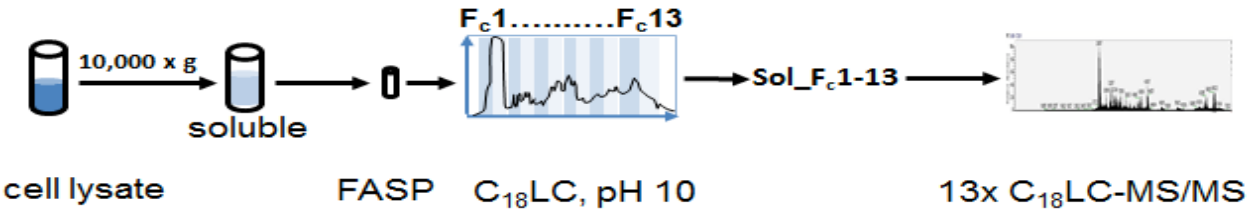


Figure 4-P1. The workflow for the analysis of peptides using 2-DLC-MS/MS

A MonoMac 6 cell lysate is generated from a specific infectious stage. The soluble fraction is enriched for proteins and digested with trypsin using the FASP method (Wisnewski, 2009). Peptide mixtures are separated by one reversed phase (C_{18}) separation at neutral to alkaline pH, whose fractions are subjected to online C_{18} separation prior to electrospray ionization ion trap tandem mass spectrometry (LC-MS/MS) using an EasyLC Velos Pro instrument (Thermo-Fisher). Metal chelate chromatography will allow isolation of an enriched phosphopeptide fraction (after the FASP step), TiO_2 resin will enrich phosphopeptides further in batch mode prior to LC-MS/MS (these steps are not shown in the schematic).

The depicted workflow results in more than 1,500 protein identifications from a cell lsyate. Phosphoproteomics requires enrichment of phosphopeptides. A recently published workflow that reports the feasibility to identify 4,700 unique phosphopeptides from a relatively small amount of protein sample (400 µg) will be applied [12]. Step-wise sample preparation is depicted in Figure 4-P1. The phosphoproteomic analysis workflow will integrate a step to pre-fractionate phosphopeptides into a mono-phosphorylated and a multiply phosphorylated peptide mixture using selected immobilized metal chelate chromatography and application of all C_{18} fractions to TiO_2 resin which further enriches for phosphopeptides. All fractions are subsequently analyzed by LC-MS/MS, followed by database searches including phosphate-specific mass modifications (for the Toxoplasma and human proteome) e.g. MS^2 neutral loss triggering MS^3 scans and use of hierarchical MS^2/MS^3 search modes [13]. The Velos Pro mass spectrometer has electron transfer dissociation (ETD) as a fragmentation option which will be explored for this project. ETD is specifically effective in identifying phosphosites in peptides with multiple candidate amino acids. The scope of work includes four separate analyses (two infection time points, two *Toxoplasma* strains) with two technical replicates of each condition. The data matrix will be integrated with transcriptomics data.

Protein-protein interaction project. A key host gene region has been associated with *Toxoplasma* pathogenicity in a rat model and human cells. A *T.gondii* protein *GRA10* conditional knockout can escape the cell death mechanism in MonoMacs-6 cells, as does a set of chemical and insertional mutants both in the rat model resistant cells and in MonoMacs-6 cells creating a similar phenotype in the MonoMacs-6 Cells. Protein networks have relied on transcriptional studies and those targeting specific Toxoplasma (or host) genes. We propose a study to experimentally verify protein network predictions, focusing on *Toxoplasma* GRA10, GRA1, GRA15, Rop16, Adaptin 3 Beta, Nalp1 and Ap2 genes vs host genes. JCVI has a laboratory with extensive knowledge in yeast-two-hybrid screens, a genetic screen to analyze protein-protein interactions that also uses Sanger sequencing **[14-16]**.

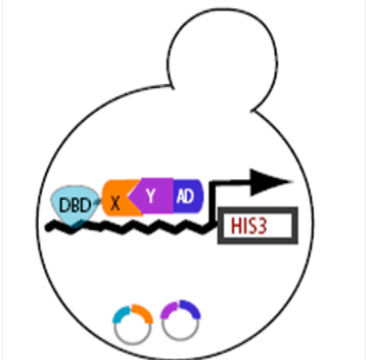


Figure 4-P2. General principle of the yeast two-hybrid system with the DNA-binding domain (DBD) of yeast Gal4 activator fused to a "bait" protein (X) and the activation domain (AD) of Gal4 fused to some "prey" Y of interest. If X interacts with Y, a reporter gene is activated (here: His3) that allows the cell to grow on selective media (here: Histidine-deficient media)

The methodology which detects binary protein interactions in yeast cells is illustrated in schemati in Figure 4-P2.

Haploid yeast strains expressing a single *Toxoplasma* (bait) protein at a time as a DBD fusion are mixed with yeast haploid strains expressing a human cDNA library (prey library). Mating and growth on selective media for mated yeast cells is followed by identification of the unknown interacting prey genes by yeast

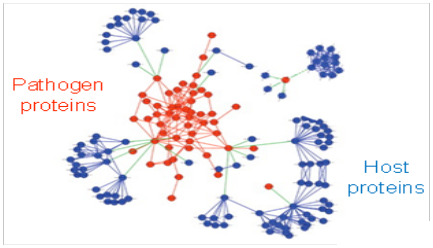


Figure 4-P3. Illustration of host-pathogen protein-protein interactions. Shown is a partial network of Kaposi sarcoma-associated herpes virus proteins (red nodes) and human host proteins (blue nodes)[1]

colony PCR and DNA sequencing of the PCR product [1]. A binary protein interaction matrix is established (similar but less complex than that depicted in Figure 4-P1) that will functionally characterize the bait protein and facilitate data integration in the transcriptome-based molecular network. An example is in in Figure 4-P3.

5. Community Support and Collaborator Roles:

7. *Provide evidence of the relevant scientific community's size and depth of interest in the proposed sequencing or genotyping data for this organism or group of organisms. Please provide specific examples.*

8. *List all project collaborators and their roles in the project*

9. *List availability of other funding sources for the project.*

Community size and involvement

The *Toxoplasma* research community is highly collaborative. The community has worked closely to advance common technology platforms and biological resources to improve research opportunities. Consistent with these previous efforts, this white paper has been assembled with input from many in the community. We recently held the *10th International Toxoplasmosis* meeting, and aspects of this project were discussed with scientists listed below, among others who were not able to be present at this meeting. The participants at the meeting represented more than 200 worldwide laboratories that are engaged in basic research in *T. gondii*. Nearly all of these laboratories are active users of genomic data obtained through previous sequencing efforts, which are housed in GenBank and ToxoxDB. There is overwhelming enthusiasm for this project, which utilizes techniques and will provide data which will be of interest to many. These data will be critical for understanding *T. gondii* virulence, pathogenesis and interaction of host and parasite in human infection. To meet these needs of the research

community, the following collaborations were established. They will guide the proposed project. Data and advice will be shared with all and placed into public repositories available for everyone in the *Toxoplasma* community promptly for the transcriptomes, proteomes and interactomes. This project was enthusiastically received by NIAID Program Officers F. Lee Hall, MD, PhD, S. Rosenthal MD, and D. Joy PhD. They suggested the GSC program.

The following persons have a special interest in this work. Those with an asterisk will be asked or have already been asked to comprise the advisory board.

Dr. Kenneth Boyer, Rush University, USA
Dr. Ying Zhou, University of Chicago, USA*
Dr. Craig W. Roberts, University of Strathclyde, UK*
Dr. Fiona Henriquez-Mui, University of West Scotland, USA*
Dr. Kamal El Bissati, University of Chicago, USA*
Dr. William Witola, Tuskeegee University, USA*
Dr. Marie-France Cesbron-Delauw, CNRS, University of Graenoble, France*
Dr. Cordelia Bisanz, CNRS, University of Grenoble, France
Dr. Gilbert Fournie, Toulouse University, France
Dr. Hernan Lorenzi, JCVRI, USA*
Dr. Rempert Pieper, JCVRI, USA*
Dr. James W. Ajioka, Cambridge University, UK*
Dr. Daniel Ajzenberg, University of Limoges, France
Dr. John C. Boothroyd, Stanford University, USA
Dr. J.P. Dubey, USDA, USA*
Dr. Marie Laure Dardé, University of Limoges, France*
Dr. Michael E. Grigg, NIH, USA*
Dr. David S. Roos, University of Pennsylvania, USA
Dr. L. David Sibley, Washington University, USA
Dr. Jeroen Saeij, Massachusetts Institute of Technology, USA*
Dr. Wilma Buffalano, University of Naples, Italy
Dr. Liliana Soreaceanu, California Pacific Medical Center Foundation, USA
Dr. Charles Cobb, California Pacific Medical Center Foundation, USA
Dr. Huan Ngo, Northwestern University, USA
Dr. Justine Smith, University of Oregon, USA
Dr. Kameran Lashkari, Harvard University, USA
Dr. Fiona Roberts, Glasgow University, UK
Dr. Yves Lussier, University of Illinois at Chicago, USA*
Dr. Kevin White, University of Chicago, USA*
Dr. Leroy Hood, Systems Biolog Insitute, Seattle,WA, USA
Dr. Alexandre Monpetit, McGill University, Genome Quebec, USA
Dr. Jennifer Blackwell, University of Western Australia, Australia

Dr. Rima McLeod will oversee the entire project, will coordinate selection of *T. gondii* strains and human cell lines and will carry out parasite infections. She will also prepare all RNA and protein samples to be submitted to JCVI for sequencing and proteomics analysis.
Drs Jeroen Saeij, John Boothroyd, David Sibley, David Roos, Wilma Buffalano will provide some of the strains required in this white paper.
Dr. Rembert Pieper will be in charge of proteomics and Y2H analyses.
Dr. Hernan Lorenzi will be responsible for coordinating transcriptomics sequencing and analysis and data submission to public repositories specified by NIAID,

At the last International Toxoplasmosis meeting there were ~200 investigators and ~40 laboratories worldwide. Level of interest in the data set will be high.

Other funding Sources.
NIH R01s & UO1s in our laboratory are parent grants U01 AI077887 and 2R01 AI027530. We are also applying for an NIH grant for laboratory support for the work proposed that will take place in our research laboratory at the University of Chicago. Our collaborators also have their own distinct research support. All work will be distinct, non-overlapping, and complimentary.

6. Availability & Information of Strains:

10. Indicate availability of relevant laboratory strains and clinical isolates. Are the strains/isolates of interest retrospectively collected, prepared and ready to ship?

Note: If samples are prospectively prepared the GSC can provide protocols and recommendation based on the Centers past experiences. The samples must however meet minimum quality standards as established by the Center for the optimal technology platform (sequencing/ genotyping) to be used in the study.

Location of materials.
These are in our laboratory, in repositories, or in the laboratories of our collaborators. These include cell lines, parasites, cross over parasites, other parasites, cells of each of the lineages we mention, both primary and relevant tissue specific.

Parasites. Fifteen parasites lineages that have been sequenced at JCVI as part of a White Paper from David Sibley on behalf of the *Toxoplasma* research community, are in our lab or available from Michael Grigg at NIH or in a repository. Those that we will study include RH, GT1, ME49, Prugneand, Veg, Ray and TgPBr. Later we intend to study a highly virulent strain isolated from persons living along the Moroni River in Guyana (GUY [RUB]).

Other strains with unusual phenotypes are available from JP Dubey.

Over expressing and conditional knockout or gene swap parasites are available from our collaborators Jeroen Saeij, John Boothroyd, David Sibley, David Roos, Wilma Buffalano.

Isolates: see list of *T. gondii* strains being sequenced at JCVI as part of the current GSC contract (https://sites.google.com/site/toxoplasmagondiigscidproject/home). These are being sequenced now and include Type I, II, III, IV Atypical and Brazilian and Guyana strains. We will add to these parasites with phenotypes of interest such as markedly decreased virulence in our oocyst model present in a Type II guineafowl strain (TgBrNmBr1).

Knockout, complemented, and overexpressing parasites

Sources for transgenic parasites proposed for this study are shown in Table 4 below:

Protein	Parasite Type	Host Target/cell	Investigator to provide
GRA1	RH (I)	TGFβ/Human monocytes	Wilma Buffolano,
GRA10	RH (I)	TAF1B/HFF	Rima McLeod,Wm Witola
GRA15	ME49 (II)	IKB(NFκB),UHFR/MouseMϕ	Jeroen Saeij
ROP5	GT1/ME49 (I/II)	ROP18/HFF	John Boothroyd,M Reese
ROP16	RH (I)	STAT3,6/HFF	John Boothroyd,,J Saeij
ROP18	GT1 (I)	GTPAse/MouseMϕ	David Sibley
ROP38	VEG (III)	MAPK/Human monocytes	David Roos

Table 4. ***Transgenic parasites and their sources*** Abreviations: GRA=dense granule protein,

ROP=rhoptry protein. These are secreted proteins that modulate expression or function of host cell genes or gene products. Target refers to those host genes or gene products that interact with the corresponding parasite protein (column 1) and cell refers to the host cell type in which the interaction has been studied. The Investigator is the person who has done that work and agreed to provide the transgenic parasite for the work described in this White paper. These host-cell signaling factors or transcription factors control key biological functions of the host cell and result in differing phenotypes. Some of those phenotypes, e.g. the phenotypes modulated by GRA15, ROP5, ROP16, and ROP18 differ between parasites of differing lineages and therefore are key virulence determinants for the parasites. These genes have been variably knocked out and in some complemented in the knockout, or the domains have been swapped from strains of differing lineages in addition to being knocked out or conditionally modified using tetracycline responsive promoter elements. This makes the study of their functions robust. None have yet been studied in the human cell line lineages described herein to date.

Human Cells:

Neuronal stem to differentiated, McLeod Lab

Monomacs cells-with lentivirus knockdowns for ALOX12 and NALP1 and others-McLeod lab

Tissues

We have banked brain and eye tissues from congenitally infected and immune compromised persons, but these are in formalin.

Frozen tissues from humanized mice are available for laser-capture microscopy of cysts and single cell transcriptomics.

11. Attach relevant information, if available in an excel spreadsheet for multiple samples: e.g

- *Name*
- *Identifier*
- *Material type (DNA/RNA/Strain)*
- *Genus*
- *Species*
- *Specimen / Strain*
- *Isolation source*
- *Isolated from*
- *Select agent status*
- *International permit requirement*
- *BEIR/ATCC repository accession number*
- *Other public repository location*
- *Other public repository identifier*
- *Sample provider's name*
- *Sample provider's contact*

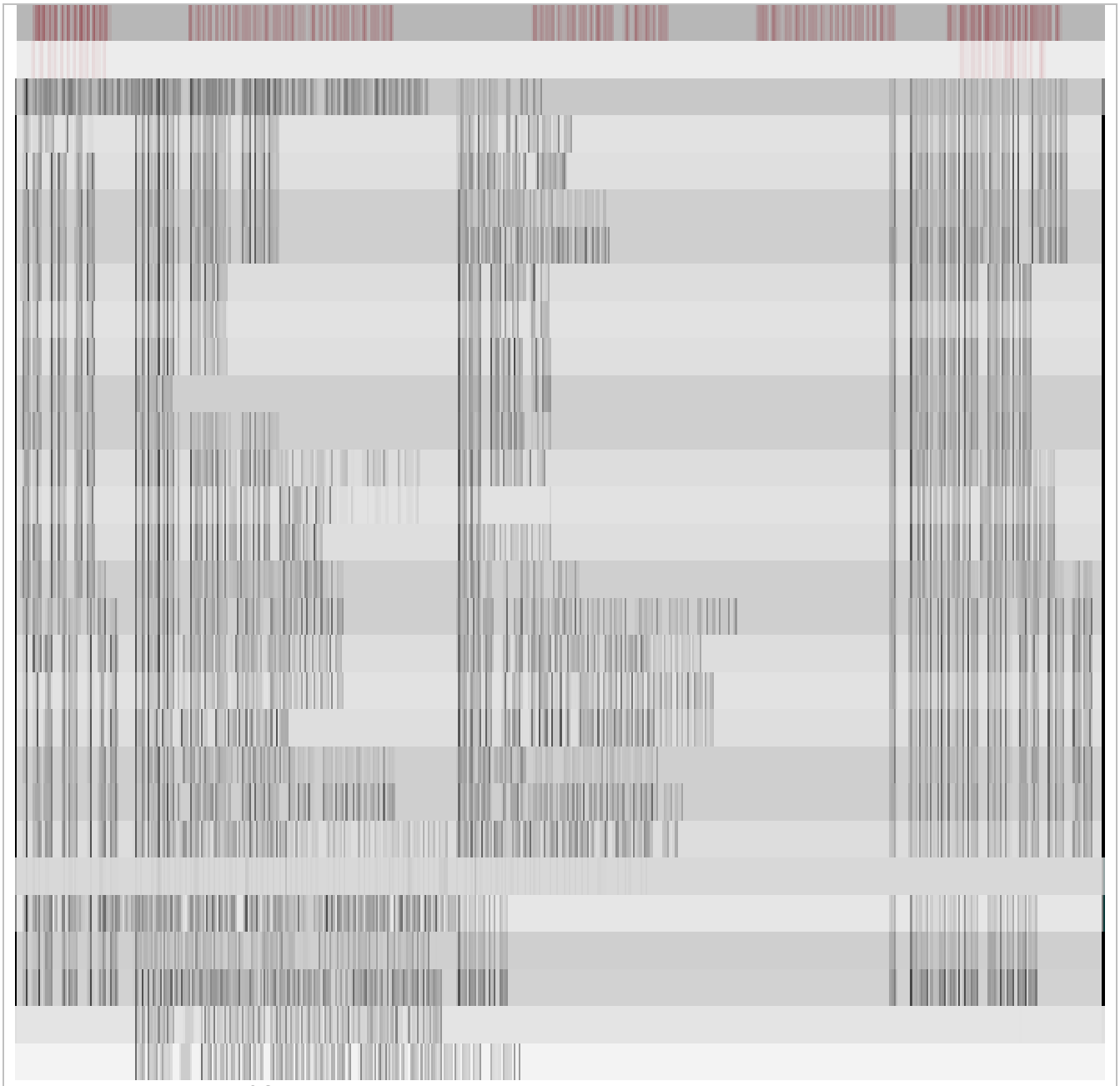


Table 5: Description of Samples to be used in this White Paper.

Note: The Atypical RAY is a haplogroup XII (type IV) parasite described earlier. The Guyana (Maroni) parasite GUY-(RUB) is a hypervirulent isolate that was obtained during an epidemic in Guyana along the Maroni River.It is one of the wild sylvatic Amazonian species of *T.gondii*. The knockout and mutant parasites have genetically engineered parental parasites.

GRA 1 is a knockout in a type 1 background. It modulates TGFβ; GRA10 is a conditional knockout in a type 1 RH background and it modulates TAF1β; GRA 15 is a domain swapped Type II parasite which modulates the NFκβ pathway; ROP5, 16, & 18 are virulence factors available as domain swapped or type I complemented parasites; ROP 38 is a virulence factor which is available as an overexpressing parasite on a type III background.

Enoyl reductase (ENR) is a conditional mutant available on type I and type II backgrounds. It is key in synthesis of unique type II fatty acid biosynthesis pathway associated with plastid function, replication and parasite survival.

Ribosomal protein S13 (RPS 13) is on a type I background as a conditional knockout. The knockout parasite has a G0 arrest phenotype in fibroblasts which is an escape from the parasite cell cycle that allows the parasite to persist for prolonged times.

ALOX12 and *NALP1* are genes of interest in the human genome. They will be characterized with RH and insertional mutant parasites. NFKβ will be characterized with prugneaud and domain swapped parasites as GRA15 modulates this pathway in other cell types and likely will do so in these human cell lineages also. TGFβ is modulated by GRA1 and will be characterized with type I parasites.

12. What supporting metadata and clinical data have been collected or are planned on being collected that could be made available for community use?

Phenotypes caused by the different transgenic/wt parasites used in this study and other associated metadata (isolation date / place / etc) will be also submitted.

7. Compliance Requirements:

7a. Review NIAID's Reagent, Data & Software Release Policy:

NIAID supports rapid data and reagent release to the scientific community for all sequencing and genotyping projects funded by NIAID GSC. It is expected that projects will adhere to the data and reagent release policy described in the following web sites.

http://www.niaid.nih.gov/labsandresources/resources/mscs/data.htm
http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-013.html

Once a white paper project is approved, NIAID GSC will develop with the collaborators a detailed data and reagent release plan to be reviewed and approved by NIAID.

Accept ☒ Decline ☐

7b. Public Access to Reagents, Data, Software and Other Materials:

13. State plans for deposit of starting materials as well as resulting reagents, resources, and datasets in NIAID approved repositories. Sequencing projects will not begin until the strain is deposited into NIAID funded BEI repository (http://www.beiresources.org/). This includes web based forms are completed by the collaborator and received by the NIAID BEI (http://www.beiresources.org/).

All the data generated by this project will be deposited in public repositories approved by NIAID such as ToxoDB and GenBank, including signatures for the parasites for each type of cell, each type of parasite and in each condition for knockdown or overexpression of host or parasite genes. Also for each life cycle stage assessed in this white paper.

Proteomics and interactomics data will be deposited as soon as they are generated in NCBI and ToxoDB plus any other public repository required by NIAID.

T.gondii wild type strains used for transcriptomics data generation will be deposited in BEI if they are not already there.

All publications will be deposited in pubmed in accordance with NIH policy.

References for the entire document

1. Albers M, Kranz H, Kober I, Kaiser C, Klink M, Suckow J, Kern R, Koegl M: **Automated yeast two-hybrid screening for nuclear receptor-interacting proteins**. *Mol Cell Proteomics* 2005, **4**(2):205-213.
2. Yolken RH, Dickerson FB, Fuller Torrey E: **Toxoplasma and schizophrenia**. *Parasite Immunol* 2009, **31**(11):706-715.
3. Peixoto L, Chen F, Harb OS, Davis PH, Beiting DP, Brownback CS, Ouloguem D, Roos DS: **Integrative genomic approaches highlight a family of parasite-specific kinases that regulate host responses**. *Cell Host Microbe* 2010, **8**(2):208-218.
4. Rosowski EE, Lu D, Julien L, Rodda L, Gaiser RA, Jensen KD, Saeij JP: **Strain-specific activation of the NF-kappaB pathway by GRA15, a novel Toxoplasma gondii dense granule protein**. *J Exp Med* 2011, **208**(1):195-212.
5. Yamamoto M, Standley DM, Takashima S, Saiga H, Okuyama M, Kayama H, Kubo E, Ito H, Takaura M, Matsuda T *et al*: **A single polymorphic amino acid on Toxoplasma gondii kinase ROP16 determines the direct and strain-specific activation of Stat3**. *J Exp Med* 2009, **206**(12):2747-2760.
6. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data**. *Bioinformatics* 2010, **26**(1):139-140.
7. Klipper-Aurbach Y, Wasserman M, Braunspiegel-Weintrob N, Borstein D, Peleg S, Assa S, Karp M, Benjamini Y, Hochberg Y, Laron Z: **Mathematical formulae for the prediction of the residual beta cell function during the first two years of disease in children and adolescents with insulin-dependent diabetes mellitus**. *Med Hypotheses* 1995, **45**(5):486-490.
8. Saeij JP, Boyle JP, Coller S, Taylor S, Sibley LD, Brooke-Powell ET, Ajioka JW, Boothroyd JC: **Polymorphic secreted kinases are key virulence factors in toxoplasmosis**. *Science* 2006, **314**(5806):1780-1783.
9. Kuntumalla S, Braisted JC, Huang ST, Parmar PP, Clark DJ, Alami H, Zhang Q, Donohue-Rolfe A, Tzipori S, Fleischmann RD *et al*: **Comparison of two label-free global quantitation methods, APEX and 2D gel electrophoresis, applied to the Shigella dysenteriae proteome**. *Proteome Sci* 2009, **7**:22.
10. Kuntumalla S, Zhang Q, Braisted JC, Fleischmann RD, Peterson SN, Donohue-Rolfe A, Tzipori S, Pieper R: **In vivo versus in vitro protein abundance analysis of Shigella dysenteriae type 1 reveals changes in the expression of proteins involved in virulence, stress and energy metabolism**. *BMC Microbiol* 2011, **11**:147.
11. Pieper R, Zhang Q, Clark DJ, Huang ST, Suh MJ, Braisted JC, Payne SH, Fleischmann RD, Peterson SN, Tzipori S: **Characterizing the Escherichia coli O157:H7 proteome including protein associations with higher order assemblies**. *PLoS One* 2011, **6**(11):e26554.
12. Engholm-Keller K, Hansen TA, Palmisano G, Larsen MR: **Multidimensional strategy for sensitive phosphoproteomics incorporating protein prefractionation combined with SIMAC, HILIC, and TiO(2) chromatography applied to proximal EGF signaling**. *J Proteome Res* 2011, **10**(12):5383-5397.
13. Xu H, Wang L, Sallans L, Freitas MA: **A hierarchical MS2/MS3 database search algorithm for automated analysis of phosphopeptide tandem mass spectra**. *Proteomics* 2009, **9**(7):1763-1770.
14. Uetz P, Dong YA, Zeretzke C, Atzler C, Baiker A, Berger B, Rajagopala SV, Roupelieva M, Rose D, Fossum E *et al*: **Herpesviral protein networks and their interaction with the human proteome**. *Science* 2006, **311**(5758):239-242.
15. Rajagopala SV, Titz B, Goll J, Parrish JR, Wohlbold K, McKevitt MT, Palzkill T,

Mori H, Finley RL, Jr., Uetz P: **The protein network of bacterial motility**. *Mol Syst Biol* 2007, **3**:128.
16. Chen YC, Rajagopala SV, Stellberger T, Uetz P: **Exhaustive benchmarking of the yeast two-hybrid system**. *Nat Methods* 2010, **7**(9):667-668; author reply 668.

7c. Research Compliance Requirements

Upon project approval, NIAID review of relevant IRB/IACUC documentation is required prior to commencement of work. Please contact the GSC Principal Investigator(s) to ensure necessary documentation are filed for / made available for timely start of the project.

IBC Protocol #: 737
IBC Approval Date:07/16/2011
IACUC Protocol #: 71734
IACUC Approval Date: 08/28/2011
IRB Protocols (Approval Dates): 8792 (3/1/11), 8793 (3/1/11), 8794 (3/1/11), 8795 (3/1/11), 8796 (3/16/11), 8797 (3/1/11), 8798 (3/15/11), 15408A (3/16/11), 16708A (3/7/11)

Investigator Signature:

Investigator Name: Rima McLeod, MD **Date:** 6 April 2012