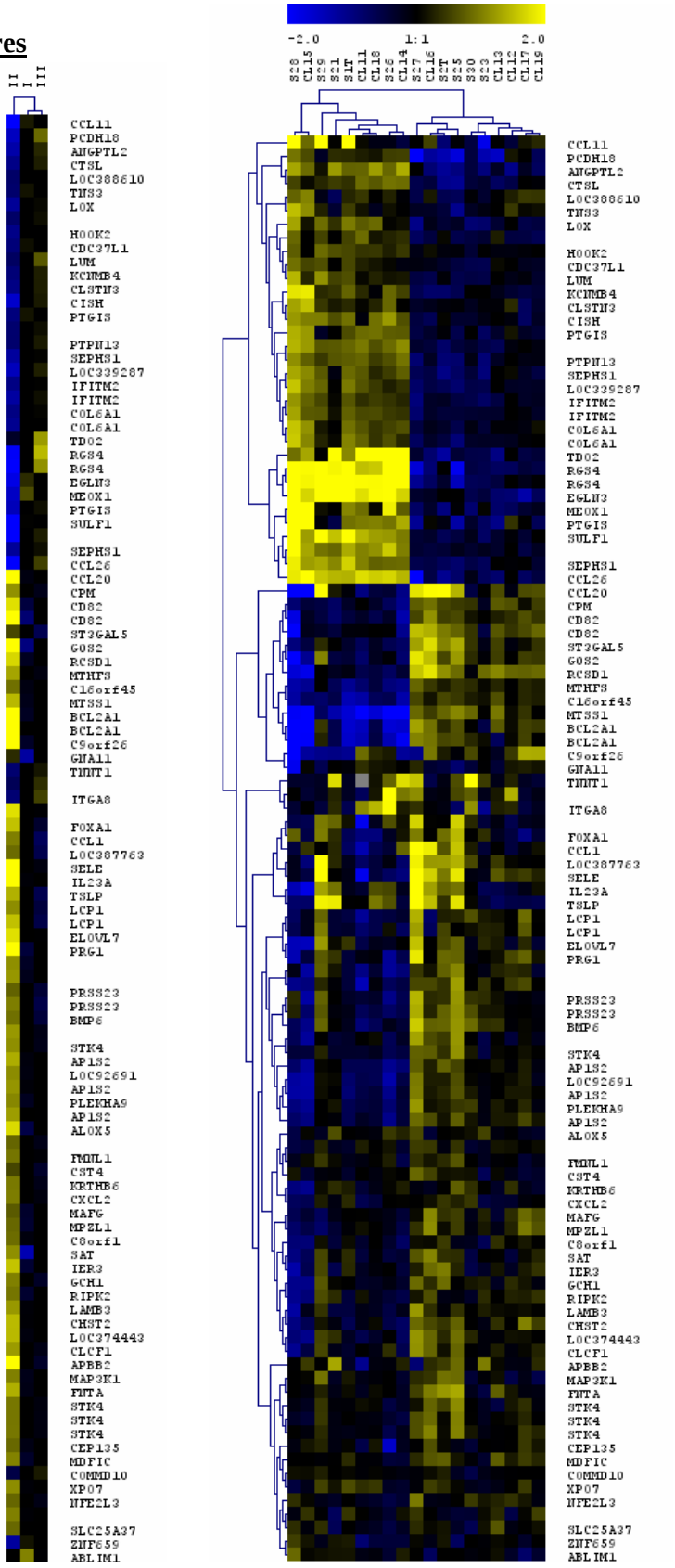
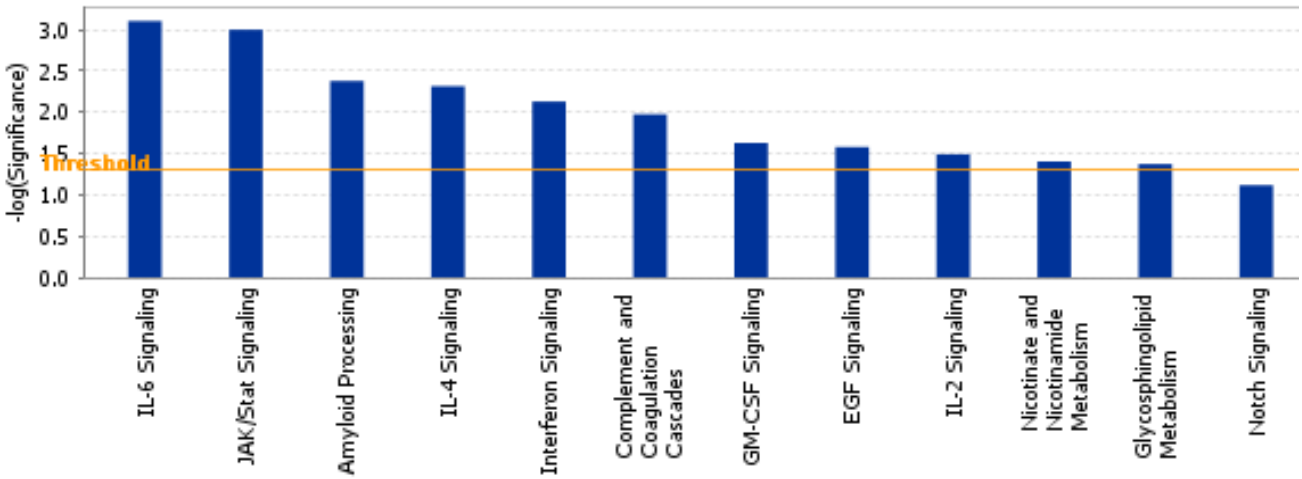
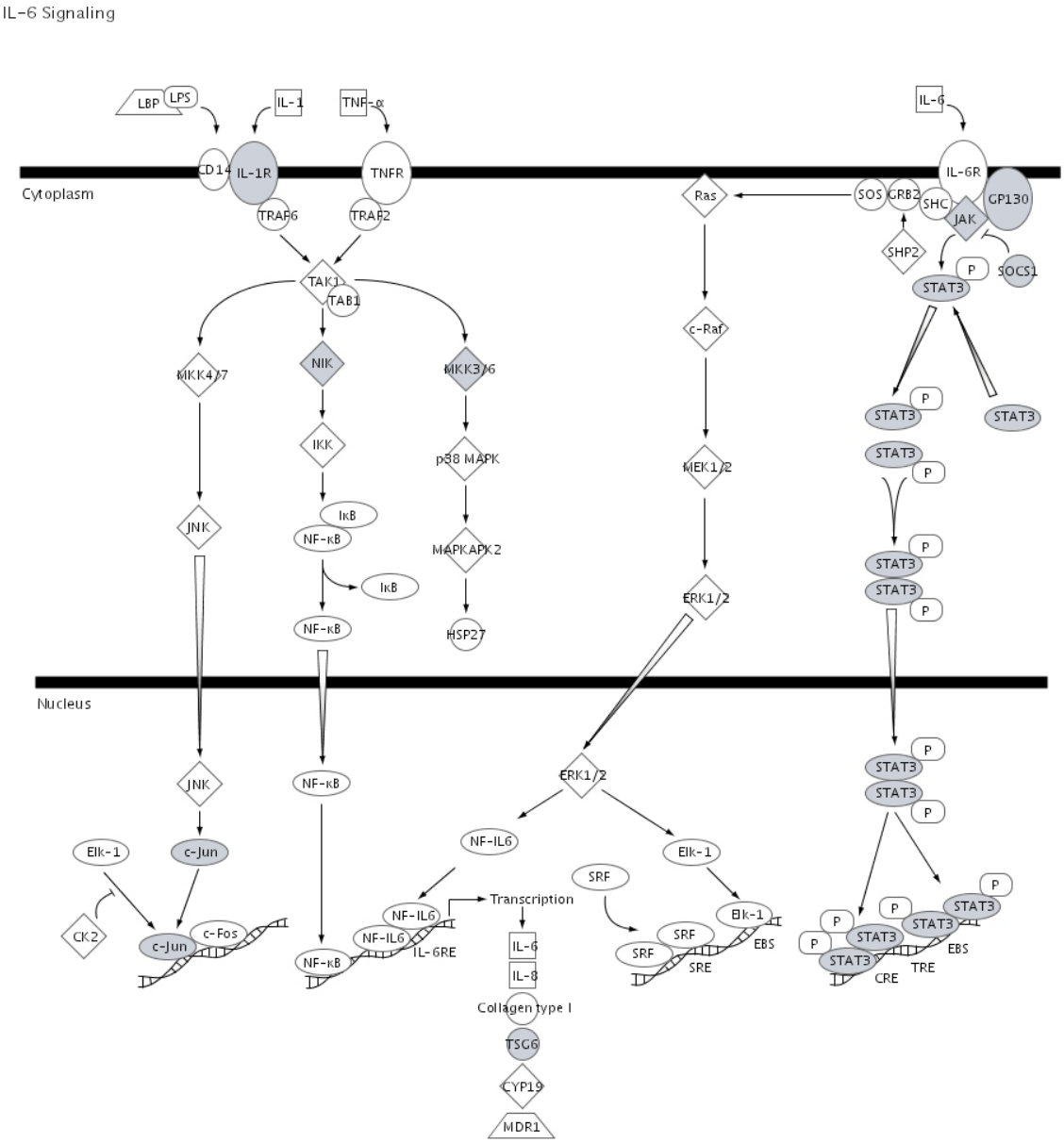
Supplementary Figures



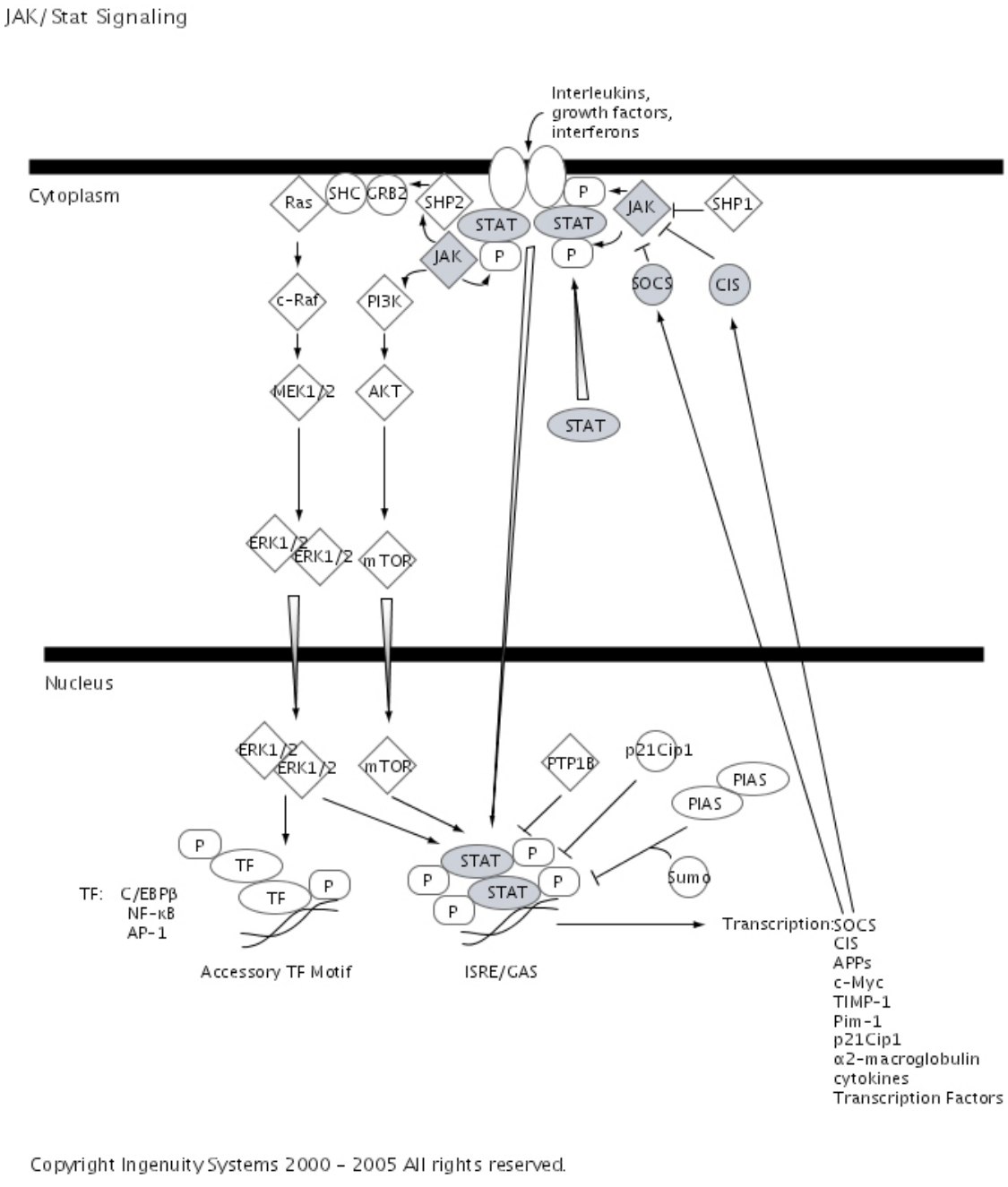
Supplementary Figure. 1. This is an enlargement of Fig. 1. For details see Figure legend 1 and supplementary table 1.

Supplementary Figure 2a

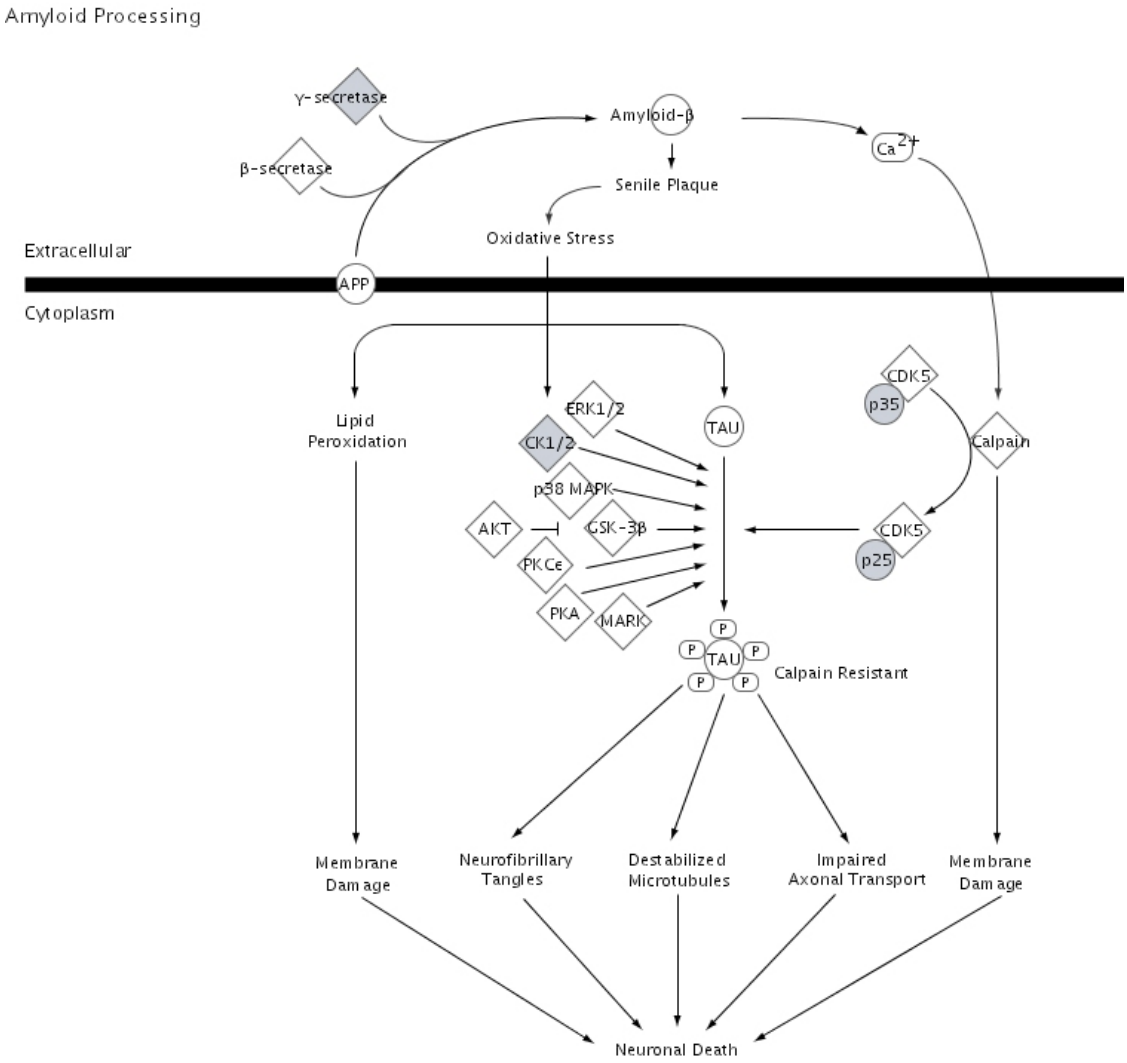




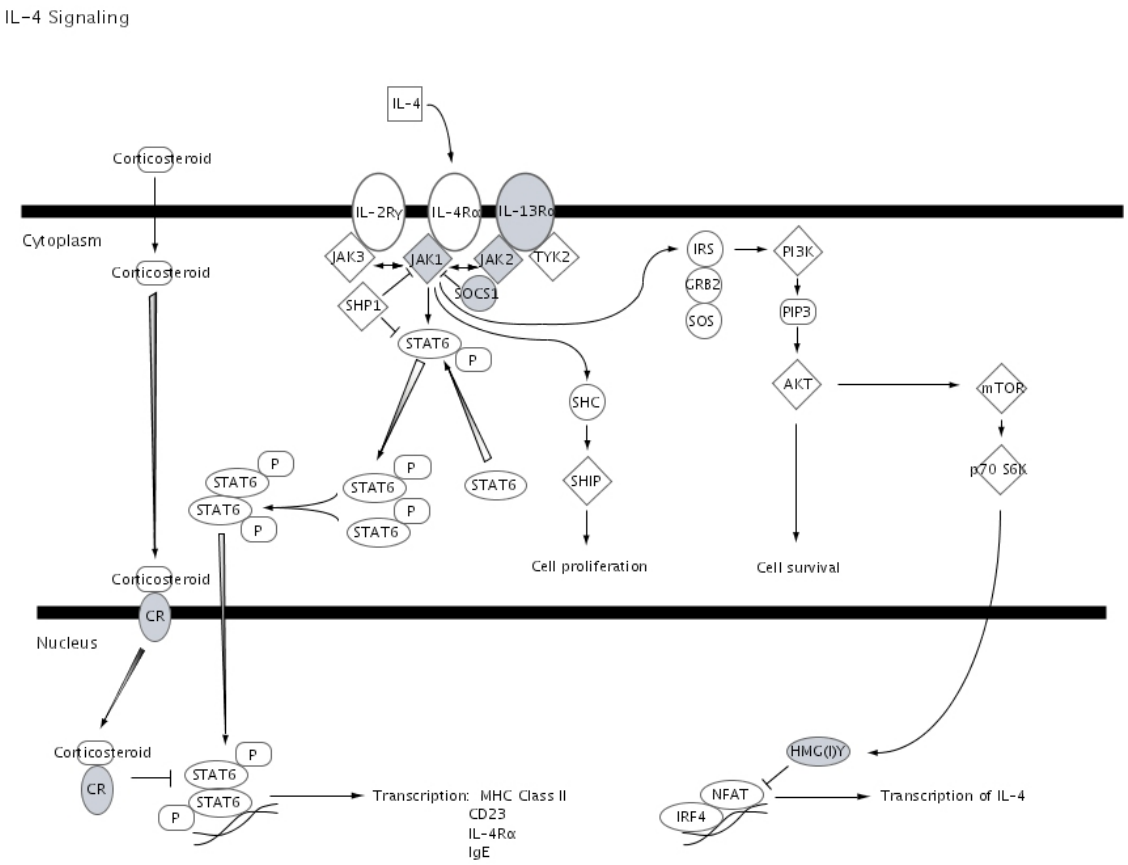
Supplementary Figure 2b



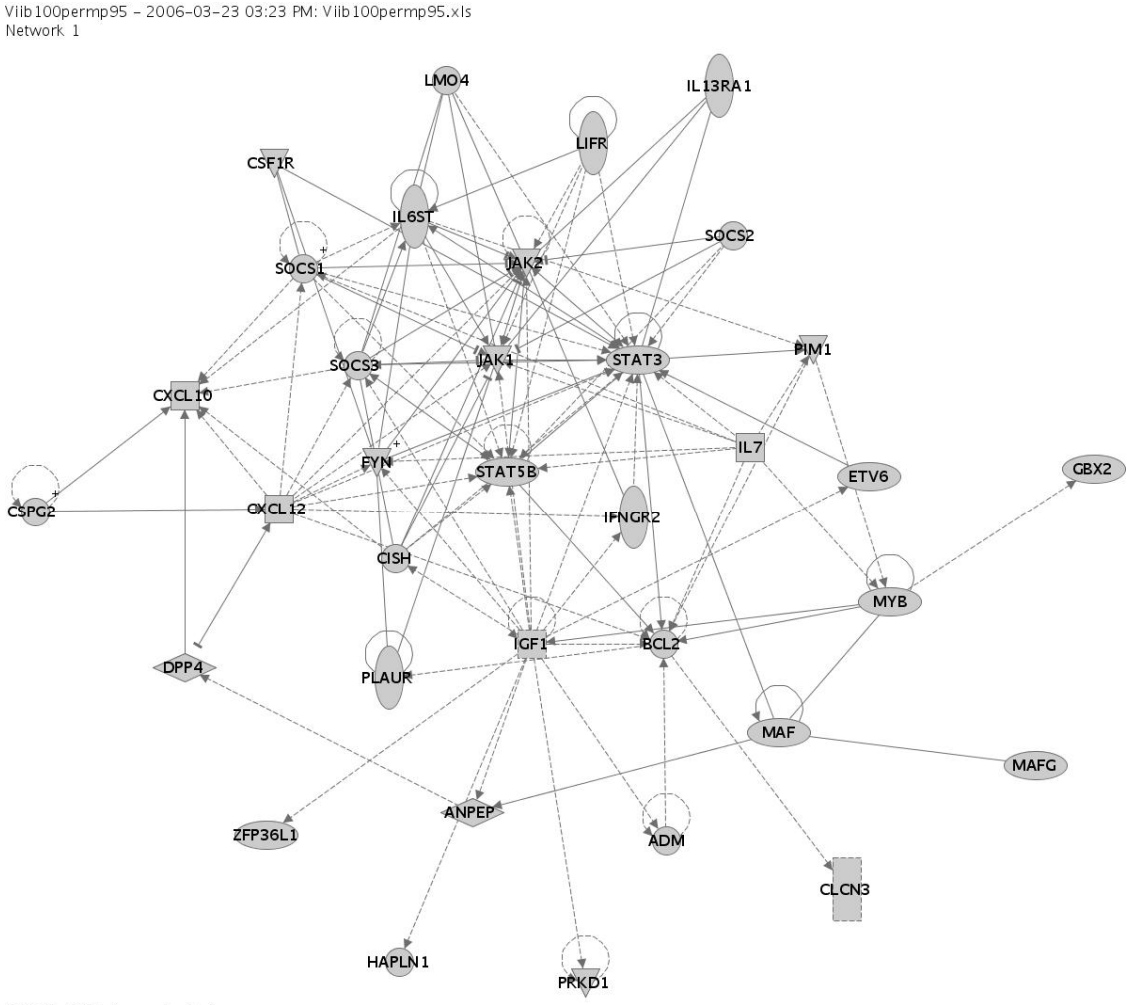
Supplementary Figure 2c



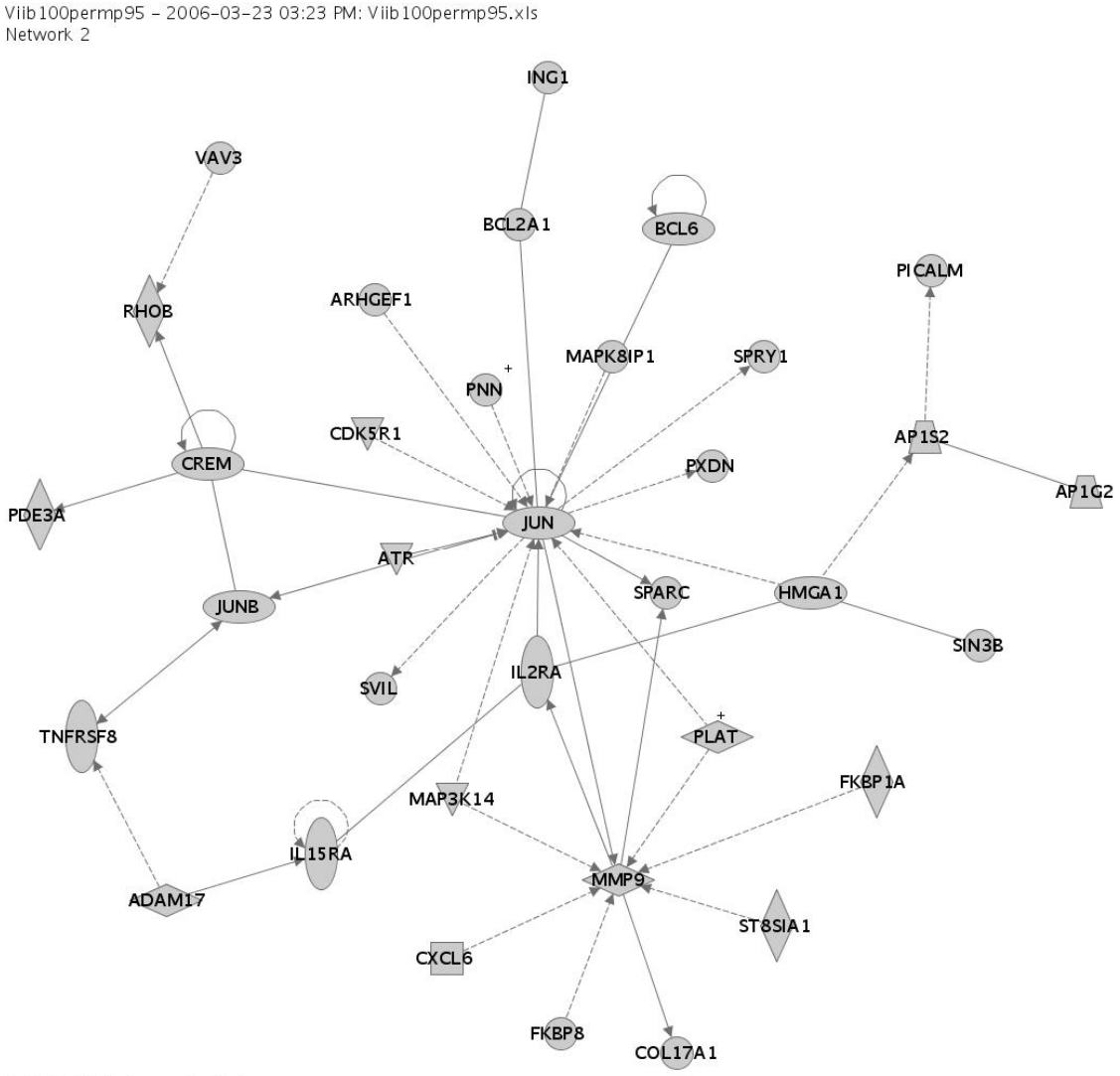
Supplementary Figure 2d



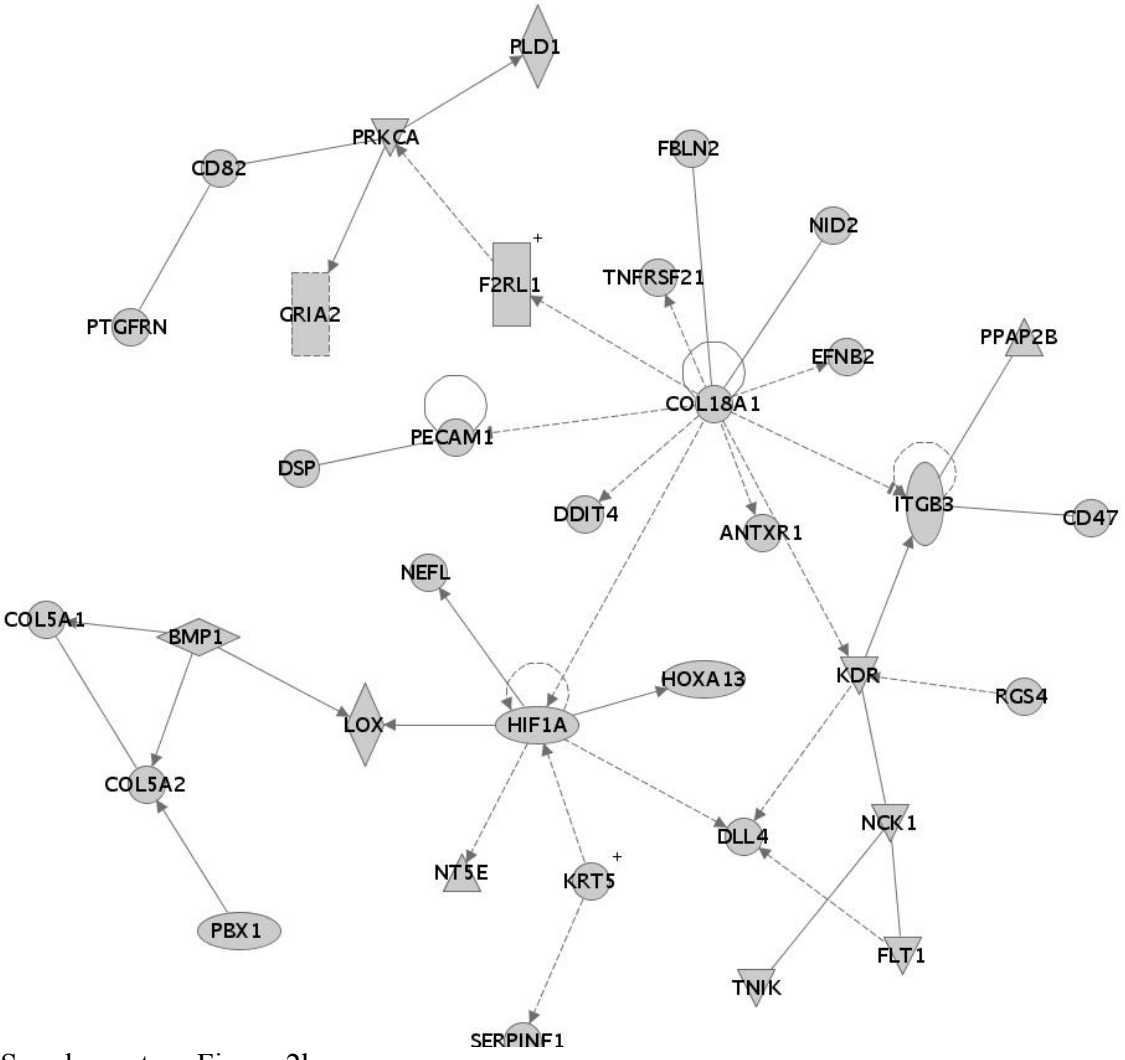
Supplementary Figure 2e



Supplementary Figure 2f



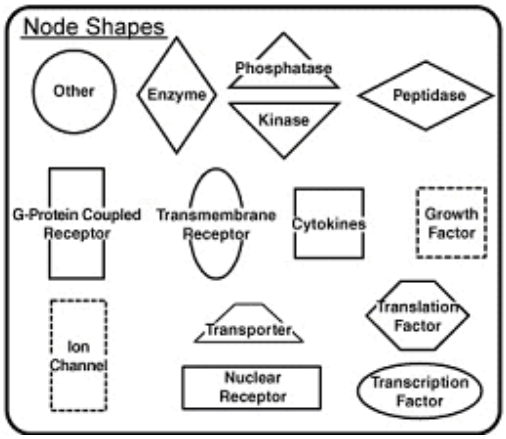
Supplementary Figure 2g



Supplementary Figure 2h

Supplementary Figure 2. a, Canonical pathways analysis identified the pathways from the Ingenuity Pathways Analysis library that were most significant to the dataset (www.ingenuity.com). Genes from the dataset that mapped to *Toxoplasma* chromosome VIIb (genome-wide P-value < 0.05) and were associated with a canonical pathway in the Ingenuity Pathways Knowledge Base were considered for the analysis. Fischer's exact test was used to calculate a P-value determining the probability that the association between the genes in the dataset and the canonical pathway is explained by chance alone. **b-e,** The human genes that mapped to *Toxoplasma* chromosome VIIb and contributed to the four pathways that appeared most significant in Suppl. Fig. 2a are drawn in grey. **f-h,** A data set containing the gene identifiers for the genes that mapped significantly to chromosome VIIb was uploaded and each gene identifier was mapped to its corresponding gene object in the Ingenuity Pathways Knowledge Base. These genes, called Focus Genes, were overlaid onto a global molecular network developed from information contained in the Ingenuity Pathways Knowledge Base. Networks of these

Focus Genes were then algorithmically generated based on their connectivity. A network/My Pathways is a graphical representation of the molecular relationships between genes/gene products. Genes or gene products are represented as nodes, and the biological relationship between two nodes is represented as an edge (line). All edges are supported by at least 1 reference from the literature, from a textbook, or from canonical information stored in the Ingenuity Pathways Knowledge Base. Human, mouse and rat orthologues of a gene are stored as separate objects in the Ingenuity Pathways Knowledge Base, but are represented as a single node in the network. Nodes are displayed using various shapes that represent the functional class of the gene product. Edges are displayed with various labels that describe the nature of the relationship between the nodes (e.g., P for phosphorylation, T for transcription). The three networks with the highest score are indicated.

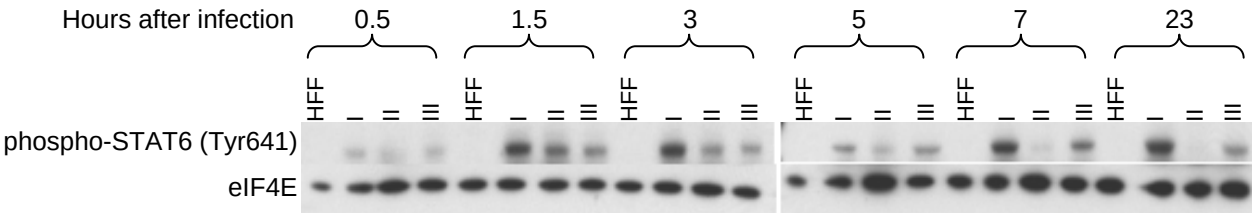


		Strain						
Marker	Position	S23	CL16	S28	S30	D3XI	JD4	KD11
AK103	4996718	2	3	3	2	2	3	3
361*	4779775	2	2	3	2	2	3	3
615*	4724768	2	2	3	2	2	3	3
76*	4621257	2	2	3	2	2	3	3
920*	4527248	2	2	3	2	2	3	3
L339	4466884	2	2	3	2	2	3	3
L363	4158817	2	2	3	2	2	3	3
398*	3956443	2	2	3	2	2	2	3
Rop16*	3954398	2	2	3	2	2	2	3
1060*	3792522	2	2	3	2	2	2	3
261*	3718397	2	2	3	2	2	2	3
AK104	3603126	2	2	3	2	3	2	3
244*	3602023	2	2	3	2	3	2	3
contig1*	3456578	2	2	3	2	3	2	3
285*	3198244	2	2	3	2	3	2	3
377*	2893485	2	2	2	2	3	2	3
AK105	2892714	2	2	2	2	3	2	3
cA5-2	2497749	2	2	2	3	3	2	3
STAT6 Phenotype:		-	-	+	-	-	-	+

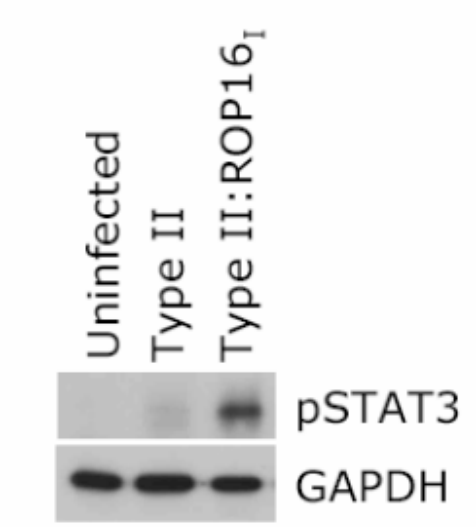
Supplementary Figure 3. Fine mapping of the *Toxoplasma* locus involved in strain-specific regulation of STAT6. Genotyping of recombinant progeny at 18 markers. Markers with asterisks were made in our laboratory (sequences available upon request), and the rest of the markers have been used previously and are available at http://toxomap.wustl.edu/. All markers are present on the largest scaffold on chromosome VIIb (TGG_995366; http://www.toxodb.org) and are presented in known order. Progeny were from crosses between a type II and a type III strain (S23, CL16, S28, S30, D3XI) or a backcross between S23 and a type III strain (JD4, KD11). STAT6 was detected in the nucleus of cells infected with the different strains as described in Materials and Methods. The limits of the region implicated by this analysis are denoted by a box.

```
#I   MKVTTKGLAFALALLFCTRCATARYMSFEEAQKASEAAKRQIATLPSPDSPLSNPGSRHRNRGGSPTAGQPSQSTLQPEQAAAEVGLGAG
#III ................................................................................................
#II  ..................................................T......K........A.....................

#I   GSTQGQGRTGGSAGAREERRSPSPESAYPATSSASLRGYQTQLSPSHLPPHSSGPGGWFPTESIYTLWSSPPQRLTHRKPSLSGVVVTEF
#III ........................Q.......................................................................
#II  ........................Q.........................R.............F.P......P..Q.............

#I   QEPQEQYGAASSLASSPKGYVGGASSSALSGKAVPTPASLGQENPLFPGQSATLDSGIQSPAQKRRGSPQRQSAMPTGNPADSGASQLAF
#III ................................................................................................
#II  ..................R..S.........................V..............E........I..S.E...........S

#I   SHSSYVSVQASLAKRSERIRRVRLSEEGLEEVQQLKAAAAQLLVAVPDYEAMRAVLQEAVLSEQRVAARKRKRKQPPGAVESAVDEVFPP
#III ................................................................................................
#II  .V....A..TPHV.....................................................T.....................

#I   NERVMMINANGVPIALYNRGHLGSGHFGAVIKASLDDGTLYAAKVPYSQIVPNADATSAELEAGISSARAELVKTIRQELDVRDKLVAKG
#III ................................................................................................
#II  ..........................................................E.....................

#I   LTLTETVSQYGLPLCQMTLTLPENKATVVRRGSRLFVVSKEVMLLPLIDGSALNSLVQSQPPFLFQRAVAREAILALAKLHELGFAHGDV
#III ..........................................................L.....................
#II  ......AE.........................V..............PS.....................I..............

#I   KLNNMMIDVHGFGHMLDMGSVRPVDSCVSEEDKYYLRLWAPELAKSQHTSQKTCLKRGALDVWALGLAIFEFVCFNRLPYSLSNLPSSFW
#III ................................................................................................
#II  ..................................................Q.................................L.

#I   SRVEHLSRLRLSDFSVKDCNESDPAVMGIVVQFLNPDPQERPELPKFVNSYTFFQQAPGVTSHLTRIPTTELSSHRM
#III .............................A..................................................
#II  ..............A..............A.....N.E.........S.....R.....................
```

Supplementary Figure 4. The predicted amino acid sequence for the primary translation product of the *ROP16* gene is shown for type I parasites (top line). The sequences for types III and II are shown on the second and third lines, respectively with a dash to indicate identity to the type I sequence. The putative signal peptide is shown in blue (the cleavage site is predicted by SignalP[28]). A putative nuclear localization signal is in red (as predicted by PredictNLS[29]). Regions shown in magenta and light blue are the predicted ATP-binding and catalytic sites, respectively.



Supplementary Figure 5. Serum-starved HFFs were infected with a type I, type II or type III *Toxoplasma* strain (MOI 10) and at the indicated time-points after infection cells were lysed and analyzed by Western blotting with an antibody against phospho-STAT6 (Tyr641) (from Cell Signaling). An antibody against eIF4E was used as a loading control.



Supplementary Figure 6. RAW 264.7 macrophages were infected as described in the Materials and Methods for the HFF infections. These results demonstrate that in a similar manner as observed for HFFs, Type II parasites alone do not induce STAT3 phosphorylation in macrophages, but upon addition of the Type I ROP16 allele, infection induces activation of STAT3.

Supplementary Methods

Parasites.

Strains were maintained in vitro by serial passage on monolayers of HFFs as previously described[30]. The RH (type I), ME49 and Pru (both type II), CEP (type III) strains and the F1 recombinant progeny used in the microarray analyses, derived from crosses between ME49 and CEP[31,32], have all been described previously.

Genetic crosses. Genetic crosses were carried out in cats as described previously[31]. Briefly, *T. gondii* cysts were harvested from the brains of infected mice 24-30 days p.i. and fed in equal amounts to 6-10 week-old kittens. Feces were collected daily during days 3-8, and oocysts were floated on sucrose, washed, and allowed to sporulate for 3-8 days in 0.2% sulfuric acid. Sporozoites were excysted using standard protocols. For the backcross between the S23 strain (an F1 progeny clone of one of the previous type II x type III crosses[31]) and the type III strain (CL14), S23 was engineered to be resistant to FUdR by knocking out the gene for uracil phosphoribosyltransferase, and recombinants were selected by growth in FUdR (5 x 10-5 M) and AraA (1.3 x 10-4 M).

Transgenic parasites. Types I (RH) and II (Pru) parasites carrying a null mutant of the *HXGPRT* gene were transfected with the type I *ROP16* gene (from 2000 nt upstream of the predicted ATG start codon to the base pair immediately preceding the stop codon which was replaced with a single HA tag) cloned into the pHEX2 vector which provided the *SAG1* 3' UTR from the type I parasites and the *HXGPRT* gene. Transgenic parasites expressing *HXGPRT* were selected using normal cDMEM plus mycophenolic acid and xanthine as previously described[33]. A version of ROP16 carrying a mutated NLS was constructed by PCR amplification of type I ROP16 using primers with nucleotides encoding methionine (ATG) to replace the lysine (AAG) in the NLS sequence RKRKRK to RMRMRM (CGT A**T**G CGG A**T**G AGA A**T**G). The resulting ROP16 gene was cloned into the pHEX2 vector as described above.

IFA. Parasites were allowed to invade confluent HFF monolayers on coverslips for 0.17, 0.5, 1, 4, 6, 8, 18 or 24 h. The cells were then fixed (for nuclear localization, 4.0 % methanol-free formaldehyde for 20 min; for phospho-STAT analysis, 4.0 % methanol-free formaldehyde for 20 min followed by 100 % ice-cold methanol for 5 min), blocked in PBS supplemented with 3% BSA and permeabilized with 0.2% triton-X100 overnight at 4°C. Coverslips were then incubated with 3F10 (anti-HA) antibody (Roche, Palo Alto, CA), antibodies specific for the phosphorylated forms of STAT3 (phospho-Tyr705) and STAT6 (phospho-Tyr641) (Cell Signaling Technologies, Danvers, MA), or a mouse monoclonal antibody to SAG1 (DG52[34]), for 1-3 h at room temp. Fluorescent secondary antibodies (Invitrogen/ Molecular Probes, Carlsbad, CA) and Hoechst dye (Polysciences, Inc., Warrington, PA) were used for antigen and DNA visualization, respectively.

Microarrays: MIAME Checklist

Experiment Design

The goal of the experiment: Genome-wide identification of *Toxoplasma* genomic loci mediating strain-specific modulation of host gene expression identifies ROP16 as a polymorphic *Toxoplasma* kinase that targets the host nucleus and co-opts transcription.

Brief description of the experiment: *Toxoplasma gondii*, an obligate, intracellular parasite of the phylum Apicomplexa, can cause severe disease in persons with an immature or

suppressed immune system. The majority of *Toxoplasma gondii* isolates so far identified within Europe and North America belong to three distinct clonal lines that differ in their fundamental virulence and disease presentation in mice and humans. We have determined that these three strains also differ substantially in how gene expression in the host cell is altered by infection. Using genetic crosses between a type II and type III strain, we have mapped the *Toxoplasma* loci involved and for one we have identified a polymorphic, secreted protein kinase (ROP16) as the responsible molecule involved. We show that upon invasion, this parasite kinase (ROP16) is injected into the host cytosol from the apical, secretory organelles known as rhoptries and that it then rapidly traffics to the host cell nucleus. Using transgenic parasites, we show that ROP16 affects activation of the STAT3 and STAT6 signalling pathways. Our findings provide a completely new mechanism for how an intracellular eukaryotic pathogen can interact with its host and co-opt gene expression. They also reveal major differences in how different Toxoplasma lineages have evolved to exploit this interaction.

Keywords: *Toxoplasma gondii*, strain comparison, human foreskin fibroblast, microarray, STAT, parasite

Experimental factors: *Toxoplasma gondii* genetic variation

Experimental design: Human Foreskin Fibroblasts (passage 10- 12) were infected with different *Toxoplasma gondii* strains and 24 h post-infection RNA was isolated. Indicated is which *Toxoplasma gondii* strains was used on each array and the multiplicity of infection (MOI) used (Strain_Array_MOI)

RH_SHEM171_MOI10 (Type I strain)
RH_SHEM173_MOI5 (Type I strain)
RH_SHEM38_MOI10 (Type I strain)
RH_SHEM86_MOI10 (Type I strain)
PDS_SHEM137_MOI10 (Type II strain)
PDS_SHEM139_MOI10 (Type II strain)
PDS_SHEM82_MOI10 (Type II strain)
CTG_SHEM140_MOI10 (Type III strain)
CTG_SHEM175_MOI10 (Type III strain)
CTG_SHEM83_MOI10 (Type III strain)
CTG_SHEM84_MOI10 (Type III strain)
CL12_SHFB243_MOI10 (F1 recombinant progeny from Type II X Type III cross)
CL13_SHFB123_MOI10 (F1 recombinant progeny from Type II X Type III cross)
CL16_SHFB170_MOI10 (F1 recombinant progeny from Type II X Type III cross)
CL17_SHFB172_MOI10 (F1 recombinant progeny from Type II X Type III cross)
CL19_SHFB173_MOI10_regrid (F1 recombinant progeny from Type II X Type III cross)
S23_SHFB245_MOI10 (F1 recombinant progeny from Type II X Type III cross)
S25_SHFB215_MOI10 (F1 recombinant progeny from Type II X Type III cross)
S27_SHFB213_MOI10 (F1 recombinant progeny from Type II X Type III cross)
S2T_SHFB171_MOI10 (F1 recombinant progeny from Type II X Type III cross)
S30_SHFB242_MOI10 (F1 recombinant progeny from Type II X Type III cross)

CL14_SHFB244_MOI10 (F1 progeny from Type II X Type III cross, is identical to type III parent)
CL15_SHFB124_MOI10 (F1 recombinant progeny from Type II X Type III cross)
CL11_SHFB174_MOI10 (F1 recombinant progeny from Type II X Type III cross)
S1T_SHFB120_MOI10 (F1 recombinant progeny from Type II X Type III cross)
CL18_SHFB211_MOI10 (F1 recombinant progeny from Type II X Type III cross)
S21_SHFB214_MOI10 (F1 recombinant progeny from Type II X Type III cross)
S26_SHFB212_MOI10 (F1 recombinant progeny from Type II X Type III cross)
S28_SHFB121_MOI10 (F1 recombinant progeny from Type II X Type III cross)
S29_SHFB122_MOI10 (F1 recombinant progeny from Type II X Type III cross)

cDNA obtained from HFFs infected with the strains above was co-hybridized with cDNA made from Universal Human Reference RNA.
Gene expression of HFFs infected with a type II strain or a type II:ROP16$_1$ was compared by co-hybridization of cDNA derived from HFFs infected with each strain:
SHFB30 and SHFB32 are the corresponding arrays. The type II strains used were: PrugniaudΔHXGPRT and transgenic PrugniaudΔHXGPRT complemented with a plasmid transfected with the type I ROP16 gene (2000 nt upstream of the predicted ATG start codon up to but not including the stop codon which was replaced with a single HA tag (TAC CCG TAC GAC GTC CCG GAC TAC GCG TAA) cloned into the pHEX2 vector which provided the SAG1 3' UTR from the type I parasites and the HXGPRT gene.

Quality control steps taken: For the Type I, II and III strain at least three independent biological replicates were performed.

Samples used, extract preparation and labelling:
Origin of each biological sample: Detailed information about the *Toxoplasma* strains used in this experiment can be found at: http://www.toxomap.wustl.edu/
The genotypes of the parasites used in this paper can be found at:
http://www.toxomap.wustl.edu/IIxIII_Typing_Table.html

Manipulation of biological samples and protocols used:
Strains were maintained *in vitro* by serial passage on monolayers of human foreskin fibroblasts (HFFs) at 37°C in the presence of 5% CO_2. HFFs were grown in Dulbecco modified Eagle medium (GIBCO BRL) supplemented with 10% NuSerum (Collaborative Biomedical Products), 2 mM glutamine, 50 µg/ml penicillin, 50 µg/ml streptomycin, and 20 µg/ml gentamicin (cDMEM).
Intracellular *Toxoplasma* from one T175 were isolated by scraping the monolayer. Infected cells were spun down at 250 g for 10 min, supernatant was removed and cells resuspended in 4 ml of cDMEM. *Toxoplasma* were isolated from infected cells by syringe lysing them first through a 25 gauge needle and then through a 27 gauge needle. The isolated parasites were washed twice with 50 ml of PBS and resuspended in 5 ml of complete medium. After counting parasites, parasites (MOI 10) were added to confluent monolayers of HFFs in two separate T175 and infection proceeded for 24h.
The T175s that had most similar infection rates among the different F1 progeny were used for RNA extraction.

Experimental factor value for each experimental factor, for each sample: All samples were prepared 24 h after infection of HFFs with the different *Toxoplasma* strains.

Technical protocols for preparing the hybridization extract:
Total RNA was isolated from approximately 10^7 *Toxoplasma* infected human foreskin fibroblasts (HFFs) (one confluent T175) using Trizol (Invitrogen) according to the manufacturer's instructions, quantified by absorption spectroscopy and the integrity assessed by agarose gel electrophoresis. From this total RNA polyadenylated RNA was purified using Oligotex (QIAGEN, Valencia, CA). Microarray probes were prepared in two steps. First, cDNA was synthesized as follows: the total amount of polyadenylated RNA from one confluent T175 was mixed with 4 μg of oligo-dT18VN primer in a total volume of 25 μl, heated to 70°C for 10 min and cooled on ice. To this sample, we added dNTP, DTT, 5x reaction-mix and 1 μl of superscript II reverse transcriptase. After a 1-h incubation at 42 °C another 1 μl of superscript II RT was added and the reaction was incubated for another hour at 42 °C. RNA was degraded using Rnase H and RNAse cocktail. The concentration of cDNA was estimated by running 2.5 μl on an agarose gel next to a known amount of DNA ladder. Subsequently, 300 ng cDNA was labeled with Cy5-dUTP using random nonamers and DNA pol I Klenow. Universal human reference RNA was processed in the same manner and 250 ng cDNA was labeled with Cy3-dUTP. We added Cot1 Human DNA (Gibco-BRL), yeast tRNA and poly-A RNA (all, 20 μg), combined both probes, and purified the probe by centrifugation in a Centricon-30 micro-concentrator (Amicon) tubes.

Hybridization procedures and parameters
Arrays were rehydrated by placing them face down in a humid chamber with 100 ml 1X SSC covered chamber with lid. They were rehydrated until array spots glisten, approximately 5-15 minutes. Subsequently, they were snap-dried (DNA side up) on a 70-80°C inverted heat block for 3 seconds and the DNA was UV crosslinked to glass with Stratalinker set for 65 mJ. Prehybridization was done in 5x SSC, 0.1 mg/ml BSA, 0.1% SDS at 42°C for 60 min. (300 ml MQ H2O, 100 ml filtered 20 x SSC, 4 ml 10mg/ml BSA stock solution (filtered), 4 ml 10% SDS, prewarmed in 42°C waterbath for 60 min). After 60 min. arrays were washed in H_20 (three times), placed for 5 min in isopropanol and dried by centrifugation.

Hybridization solution (5.95 μl 20 X SSC, 1.05 μl SDS) was added to the probe and H_2O added to a total volume of 35 μl. Then this solution was put at 100 °C for 2 min, centrifugated and added on top of a microarray. Sequence-verified human cDNA arrays (SHFB and SHEM series, 42,000 spots) were produced at Stanford. Hybridizations were performed under 22x60 mm cover slips at 65 °C in a humidified chamber for 16-20 h (with 4 drops of 3X SSC added to each corner to prevent dry-out). The coverslip was removed in 2x SSC, 0.03% SDS (40 °C) and Hybridized arrays were washed vigorously as follows: one wash for 5 min with 2X SSC (40 °C), once for 5 min with 1X SSC (37°C), and 1 times for 5 min with 0.1X SSC at room temperature. Arrays were dried by centrifugation for 2 min in a slide-rack in a Sorval RT7 plus centrifuge at 700 rpm.

Measurement data and specifications

Data: Raw data will become available at the Stanford Microarray Database (SMD) (http://genome-www5.stanford.edu/)

Normalized and summarized data: The normalized and summarized data is attached in supplementary data 1,2,3

Data extraction and processing protocols: Arrays were scanned using a GenePix 4000A microarray scanner, controlled by GenePix Pro 5.1 software with a pixel size of 10 µm and two-pass sequential line averaging. Laser power was set to 100% and PMT gains were adjusted during prescan to maximize effective dynamic range and to limit image saturation. To extract data from microarray scans, previously stored image files were gridded using Genepix 5.1 and uploaded into Stanford Microarray Database[35] (SMD).

Normalization, transformation and data selection procedures and parameters:
Although changing the normalization method or the parameters for data extraction can change the exact list of genes mentioned in this paper to some extent, this did not lead to substantial alterations in the results of the analyses.

The significance analysis of microarrays (SAM) algorithm with a false discovery rate (FDR) < 15% was used to identify significantly regulated genes. We did the following SAM analyses, one three class comparing I, II and III and all three possible two class comparisons (Type I vs II/III; Type II vs I/III, Type III vs II/I). cDNAs that had a FDR <15% were pooled and the microarray data were analyzed using the 'Multi Experiment Viewer' – MeV version 3.1, included in the TM4 software package from The Institute of Genomic Research (http://www.tigr.org/).
To identify *Toxoplasma* genomic loci involved in strain-specific modulation of host expression a genome-wide scan for association of *Toxoplasma* genetic markers and the expression level of each of the 42,000 cDNAs on the microarray was performed using the program R and the package R/qtl. Genome-wide significance levels for each cDNA were determined by 100 permutations.

Array design
General array design: SHEM and SHFB human cDNA arrays produced at Stanford were used for all the experiments. The arrays are printed with print tip type Majer 17-4 ss on Ultra GAPS slides coated with Corning amino-silane.

Supplementary Notes

J.P.J.S was supported by a grant from the University of California Universitywide AIDS Research Program (F04-ST-216).

28. Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340(4)**, 783-95 (2004).
29. Cokol, M., Nair, R. & Rost, B. Finding nuclear localization signals. *EMBO Rep* **1(5)**, 411-5 (2000).
30. Roos, D. S., Donald, R. G., Morrissette, N. S. & Moulton, A. L. Molecular tools for genetic dissection of the protozoan parasite Toxoplasma gondii. *Methods Cell Biol* **45**, 27-63 (1994).
31. Sibley, L. D., LeBlanc, A. J., Pfefferkorn, E. R. & Boothroyd, J. C. Generation of a restriction fragment length polymorphism linkage map for Toxoplasma gondii. *Genetics* **132**, 1003-15 (1992).
32. Khan, A. et al. Composite genome map and recombination parameters derived from three archetypal lineages of Toxoplasma gondii. *Nucleic Acids Res* **33**, 2980-92 (2005).
33. Donald, R. G., Carter, D., Ullman, B. & Roos, D. S. Insertional tagging, cloning, and expression of the Toxoplasma gondii hypoxanthine-xanthine-guanine phosphoribosyltransferase gene. Use as a selectable marker for stable transformation. *J Biol Chem* **271**, 14010-9. (1996).
34. Burg, J. L., Perelman, D., Kasper, L. H., Ware, P. L. & Boothroyd, J. C. Molecular analysis of the gene encoding the major surface antigen of Toxoplasma gondii. *J Immunol* **141**, 3584-91 (1988).
35. Ball, C. A. et al. The Stanford Microarray Database accommodates additional microarray platforms and data formats. *Nucleic Acids Res* **33(Database issue)**, D580-2 (2005).