Immunological Reviews

Daniel P. Beiting
David S. Roos

# A systems biological view of intracellular pathogens

Authors' address
*Daniel P. Beiting*[1], *David S. Roos*[1]
[1]Department of Biology, Penn Genome Frontiers Institute, University of Pennsylvania, Philadelphia, PA, USA.

Correspondence to:
*David S. Roos*
304B Caroline Lynch Laboratories
Department of Biology
University of Pennsylvania
Philadelphia, PA 19104, USA
Tel.: +1 215 898 2118
Fax: +1 215 898 8780
e-mail: droos@sas.upenn.edu

Acknowledgement
D. P. B and D. S. R are supported by NIH grants AI075846 and AI028724, respectively. The authors declare no conflicts of interest.

**Summary:** As biomedical research becomes increasingly data-intensive, it is increasingly essential to integrate genomic-scale datasets, so as to generate a more holistic picture of complex biological processes. The systems biology paradigm may differ in strategy from traditional reductionist scientific methods, but the goal remains the same: to generate tenable hypotheses driving the experimental elucidation of biological mechanisms. Intracellular pathogens provide an excellent opportunity for systems analysis, as many of these organisms are amenable to genetic manipulation, allowing their biology to be played off against that of the host. Moreover, many of the most fundamental biological properties of these microbes (host cell invasion, immune evasion, intracellular replication, long-term persistence) are directly linked to pathogenesis and readily quantifiable using genomic-scale technologies. In this review, we summarize and discuss some of the available and foreseeable functional genomics datasets pertaining to host–pathogen interactions and suggest that the host–pathogen interface represents a promising, tractable challenge for systems biological analysis. Success will require developing and leveraging new technologies, expanding data acquisition, and increasing public access to comprehensive datasets, to assemble quantitative and testable models of the host–pathogen relationship.

Keywords: *host–pathogen, genomics, systems biology*

## Pathogens as biological 'systems': open, robust, modular, and stochastic

Systems biology has enjoyed a recent surge in attention, but the discipline is rooted in general systems theory, exemplified by the work of Ludwig von Bertalanffy, who noted that biological systems are open, i.e., they exchange information with the surrounding environment (1). Intracellular pathogens can be viewed as the ultimate open system: they are constantly interacting with the infected cell and modulating both their own state, as well as that of the host, so as to establish a viable physical/temporal niche. These organisms typically exhibit specific auxotrophies, relying on host metabolic processes for essential nucleotides, amino acids, lipids, sugars, vitamins, etc. In addition to taking in host nutrients for their own use, bacterial, fungal, and protozoan pathogens deploy an arsenal of secreted proteins (the 'secretome'), including many factors that interact with host cell components. This openness means that biological

*Immunological Reviews 2011*
*Vol. 240: 117–128*

Immunological Reviews
0105-2896

systems cannot simply be described as the sum of their parts (a characteristic termed 'emergence' by Bertalanffy).

The survival of both pathogens and their host cells depends on the ability to buffer extreme environmental perturbation, permitting comparable biological outcomes to emerge even in the face of differing genetic backgrounds and environmental conditions (the phenomenon of convergence, which Bertalanffy called 'equifinality'). Yet, pathogens and host cells must also remain sensitive to *bone fide* cues warranting a change in behavior. This robust quality of biological systems is achieved through a combination of functional redundancy and feedback regulation (both positive and negative). Bacterial chemotaxis provides a classic example of robust behavior, responding to relative rather than absolute levels of soluble signal, using methylation-mediated chemokine receptor desensitization to maintain constant steady-state movement over many orders of magnitude in stimulus concentration (2, 3). Disrupting the robustness of regulatory responses can have a dramatic impact on cell signaling (4, 5), gene transcription (6, 7) and organismal phenotype (8, 9), and may constitute an important mechanism for tipping the scales in favor of the pathogen versus the host.

Cellular processes are rarely dictated by single genes acting in isolation. More commonly, they are governed by functional modules defined by the association of coordinately expressed genes (and protein–protein, protein–DNA complexes, etc.). In an elegant experimental demonstration of modularity, Geut *et al.* (10) combined three bacterial transcriptional regulators with five promoters (each responsive to one of the three regulators) in various permutations and combinations, generating a library of >100 promoter-gene combinations. The resulting networks of elements previously thought to be well understood exhibit a striking diversity of signal output. Similar complexity is evident in other examples of host–pathogen interactions such as cell invasion and tissue specificity, where specific interactions of encapsidated viruses with specific host receptors (11) are replaced by a diversity of interacting receptors and ligands for bacterial and eukaryotic pathogens (12, 13). The networks of genes and functional interactions underlying basic biological properties of pathogens (i.e. interactions with infected cells) are not well understood, in part because of their complexity, and in part because we lack the statistical/computational and biochemical/biophysical/genetic/cell biological tools required to assess the relative contributions of multiple, redundant, low-specificity pathways, involving contributions from both the host and the pathogen.

Despite their robust nature, biological pathways often exhibit a considerable degree of cell-to-cell variation or stochasticity. The processes that govern cell behavior, such as transcription and translation, are inherently noisy, due to the low copy number of molecules involved (transcription factors, coactivators, etc.), the unpredictability of when components will come together to perform a function (14), unequal partitioning of components during cell division (15), and other factors. The end result of this noise is that a single cell can give rise to a population exhibiting heterogeneous properties (16), as has been known for decades from studies on the development of spontaneous resistance to bacteriophage infection in *Escherichia coli* (17). Recent observations at single cell and single molecule resolution have shown that stochasticity can have a profound impact on organismal biology, including microbial persistence and pathogenicity in the face of selective pressure from drug treatment or host immune responses. Using microfluidic channels to track bacterial replication at single cell resolution during antibiotic treatment demonstrates that those bacteria that happen to be growing most slowly in a population prior to drug treatment (the tail of the curve) are also those that persist after drug treatment (18). Further studies have used an inducible fluorescent reporter to track the fate of individual *Salmonella* for up to 10 generations during *in vitro* or *in vivo* infection (19, 20), revealing that a sub-population of bacteria rapidly enter a non-dividing dormant state and remain viable. The development of chronic infections resistant to innate and adaptive immune defenses is a defining feature of *Mycobacterium tuberculosis*, *Chlamydia trachomatis*, *Toxoplasma gondii*, *Plasmodium vivax*, *Trypanosoma cruzi*, and numerous other pathogens; stochasticity may play an important role in determining which pathogens are cleared, and which persist over the long-term.

## Expression profiling of intracellular pathogens

Systems approaches require a wealth of data, integrating information on genetic diversity, gene regulation, transcript abundance, protein synthesis and turnover, metabolite levels, etc. Recent years have seen a surge in the availability of such datasets, generated on various platforms, and this flood in information will undoubtedly grow, with the advent of new sequencing technologies (21), improvements in quantitative label-free proteomics (22, 23), and the rise of metabolomic analysis (24) and other approaches. Complete genome sequence and gene expression profiling data – often from multiple isolates – are available for all of the pathogens discussed in this issue, in many cases including both axenic (free-living) organisms and isolates obtained during the course of *in vitro* and/or *in vivo* infection.


 

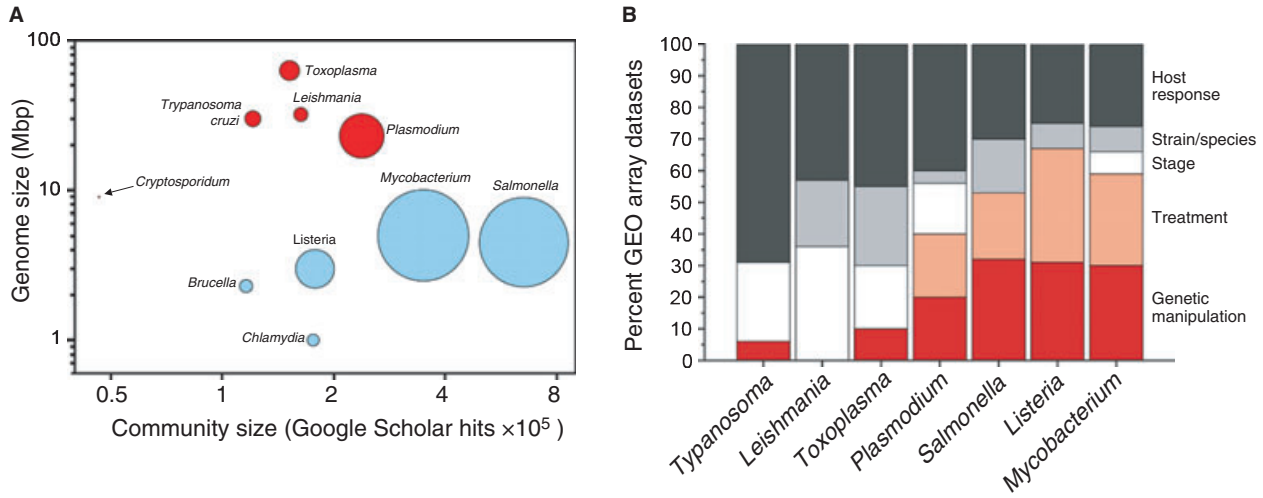


***Fig. 1.* Genomes and gene expression datasets for intracellular pathogens.** (A) Bacterial (blue) and protozoan (red) pathogens plotted to indicate genome size and the size of the research community, estimated based on citations in Google Scholar. Bubble size indicates the number of gene expression datasets (pathogen and host) accessible through NCBI GEO and/or ArrayExpress. (B) Gene expression datasets for three bacterial pathogens and four protozoan parasites, classified according to the kind of experiments they represent. See Table S1 for details and text for further discussion.

To assess the extent to which these tools have been exploited by the research community, we selected ten intracellular pathogen species, including both bacteria and protozoa, and extracted from the Gene Expression Omnibus (GEO) (http://ncbi.nlm.nih.gov/geo) (25) and ArrayExpress (http://www.ebi.ac.uk/arrayexpress) (26) databases all microarray datasets publically deposited for each species. The size of bubbles shown in *Fig. 1A* indicates the number of datasets available for each pathogen at the time of submission of this manuscript, plotted as a function of genome size, and the scale of the research community (estimated from the number of Google Scholar hits). As one might expect, the largest research communities (*Salmonella*, *Mycobacterium*, *Plasmodium*) have typically produced the most extensive array datasets (90, 92, and 45, respectively), but expression profiling is clearly underutilized as tool to dissect gene function in some species (*Chlamydia*, *Leishmania*, *Trypanosoma*) relative to others (*Listeria*, *Toxoplasma*). The number of expression profiling datasets for even the most intensively studied pathogens pales, however, by comparison with single-celled model organisms, such as *Saccharomyces cerevisiae*, for which >1500 microarray datasets are available – far more than for all pathogen species combined.

While the absolute number of available expression profiling datasets may be of some interest, the nature of the experiments from which they were generated is a more useful indicator of pathogen research. We therefore classified these database entries as focused on treatments (e.g. exposure to drugs, or incubation under differing conditions), genetic manipulation (e.g. comparison of wildtype with mutants or knockouts), stage (e.g. cell-cycle progression, pathogen differentiation), strain/species (e.g. field isolates versus culture-adapted strains), or host response (transcript profiling of the host rather than the pathogen), as indicated in *Fig. 1B* (Table S1 Supporting information). Several interesting themes emerge from this analysis. For example, significantly more genetic manipulation studies have been carried out in bacterial than protozoan pathogen systems (red bars in *Fig. 1B*), highlighting an area in need of further emphasis, particularly in light of technological breakthroughs improving the efficiency of mutagenesis, targeted gene disruption and regulatable gene expression in eukaryotic pathogens (27–33).

Another interesting theme to emerge from the classification of expression profiling datasets is the focus of certain research communities on metabolic pathways, using specific drug and chemical treatments (pink bars in *Fig. 1B*). This emphasis undoubtedly reflects high interest in drug discovery (particularly for tuberculosis and malaria), but it is important to consider that such studies also help elucidate metabolic pathways and regulatory gene networks. The most extensive expression profiling study reported for any intracellular pathogen includes 430 microarray experiments examining the effects of 75 different drugs and compounds on gene expression in *M. tuberculosis* (34). This work reveals a dynamic bacterial transcriptome, allowing the identification of pathogen gene modules involved in bacterial transcription, translation, DNA repair, nucleotide biosynthesis, cell wall biosynthesis, and respiration. Similarly, an extensive transcriptional map of drug-specific responses in *Plasmodium* (35) includes approximately150 microarray experiments, spanning 20 different compound treatments over multiple time-points. This work

integrated analysis of co-expression with orthology predictions and protein–protein interaction data (36), constructing a network from which gene function could be inferred, providing putative functional assignments for >90% of 'hypothetical' *Plasmodium* genes. Additional studies focused on a subset of these genes, predicted to be associated with erythrocyte invasion by *P. falciparum* merozoite, and confirmed colocalization with the apical invasion machinery for 31 of 42 candidates tested. These studies highlight the value of integrating expression profiling with other 'Omic' scale datasets in elucidating biological pathways (37, 38).

Many cells exist in different forms, as they progress through their replicative cycle, or differentiate through the various stages that characterize certain pathogen life cycles. *Plasmodium* offers an unusual take on cell-cycle progression, a phenomenon that has been extensively studied in yeast and many other organisms: hourly expression profiling through the 48 h *P. falciparum* erythrocytic cycle shows that approximately 75% of transcripts exhibit cell-cycle regulation (39, 40), in contrast with approximately 6% for *S. cerevisiae* (41). A recent study describes similar (albeit less dramatic) cell cycle-dependent transcription in *Toxoplasma* (42). These results suggest that the transcriptome of apicomplexan parasites may be less prone to perturbation (43) than bacteria (34), but offer an opportunity to exploit co-expression patterns to define probable gene function, as noted above.

Stage-specific expression is commonly associated with pathogen transmission (e.g. the insect, mammalian liver, and blood stages of *Plasmodium*) and/or disease pathogenesis, such as the latent forms of *M. tuberculosis* infecting alveolar macrophages (associated with granuloma formation, destruction of the lung parenchyma, and potential reactivation later in life), or the 'bradyzoite' tissue cyst forms of *T. gondii* (causing encephalitis in immunosuppressed patients, and recrudescent chorioretinitis following congenital infection). In an effort to dissect pathogen pathways responsible for *M. tuberculosis* persistence, Voskuil *et al.* (44) used microarrays to evaluate changes in gene expression following exposure to nitric oxide, a potent effector produced by activated macrophages. Low concentrations of nitric oxide specifically induced the expression of 48 genes, which further investigation revealed as a dormancy regulon mobilized in response to stress, driving the transition to latent infection. Similarly, several investigators have explored transcript abundance levels during the differentiation of acutely lytic *T. gondii* tachyzoites into the latent bradyzoite cyst form (http://toxodb.org/toxo/showQuestion.do?questionFullName = InternalQuestions.GenesByMicroarrayEvidence).

These studies clearly demonstrate the power of genomic approaches to elucidate genetic programs in pathogen biology. Further expression profiling (whether by microarray or high-throughput sequencing) may be expected to greatly enhance our understanding of gene function within the context of the genome as a whole, enabling any researcher working on gene X to ask: 'What other genes are co-regulated with my favorite gene?' or 'In what mutant lines, treatments, or developmental programs is the expression of my favorite gene altered?'

## Expression profiling of host responses to infection

The first experiments using microarray technology to profile host response to infection were carried out more than a decade ago (45). Many subsequent studies have examined host cell transcriptional responses to a variety of intracellular pathogens, but fundamental differences in the host cells/species and the biology of these pathogens (*in vitro* growth properties, developmental stages, etc.) make it difficult to define common versus unique components of host response to infection. Jenner & Young (46) used meta-analysis to integrate and analyze publically accessible datasets from 32 published studies, representing nearly 800 experiments and >70 host–pathogen interactions. This strategy normalizes transcript abundance across experiments to identify genes sharing common expression patterns. A core set of 511 common host response genes, highly enriched for genes known to be involved in inflammation, were identified as induced in all cell types examined, regardless of the pathogen or host species. While this report focused on viral and bacterial pathogens (only one parasitic protist was included), Zhang *et al.* (47) recently conducted a large microarray study on host response to *Leishmania* and *Trypanosoma*, using meta-analysis to compare their own data with other publicly available host response data sets for these pathogens. *L. mexicana* and *T. cruzi* induce signature host responses differing from each other, and also from the bacterial/viral signature identified previously. As in the study by Jenner & Young, this report noted that certain pathogen-specific host response signatures were conserved in all of the experiments analyzed, suggesting that certain datasets may be clustered for large-scale data analysis, regardless of the host cell type or pathogen strain/isolate used. This observation bodes well for database mining, providing a powerful way to dissect host cell responses to a broad range of pathogens, as discussed further below.

Host and pathogen gene expression studies are typically carried out as separate experiments, often in different


 

laboratories, and usually employing distinct microarray platforms specifically designed for each species. Unfortunately, microarray technologies are dependent on the quality of available genome sequence and annotation, and mandate species–specific platforms. As a consequence, relationships between the host and pathogen are usually lost, although a few recent studies have used multifunctional microarrays to probe simultaneously gene expression in *P. berghei* and its insect vector (48) or mammalian host (49). The development of new technologies based on deep sequencing (RNA-seq) (50) is therefore very exciting, offering the prospect for simultaneous profiling of pathogens and their hosts, reducing concerns about gene model accuracy and platform-dependent effects. We have recently embarked upon a study cataloging both coding mRNA and small non-coding RNA populations (in parallel with quantitative proteomics from the same samples) from human cells infected with various *Toxoplasma* strains. By generating host and parasite datasets in parallel, we hope to maintain critical relationships between host and parasite regulatory RNAs, transcripts, and proteins, providing unprecedented insight into the host–pathogen relationship. Given the growing recognition of roles played by non-coding RNAs in regulating gene expression (51, 52), it is intriguing to consider that bacterial or eukaryotic pathogen non-coding RNAs may regulate host genes, as recently demonstrated for viral pathogens (53). Simultaneous analysis of RNAs in both host and pathogen should also permit unbiased identification of co-regulated RNAs that may interact, either physically or functionally.

### Integrative approaches to elucidate host–pathogen interactions

As noted above, integrating bioinformatic analysis of genomic-scale datasets with molecular genetic tools provides many exciting research opportunities for addressing complex questions related to host–pathogen interactions. In theory, at least, computational approaches can be used to identify pathogen factors likely to be involved in salvaging nutrients from the host or disrupting host cell function, host factors likely to be involved in recognizing and controlling pathogen activity, and host and pathogen factors likely to interact with each other. For example, such integrative studies have been used to identify mechanisms by which *T. gondii* tachyzoites regulate host cell signaling. Previous work had shown that infection results in phosphorylation of the host transcription factor complex signal transducer and activator of transcription 3 (STAT3)/STAT6 and subsequent suppression of interleukin-12 (IL-12) production by macrophages, a phenotype observed only in cells infected with avirulent strains of the parasite (54, 55). Saeij *et al.* (56) isolated progeny from a genetic cross between virulent and avirulent parents, and used expression profiling data as quantitative traits for QTL mapping to define the genetic locus responsible for the regulation of host signaling. Using standard molecular genetics approaches to swap alleles between strains, they formally demonstrated that the highly polymorphic secreted kinase ROP16 is a key regulator of host STAT3/STAT6 signaling. Other studies have used genetic mapping to highlight the importance of another rhoptry kinase (ROP18) in *T. gondii* virulence (57, 58).

Given the importance of these secreted rhoptry kinases (59), the likelihood that host phosphorylation cascades will prove to be critical in regulating infection, and the wealth of kinases encoded in the parasite genome, Peixoto *et al.* (60) exploited phylogenomic approaches to define the apicomplexan kinome, focusing on those predicted to be secreted outside of the parasite. Seventeen kinases harboring a complete catalytic triad were used to generate a hidden Markov model specific for the rhoptry kinase (ROPK) family, which was applied to the *Toxoplasma* genome to identify a total of 44 genes, including many 'pseudokinases' predicted to be inactive. Integrative genomic analysis, combining differential expression data for various parasite strains and life cycle stages with evidence of evolutionary selection, highlights a limited number of genes as likely candidates for regulating host responses, including ROP16, ROP18, and the previously uncharacterized gene ROP38. Expression profiling shows that ROP38 is highly expressed in avirulent strains (e.g. VEG), but virtually absent in virulent strains (e.g. RH). This gene arose recently, *via* triplication subsequent to the divergence of *Toxoplasma* from the closely related cattle pathogen *Neospora caninum*. To assess the function of ROP38, RH strain parasites were engineered to express a ROP38 transgene to levels normally observed in avirulent strains, and the effect of this mutation was examined by expression profiling of both parasites and infected host cells. Parasites expressing the ROP38 transgene suppress the expression of over 1000 host response genes, including key transcription factors associated with mitogen-activated protein kinase (MAPK) signaling.

If we are interested in exploring host–pathogen interactions, it will not be sufficient to identify virulence factors alone; we must also define interacting partner(s). Protein–protein interactions between host and pathogen have typically been defined by affinity chromatography or precipitation, or using genetic screens such as the yeast two-hybrid assay (61). Alternatively, it is increasingly attractive to exploit informatics

approaches to identify candidate interacting partners. Davis *et al.* (62) identified host–pathogen protein pairs exhibiting similarity to protein pairs previously found to interact, and filtered this list based on structural properties of the interacting pair, subcellular localization in both host and pathogen, and expression levels during infection. In other words, they asked for host and parasite proteins present in the right place, at the right time, and with the right 'look' to interact. In another study, Dyer *et al.* (63) extracted all known virus–host protein interactions from various databases and constructed an interaction networks, revealing that pathogens typically target host factors interacting with many other proteins (hubs), or connecting many different pathways (bottlenecks). As host–pathogen interactions are prominent at the time of host cell invasion, Chen *et al.* (64) compiled a list of all proteins known to be associated with the micronemes of apicomplexan parasites, which are thought to be released in association with attachment and invasion (65). Domain signatures associated with these microneme proteins identified new microneme candidates in the genomes of 12 apicomplexan parasites, several of which were validated by subcellular localization. Domain–domain and protein–protein interactions known from other systems were also used to computationally predict host factors that might interact with microneme proteins during invasion, providing a wealth of candidates that awaits experimental validation.

## Functional genomic screens using intracellular pathogens

### Pathogen-based screens

Like the expression profiling studies reviewed above, genetic screening offers another powerful approach for interrogating the genomes of pathogens and their host cells. When Brenner (66) published his seminal paper on forward genetic screening in the free-living nematode *Caenorhabditis elegans*, he noted that while 'nematodes do not have a rich repertoire of external features for mutant selection [in contrast to *Drosophila*], this has not prevented the isolation of visible mutants. Indeed, it has had the effect of focusing selection on the behavioral characteristics of the animals.' Characterization of 'uncoordinated' mutant worms set the stage for establishing *C. elegans* as a model genetic system for studying organismal development. The ability of pathogens to invade their hosts, modulate subcellular processes, replicate intracellularly, and egress from the infected cell is central to pathogenesis, and like the movement and coordination of Brenner's worms, these traits are readily observed and measured *in vitro*. Several intracellular pathogen species are genetically tractable, making possible the same forward and reverse genetic approaches that have been used for decades to assign function to genes in model organisms.

Although the first forward genetic screens in *C. elegans* were very labor intensive, the relatively large size and hermaphroditic nature of this organism helped to simplify the process. After scrutinizing thousands of worms under a low power microscope to identify mutants of interest, individual organisms could be physically isolated for clonal propagation. Applying such approaches to pathogenic microbes raises several significant challenges, not least of which is the difficulty of identifying biologically significant changes in microscopic organisms, in the context of millions of wildtype pathogens growing in the same culture plate or infected animal.

One successful solution to this problem has been provided by genetically 'bar-coding' individuals with unique oligonucleotide sequences, and subjecting pools of tagged mutants to selective conditions (e.g. growth in minimal media, or in mice). By comparing the tags present in the starting population with those recovered after selection, one can determine which mutants failed to survive. When Hensel *et al.* (67)first described 'signature-tagged mutagenesis' (STM), they screened over 1000 tagged *Salmonella* mutants for the ability to survive during a 3-day mouse infection. Forty-three mutants present in the initial inoculum failed to survive in mice, leading to the identification of the *Salmonella* type III secretion system as a critical determinant of bacterial virulence (68). STM has been successfully adapted to numerous bacterial species (69–71), and in an effort to achieve saturation mutagenesis in *M. tuberculosis*, Sassetti *et al.* (72) have exploited microarray hybridization to facilitate target identification in complex pools of tagged mutants. Deep sequencing will undoubtedly enhance the speed and saturation of such studies in many pathogens.

These techniques have typically identified approximately 4% of non-essential genes as impacting virulence (73), but many of the genes identified by such approaches are involved in housekeeping functions or nutrient salvage, so it is not particularly surprising that their disruption produces a crippled pathogen. If such techniques are to be useful for understanding specific interactions with the host, it will be important to design secondary screens that distinguish genes specifically involved in countering or evading host immune defenses, or otherwise interfacing with host signaling pathways. In an elegant example (74), *Salmonella* type III secretion mutants were used to infect mice deficient in immune effector molecules suspected to play a role in pathogen control. Mutations in host IL-12, IFNγ, or iNOS failed to restore the


 

virulence of the mutant bacteria, but mice deficient in NADPH oxidase were susceptible to infection, suggesting that type III secretion systems might specifically target this host effector function.

The larger genome size and different gene (and genome) structure of eukaryotes complicates the use of STM approaches, especially where efficient transposon systems are lacking, as in most protozoan parasites. Insertional mutagenesis has been successfully applied for STM analysis of *T. gondii*, although not to saturation (75, 76). In other eukaryotes, including *C. elegans* (77), *Drosophila* (78), and mammalian systems (79–81), RNA interference approaches have been effectively exploited for reverse genetics, but these strategies have not yet proved practically useful for intracellular protozoan parasites, despite the apparent presence of the necessary machinery in some species of *Leishmania* and *T. gondii*, and the utility of RNAi in *T. brucei* (82). All of these parasites are amenable to stable and transient transfection, however, making forward genetics possible. Building on the pioneering genetic studies of Pfefferkorn (83), Gubbels *et al.* (29) have employed fluorescent *T. gondii* parasites and chemical mutagenesis to generate a collection of temperature sensitive mutants. Replica-plated parasite clones were monitored for growth at restrictive versus permissive temperature, allowing the identification of >65 cell-cycle mutants from an initial collection of 60 000 chemical mutants. Complementation with a genomic cosmid library permitted identification of the mutant genes responsible for the phenotype. It is exciting to consider exploiting such libraries for gain-of-function screens to identify genes enhancing virulence. Although plasmid recombination in wild-type *T. gondii* is normally dominated by non-homologous events (84), deletion of the Ku80 recombinase (28, 31) enhances homologous recombination, opening the door for high-throughput gene disruption and tagging studies.

Other strategies for forward genetic analysis have exploited fluorescent parasites to screen for genes involved in invasion and egress (85, 86), using automated microscopic imaging for analysis in 96 or 384-well format. Taking advantage of the ability to label various subcellular compartments (in various colors), in living parasites (87), and the ability to isolate living parasites clonally by fluorescence-based cell sorting, there is great potential to observe aspects of pathogen behavior ranging from cell division to protein trafficking.

While fluorescence-based assays offer the advantage of live cell and morphology-based measurements, they lack the sensitivity of luciferase assays. *Toxoplasma* parasites expressing luciferase have been used to track parasite burden *in vivo* (88) and screen for anti-parasitic drugs (89), but have not yet been fully exploited for high-throughput *in vitro* screens, in part because conventional detection requires cell lysis. Borrowing a method commonly used for circadian rhythm studies in mammalian cells (90), micromolar quantities of the non-toxic substrate D-luciferin can be added directly to living cultures (91), allowing pathogen replication or specific gene expression to be monitored longitudinally under virtually any experimental condition.

### Host cell screens

The discovery of RNAi in *C. elegans* and the subsequent realization that double stranded RNA molecules could also be used to silence genes in mammalian cells have revolutionized functional genomic analysis of eukaryotic systems (92, 93). The first genome-wide screens using RNAi to identify host genes required for infection were carried out in *Drosophila* cells infected with either *Listeria* or *Mycobacterium* (94, 95). *Drosophila* cells are well suited to large-scale RNAi screens due to their relatively low genetic redundancy, and ability to readily take up and process long gene-specific RNAs into short 21-mers, resulting in highly efficient gene silencing (96). This genetic tractability has been exploited to dissect cellular pathways required for infection by numerous viruses (97, 98).

In contrast, mammalian cells exhibit higher genetic redundancy, require transfection reagents for RNA uptake, and also require that these RNAs be provided as 21-mers targeting a specific gene of interest. As a result, mammalian RNAi studies are more prone to both false positive results (off-target effects) and false negatives (due to partial knockdown or functional redundancy). Genome-wide RNAi screens have not yet been reported in mammalian cells infected with bacterial or protozoan pathogens, although focused libraries have been screened, e.g. for cytoskeletal factors regulating the intracellular motility of Rickettsia (99), and kinases that may regulate *Plasmodium* sporozoite infection of liver cells (100). As a complementary strategy, we have exploited the Mammalian Gene Collection (MGC) library of full-length cDNAs (101, 102), to carry out gain-of-function screens in host cells infected with *Toxoplasma*. In contrast to RNAi, which often yields incomplete suppression of target genes, cDNA screens typically yield high levels of ectopic gene expression. Two separate screens surveying >18 000 cDNAs in 384-well format have identified a set of host cell genes that restrict infection by *T. gondii*, or modulate parasite regulation of IFNγ/STAT1 signaling (manuscripts in preparation).

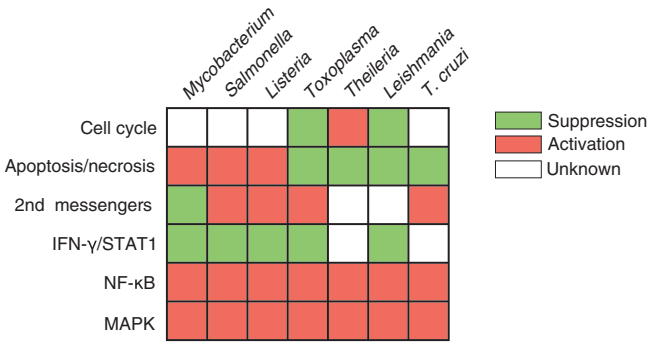


***Fig.* 2. Pathogen modulation of host signaling pathways.** The effect of bacterial or protozoan pathogens (top) on various cell-autonomous host cell pathways (left). Red indicates activation; green indicates suppression; white indicates an unknown phenotype. See Table S2 for literature references.

### Applying genomic screens to dissect pathogen control of host signaling pathways

Pathogens are pathogenic only because they impact host cells/organisms, which requires that pathogen research ultimately converge on host signaling pathways. Viruses, pathogenic bacteria, and parasitic protozoa have adapted to survive within their intracellular niches by regulating host processes in many ways, including the acquisition of nutrients, modification of the host environment and interference with immune responses. *Fig.* 2 (see Table S2 Supporting information for references) summarizes published references on pathogen regulation of key host signaling pathways. Activation of host NF-κB and MAPK pathways and suppression of IFNγ/STAT1 signaling represent common themes, for both bacterial and protozoan pathogens, and many viruses as well. In contrast, while many bacterial pathogens induce host cell apoptosis, perhaps to facilitate rapid dispersal, protozoan pathogens typically protect the infected cell from programed death, possibly reflecting a need to persist within cells for developmental progression.

Even when targeting the same pathway, modulation can be achieved by many independent means. For example, activation of NF-κB and MAPK pathways during bacterial infection is commonly mediated *via* engagement of Toll-like receptors (TLRs), but *Salmonella* can also activate these pathways by targeting Rho-family GTPases directly (103). Similarly, bacterial and protozoan pathogens employ various strategies to suppress STAT1 signaling (104–107), and an even wider range of mechanisms is known in viral systems (108). The observation that such diverse pathogens elicit similar host signaling outcomes during infection indicates that a strong selective pressure drives these specific host–pathogen interactions.

In contrast to non-autonomous traits, such as the activation of cells by diffusible cytokines, intracellular signaling pathways and other cell-autonomous phenotypes offer great potential as the basis for selective screens. For example, cells infected with *T. gondii* are resistant to apoptosis (109) raising the prospect of screening for factors – small molecules, parasite mutants, host cDNAs, siRNAs targeting the host – that re-sensitize cells to inducers of apoptosis. Screening uninfected and infected cells might be expected to identify factors that complement the targeted phenotype in infected cells, revealing genes/pathways specifically targeted by the pathogen. One might also expect to identify factors that modulate pathway activity regardless of infection status, representing genes downstream of the pathogen-induced block that improve our understanding of complex mammalian signaling networks.

Carette *et al.* (110) recently described an exciting new method for cell-based screens, using a myeloid leukemia cell line engineered to carry only a single copy of most chromosomes. Retroviral insertional mutagenesis of this effectively haploid human cell line produces a large collection of null mutants, a strategy previously available only for model organisms such as *S. cerevisiae*. Mutant populations (rather than individual clones) resistant to lysis by specific bacterial toxins or infection with influenza virus were profiled by deep sequencing of cDNAs, providing genome-wide identification of potential receptors and other targets/resistance pathways; similar strategies will undoubtedly be applied to cellular pathogens in the near future.

## Integrating, managing, and querying large data sets

With the rapid advance of technologies for high-throughput data generation, the need to handle complex datasets, derived from genome sequencing and population diversity projects, expression profiling microarrays and RNA-seq, proteomic and metabolomic analysis, large-scale screening, etc., has become a common theme in biomedical research. The ability to mine these datasets effectively hinges on tools for data integration and visualization, and a growing number of resources are available to help make sense of genomic-scale datasets (many open-source). These include tools for functional clustering (111), ontological analysis (112), gene set enrichment analysis (113), and network analysis (114); such resources have been extensively reviewed elsewhere (115, 116). As it becomes increasingly possible to exploit *in silico* experiments to drive hypothesis formulation, training in basic bioinformatics skills (such as the use of Perl to parse files, or Bioconductor and other R packages for statistical analyses) is increasingly essential for success in biomedical research.


 

Analyzing one's own genomic-scale experiments can be a daunting task, and making sense of data generated by the broader research community presents an even greater challenge. Over the past decade, the Eukaryotic Pathogen Genome Database project (http://EuPathDB.org) (117) and other Pathogen Bioinformatics Resource Centers (118) have sought to integrate diverse datasets from many researchers and experimental platforms, into queryable databases enabling users to formulate their own questions. In addition to providing encyclopedic gene and genome views (i.e. everything known about a particular gene, or chromosomal span), EuPathDB employs a graphical query interface that allows users to formulate sophisticated queries that may be saved or shared for further refinement, or downloaded for further analysis. For example, a *Plasmodium* researcher interested in identifying new vaccine candidates (119, 120) might design a query strategy seeking all *P. falciparum*-specific genes that are highly polymorphic (suggesting that they may be under immune selection), and predicted to be secreted during blood stage infection. This query might leverage genome and population genetic data, gene, and/or protein-level expression profiling, motif-finding algorithms, orthology and phylogeny tools, etc., to prioritize experimentally tractable targets for further analysis at the lab bench or in the field/clinic. We, and others, have also explored the potential of weighting criteria to facilitate prioritization further, particularly for drug targets (121–123). These resources seek to reduce the need for specialized bioinformatics expertise on the part of individual researchers, enabling them to focus on biological validation.

The queries outlined above are entirely dependent on the diverse genomic-scale datasets that have been produced over the past decade, in various laboratories around the world. If the role of scientific publication is to insure that future researchers can build on existing knowledge, large-scale datasets cannot be considered published unless they are available in electronic format. Indeed, systems-level analysis is impossible unless such data is computationally accessible. Fortunately, the development of data standards and repositories has tracked the growth of genomic-scale analysis, with resources such as GEO and ArrayExpress providing archival storage of expression profiling datasets, although these resources are not well structured for interrogation, as illustrated by the manual curation required to generate Table S1 [Supporting information]. Worse still, no tools are available for interrogating the data within these diverse datasets directly, except in cases where they have been loaded into specialized databases such as EuPathDB. For example, the summary of pathogen interactions with signaling pathways presented in *Fig.* 2 was compiled from published conclusions, rather than by interrogating the underlying data. As a result, comprehensive automated analysis of host datasets from multiple sources is not currently possible, and integrated analysis of both host and pathogen datasets in parallel is currently a genomic 'no man's land,' posing a clear challenge for the future.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

***Table S1*. Manually curated gene expression microarray datasets used to generate *Fig.* 1b.** Datasets for five bacterial pathogens and four parasitic protists were retrieved from the NCBI public repository, Gene Expression Omnibus (GEO). GEO 'Series' (GSExxxx) datasets (rows) were classified according to the kind of experiment they represent (columns). Individual pathogens are summarized on separate tabs of the spreadsheet.

***Table S2*. A summary of references that were manually curated to generate the heatmap shown in *Fig.* 2.**

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

## References


1. Bertalanffy LV. General Systems Theory: Foundations, Development, Applications. New York: George Braziller, 1968.
2. Alon U, Surette MG, Barkai N, Leibler S. Robustness in bacterial chemotaxis. Nature 1999;**397**:168–171.
3. Barkai N, Leibler S. Robustness in simple biochemical networks. Nature 1997;**387**:913–917.
4. Bhalla US, Iyengar R. Emergent properties of networks of biological signaling pathways. Science 1999;**283**:381–387.
5. Xiong W, Ferrell JE Jr. A positive-feedback-based bistable 'memory module' that governs a cell fate decision. Nature 2003;**426**:460–465.
6. Becskei A, Seraphin B, Serrano L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. EMBO J 2001;**20**:2528–2535.
7. Becskei A, Serrano L. Engineering stability in gene networks by autoregulation. Nature 2000;**405**:590–593.
8. Rutherford SL, Lindquist S. Hsp90 as a capacitor for morphological evolution. Nature 1998;**396**:336–342.

9. Queitsch C, Sangster TA, Lindquist S. Hsp90 as a capacitor of phenotypic variation. Nature 2002;**417**:618–624.
10. Guet CC, Elowitz MB, Hsing W, Leibler S. Combinatorial synthesis of genetic networks. Science 2002;**296**:1466–1470.
11. Randall-Hazelbauer L, Schwartz M. Isolation of the bacteriophage lambda receptor from *Escherichia coli*. J Bacteriol 1973;**116**:1436–1446.
12. Cowman AF, Crabb BS. Invasion of red blood cells by malaria parasites. Cell 2006;**124**:755–766.
13. Scherf A, Lopez-Rubio JJ, Riviere L. Antigenic variation in *Plasmodium falciparum*. Annu Rev Microbiol 2008;**62**:445–470.
14. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science 2002;**297**:1183–1186.
15. Christen M, Kulasekara HD, Christen B, Kulasekara BR, Hoffman LR, Miller SI. Asymmetrical distribution of the second messenger c-di-GMP upon bacterial cell division. Science 2010;**328**:1295–1297.
16. Choi PJ, Cai L, Frieda K, Xie XS. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. Science 2008;**322**:442–446.
17. Luria SE, Delbruck M. Mutations of bacteria from virus sensitivity to virus resistance. Genetics 1943;**28**:491–511.
18. Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S. Bacterial persistence as a phenotypic switch. Science 2004;**305**:1622–1625.
19. Helaine S, Thompson JA, Watson KG, Liu M, Boyle C, Holden DW. Dynamics of intracellular bacterial replication at the single cell level. Proc Natl Acad Sci USA 2010;**107**:3746–3751.
20. Roostalu J, Joers A, Luidalepp H, Kaldalu N, Tenson T. Cell division in *Escherichia coli* cultures monitored at single cell resolution. BMC Microbiol 2008;**8**:68.
21. Cloonan N, Grimmond SM. Transcriptome content and dynamics at single-nucleotide resolution. Genome Biol 2008;**9**:234.
22. Geromanos SJ, et al. The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. Proteomics 2009;**9**:1683–1695.
23. Silva JC, et al. Quantitative proteomic analysis by accurate mass retention time pairs. Anal Chem 2005;**77**:2187–2200.
24. Olszewski KL, et al. Host–parasite interactions revealed by *Plasmodium falciparum* metabolomics. Cell Host Microbe 2009;**5**:191–199.
25. Barrett T, et al. NCBI GEO: archive for high-throughput functional genomic data. Nucleic Acids Res 2009;**37**:D885–890.
26. Parkinson H, et al. ArrayExpress update – from an archive of functional genomics experiments to the atlas of gene expression. Nucleic Acids Res 2009;**37**:D868–872.
27. Balu B, Shoue DA, Fraser MJ Jr, Adams JH. High-efficiency transformation of *Plasmodium falciparum* by the lepidopteran transposable element piggyBac. Proc Natl Acad Sci USA 2005;**102**:16391–16396.
28. Fox BA, Ristuccia JG, Gigley JP, Bzik DJ. Efficient gene replacements in *Toxoplasma gondii* strains deficient for nonhomologous end joining. Eukaryot Cell 2009;**8**:520–529.
29. Gubbels MJ, et al. Forward genetic analysis of the apicomplexan cell division cycle in *Toxoplasma gondii*. PLoS Pathog 2008;**4**:e36.
30. Herm-Gotz A, et al. Rapid control of protein level in the apicomplexan *Toxoplasma gondii*. Nat Methods 2007;**4**:1003–1005.
31. Huynh MH, Carruthers VB. Tagging of endogenous genes in a *Toxoplasma gondii* strain lacking Ku80. Eukaryot Cell 2009;**8**:530–539.
32. Meissner M, Schluter D, Soldati D. Role of *Toxoplasma gondii* myosin A in powering parasite gliding and host cell invasion. Science 2002;**298**:837–840.
33. Maier AG, et al. Exported proteins required for virulence and rigidity of *Plasmodium falciparum*-infected human erythrocytes. Cell 2008;**134**:48–61.
34. Boshoff HI, Myers TG, Copp BR, McNeil MR, Wilson MA, Barry CE 3rd. The transcriptional responses of *Mycobacterium tuberculosis* to inhibitors of metabolism: novel insights into drug mechanisms of action. J Biol Chem 2004;**279**:40174–40184.
35. Hu G, et al. Transcriptional profiling of growth perturbations of the human malaria parasite *Plasmodium falciparum*. Nat Biotechnol 2010;**28**:91–98.
36. LaCount DJ, et al. A protein interaction network of the malaria parasite *Plasmodium falciparum*. Nature 2005;**438**:103–107.
37. Westenberger SJ, et al. A systems-based analysis of *Plasmodium vivax* lifecycle transcription from human to mosquito. PLoS Negl Trop Dis 2010;**4**:e653.
38. Zhou Y, et al. Evidence-based annotation of the malaria parasite's genome using comparative expression profiling. PLoS ONE 2008;**3**:e1570.
39. Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. PLoS Biol 2003;**1**:E5.
40. Le Roch KG, et al. Discovery of gene function by expression profiling of the malaria parasite life cycle. Science 2003;**301**:1503–1508.
41. Spellman PT, et al. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol Biol Cell 1998;**9**:3273–3297.
42. Behnke MS, et al. Coordinated progression through two subtranscriptomes underlies the tachyzoite cycle of *Toxoplasma gondii*. PLoS ONE 2010;**5**:e12354.
43. Ganesan K, et al. A genetically hard-wired metabolic transcriptome in *Plasmodium falciparum* fails to mount protective responses to lethal antifolates. PLoS Pathog 2008;**4**:e1000214.
44. Voskuil MI, et al. Inhibition of respiration by nitric oxide induces a *Mycobacterium tuberculosis* dormancy program. J Exp Med 2003;**198**:705–713.
45. Zhu H, Cong JP, Mamtora G, Gingeras T, Shenk T. Cellular gene expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. Proc Natl Acad Sci USA 1998;**95**:14470–14475.
46. Jenner RG, Young RA. Insights into host responses against pathogens from transcriptional profiling. Nat Rev Microbiol 2005;**3**:281–294.
47. Zhang S, Kim CC, Batra S, McKerrow JH, Loke P. Delineation of diverse macrophage activation programs in response to intracellular parasites and cytokines. PLoS Negl Trop Dis 2010;**4**:e648.
48. Lovegrove FE, Pena-Castillo L, Mohammad N, Liles WC, Hughes TR, Kain KC. Simultaneous host and parasite expression profiling identifies tissue-specific transcriptional programs associated with susceptibility or resistance to experimental cerebral malaria. BMC Genomics 2006;**7**:295.
49. Xu X, et al. Transcriptome analysis of *Anopheles stephensi–Plasmodium berghei* interactions. Mol Biochem Parasitol 2005;**142**:76–87.
50. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;**10**:57–63.
51. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 2009;**19**:92–105.
52. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell 2009;**136**:215–233.
53. Cazalla D, Yario T, Steitz J. Down-regulation of a host microRNA by a *Herpesvirus saimiri* noncoding RNA. Science 2010;**328**:1563–1566.
54. Kim L, Butcher BA, Lee CW, Uematsu S, Akira S, Denkers EY. *Toxoplasma gondii* genotype determines MyD88-dependent signaling in infected macrophages. J Immunol 2006;**177**:2584–2591.
55. Robben PM, Mordue DG, Truscott SM, Takeda K, Akira S, Sibley LD. Production of IL-12 by macrophages infected with *Toxoplasma gondii* depends on the parasite



 

genotype. J Immunol 2004;**172**:3686–3694.
56. Saeij JP, Coller S, Boyle JP, Jerome ME, White MW, Boothroyd JC. *Toxoplasma* co-opts host gene expression by injection of a polymorphic kinase homologue. Nature 2007;**445**:324–327.
57. Saeij JP, et al. Polymorphic secreted kinases are key virulence factors in toxoplasmosis. Science 2006;**314**:1780–1783.
58. Taylor S, et al. A secreted serine-threonine kinase determines virulence in the eukaryotic pathogen *Toxoplasma gondii*. Science 2006;**314**:1776–1780.
59. El Hajj H, Lebrun M, Arold ST, Vial H, Labesse G, Dubremetz JF. ROP18 is a rhoptry kinase controlling the intracellular proliferation of *Toxoplasma gondii*. PLoS Pathog 2007;**3**:e14.
60. Peixoto L, et al. Integrative genomic approaches highlight a family of parasite-specific kinases that regulate host responses. Cell Host Microbe 2010;**8**:208–218.
61. Shapira SD, et al. A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. Cell 2009;**139**:1255–1267.
62. Davis FP, Barkan DT, Eswar N, McKerrow JH, Sali A. Host pathogen protein interactions predicted by comparative modeling. Protein Sci 2007;**16**:2585–2596.
63. Dyer MD, Murali TM, Sobral BW. The landscape of human proteins interacting with viruses and other pathogens. PLoS Pathog 2008;**4**:e32.
64. Chen Z, Harb OS, Roos DS. In silico identification of specialized secretory-organelle proteins in apicomplexan parasites and in vivo validation in *Toxoplasma gondii*. PLoS ONE 2008;**3**:e3611.
65. Carruthers V, Boothroyd JC. Pulling together: an integrated model of *Toxoplasma* cell invasion. Curr Opin Microbiol 2007;**10**:83–89.
66. Brenner S. The genetics of *Caenorhabditis elegans*. Genetics 1974;**77**:71–94.
67. Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW. Simultaneous identification of bacterial virulence genes by negative selection. Science 1995;**269**:400–403.
68. Shea JE, Hensel M, Gleeson C, Holden DW. Identification of a virulence locus encoding a second type III secretion system in *Salmonella typhimurium*. Proc Natl Acad Sci USA 1996;**93**:2593–2597.
69. Mecsas J. Use of signature-tagged mutagenesis in pathogenesis studies. Curr Opin Microbiol 2002;**5**:33–37.
70. Mazurkiewicz P, Tang CM, Boone C, Holden DW. Signature-tagged mutagenesis: barcoding mutants for genome-wide screens. Nat Rev Genet 2006;**7**:929–939.
71. Saenz HL, et al. Genomic analysis of *Bartonella* identifies type IV secretion systems as host adaptability factors. Nat Genet 2007;**39**:1469–1476.
72. Sassetti CM, Boyd DH, Rubin EJ. Comprehensive identification of conditionally essential genes in *mycobacteria*. Proc Natl Acad Sci USA 2001;**98**:12712–12717.
73. Saenz HL, Dehio C. Signature-tagged mutagenesis: technical advances in a negative selection method for virulence gene identification. Curr Opin Microbiol 2005;**8**:612–619.
74. Vazquez-Torres A, et al. Salmonella pathogenicity island 2-dependent evasion of the phagocyte NADPH oxidase. Science 2000;**287**:1655–1658.
75. Frankel MB, Mordue DG, Knoll LJ. Discovery of parasite virulence genes reveals a unique regulator of chromosome condensation 1 ortholog critical for efficient nuclear trafficking. Proc Natl Acad Sci USA 2007;**104**:10181–10186.
76. Mordue DG, Scott-Weathers CF, Tobin CM, Knoll LJ. A patatin-like protein protects *Toxoplasma gondii* from degradation in activated macrophages. Mol Microbiol 2007;**63**:482–496.
77. Kamath RS, et al. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature 2003;**421**:231–237.
78. Boutros M, et al. Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. Science 2004;**303**:832–835.
79. Brass AL, et al. Identification of host proteins required for HIV infection through a functional genomic screen. Science 2008;**319**:921–926.
80. Zhou H, et al. Genome-scale RNAi screen for host factors required for HIV replication. Cell Host Microbe 2008;**4**:495–504.
81. Karlas A, et al. Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. Nature 2010;**463**:818–822.
82. Ullu E, Tschudi C, Chakraborty T. RNA interference in protozoan parasites. Cell Microbiol 2004;**6**:509–519.
83. Pfefferkorn ER, Pfefferkorn LC. *Toxoplasma gondii*: isolation and preliminary characterization of temperature-sensitive mutants. Exp Parasitol 1976;**39**:365–376.
84. Donald RG, Roos DS. Stable molecular transformation of *Toxoplasma gondii*: a selectable dihydrofolate reductase-thymidylate synthase marker based on drug-resistance mutations in malaria. Proc Natl Acad Sci USA 1993;**90**:11703–11707.
85. Carey KL, Westwood NJ, Mitchison TJ, Ward GE. A small-molecule approach to studying invasive mechanisms of *Toxoplasma gondii*. Proc Natl Acad Sci USA 2004;**101**:7433–7438.
86. Eidell KP, Burke T, Gubbels MJ. Development of a screen to dissect *Toxoplasma gondii* egress. Mol Biochem Parasitol 2010;**171**:97–103.
87. Joiner KA, Roos DS. Secretory traffic in the eukaryotic parasite *Toxoplasma gondii*: less is more. J Cell Biol 2002;**157**:557–563.
88. Saeij JP, Boyle JP, Grigg ME, Arrizabalaga G, Boothroyd JC. Bioluminescence imaging of *Toxoplasma gondii* infection in living mice reveals dramatic differences between strains. Infect Immun 2005;**73**:695–702.
89. Guiguemde WA, et al. Chemical genetics of *Plasmodium falciparum*. Nature 2010;**465**:311–315.
90. Zhang EE, et al. A genome-wide RNAi screen for modifiers of the circadian clock in human cells. Cell 2009;**139**:199–210.
91. Yamazaki S, et al. Resetting central and peripheral circadian oscillators in transgenic rats. Science 2000;**288**:682–685.
92. Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 2001;**411**:494–498.
93. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 1998;**391**:806–811.
94. Philips JA, Rubin EJ, Perrimon N. *Drosophila* RNAi screen reveals CD36 family member required for mycobacterial infection. Science 2005;**309**:1251–1253.
95. Agaisse H, Burrack LS, Philips JA, Rubin EJ, Perrimon N, Higgins DE. Genome-wide RNAi screen for host factors required for intracellular bacterial infection. Science 2005;**309**:1248–1251.
96. Ramadan N, Flockhart I, Booker M, Perrimon N, Mathey-Prevot B. Design and implementation of high-throughput RNAi screens in cultured *Drosophila* cells. Nat Protoc 2007;**2**:2245–2264.
97. Sabin LR, et al. Ars2 regulates both miRNA- and siRNA- dependent silencing and suppresses RNA virus infection in *Drosophila*. Cell 2009;**138**:340–351.
98. Shelly S, Lukinova N, Bambina S, Berman A, Cherry S. Autophagy is an essential component of *Drosophila* immunity against vesicular stomatitis virus. Immunity 2009;**30**:588–598.
99. Serio AW, Jeng RL, Haglund CM, Reed SC, Welch MD. Defining a core set of actin cytoskeletal proteins critical for actin-based motility of Rickettsia. Cell Host Microbe 2010;**7**:388–398.

 

100. Prudencio M, et al. Kinome-wide RNAi screen implicates at least 5 host hepatocyte kinases in *Plasmodium* sporozoite infection. PLoS Pathog 2008;**4**:e1000201.
101. Temple G, et al. The completion of the Mammalian Gene Collection (MGC). Genome Res 2009;**19**:2324–2333.
102. Warzecha CC, Sato TK, Nabet B, Hogenesch JB, Carstens RP. ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. Mol Cell 2009;**33**:591–601.
103. Bruno VM, Hannemann S, Lara-Tejero M, Flavell RA, Kleinstein SH, Galan JE. *Salmonella Typhimurium* type III secretion effectors stimulate innate immune responses in cultured epithelial cells. PLoS Pathog 2009;**5**:e1000538.
104. Forget G, Gregory DJ, Olivier M. Proteasome-mediated degradation of STAT1alpha following infection of macrophages with *Leishmania donovani*. J Biol Chem 2005;**280**:30542–30549.
105. Matte C, Descoteaux A. *Leishmania donovani* amastigotes impair IFN-γ-induced STAT1α nuclear translocation by blocking the interaction between STAT1α and importin-α5. Infect Immun 2010;**78**:3736–3743.
106. Rayamajhi M, Humann J, Penheiter K, Andreasen K, Lenz LL. Induction of IFN-alphabeta enables *Listeria monocytogenes* to suppress macrophage activation by IFN-gamma. J Exp Med 2010;**207**:327–337.
107. Zhong G, Fan T, Liu L. *Chlamydia* inhibits interferon gamma-inducible major histocompatibility complex class II expression by degradation of upstream stimulatory factor 1. J Exp Med 1999;**189**:1931–1938.
108. Garcia-Sastre A. Mechanisms of inhibition of the host interferon alpha/beta-mediated antiviral responses by viruses. Microbes Infect 2002;**4**:647–655.
109. Hippe D, Weber A, Zhou L, Chang DC, Hacker G, Luder CG. *Toxoplasma gondii* infection confers resistance against BimS-induced apoptosis by preventing the activation and mitochondrial targeting of pro-apoptotic Bax. J Cell Sci 2009;**122**:3511–3521.
110. Carette JE, et al. Haploid genetic screens in human cells identify host factors used by pathogens. Science 2009;**326**:1231–1235.
111. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998;**95**:14863–14868.
112. Dennis G Jr, et al. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 2003;**4**:P3.
113. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 2005;**102**:15545–15550.
114. Cline MS, et al. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc 2007;**2**:2366–2382.
115. Nielsen CB, Cantor M, Dubchak I, Gordon D, Wang T. Visualizing genomes: techniques and challenges. Nat Methods 2010;**7**:S5–15.
116. Gehlenborg N, et al. Visualization of omics data for systems biology. Nat Methods 2010;**7**:S56–68.
117. Aurrecoechea C, et al. EuPathDB: a portal to eukaryotic pathogen databases. Nucleic Acids Res 2010;**38**:D415–419.
118. Greene JM, et al. National Institute of Allergy and Infectious Diseases bioinformatics resource centers: new assets for pathogen informatics. Infect Immun 2007;**75**:3212–3219.
119. Tongren JE, Zavala F, Roos DS, Riley EM. Malaria vaccines: if at first you don't succeed. Trends Parasitol 2004;**20**:604–610.
120. Ochola LI, Tetteh KK, Stewart LB, Riitho V, Marsh K, Conway DJ. Allele frequency-based and polymorphism-versus-divergence indices of balancing selection in a new filtered set of polymorphic genes in *Plasmodium falciparum*. Mol Biol Evol 2010;**27**:2344–2351.
121. Crowther GJ, et al. Identification of attractive drug targets in neglected-disease pathogens using an in silico approach. PLoS Negl Trop Dis 2010;**4**:e804.
122. Aguero F, et al. Genomic-scale prioritization of drug targets: the TDR Targets database. Nat Rev Drug Discov 2008;**7**:900–907.
123. Prathipati P, Ma NL, Keller TH. Global Bayesian models for the prioritization of antitubercular agents. J Chem Inf Model 2008;**48**:2362–2370.


 